STEREOSELECTIVE BIOCATALYTIC C–C BOND FORMATION FOR THE
GENERATION OF γ-HYDROXY AMINO ACIDS AND EXPLORATION OF
PROMISCUITY-GUIDED ENGINEERING METHODS FOR THE DEVELOPMENT OF
GENERALIST BIOCATALYSTS.

by

Meghan E. Campbell

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 05/07/2024

The dissertation is approved by the following members of the Final Examination Committee:
 Andrew R. Buller, Professor, Chemistry
 Tina Wang, Professor, Chemistry
 Jeff Martell, Professor, Chemistry
 Clark Landis, Professor, Organic Chemistry

ACKNOWLEDGMENTS

Thank you to my committee members, all of whom have provided me with support and encouragement over the last five years. It has been a true privilege to benefit from your mentorship and scholarship.

Thank you to Professor Andrew Buller for being an excellent mentor. You have taught me so much about professionalism, scholarship, and mentorship. Your constant support and enthusiasm have carried me through the last five years. I can only hope that wherever I go next, I will work for people as dedicated to their employees and work as you.

To all the former Buller Buddies, I want to thank you for accepting me into the lab and taking the time to teach me all about biology (because I seriously knew nothing!). Particular thanks to Dr. Jon Ellis for all his mentorship and laying the foundation for the UstD project. Without you there would be no UstD project and I am very grateful to you for your tireless effort. To Dr. Peyton Higgins and Dr. Tony Meza, I am so grateful I joined the lab with you two wonderful individuals! We went through some pretty crazy times together (building shutdowns, COVID shutdowns, phage outbreaks), but we made it out with our sanity mostly intact. Thanks for being my voices of reason when I needed them. To Dr. Tyler Doyon, thank you for encouraging me to advocate for myself and providing steadfast support. To the current Buller Buddies, I want to say thanks for being my labmates. The last few years have been extremely rough for our lab, but your tenacity and team-spirit have made work bearable in an otherwise un-bearable situation. The friendships I have made here will last a lifetime, you all are my ride or die! To Amanda Ohler, I know I leave the UstD project in your very capable hands. I cannot wait to see where you take the project! It has been a true pleasure to mentor you.

I would like to thank my family for supporting me throughout my graduate school career. Special thanks to my mom, Stacey, for being my number one supporter and cheerleader, especially during the last few weeks of writing. Thanks to my sisters, Cassie and Keely, for constantly reminding me that I am a girl boss who is capable of anything. To my dad, Doug, thank you so much for encouraging me to pursue chemistry. I never would have done it without your encouragement, and I wish with all my heart you would have lived to see me follow in your footsteps.

To my husband, Dr. Eric Weeda, thank you for everything. I constantly thank whatever luck I live by that I managed to find and keep you. Most people go their whole lives never knowing the kind of unconditional love, support, and partnership you provide me. You are my best friend, my intellectual match, the calm and collected to my fire and chaos. We started this crazy graduate school endeavor together five years ago and now we are finally here at the finish line together. Thank you for keeping me sane. I never would have gotten here without you, and I cannot wait to see where our lives take us.

ABSTRACT

Increasingly, the demand for high quality enzymes for industrial biocatalytic processes is rising. In order to meet these demands, many enzymes have been evolved using directed evolution to improve on one or more enzyme properties. However, a long-standing challenge in biocatalysis is generation of new carbon-carbon bond frameworks using aldolases. Compared to synthetic methods, biocatalytic methods to form carbon-carbon bonds remain limited because aldolase reactions tend to be highly reversible, reducing yield, selectivity, and limiting synthetic utility. UstD, a PLP-dependent aldolase in the Ustiloxin B biosynthesis pathway, purportedly proceeds through an unusual decarboxylation step which acts as a thermodynamic sink driving the reaction in the synthetic direction. Initial characterization of the enzyme with aldehyde electrophiles revealed the native enzyme has broad substrate promiscuity for many aldehyde electrophiles. Using traditional directed evolution techniques, UstD was engineered to increase activity with aldehyde electrophiles leading to two activated variants, UstD^{TLM} and UstD^{FVF}. An additional round of surface cysteine mutagenesis using linear regression modeling led to a computationally identified variant, UstD$^{2.0}$ which had ~5-fold increase in soluble expression. Whole cell reactions using UstD$^{2.0}$ were capable of synthesizing γ-hydroxy amino acids on gram scale. Additionally, the first crystal structure of UstD was obtained from UstD$^{2.0}$ at 2.25-Å. We sought to increase the electrophile scope of UstD$^{2.0}$ to ketones, providing access to chiral tertiary alcohol non-standard amino acids (nsAAs). An alternative directed evolution strategy using a substrate multiplexed screening (SUMS) approach was used to leverage variant promiscuity and activity to traverse the protein fitness landscape. This promiscuity-guided engineering strategy allowed for identification of allosteric sites that shifted enzyme promiscuity towards ketone electrophiles. Subsequent mutagenesis (SSM) and recombination both in and out of the active

site resulted in two variants 7G11 and 7B05 with distinct activity profiles toward ketone electrophiles. Exploration of the ketone scope for each enzyme revealed 7G11 and 7B05 have broad reactivity with ketones providing access to nsAAs with chiral tertiary alcohols. Altogether, these results demonstrate the power of promiscuity-guided evolution for the development of generalist biocatalysts even for challenging reactions.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

1.1 Natural evolution of enzymes

1.1.1 The origins of promiscuity found in modern enzymes

The classical model of enzyme catalysis is described as a lock and key mechanism, where each enzyme is highly specific for a single substrate. However, over time, it has been empirically determined that many enzymes can carry out promiscuous functions, either by using non-native substrates (substrate promiscuity) or supporting alternative catalytic mechanisms (catalytic promiscuity).[1–7] For example, fructose-6-phosphate aldolase, a class I aldolase which catalyzes conversion of fructose-6-phosphate to dihydroxyacetone and glyceraldehyde-3-phosphate, displays substrate promiscuity for both its nucleophile and electrophile substrates.[8] Likewise, a pyridoxal 5'-phosphate (PLP) dependent threonine transaldolase, ObiH, displays broad substrate promiscuity for aldehyde electrophiles.[9]

Beyond substrate promiscuity, enzymes are also capable of catalytic promiscuity, where multiple similar reaction mechanisms are catalyzed by the same enzyme. For example, *Candida antarctica* lipase B (CAL-B) is a native esterase that can also catalyze aldolase activity, promiscuously.[10] The class II aldolase ketopantoate hydroxymethyl transferase (KPHMT) is capable of catalyzing aldol reactions without its native tetrahydrofolate cofactor, achieving the same transformation using an alternative mechanistic pathway.[11] The widespread observation of promiscuous functions across enzyme classes has dispelled the notion that enzymes with high native activity must be highly specialized. Instead, it is now generally accepted that many enzymes have promiscuous functions while some are highly specific.[12] Still, questions remain about how and why these promiscuous functions exist in modern enzymes, sparking inquiries into the evolutionary biology of proteins.

1.1.2 The Patchwork Theory

The predominating evolutionary theory for explaining the presence of promiscuous functions is known as the Patchwork Theory.[2] This theory posits that primitive enzymes, due to lack of resources, had broad promiscuity which allowed them to rapidly adapt in changing conditions. Over time gene duplication, subsequent mutation, and evolutionary divergence allowed for specialization of genes leading to increases in efficiency and narrowing specificity.[2] The divergence processes eventually resulted in the protein families and superfamilies known to us today. The conserved features of these families (i.e. native functions, reactive intermediates, conserved active sites, etc.) are the remnants of the divergence processes.[13]

In a fundamental study performed by Tawfik *et al.*, three different enzymes were evolved to increase their promiscuous functions using random mutagenesis. In evolutionary terms, this is known as adaptive evolution or positive selection and can be depicted on a protein fitness landscape as an uphill climb to increase fitness (**Fig 1.1A**).[14] For this study, the enzymes included were serum paraoxonase (PON1) which natively catalyzes the reversible formation of five and six membered rings, a bacterial phosphotriesterase (PTE) which natively hydrolyzes phosphates, and carbonic anhydrase II (CAII) which natively catalyzes hydration of carbon dioxide. Each enzyme demonstrated promiscuous activities which Tawfik *et al.* engineered towards during evolution.
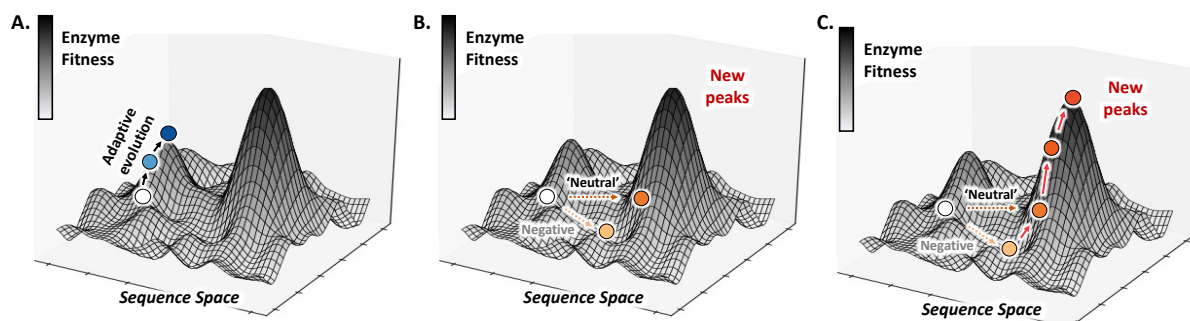
Figure 1.1 Evolution trajectories across protein fitness landscapes using different selection strategies. A. Adaptive evolution trajectory using positive selection with the starting enzyme fitness depicted by the white circle. **B.** Non-adaptive evolution trajectories using either neutral selection or negative selection **C.** Combination of non-adaptive and adaptive evolution trajectories to access new fitness peaks.

The PON1 enzyme was engineered to increase its hydrolysis activity with aryl esters and phosphoesters, PTE was engineered to increase its promiscuous lactonase and esterase activities, and CAII was engineered to increase its promiscuous esterase activity with activated esters. Following evolution, variants displayed a 10-150-fold increase in activity for their promiscuous function compared to parent activity. Despite the changes in the promiscuous activity, the native reactions were only modestly affected, 1.4-5-fold decrease in activity. Exploration of the literature found that among 18 studies where promiscuous activity was increased by >1000-fold on average, the native function only decreased by an average 3.2-fold. Therefore, Tawfik hypothesized that enzymes can be evolved for large increases in promiscuous activity while maintaining native activity.[14] Additionally, it was proposed that the variants reached a more 'generalist' state capable of catalyzing both the native and promiscuous reactions. These results provided early evidence for the Patchwork Theory by empirically showing that protein evolution can proceed through a generalized intermediate state before re-specializing for the function under selection.[14]

1.1.2 Nonadaptive evolution

Tawfik's observations contributed to a related hypothesis proposed by Wagner regarding the role of neutral/epistatic drift in enzyme evolution, also referred to as non-adaptive evolution. Wagner argued that in systems that are robust to mutation (maintain native activity), neutral mutations, which neither improve nor degrade activity, can drive innovation of new functions.[15] He also redefined neutrality to state that a neutral mutation is one that does not change one aspect of a biological system *in a specific environment or genetic background*, meaning a mutation that appears neutral under one selection can cause changes in fitness under other selections. The first empirical study to provide support for the idea that neutral mutations are the driver of innovation was performed by Tawfik *et al.*[16] Using the enzyme PON1, random mutation and selection for variants whose activity and expression were within parent levels (neutral selection) led to a collection of neutral variants whose promiscuous functions were studied. Overall, they found that neutral mutations typically changed promiscuity by changing the specificity of the enzyme when assayed on pairs of substrates in competition. Additionally, they found that some of the neutral variants had lower levels of inhibition than parent when exposed to the same known inhibitor. Therefore, they reasoned these specificity changes are a form of latent adaptation that could provide the enzymes with a selective advantage by expanding reactivity. Additionally, they asserted that neutral movement across a protein fitness landscape brings different genotypes within closer proximity of a transition point between phenotypes (**Fig. 1.1B**).[16] Therefore, neutral drift could provide diverse starting points for further evolution.

Another landmark study in this area, done by Ostermeier *et al.*, focused on discovering potential evolutionary mechanisms for enzymes to traverse across fitness valleys when

confronted with rugged fitness landscapes.[17] They hypothesized that by maintaining sequence,

but changing environment, the fitness landscapes may be altered such that low fitness sequences

may have high fitness in the new environment. To test the theory, comparison of four

evolutionary strategies (positive, neutral, negative, and oscillating) were used to interrogate how

effective each is at producing high fitness variants in changing environments.[17] Each selection

strategy would lead to different trajectories across the fitness landscape (**Fig. 1.1A-C**), providing

information about how natural evolution navigates across fitness landscape. For this study, the β-

lactamase TEM-15, was chosen as the parent enzyme, and clones bearing variants of TEM-15

were selected for cefotaxime resistance.[17] Eight total rounds of evolution for cefotaxime

resistance were performed for all selections, with the first three rounds under the specified

selection and the following five under positive selection. Under positive selection, variants were

selected for increased resistance only. Under neutral selection, variants with the same resistance

as parent were selected. Under negative selection, variants were selected for resistance just above

background *E. coli* resistance. The first three rounds of mutagenesis for the oscillating selection

had two rounds of mutagenesis under negative selection separated by one round under positive

selection followed by five rounds under positive selection.

 All the selections led to variants with higher resistance than parent, but the variants

identified through neutral and negative selection displayed higher resistance than the variants

under positive and oscillating selection.[17] Surprisingly, variants generated under negative

selection conferred the greatest increase in cefotaxime resistance. Deep sequencing and analysis

of the variants under all selection types indicated that the variants under negative selection had a

mutation F230S that was not present in the other selections. When F230S is added as a single

mutation to TEM-15, it is highly deleterious. However, retrospective analysis of other

evolutionary intermediates with and without F230S present indicated that the residue contributes to higher order epistatic effects.[17] The average fitness of the proteins containing F230S was five times higher than expected based on the effects of the individual mutations in TEM-15. Therefore, F230S alleviated some negative interactions between the other beneficial mutations leading to the highest fitness variant.[17] Altogether these data show how negative and neutral selection can diversify genotypes providing valuable starting points for adaptive evolution, especially under changing selection (**Fig. 1.1C**). Natural evolution may follow similar evolutionary trajectories especially when confronted with a rugged fitness landscape as a mechanism to move between fitness peaks.

1.1.3 Evolvability lessons from evolutionary biology

Increasingly, the demand for high quality enzymes for industrial biocatalytic processes is rising.[18] The properties required of these industrial biocatalysts include high catalytic efficiency, tolerance to high substrate loading, high expression titer, chem- regio- and stereoselectivity.[19] In order to meet these demands, many enzymes have been evolved using directed evolution to improve on one or more enzyme properties.[18,20] Therefore, understanding the underlying forces behind natural evolution can provide insight for directed evolution strategies. The selected examples discussed in this section describe several key concepts of natural evolution:

1. Nature evolves enzymes using iterative cycles of mutagenesis and selection.

2. Early variants in an evolutionary lineage might have more generalized function prior to specialization for the reaction under selection.[14]

3. Highly evolved enzymes may nevertheless maintain features of substrate and catalytic promiscuity.

4. Neutral/deleterious mutations can increase the evolutionary adaptive potential of a variant by changing latent functions not under selection.[15–17]

Taken altogether, these ideas provide a framework to think more broadly about the underlying driving forces of natural evolution and how they may be harnessed for *in vitro* enzyme evolution.

1.2 Directed evolution of proteins

1.2.1 Protein engineering and directed evolution

Over the last several decades, enzymes have caught the attention of chemical industry as potential catalysts for complex synthesis under safer and more environmentally friendly conditions.[18–21] As a result, development of numerous strategies to increase desirable promiscuous functions through mutation of the native sequence have been explored.[22] The most successful of all these strategies is directed evolution, which uses iterative rounds of mutagenesis and artificial selection to generate new enzymes.[23] The method, pioneered by Professor Frances Arnold, won the Nobel Prize in 2018. The iterative nature of directed evolution expertly mimics natural evolution on a laboratory timescale allowing for quick identification of improved catalysts. Crucially, ***the method requires no detailed mechanistic or structural information*** and can be successfully used to enhance most researcher defined parameters if an appropriate selection can be developed. A typical experimental workflow (**Fig. 1.2**). follows a general pattern of parent enzyme identification, mutagenesis of parent DNA, expression, screening of variants for increases in fitness, and forwarding of high fitness variants to use as the new parent enzyme. Each stage of this process is customizable, allowing flexibility over choosing a starting enzyme, as well as the mutagenesis and screening strategies.
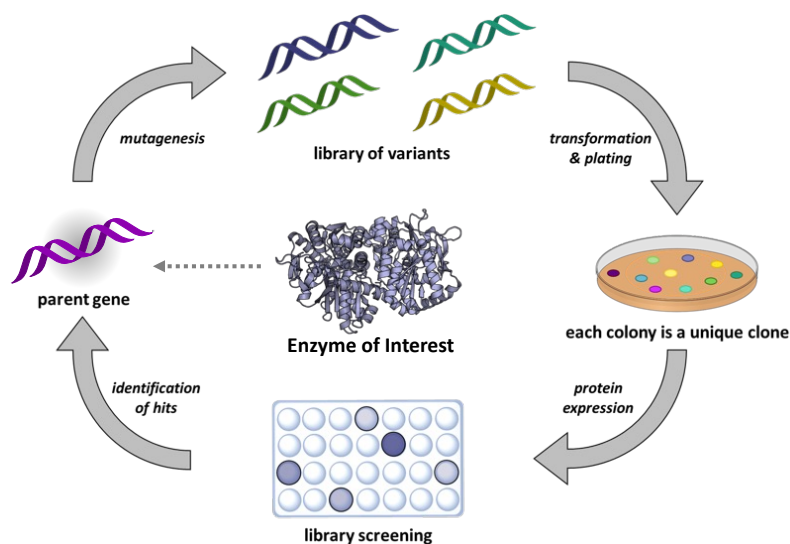
Figure 1.2 Directed evolution cycle.

1.2.2 Considerations for selecting a parent enzyme for evolution

Choosing a starting enzyme is usually straightforward, but some forethought can improve the ease of engineering. First and foremost, choosing an enzyme that performs some base level of desired reactivity can reduce the engineering burden since this process depends on adaptive evolution. It is very rare to perform successful evolution without some small amount of detectable initial activity.[23] Alternative factors to consider are the stability of the enzyme since this property has been linked to evolvability.[24] For example, an enzyme from a thermophilic organism might be more robust to mutations because it is more stable than a mesophilic protein in order to resist unfolding in extreme environments.[24] Once a starting point is chosen, the more complex considerations of mutagenesis and screening can be considered.

1.2.3 Mutagenesis techniques for directed evolution

Many strategies have been developed to diversify a protein's sequence during directed evolution. Three common strategies are global random mutagenesis, site directed mutagenesis, and rational design. Global random mutagenesis only requires knowledge of the genetic sequence of the enzyme of interest to use and is easily implemented through error-prone PCR.

For these reasons, it is a popular diversification strategy. Site directed mutagenesis combines aspects of naïve and rational design but can be used effectively with minimal prior information. Comparatively, rational design approaches require more detailed structural and/or mechanistic information usually obtained by a combination of empirical and computational methods. While rational design can be very powerful, it is a more specialized strategy because it requires both enzyme characterization and knowledge of computational methods.

Global random mutagenesis is typically used to identify distal sites or regions that improve enzyme fitness (**Fig. 1.3A**). Libraries generated through error-prone PCR can tune the mutational loading by changing the concentration of Mn^{2+}.[25] A combinatorial approach to global random mutagenesis encompasses DNA shuffling or staggered extension process (StEP).[26,27] DNA shuffling entails pooling of variant or homologous DNA, digestion of the pooled DNA into small fragments then re-ligation of the fragments using a PCR to create a new mutant gene.[26] StEP is a very similar process, but instead of digesting the DNA, it is used as a template in a PCR reaction with very short extension and annealing times which generates the fragment DNA.[27]

Site directed mutagenesis is targeted mutagenesis of pre-selected residues (**Fig. 1.3B**). Site directed mutagenesis is usually a mixture between rational design and naïve mutagenesis as it is most effective when there is some rationale behind choosing the targeted sites, but it is not required. A highly effective strategy for site directed mutagenesis is targeting the first shell residues of the active site. Since the first shell encompasses residues in direct contact with the reactive substrate, cofactor, or product, they often provide the biggest changes in fitness when mutated. Point mutants are typically made by site saturation mutagenesis (SSM) at a single site

using the '22-codon trick' to encode all possible amino acids at that site with minimal

redundancy.[28]
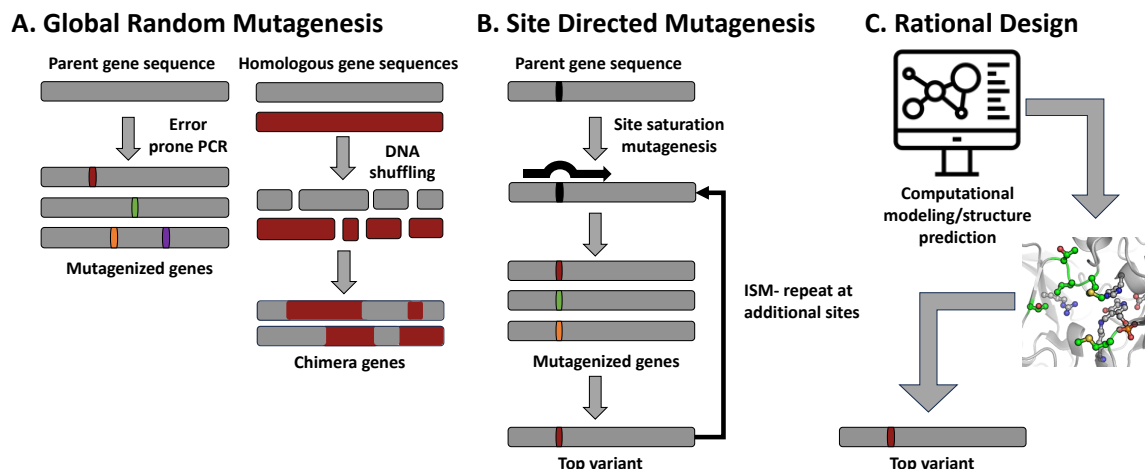
Figure 1.3 Selected examples of mutagenesis strategies. A. Global random mutagenesis to identify individual sites using error prone PCR and recombination of regions using DNA shuffling. **B.** Site directed mutagenesis strategies for targeting single site using site saturation mutagenesis and recombining multiple targeted sites using iterative site saturation mutagenesis (ISM). **C.** Rational design workflows for generating activated variants. Credit for icons: Computer icons created by FBJan – Flaticon https://www.flaticon.com/free-icons/computer

Combinatorial methods include iterative site saturation mutagenesis (ISM) and combinatorial

active-site saturation test (CASTing).[29,30] ISM requires fixing of the best mutation after each

SSM library before screening the next SSM library. While there could be some bias of epistatic

effects by fixing mutations after each library, this approach reduces the library size (number of

possible mutants) compared to a library where multiple sites are mutated simultaneously.[29]

CASTing involves simultaneous randomization of pairs of active site residues that are

structurally close by SSM. By limiting the combinatorial space to two sites, the library size is

lowered so all possible combinations can reasonably be screened and all possible beneficial

epistatic interactions would be identified for that pair.[30] With all combinatorial libraries the

library size is a critical feature to manage, as statistically three times the number of possible

mutants must be screened to achieve 95% confidence that the entire library was sampled.[30]

Therefore, a balance must be struck between the practical limitations of screening and thorough exploration of sequence space.

Rational design is an extremely broad term to define any evolution strategy where specific changes are made on the basis of a structure/activity hypothesis (**Fig. 1.3C**). Examples of rational design range from using structural information to identify relevant residues for mutation to *de novo* protein design.[31,32] Structure prediction has seen major increases in the number of available tools, such as AlphaFold and RosettaDesign, capable of delivering accurate structural models.[44,45] These tools alleviate the burden of protein crystallography or cryo-electron microscopy to obtain structural information as many proteins are difficult or unable to be structurally characterized. Without some form of structural information, application of rational design is severely limited making these structure design tools all the more crucial for gaining information. While these tools are quite robust, not all structure designs are functional. Recent examples have emerged where a sequence re-design algorithm, ProteinMPNN, can successfully rescue failed designs.[46] All together these tools lay the foundation for robust sequence/structure prediction for rational design strategies.

A fairly standard mutagenesis technique for rational design is mutational scanning where SSM is performed at every or every other residue in the protein. While the required cloning and screening are more extensive this strategy has proven to be effective at identifying sites amenable to mutation including distal residues.[31,33–35] This approach is especially valuable since distal residues have been shown to confer stability, activity, and selectivity benefits.[36,37] However, streamlined methods to access this information would be highly desirable. The ability of computation-based rational design to comprehensively analyze vast sequence spaces *in silico* can significantly reduce screening efforts to identify distal residues that effect catalytic activity.[36]

While computation is currently used on a case-by-case basis, recent successful engineering campaigns using computationally-guided methods to increase enzyme activity indicate that in time generalized and robust methods will follow.[38–43]

1.2.4 Screening and selection strategies for directed evolution

1.2.4.1 Adaptive walking using model substrate screening

Evaluation of variant fitness is usually based on increases in activity for a single reaction using a model substrate (**Fig 1.4A**). This is often described as an "adaptive walk" or "positive selection" where only upward trajectory on a fitness landscape is considered.[23] This contrasts with the more circuitous routes of sequence sampling performed during natural evolution which notably incorporate more neutral and slightly deleterious mutations in complex mixtures of biomolecules and metabolites before adapting new functions.[2,15,17,23] Nevertheless, the literature is replete with successful examples of enzyme engineering using model substrate screens.
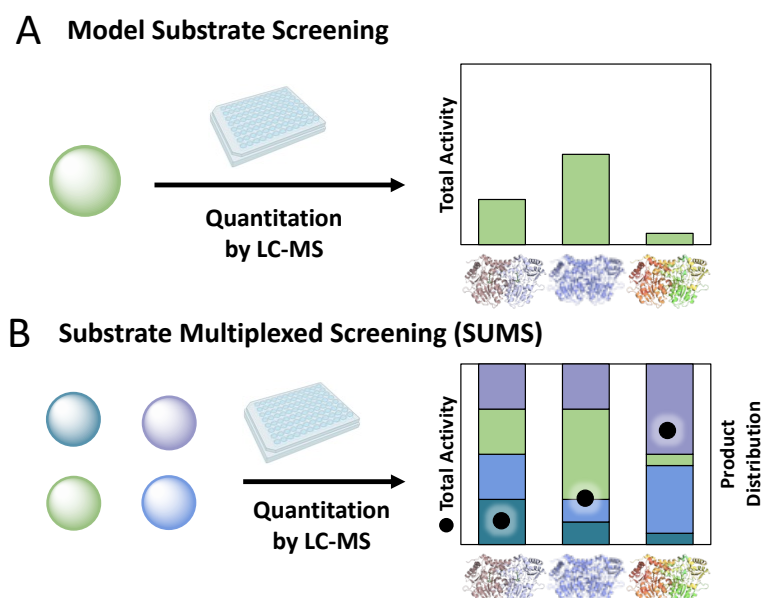
Figure 1.4 A. Model substrate screening and subsequent visualization of variant activity against the substrate depicted by the green bars. **B.** Substrate multiplexed screening (SUMS) where multiple substrates are screened in competition. Subsequent data visualization shows total activity on all substrates with black dots while the distribution in products (promiscuity) is depicted by the colored bars.

In theory, methods capable of leveraging neutral and slightly deleterious mutations could allow movement across fitness landscapes as Tawfik and Ostermeier have both demonstrated the utility of neutral and negative selection for evolution (**Fig. 1.1B-C**).[15–17] Adaptive walks can only move effectively by increasing fitness, making traversing the fitness landscape across fitness valleys and to new peaks difficult (**Fig. 1.1A**).[23] However, there are two principle arguments that alleviate some of these concerns. While sequence space is vast, most mutations are deleterious and therefore most protein sequences do not encode functional proteins, limiting fitness peaks.[47,48] Additionally, it has been noted that functional sequences cluster together closely on the fitness landscape.[49] Taken together, we can interpret this observation to mean that on smoother fitness landscapes, the chances of reaching a local maxima trap is less likely than what might be predicted given the vastness of sequence space.

A separate, but related consideration concerns evolving for enzyme generality. The adage "you get what you screen for" is a common mantra in directed evolution.[23] As most directed evolution campaigns use a single model substrate for evolution, this can lead to generation of enzymes with specialist activity for the model substrate limiting their synthetic utility.[50–52] Therefore, development of methods for intentional selection of generalist enzymes remains an active area of research.

1.2.4.2 Adaptive walking using substrate multiplexed screening (SUMS)

Developing synthetically relevant enzymes is a standard objective in many protein engineering campaigns. However, the promiscuity of variants is rarely tested directly during the engineering phase of a project. This approach can lead to the inadvertent generation of specialist enzymes which could require evolution to be repeated if activity on an alternative substrate is desired. Additionally, when promiscuity is tested, the libraries are typically re-screened with

different substrates to generate promiscuity profiles for each variant.[30,53] This is a very tedious and resource intensive process. Recently, our group proposed a substrate multiplexed screening (SUMS) approach to directed evolution in which multiple substrates are screened in competition so that activity and promiscuity of each variant are assessed simultaneously (**Fig 1.4B**).[54] The goal of SUMS is to identify synthetically useful biocatalysts during engineering by screening with a diverse substrate scope.[54] Besides our study, only a few examples of competition screening to develop synthetically relevant biocatalysts have been reported using small, rationally designed active site libraries.[55–57] We hypothesize that broad adoption of SUMS for protein engineering is lacking due to uncertainties associated with the underlying factors governing these competition screens and their data interpretation. To expand knowledge in the field and alleviate these concerns, we engineered two separate enzymes, a L-tryptophan decarboxylase (TDC) and 2B9 (a variant of the tryptophan synthase β subunit), using SUMS for increased synthetic utility to provide guidance on how SUMS could be used in protein engineering contexts.[54]

Beginning with TDC, the enzyme natively catalyzes the unimolecular decarboxylation of L-tryptophan into tryptamine and has been shown by our lab to have broad substrate promiscuity, accepting many Trp analogues.[58] However, we noted TDC had poor activity with 4- and 5-substituted tryptophan analogues requiring engineering to expand the scope. Using targeted active site SSM and mixtures of substrates, several variants were identified with distinct promiscuity profiles.[54] Upon validation of these variants, several different point mutants were identified that increased activity with substrate classes beyond the initially desired ones.[54]

The engineered tryptophan synthase β-subunit, 2B9, enzyme catalyzes the bimolecular condensation of serine and indole and we sought to increase its synthetic utility by improving

activity with indole analogues.[37,59] Since this reaction is biomolecular, the kinetics of a competition screen are much more complex since the promiscuity profiles can become decoupled from catalytic efficiencies under pseudo-first order conditions.[54] Nevertheless, screening a single round of global random mutagenesis using SUMS identified the distal residue His275 which changed promiscuity, but did not increase activity.[54] Subsequent SSM at this sight led to the variant H275E which had increased activity on substrates included in the screen as well as those that were not. In this way, SUMS can be used to identify 'hotspot' residues that shift promiscuity and use them as evolutionary starting points to engineer for activated variants with broadened substrate scopes.[54]

The advantage of using SUMS screening over single substrate screening is that changes in promiscuity can be observed even for neutral and deactivated variants. Consequently, changes in promiscuity for low activity variants can be leveraged to find activated variants with altered promiscuity.[54] This method aims to utilize both the power of adaptive walks and the latent evolutionary potentials of neutral/deleterious mutations described by Tawfik and Ostermeier.[16,17] While the full extent of advantages and disadvantages have yet to be explored, SUMS provides an intriguing alternative to traditional model substrate screening.

1.2.5 Conclusions and perspectives on directed evolution

Directed evolution has been integral in developing high quality enzymes for biocatalytic applications. By mimicking natural evolution on the laboratory timescale, quick diversification and enhancement of desirable qualities are imbued into enzymes. While there are many effective mutagenesis and screening strategies to increase promiscuous functions, these methods struggle to identify distal residues that modulate activity. Additionally, standard adaptive methods make traversing the protein fitness landscape to new fitness peaks extremely challenging. While

computational aided directed evolution has seen significant advances in recent years for reducing screening load and traversing fitness landscapes effectively, this is still a specialized technique that requires understanding of computational methods. An alternative screening method, SUMS, is beginning to gain some traction as a straightforward way to apply standard engineering techniques to competition screening for the development of synthetically useful biocatalysts.

1.3 Perspective on underdeveloped biocatalytic C–C bond formation

With the wealth of well-developed enzyme catalyzed transformations, identifying underdeveloped biocatalytic transformations can be challenging. In general, directed evolution has improved many enzymes of diverse function to facilitate functional group interconversions.[18,19] Additionally, current research focuses on the introduction of new reaction pathways into known enzymes.[60–63] However, a long standing challenge in the field is generation of new carbon-carbon bond frameworks using aldolases.[64–66]

There are three key classes of aldolases Class I, Class II, and PLP dependent (**Fig. 1.5**). Class I aldolases form Schiff-bases with an active site residue to covalently bind their nucleophilic substrates for attack into an electrophile. Class II aldolase use divalent cations to generate metal bound enolates to perform nucleophilic addition into electrophiles. Lastly, PLP-dependent aldolases use pyridoxal-5'-phosphate (PLP) as a cofactor to covalently bind their substrate via Schiff-base, then proceed through mechanistically diverse pathways to perform aldol type reactions. Compared to synthetic methods, biocatalytic methods to form carbon-carbon bonds remain limited because aldol reactions rely on the formation of resonance-stabilized carbanions. Formation of carbanions in water is challenging due to protonation of the reactive intermediate. Additionally, aldolase reactions tend to be highly reversible, reducing

yield, selectivity, and limiting synthetic utility.[64–66] Nevertheless, the demand for stereoselective enzymatic C–C bond forming reactions continues to promote development of aldolases.
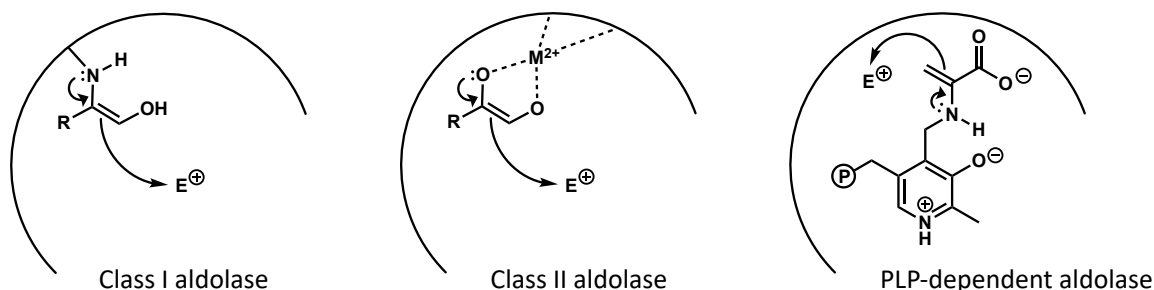
Figure 1.5 The key intermediate for the carbon bond formation step of each aldolase class. The semi-circle represents a generic enzyme scaffold. The PLP-dependent intermediate comes from the proposed key intermediate of UstD.[67]

In particular, our lab has focused on the study of PLP-dependent C–C bond forming enzymes because the breadth of chemistry the cofactor enables.[65,68] PLP dependent aldolases have shown promise as capable biocatalysts for C–C bond formation.[65,68,69] For example, our lab showed that ObiH, a PLP dependent threonine transaldolase, has broad native activity on a variety of aldehyde substrates.[9,70] Additionally, identification of a PLP-dependent aldolase, UstD, in the Ustiloxin B biosynthesis pathway proceeds through an unusual decarboxylation step.[67,71] This enzyme could show excellent promise as a biocatalyst since the loss of CO_2 acts as a thermodynamic sink driving the reaction in the synthetic direction. Continued development of aldolases using the directed evolution techniques described in this chapter will provide new synthetic disconnections not currently available to chemists.

1.4 References

1.   Tawfik, O. K. and D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79**, 471–505 (2010).

2.   Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**,

409–425 (1976).

3. Li, G. *et al.* Substrate profiling of cyclohexylamine oxidase and its mutants reveals new biocatalytic potential in deracemization of racemic amines. *Appl. Microbiol. Biotechnol.* **98**, 1681–1689 (2014).

4. Huang, H. *et al.* Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1974–E1983 (2015).

5. Zhang, X. *et al.* Biochemical characterization and substrate profiling of a reversible 2,3-dihydroxybenzoic acid decarboxylase for biocatalytic Kolbe-Schmitt reaction. *Enzyme Microb. Technol.* **113**, 37–43 (2018).

6. Andorfer, M. C. *et al.* Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal.* **7**, 1897–1904 (2017).

7. Schülke, K. H., Ospina, F., Hörnschemeyer, K., Gergel, S. & Hammer, S. C. Substrate Profiling of Anion Methyltransferases for Promiscuous Synthesis of S-Adenosylmethionine Analogs from Haloalkanes. *ChemBioChem* **23**, e202100632 (2022).

8. Concia, A. L. *et al.* D-Fructose-6-phosphate Aldolase in Organic Synthesis: Cascade Chemical-Enzymatic Preparation of Sugar-Related Polyhydroxylated Compounds. *Chem. – A Eur. J.* **15**, 3808–3816 (2009).

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. Branneby, C. *et al.* Carbon-carbon bonds by hydrolytic enzymes. *J. Am. Chem. Soc.* **125**, 874–875 (2003).

11. Sugantino, M., Zheng, R., Yu, M. & Blanchard, J. S. Mycobacterium tuberculosis

ketopantoate hydroxymethyltransferase: Tetrahydrofolate-independent hydroxymethyltransferase and enolization reactions with α-Keto acids. *Biochemistry* **42**, 191–199 (2003).

12.   Copley, S. D. Shining a light on enzyme promiscuity. *Curr. Opin. Struct. Biol.* **47**, 167–175 (2017).

13.   Gerlt, J. A. & Babbitt, P. C. Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* **70**, 209–246 (2001).

14.   Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nat. Genet. 2004 371* **37**, 73–76 (2004).

15.   Wagner, A., Russell, R. & Superti-Furga, G. Robustness, evolvability, and neutrality. *FEBS Lett.* **579**, 1772–1778 (2005).

16.   Amitai, G., Gupta, R. D. & Tawfik, D. S. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **1**, 67–78 (2007).

17.   Steinberg, B. & Ostermeier, M. Environmental changes bridge evolutionary valleys. *Sci. Adv.* **2**, (2016).

18.   France, S. P., Lewis, R. D. & Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au* **3**, 715–735 (2023).

19.   Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chemie Int. Ed.* **60**, 88–119 (2021).

20.   Radley, E. *et al.* Engineering Enzymes for Environmental Sustainability. *Angew. Chemie*

Int. Ed. **62**, e202309305 (2023).

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

22. Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angew. Chemie Int. Ed.* **50**, 138–174 (2011).

23. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

24. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5869–5874 (2006).

25. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci.* **90**, 5618–5622 (1993).

26. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nat. 1994 3706488* **370**, 389–391 (1994).

27. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol. 1998 163* **16**, 258–261 (1998).

28. 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).

29. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc. 2007 24* **2**, 891–903 (2007).

30. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the Range of

Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chemie Int. Ed.* **44**, 4192–4196 (2005).

31.　Honda Malca, S. *et al.* Effective engineering of a ketoreductase for the biocatalytic synthesis of an ipatasertib precursor. *Commun. Chem. 2024 71* **7**, 1–11 (2024).

32.　Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387 (2008).

33.　Guo, C., Biewenga, L., Lubberink, M., van Merkerk, R. & Poelarends, G. J. Tuning Enzyme Activity for Nonaqueous Solvents: Engineering an Enantioselective "Michaelase" for Catalysis in High Concentrations of Ethanol. *ChemBioChem* **21**, 1499–1504 (2020).

34.　Guo, C. *et al.* Using Mutability Landscapes To Guide Enzyme Thermostabilization. *ChemBioChem* **22**, 170–175 (2021).

35.　van der Meer, J. Y., Biewenga, L. & Poelarends, G. J. The Generation and Exploitation of Protein Mutability Landscapes for Enzyme Engineering. *ChemBioChem* **17**, 1792–1799 (2016).

36.　Gu, J., Xu, Y. & Nie, Y. Role of distal sites in enzyme engineering. *Biotechnol. Adv.* **63**, 108094 (2023).

37.　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).

38.　Ma, E. J. *et al.* Machine-Directed Evolution of an Imine Reductase for Activity and Stereoselectivity. *ACS Catal.* **11**, 12433–12445 (2021).

39.　Büchler, J. *et al.* Algorithm-aided engineering of aliphatic halogenase WelO5* for the

asymmetric late-stage functionalization of soraphens. *Nat. Commun. 2022 131* **13**, 1–11 (2022).

40. Saito, Y. *et al.* Machine-Learning-Guided Library Design Cycle for Directed Evolution of Enzymes: The Effects of Training Data Composition on Sequence Space Exploration. *ACS Catal.* **11**, 14615–14624 (2021).

41. Lu, H. *et al.* Machine learning-aided engineering of hydrolases for PET depolymerization. *Nat. 2022 6047907* **604**, 662–667 (2022).

42. Vornholt, T. *et al.* Systematic engineering of artificial metalloenzymes for new-to-nature reactions. *Sci. Adv.* **7**, 4208–4230 (2021).

43. Greenhalgh, J. C., Fahlberg, S. A., Pfleger, B. F. & Romero, P. A. Machine learning-guided acyl-ACP reductase engineering for improved in vivo fatty alcohol production. *Nat. Commun. 2021 121* **12**, 1–10 (2021).

44. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science (80-.).* **373**, 871–876 (2021).

45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* **596**, 583–589 (2021).

46. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science (80-.).* **378**, 49–56 (2022).

47. Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nat. 2001 4106829* **410**, 715–718 (2001).

48. Axe, D. D. Estimating the Prevalence of Protein Sequences Adopting Functional Enzyme Folds. *J. Mol. Biol.* **341**, 1295–1315 (2004).

49.   Taverna, D. M. & Goldstein, R. A. Why are proteins so robust to site mutations? *J. Mol. Biol.* **315**, 479–484 (2002).

50.   Romney, D. K., Sarai, N. S. & Arnold, F. H. Nitroalkanes as Versatile Nucleophiles for Enzymatic Synthesis of Noncanonical Amino Acids. *ACS Catal.* **9**, 8726–8730 (2019).

51.   Blikstad, C., Dahlström, K. M., Salminen, T. A. & Widersten, M. Substrate scope and selectivity in offspring to an enzyme subjected to directed evolution. *FEBS J.* **281**, 2387–2398 (2014).

52.   Matsumura, I. & Ellington, A. D. In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds Through Non-specific Intermediates. *J. Mol. Biol.* **305**, 331–339 (2001).

53.   Chen, W. *et al.* Promiscuous enzymatic activity-aided multiple-pathway network design for metabolic flux rearrangement in hydroxytyrosol biosynthesis. *Nat. Commun. 2019 101* **10**, 1–12 (2019).

54.   Mcdonald, A. D., Higgins, P. M. & Buller, A. R. Substrate multiplexed protein engineering facilitates promiscuous biocatalytic synthesis. doi:10.1038/s41467-022-32789-w

55.   Jakoblinnert, A. *et al.* Reengineered carbonyl reductase for reducing methyl-substituted cyclohexanones. *Protein Eng. Des. Sel.* **26**, 291–298 (2013).

56.   Junker, S., Roldan, R., Joosten, H. J., Clapés, P. & Fessner, W. D. Complete Switch of Reaction Specificity of an Aldolase by Directed Evolution In Vitro: Synthesis of Generic Aliphatic Aldol Products. *Angew. Chemie Int. Ed.* **57**, 10153–10157 (2018).

57.   Knorrscheidt, A. *et al.* Simultaneous screening of multiple substrates with an unspecific

peroxygenase enabled modified alkane and alkene oxyfunctionalisations. *Catal. Sci. Technol.* **11**, 6058–6064 (2021).

58. McDonald, A. D., Perkins, L. J. & Buller, A. R. Facile in Vitro Biocatalytic Production of Diverse Tryptamines. *ChemBioChem* **20**, 1939–1944 (2019).

59. Herger, M. *et al.* Synthesis of β-Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase. *J. Am. Chem. Soc.* **138**, 8388–8391 (2016).

60. Bender, S. G. & Hyster, T. K. Pyridylmethyl Radicals for Enantioselective Alkene Hydroalkylation Using "Ene"-Reductases. *ACS Catal.* **13**, 14680–14684 (2023).

61. Gu, Y., Natoli, S. N., Liu, Z., Clark, D. S. & Hartwig, J. F. Site-Selective Functionalization of (sp3)C−H Bonds Catalyzed by Artificial Metalloenzymes Containing an Iridium-Porphyrin Cofactor. *Angew. Chemie* **131**, 14092–14098 (2019).

62. Cheng, L. *et al.* Stereoselective amino acid synthesis by synergistic photoredox-pyridoxal radical biocatalysis. *Science (80-.).* **381**, 444–451 (2023).

63. Thorpe, T. W., Marshall, J. R. & Turner, N. J. Multifunctional Biocatalysts for Organic Synthesis. *J. Am. Chem. Soc.* (2024). doi:10.1021/JACS.3C09542

64. Schmidt, N. G., Eger, E. & Kroutil, W. Building Bridges: Biocatalytic C-C-Bond Formation toward Multifunctional Products. *ACS Catalysis* **6**, 4286–4311 (2016).

65. Zetzsche, L. E. & Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nat. Rev. Chem.* **4**, 334–346 (2020).

66. Fesko, K. & Gruber-Khadjawi, M. Biocatalytic Methods for C-C Bond Formation. *ChemCatChem* **5**, 1248–1272 (2013).

67. Ye, Y. *et al.* Unveiling the Biosynthetic Pathway of the Ribosomally Synthesized and

Post-translationally Modified Peptide Ustiloxin B in Filamentous Fungi. *Angew. Chemie Int. Ed.* **55**, 8072–8075 (2016).

68.   Du, Y. L. & Ryan, K. S. Pyridoxal phosphate-dependent reactions in the biosynthesis of natural products. *Nat. Prod. Rep.* **36**, 430–457 (2019).

69.   Fesko, K. Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Appl. Microbiol. Biotechnol.* **100**, 2579–2590 (2016).

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

71.   Umemura, M. *et al.* Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in Aspergillus flavus. *Fungal Genet. Biol.* **68**, 23–30 (2014).

CHAPTER TWO: BIOCATALYTIC SYNTHESIS OF NON-STANDARD AMINO ACIDS BY A DECARBOXYLATIVE ALDOL REACTION

2.1 Introduction

Major advances have been made in the practical use of enzymes for enantioselective

functional group manipulations[1]. For example, asymmetric reduction of ketones and

enantiospecific hydrolysis of racemic esters are now routine in process chemistry. There have

also been impressive strides made in enzymatic C-H activation[2]. However, the development of

enzymes to form C–C bonds on preparative scale lags far behind traditional synthetic organic

methodology[3]. While nature is rife with C–C bond forming enzymes,[4,5] these catalysts often have

significant limitations, such as limited substrate scope or poor heterologous expression.[6] While

engineering can overcome these challenges, a more severe limitation is thermodynamic in

nature: reactions that form carbon nucleophiles via C-H deprotonation, such as classic aldol

transformations, are typically reversible.[7] In nature, metabolic flux drives reactions and preserves

the stereochemical purity of products. Laboratory approaches mimic nature by coupling

reversible biocatalytic C–C   bond forming reactions to a thermodynamic sink, such as a

subsequent transformation or selective crystallization.[8–11] While these advances are substantial,

the potential of biocatalytic enzymes in assembling carbon chains is still hindered by the simple

lack of high-quality, exergonic transformations.[12] Hence, development of scalable and

thermodynamically favorable C–C bond forming reactions may open avenues of biocatalytic

synthesis.

To fill this gap, we were drawn to a recently described pyridoxal phosphate (PLP)

dependent enzyme involved in the biosynthesis of Ustiloxin B, an inhibitor of microtubilin

polymerization (**Fig. 2.1A**)[13]. This enzyme, UstD, decarboxylates the side chain of L-aspartate

(**2.1**), forming a putative nucleophilic enamine intermediate (**Fig. 2.1B**). This enamine then

attacks an aliphatic aldehyde appended to a cyclic tetrapeptide, resulting in the formation of a γ-hydroxy amino acid side chain. The loss of CO_2 renders this enantioselective C–C bond forming reaction effectively irreversible. This decarboxylative aldol addition mechanism is distinct from the classic aldolases, transketolases, and PLP-dependent Thr aldolases, which catalyze tautomerization of an imine to form an enamine nucleophile.[14,15]
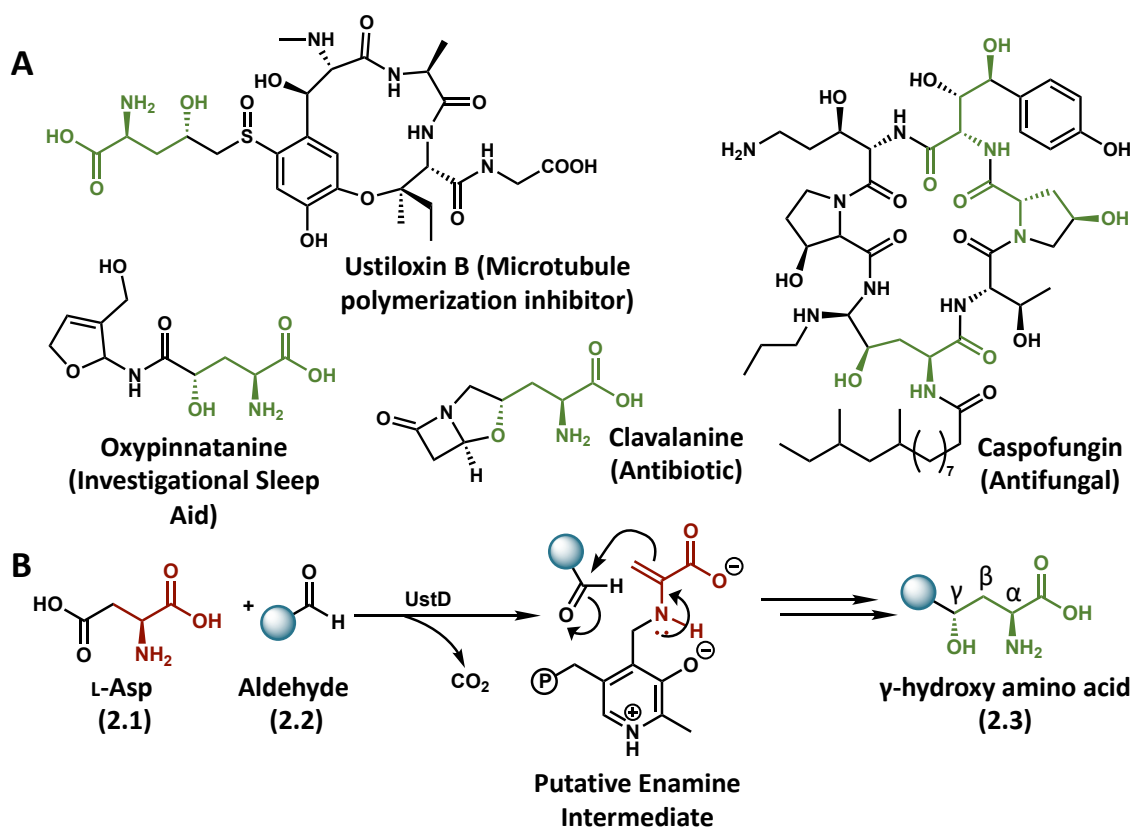
Figure 2.1 A. The generalized decarboxylative aldol reaction of UstD showing the putative enamine nucleophilic intermediate. **B.** Bioactive molecules with a 1,3-amino alcohol motif shown in purple.

It has been shown that the transketolase catalytic cycle can be non-natively entered through decarboxylation, and that reactions initially proceed to high conversion. However, the native proton transfer machinery eventually breaks down the product into an equilibrium mixture with starting materials.[16] Although the detailed mechanism of UstD has not yet been explored,

Ye et al. reported that the UstD reaction cannot be initiated from L-Ala, indicating enamine formation through tautomerization is not viable. Therefore, UstD is mechanistically distinct from classic aldolases and may have unique properties as a biocatalyst.

The native substrate for UstD is a complex, cyclic peptide, and it was unknown if this enzyme would react promiscuously with alternative substrates. If so, the enzyme would directly produce γ-hydroxy amino acids (**Fig. 2.1B**). Such non-standard amino acids (nsAAs) are found in bioactive natural products, such as caspofungin and clavalanine (**Fig. 2.1A**)[17]. While nature employs side chain hydroxylation to tune bioactivity, these nsAAs are virtually absent from medicinal chemistry[18] because they require multistep synthesis[17]. The need for multistep synthesis to prepare these nsAAs has begun to be addressed by biocatalysis, where an elegant multi-enzyme cascade was recently developed by Clapés et al. to access gamma-hydroxy nsAAs[19,20]. However, the ability to use a single enzyme to produce the same motif would offer greater practical utility and versatility. Beyond their use in pharmaceuticals, nsAAs can be enabling for a host of synthetic and chemical biology applications[21,22]. Therefore, the development of UstD for organic synthesis would introduce a valuable and much-needed enantioselective C–C bond-forming enzyme into the biocatalytic toolbox and provide direct access to a structurally complex synthon.

2.2 Results

2.2.1 Initial characterization of UstD

We expressed *C*-His-UstD (wt-UstD) in *Escherichia coli* (**Fig. 2S.1**) but were uncertain whether molecular recognition for the structurally complex native substrate would be required for catalytic activity. We therefore assessed the reactivity of wt-UstD with benzaldehyde (**2.2a**) and were pleased to observe a successful decarboxylative aldol addition to afford the γ-hydroxy

nsAA **2.3a** by UPLC-MS. A preparative scale reaction with 0.125 mol % catalyst gave the product in 43% yield, and analysis by nuclear magnetic resonance (NMR) spectroscopy indicated a single diastereomer predominated (dr >98:2). To determine the absolute stereochemical preference for the enzyme, we analyzed the product from a reaction with 4-bromobenzaldehyde (**2.2b**). The crystal structure of the product (**2.3b**) revealed the aldol addition occurred with the same stereochemical outcome as the native reaction (**Fig 2S.2**). These transformations indicated wt-UstD has potential for organic synthesis, but the comparatively modest activity (<1,000 turnovers with initial reaction conditions) and low catalyst expression would hinder routine use of the natural enzyme. Given the inherent structural differences between the native tetrapeptide substrate and simpler commercially-available aldehydes (such as **2.2a**), we hypothesized that directed evolution and reaction condition optimization could be used to increase the catalytic efficiency of UstD toward non-native substrates.

2.2.2 Directed evolution of UstD for improved catalytic activity

To inform our engineering process, we used a homology model of wt-UstD derived from a distantly related cysteine desulfurase (27% identity)[23,24]. Six residues in the predicted active site were chosen for saturation mutagenesis, and we used benzaldehyde (**2.2a**) as a model substrate for directed evolution (**Fig. 2.2A**).
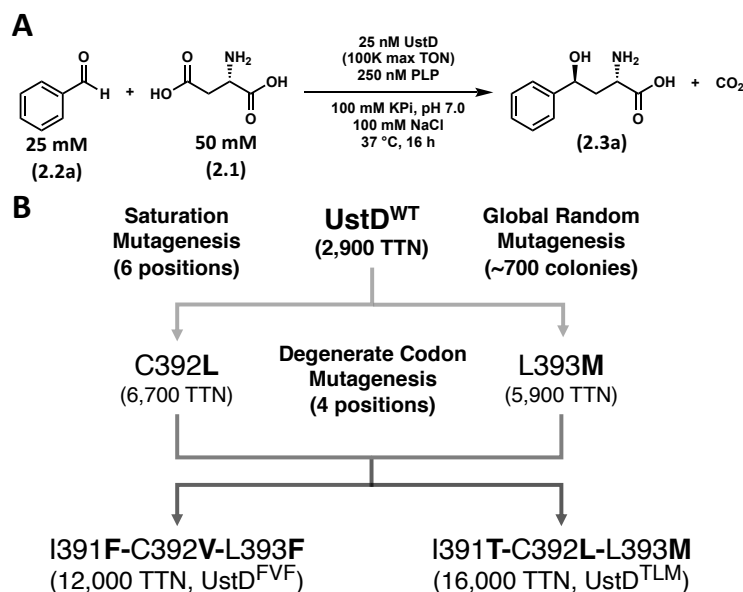
Figure 2.2 Directed evolution of UstD. A. Reaction under selective pressure and re-screening conditions used to validate hits. **B.** Lineage of independently activated variants. Catalyst activity measured by total turnover number (TTN).

Mutation at positions predicted to form direct contacts with the cofactor resulted in inactivation of the catalyst, a common trend amongst PLP-dependent enzymes[25]. Nevertheless, these libraries yielded a single variant in a putative loop region flanking the substrate binding site, C392L, with a 2.3-fold boost in activity (**Fig. 2.2B**). Concurrently, we employed global random mutagenesis on wt-UstD to search throughout the protein sequence for activating mutations. A second activating mutation was discovered, L393M, immediately adjacent to Cys392. We combined these mutations to yield the double variant UstD[C392L, L393M], which had a further increase in activity to 4.9-fold above wild-type (**Fig. 2.3A**).
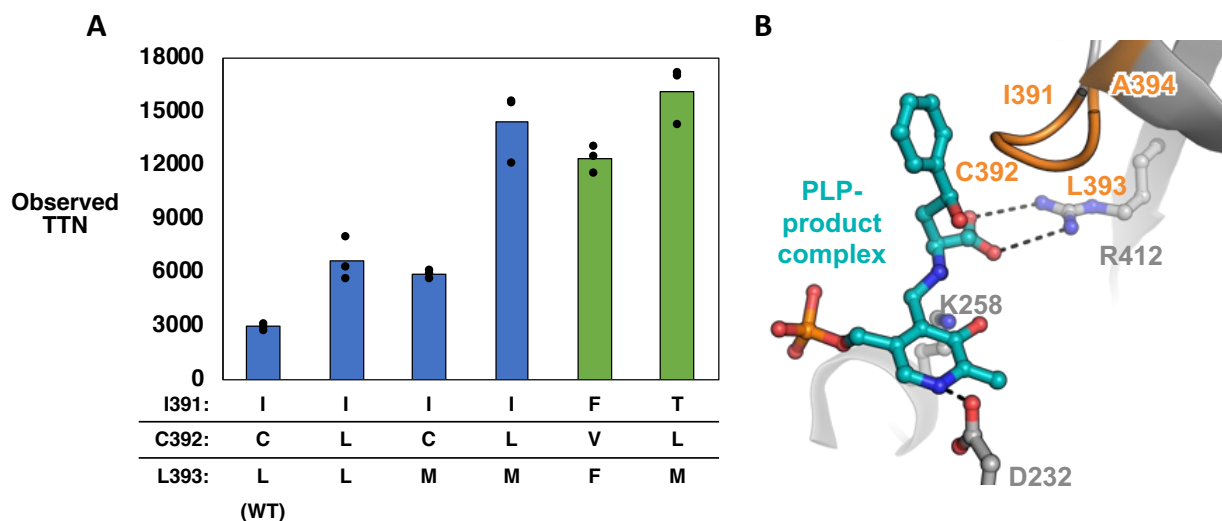
Figure 2.3 A. Observed turnover numbers for key variants of the UstD evolutionary lineage. The standard deviations represent triplicate or duplicate technical replicates. **Conditions:** 25 mM benzaldehyde, 50 mM L-aspartate, 10x PLP to catalyst concentration, 5% DMSO, buffer (100 mM KPi, pH = 7.0, 100 mM NaCl), max TON = 15,000, 37 °C, 16 h. **B.** Structure of UstD bound to γ-hydroxy-amino acid product, **2.3a** derived through homology modeling. Active site residues are shown as sticks and activating loop residues are colored in orange. Hydrogen bonds are shown as black dashes.

It is common for mutation of neighboring residues to display cooperativity[26,27], and we chose to test additional mutations in this region of the sequence (**Fig. 2.3B**). We used a degenerate codon mutagenesis strategy on four contiguous residues from Ile391-Ala394. We restricted the sequence space to residues commonly found among UstD homologs, which provided good structural diversity in a focused set of mutations (see SI for details). Screening this library revealed that mutation of Ala394 was generally deleterious. However, multiple highly active variants retained Ala394 and contained mutations at Ile391, Cys392, and Leu393. To best capture relative rate effects of mutations, catalysts were compared under dilute conditions. Variants UstD^{TLM} and UstD^{FVF} (the superscript refers to the identity of the residues at positions 391-393) had a 5.1-fold and 4.1-fold increase in activity relative to wt-UstD, respectively.

We next optimized reaction conditions for the most active variant, UstD[TLM]. Reaction mixtures were initially colored yellow (**Fig. 2S.1**) by the presence of PLP that co-purified with the enzyme but became colorless over time, suggesting the cofactor is degraded during the reaction. Gratifyingly, supplementation of PLP led to a large increase in product formation (**Fig. 2.4A**).
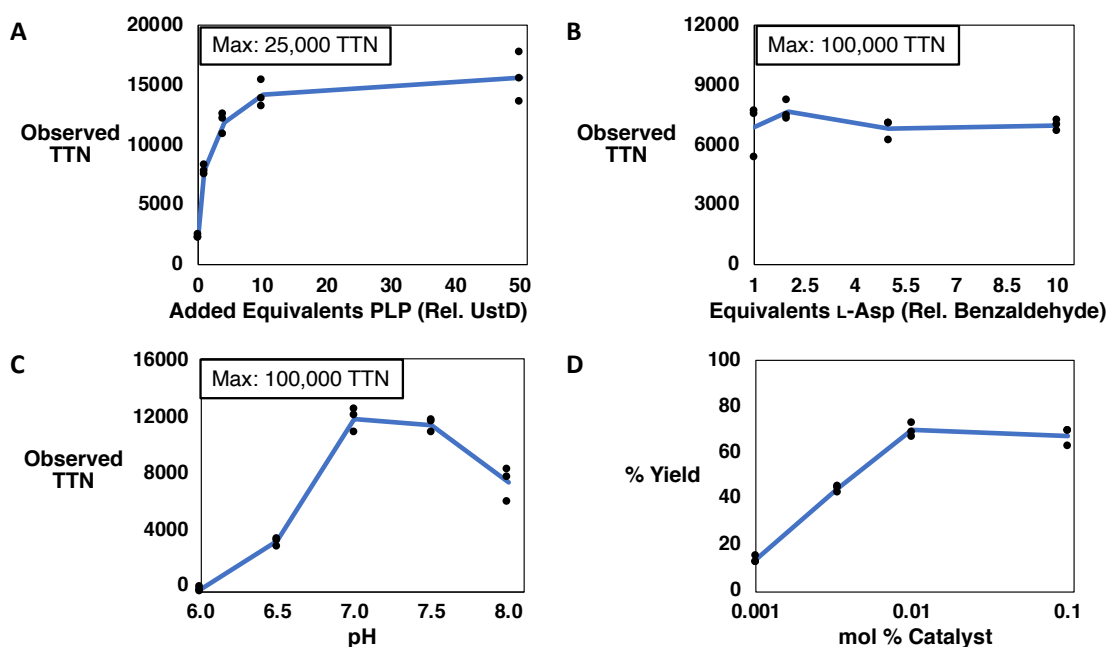
Figure 2.4 Reaction condition optimization of UstD[TLM]. Initial reaction conditions are depicted in the top scheme. Conditions were optimized sequentially: **A.** 10 equivalents PLP relative to catalyst, **B.** 2 equivalents of L-aspartate relative to benzaldehyde, **C.** pH 7.0. The dependence of observed yield on catalyst loading **D.** was conducted with optimized reaction conditions.

We did not observe a significant change when the concentration of **2.1** was increased (**Fig. 2.4B**). However, we observed formation of L-alanine in reactions, indicating some **2.1** is lost to a non-productive protonation of the nucleophilic enamine intermediate[13]. We therefore used aldehyde as the limiting reagent and two equivalents of **2.1** for subsequent experiments, which identified an optimal initial pH of 7.0 (**Fig. 2.4C**). Lastly, we varied the catalyst loading and found that UstD[TLM] was capable of high conversion (~70%) with just 0.01 mol % catalyst

loading (**Fig. 2.4D**). With these optimized conditions, we evaluated the performance of wt-UstD and both activated variants, UstD[TLM] and UstD[FVF], with a more diverse set of aldehyde substrates. We anticipated that the striking sequence divergence in the putative loop would lead to distinct trends in substrate selectivity.

2.2.3 Performance analysis of UstD and its variants

Engineering enzymes for activity on a model substrate often leads to specialist catalysts with diminished activity on substrate analogs[28,29]. Initial comparisons among wt-UstD, UstD[FVF], and UstD[TLM] with a small panel of aldehydes suggested that both variants had evolved towards improved overall activity (**Fig. 2.5**).
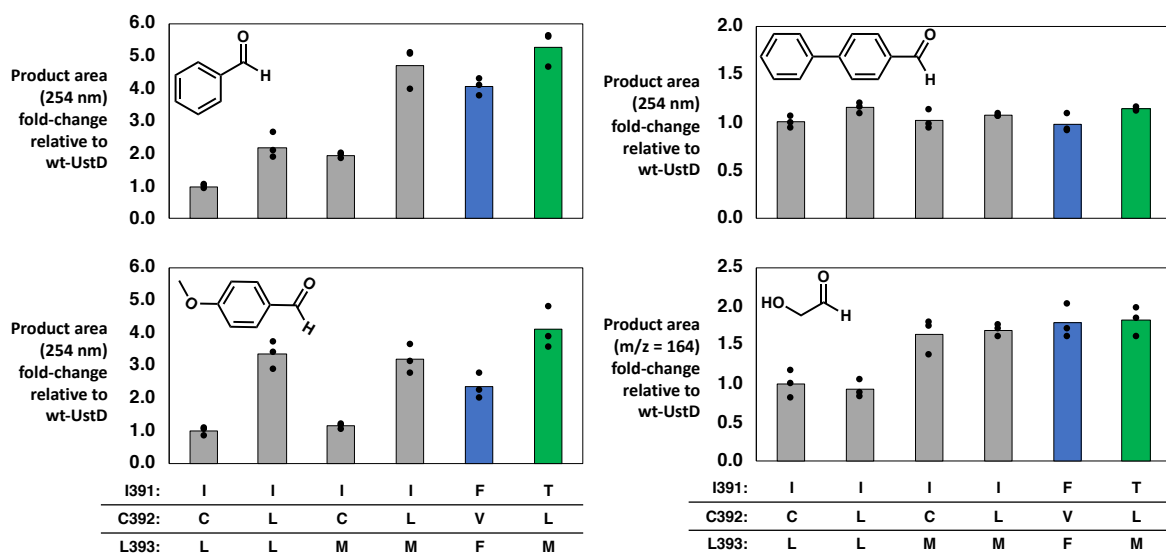
Figure 2.5 Comparison of UstD variants with a variety of aldehyde substrates. Reactions conducted with the optimized conditions described in Figure 2.4. Each bar represents a unique variant. Activity is reported as the fold-change relative to wild-type (left). The variants UstD[FVF] and UstD[TLM] are in blue and green, respectively.

We therefore expanded the substrate scope. Marfey's reagent cleanly derivatized the diverse products, providing a uniform chromophore for quantitative measurement of turnover and selectivity via UPLC-MS[30]. Product formation was observed with virtually every substrate

tested, from the large and hydrophobic biphenyl aldehyde (**2.2g**) to the small and hydrophilic

glycolaldehyde (**2.2p**) (**Fig. 2.6**).
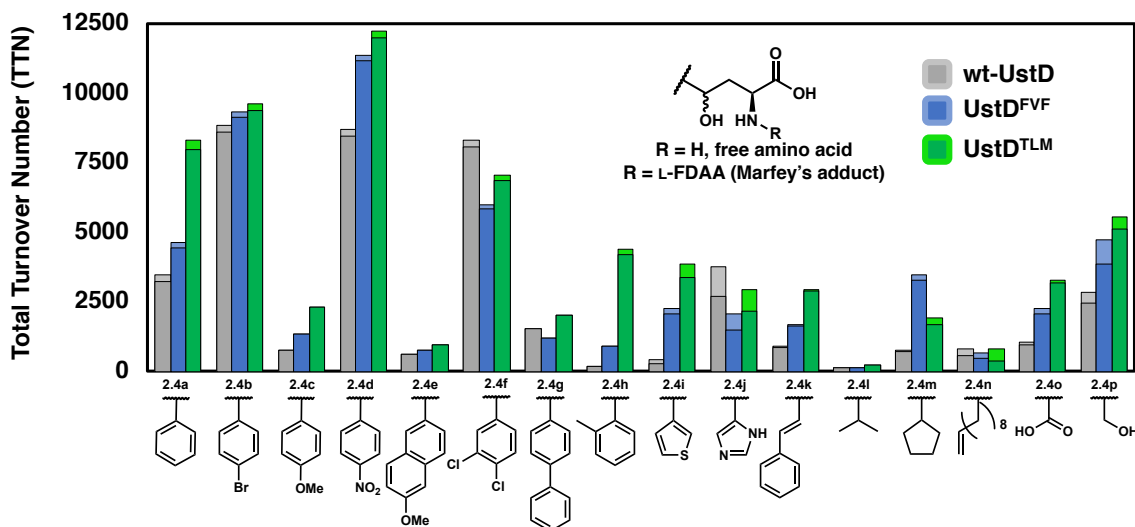
Figure 2.6 Analytical substrate scope quantified through Marfey's derivatization. The enzymatic reaction conditions are 25 mM aldehyde, 50 mM L-aspartate, 1.67 μM UstD/UstDFVF/UstDTLM (0.007 mol% cat., 15000 max TON), 16.7 μM PLP (10 equiv. compared to catalyst), buffer (100 mM KPi, pH = 7.0, 100 mM NaCl), 5% DMSO, 37 °C, 16 h. Reactions were quenched with 1 vol. equiv. of ACN. The derivatization reaction conditions are 0.5 mM of total amines from the quenched enzymatic, 5 mM L-FDAA, 10 mM NaHCO3, 1:1 H2O:ACN, 37 °C, 16 h. Derivatization reactions were quenched with 1 vol. equiv. of 1:1 ACN: 60 mM HCl and analyzed by UPLC-MS within 24 h.

Generally, the variant UstD^{TLM} performed the most turnovers and displayed excellent

diastereoselectivity, typically forming a 95:5 ratio of diastereomers (dr). While UstD^{FVF} typically

performed fewer turnovers than UstD^{TLM} with most substrates, UstD^{FVF} generally had higher

selectivity than wt-UstD or UstD^{TLM} (**Table 2S.1**). Reactions with *p*-substituted aromatic

aldehydes exhibited a Hammett-like reactivity trend: more product was formed as aldehyde

electrophilicity increased. Activity was lowest with the electron rich *p*-anisaldehyde (**2.2c**), but

high activity was observed for the electron deficient *p*-NO$_2$-benzaldehyde (**2.2d**) with both

engineered enzymes. To better capture the maximum turnover number with **2.2d**, we repeated

the reactions at lower catalyst loadings, which revealed that the engineered variants can perform

~34,000 turnovers (**Fig. 2.7**).
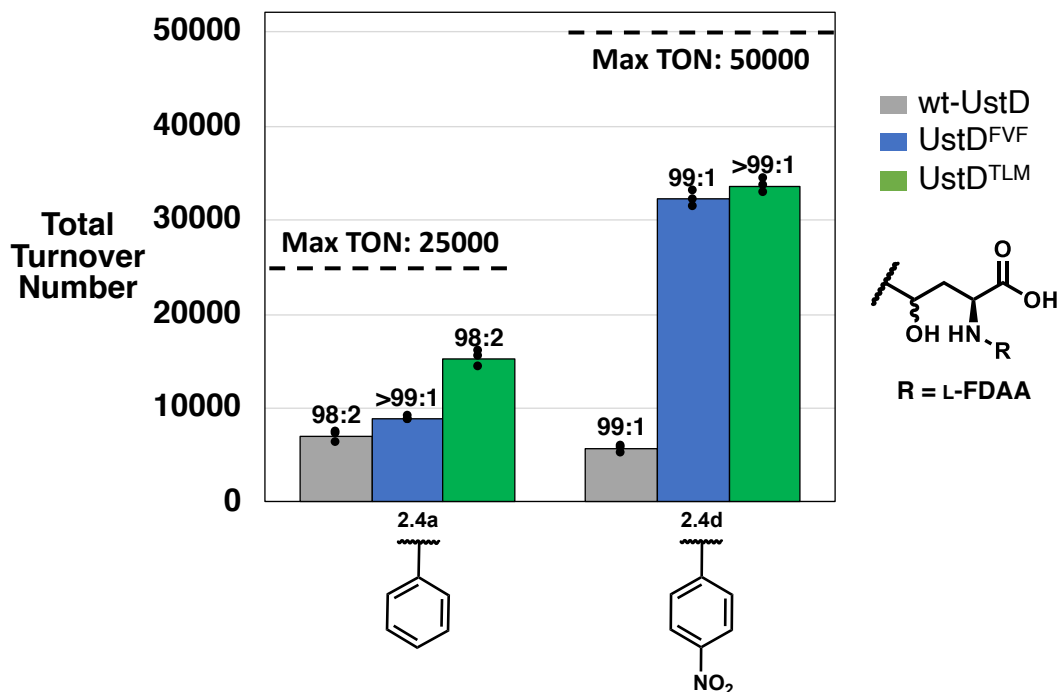
Figure 2.7 Additional triplicate replicates of total turnover number analyses for 4-nitrobenzaldehyde and benzaldehyde at reduced catalyst loading. Conditions: 25 mM aldehyde, 50 mM L-aspartate, 10x PLP to catalyst concentration, 5% DMSO, buffer (100 mM KPi pH = 7.0, 100 mM NaCl), 37 °C, overnight. The max TON for benzaldehyde is 25,000 and the max TON for 4-nitrobenzaldehyde is 50,000, as indicated by black hash marks. Product quantitation was performed via Marfey's analysis, as in the original performance evaluation. Measured dr values are shown above each bar.

Active site mutagenesis had little apparent impact on reactions with some highly

hydrophobic substrates, such as the methoxynaphthyl (**2.2e**), 3,4-dichlorobenzyl (**2.2f**), and

biphenyl (**2.2g**) aldehydes; reactivity in these cases may be limited by poor aqueous solubility

(**Fig. 2.6**). In contrast, reactivity on *o*-tolualdehyde (**2.2h**) and thiophene-3-carboxaldehyde (**2.2i**)

increased dramatically during evolution. UstD^{TLM} displayed a nine-fold increase in activity on

2.2i and a remarkable 23-fold increase in turnovers with **2.2h** compared to wt-UstD. Activity

with the imidazole substrate **2.2j** was demonstrated and was one of the few substrates for which

wt-UstD had the highest activity. To the best of our knowledge, the product is a previously

unreported analog of histidine. Reactivity with the cinnamaldehyde (**2.2k**) improved with both

variants relative to wt-UstD. Reactions proceeded smoothly with several aliphatic substrates,

including isobutyraldehyde (**2.2l**), cyclopentylaldehyde (**2.2m**), and even 10-undecenal (**2.2n**); in

this last case reactivity appeared to be limited by solubility. Pivaldehyde, however, was

unreactive with all three enzymes, an observation we attribute to steric bulk near the carbonyl.

The engineered UstD enzymes were active with glyoxylic acid (**2.2o**), which resulted in

formation of γ-hydroxy-glutamate, an intermediate in hydroxyproline metabolism[31]. Lastly, we

observed good reactivity with glycolaldehyde to yield the di-hydroxylated amino acid **2.3p**.

Previously, a protected form of **2.3p** was identified as a key intermediate in the synthesis of

clavalanine (**Fig. 2.1A**)[17], an antibiotic that inhibits the biosynthesis of methionine[32]. Activity on

2.2p increased two-fold, with improved diastereoselectivity and pristine enantioselectivity, for

UstD[TLM] relative to the wild-type enzyme. These substrates collectively demonstrate that the

active site of UstD is remarkably permissive of diverse functional groups and that catalytic

activity and selectivity can be rapidly optimized by mutation at residues 391-393.

These engineered enzymes enable a stereoselective synthesis of γ-hydroxy nsAAs in a

single step from cheap, commercially available starting materials. The production of unprotected

amino acids affords complete flexibility with regards to subsequent manipulation, but isolation

of free amino acids themselves is challenging due to their hydrophilic, zwitterionic nature.

Therefore, we selected a representative set of products to demonstrate isolation strategies (**Fig.
2.8**). Sufficiently hydrophobic products were isolated as the free amino acid, while others

utilized protection with fluorenylmethoxycarbonyl (Fmoc) to increase hydrophobicity,

simultaneously adding a handle commonly used in solid phase peptide synthesis. Diverse

manipulations, such as lactonization with the γ-hydroxy group, can also be employed to facilitate isolation and downstream manipulation[19]. Throughout these reactions a second, minor diastereomer was observed. The mixture of configurations at Cγ arises through imperfect selectivity with the aldol addition and could be aggravated by reversible retro-aldol cleavage of the major diastereomer.
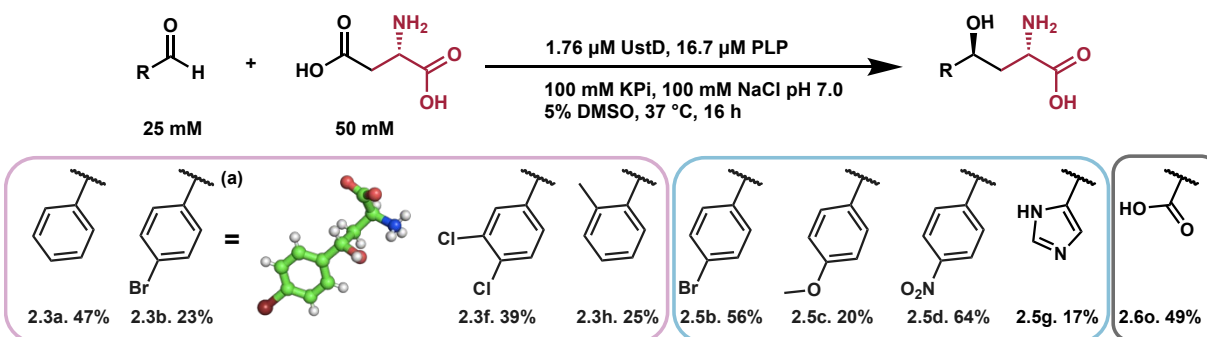
Figure 2.8 Synthesis of select products at a 0.2 mmol scale with isolated yields. The different purification strategies are denoted by the different colors, free amino acid (purple), Fmoc-protected amino acid (blue) and lactonization with Fmoc protection (grey). The letter (a) denotes that the reactions from which **2.3b** was purified used wt-UstD.

We tested the latter possibility by re-subjecting products **2.3a** and **2.3d** to reaction conditions and observed no change in the diastereomeric ratio by Marfey's analysis (**Fig. 2S.3**).However, in the case of **2.3a**, formation of alanine (Ala) was observed concomitant with a decrease in product peak area. This observation is consistent with slow product re-entry into the catalytic cycle via retro-aldol cleavage of **2.3a** to reform **2.2a** and Ala.

2.2.4 Linear regression guided protein engineering

The above studies relied on purified protein for preparative scale reactions. However, access to enzymes in sufficient quantity is a common and often under-appreciated limitation of biocatalysis. As is observed for many proteins, UstD had relatively low expression titers in *E. coli* (8 mg L[-1] culture) due to poor solubility (**Fig. 2S.1**). While enzyme immobilization can be used to increase the utility of purified protein catalysts[33], a complementary synthetic

methodology would use whole-cell preparations of UstD; this latter approach is attractive to process chemists[34]. Whole-cell catalysts are operationally simple to generate, stable over long periods, and obviate the need for expensive protein purification.

We sought to further engineer UstD^{TLM} to increase soluble heterologous expression in *E. coli* for whole-cell biocatalysis. This enzyme contains nine Cys residues, and our homology model suggested five are surface exposed (**Fig. 2.9**).
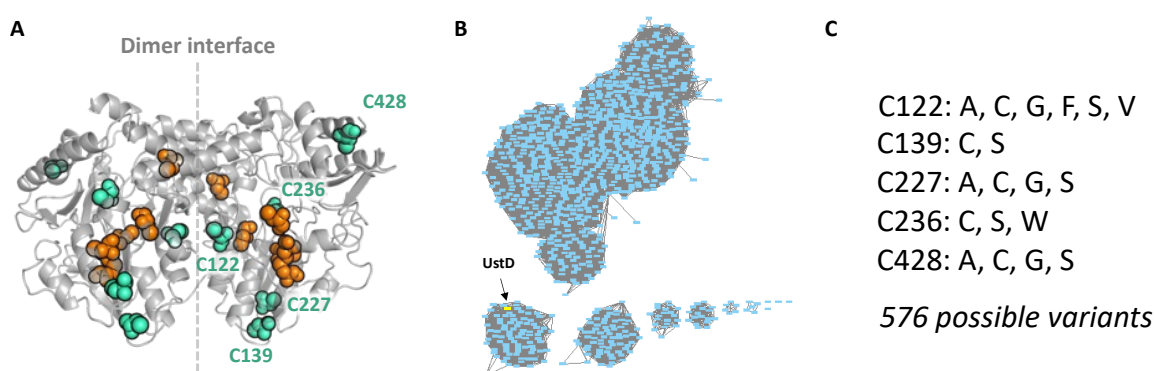
Figure 2.9 UstD Cys shuffle library construction process. A. Homology model of UstD^{TLM} with cysteine residues shown as spheres. Positions selected for mutagenesis are highlighted as cyan spheres. **B.** Sequence similarity network with UstD^{TLM} containing node highlighted in yellow. **C.** Possible mutations for each chosen position in the Cys shuffle library.

It is well known among protein crystallographers that removing surface Cys residues can increase soluble expression and increase the probability of crystallization[35]. However, we found that mutation of all five putative surface Cys residues to Ala eliminated catalytic activity. To identify mutations that would retain activity while increasing soluble expression, we performed sequence-similarity network analysis to identify non-Cys residues at these positions common among UstD homologs. Based on this analysis, we constructed a five-site degenerate codon library (**Fig. 2.9B,C**).

To efficiently navigate this sequence space, we employed linear regression modelling to predict sequence-activity relationships[36]. We hypothesized this simple computational approach

would be effective because the target residues are dispersed throughout the protein, which should make non-linear, pair-wise mutational effects unlikely. We screened and sequenced 176 random clones from this library for increased activity in lysate, which is sensitive to changes in both soluble enzyme expression and enzymatic efficiency. Although most variants in this library were inactive, we were heartened to observe several apparently improved variants (**Fig. 2.10A**). Linear regression model testing using leave-one-out cross-validation (LOOCV) of the full dataset indicated poor predictive behavior of the model for high-activity variants (**Fig. 2S.4**). We suspected that the model quality was diminished by the abundance of inactive variants, for which activity measurements are indistinguishable from experimental noise. We therefore restricted our analysis to variants for which bonafide activity could be measured, leaving just 26 sequence-activity relationships. Despite the sparsity of these data (~5% of the sequence space), LOOCV showed the model was dramatically improved (See SI for details).
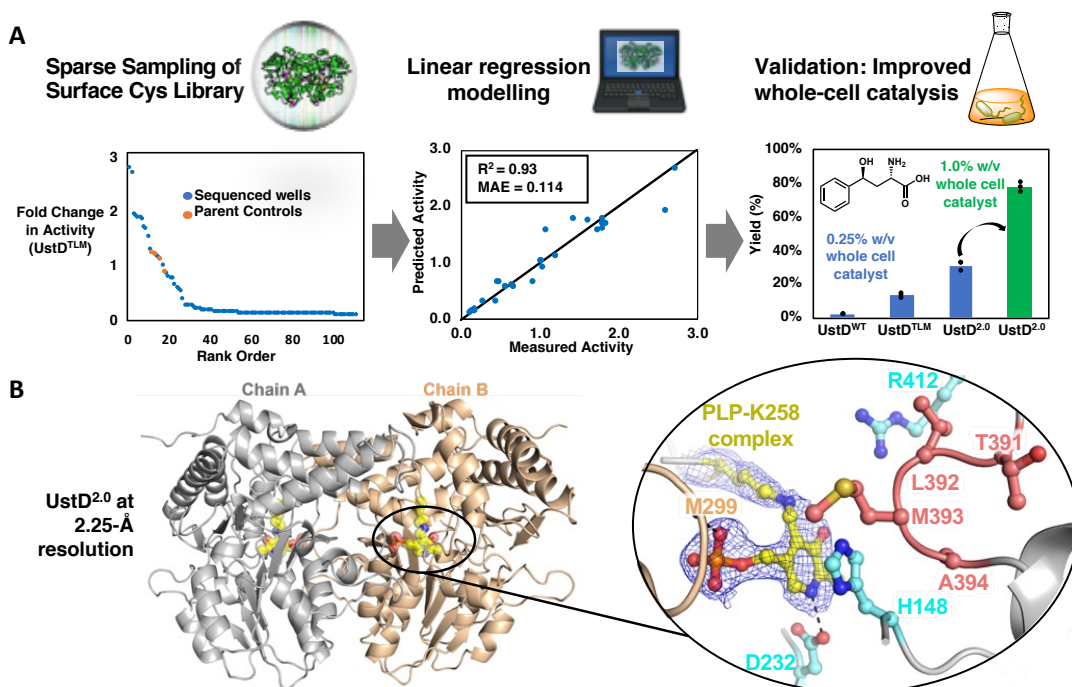
Figure 2.10 Engineering UstD for increased crystallizability and activity in whole-cell catalysis. A. Experimental process for bioinformatic and regression-guided mutagenesis of UstD. In the first stage, a small mutagenesis library is sampled to collect sequence and/or activity data. The second stage builds a linear regression model to correlate sequences to activity. This regression model is then used to predict the activated sequences, which are validated in the last stage using whole-cell catalysis. The dots in the bar graph represent the individual measurements of triplicate technical replicates. **B.** Representation of the overall structure of UstD$^{2.0}$. Individual monomers are colored grey (chain A) and brown (chain B). The PLP-K258 complex is shown as yellow spheres and sticks. Inset: active-site residues superimposed on the 2mFo–DFc electron density map (blue mesh, $\sigma = 1.2$) are shown as sticks. The TLMA loop residues are colored pink. Hydrogen bonds are shown as black dashes. MAE, mean absolute error.

We evaluated the three most active variants predicted by the model, UstD$^{TLM-ACASC}$, UstD$^{TLM-ASCSC}$, and UstD$^{TLM-ASASC}$. Comparisons of expression and whole cell activity were made between these variants, the parent enzyme, and most active variant identified from screening, UstD$^{TLM-SCASC}$. We were delighted to find the expression titer was increased relative to UstD^{TLM} for all variants, up to 48 mg protein L^{-1} culture (**Fig. 2.11**). While purified enzyme activity is slightly decreased for the new variants, their overall activity in whole cells is significantly improved (**Fig. 2.10A, 2.11**).
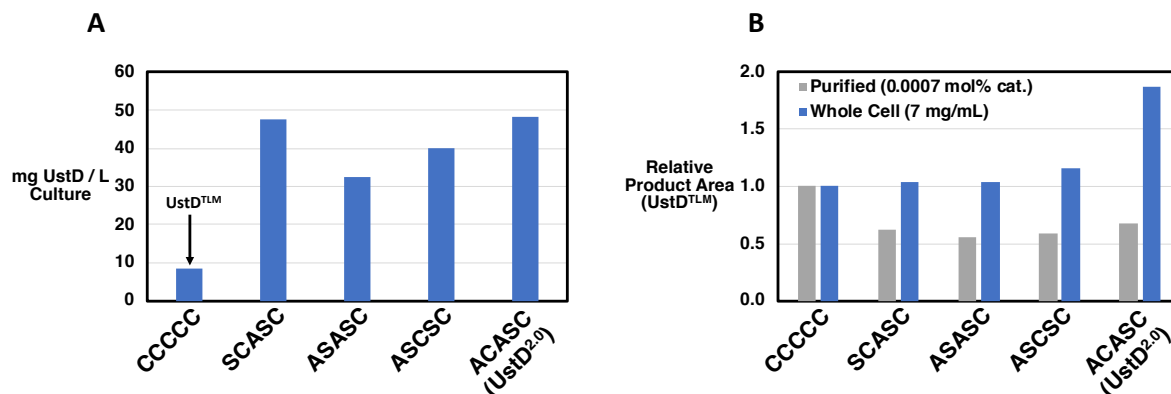
Figure 2.11 Comparison of Cys shuffle variants. A. Isolated soluble protein expression, as measured by Bradford assay. **B.** Relative product **2.3a** formation of variants in purified enzyme and whole cell biocatalyst contexts. Both reaction types contained 25 mM **2.2a** and 50 mM **2.1** in 100 mM KPi, pH 7.0, with 100 mM NaCl. For purified enzyme reactions, 10 eq. of PLP relative to catalyst concentration was added, while no additional PLP was added in the whole cell reactions.

Tests at analytical scale showed, at 0.25% w/v cell loading, that UstD^{TLM} formed **2.3a** in just 13% yield, highlighting the challenges associated with translating in vitro activity to large-scale reaction formats. In contrast, the variant with the highest whole-cell activity, the computationally-predicted UstD$^{TLM-ACASC}$ (designated UstD$^{2.0}$), produced **2.3a** in 31% yield, a 2.4-fold boost over UstD^{TLM} and a cumulative 15-fold boost over wild type. Higher conversions were achieved by increasing the cell loading of UstD$^{2.0}$ to 1% w/v, which afforded **2.3a** in 78% yield on analytical scale (**Fig. 2.10A**). To demonstrate the utility of UstD$^{2.0}$, large-scale reactions were carried out with **2.2a** and **2.2d**. Reaction with **2.2a** at 0.5% w/v catalyst loading afforded 0.80 g **2.3a** in 77% isolated yield with pristine stereoselectivity following purification by reverse-phase chromatography. Reaction with **2.2d** at just 0.1% w/v catalyst loading provided 1.4 g **2.3d** in 98% isolated yield with high stereoselectivity (see SI for details). Notably, these cell loadings are sufficient for process-scale biocatalytic reactions[37], illustrating that UstD$^{2.0}$ can operate on the scale needed to meet the demands of practical organic synthesis.

2.2.5 Crystallography of UstD$^{2.0}$

While the engineering we report here produced a generalist variant of UstD, structural information could guide more targeted engineering to produce specific γ-hydroxy nsAAs. Despite extensive efforts, we were unable to produce crystals of wt-UstD. In contrast, UstD$^{2.0}$ readily crystallized, which we attribute to the decrease in surface Cys residues. The 2.25-Å crystal structure of UstD$^{2.0}$ was determined using experimental phases from a Au(III) derivative (**Fig. 2.10B**, PDB ID: 7MKV). This structure revealed an active site at the dimer interface, which is common among fold-type I PLP-dependent enzymes [38]. The internal aldimine involving a Schiff base linkage to Lys258 and a salt bridge between the pyridinium N1 and Asp232 were clearly resolved in the active site. The 391-393 loop harboring the activating TLM mutations projects over the top of the active site forming part of the substrate binding pocket. The remainder of the pocket appears to be solvent exposed, explaining the tolerance of UstD for diverse aldehyde substrates (**Fig. 2S.5**).

In the future, we envision engineering UstD for increased activity with non-aldehyde substrates. As an initial demonstration, we showed that purified UstD$^{2.0}$ performs ~50 turnovers with the ketone substrate trifluoroacetone to produce a nsAA bearing a tertiary alcohol side chain (**Fig 2S.6**). The comparatively low turnover highlights the challenges associated with aldol addition into ketones. When nucleophilic attack is sufficiently slow, irreversible protonation of the enamine can quench the reactive intermediate and, indeed, we observed significant accumulation of L-alanine in this reaction. A similar scenario was observed with hydrolysis of an electrophilic PLP intermediate formed by TrpB and reactions with attenuated substrates were enabled by directed evolution that increased the lifetime of the reactive intermediate.[39,40] Hence,

future engineering to decrease the rate of enamine protonation in UstD$^{2.0}$ may further expand the substrate scope.

2.3 Conclusions

Here, we improved a C–C bond forming enzyme, UstD, that catalyzes a decarboxylative aldol addition using the loss of CO_2 from L-aspartate as a thermodynamic driving force to produce γ-hydroxy amino acids. This mechanism of action and innate tolerance of diverse aldehydes marked UstD as a candidate for directed evolution into a versatile catalyst for organic synthesis. To screen for improved catalysts, we used a combination of globally random, site-saturation, and degenerate codon mutagenesis libraries. We illustrate the engineering potential of the active site with two variants, UstD^{FVF} and UstD^{TLM}, that share no mutations in common and display commensurate or superior activity to wt-UstD with the vast majority of aldehydes tested. We demonstrated how a simple regression-modeling approach to protein engineering can increase protein soluble expression and crystallizability. The evolved variant, UstD$^{2.0}$, is poised to deliver a desirable nsAA precursors for medicinal chemistry, and the crystal structure will facilitate future work to explore the mechanism and reactivity of this intriguing enzyme.

2.4 Supplementary Figures, Tables, Information

Figure 2S.1. Purification of UstD. A. 12% SDS-PAGE Gel following steps of purification. **B.** Picture of purified UstD as snap-frozen catalytic droplets. Yellow color results from the binding of PLP cofactor to UstD during purification. **C.** Picture comparing a no enzyme control reaction with cofactor added (left) and an enzyme-containing reaction (right) demonstrating the loss of yellow color associated with enzymatic degradation of the PLP cofactor.

Figure 2S.2. X-ray crystal structure of 2.3b. Product formed via the reaction of UstD with L-aspartate and p-bromobenzaldehyde. A molecule of water crystallized with **2.3b** which coordinates with the hydroxy group of the amino acid.

Figure 2S.3. Product stability **a.** Marfey's analysis spectrum of **2.3d** when subjected to reaction conditions in the presence and absence of enzyme. **b.** Marfey's analysis spectrum of **2.3a** when subjected to reaction conditions in the presence and absence of enzyme. For both experiments, the dr for each experiment was the same, ~98:2. Extracted ion chromatograph traces verify the positions of the derivatized products, as well as the peak attributed to L-alanine formation via retro-aldol cleavage in reactions with **2.3a**.
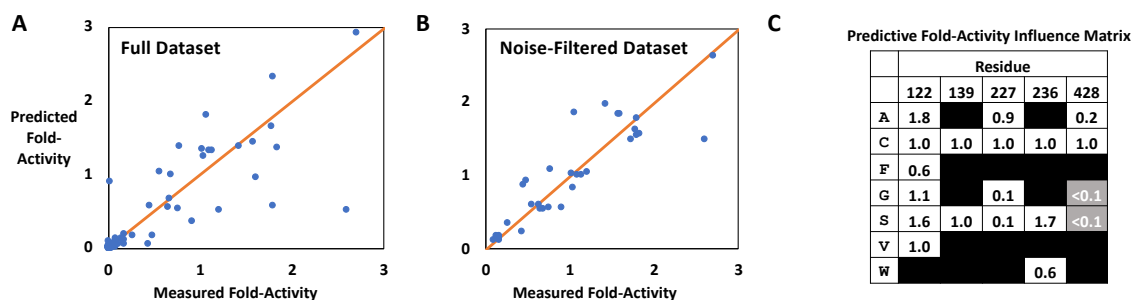
Figure 2S.4. Leave-one-out cross-validation (LOOCV) of UstD Cys shuffle library linear regression model. **A.** Comparison of measured and predicted fold activity changes when the regression model is analyzed using LOOCV. **B.** Comparison of measured and predicted fold activity changes when a subset of the sequences for which bonafide activity measurements where made. **C.** Fold-activity matrix determined from the processing in panel B. See SI methods for further detail.
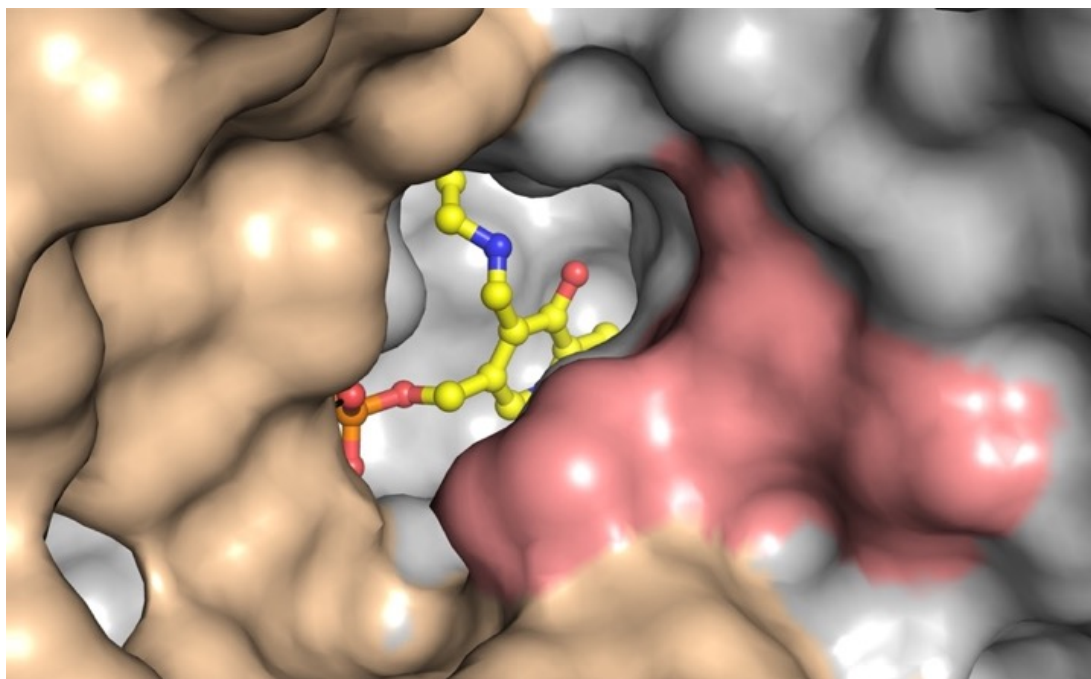
Figure 2S.5. Substrate access channel in UstD$^{2.0}$. Individual monomers are shown as grey and brown surfaces. PLP-K258 complex is shown as yellow spheres and sticks. TLMA loop residues are colored in salmon.
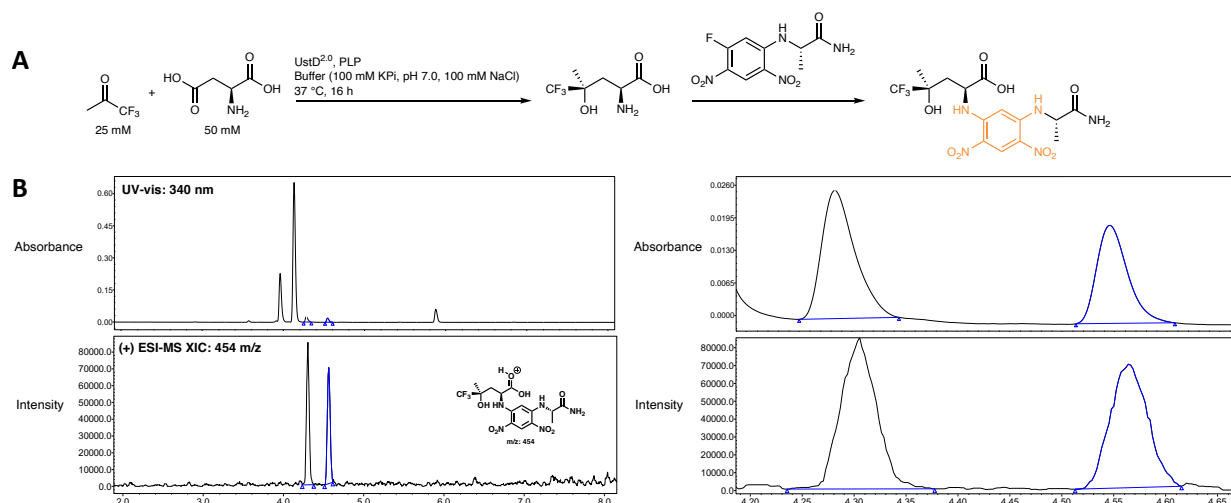
Figure 2S.6. UstD$^{2.0}$ activity with trifluoroacetone. **A.** Reaction scheme showing enzymatic reaction conditions followed by Marfey's derivatization (see section SI for procedure). **B.** Representative UPLC-MS trace showing the two diastereomeric Marfey's products.

Characterization of γ-hydroxy Amino Acids by Marfey's Analysis

All products derivatized according to according to the General Marfey's Procedure. Quantitative analysis of the data is provided in **Table 2S.1**.
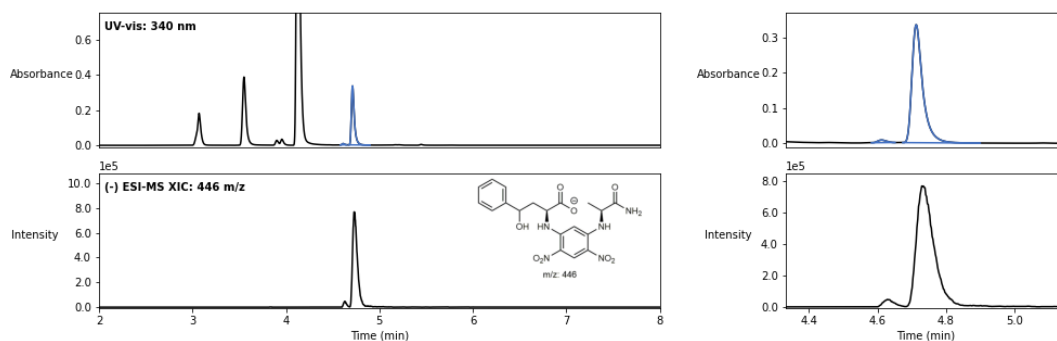
Figure 2S.7 UPLC trace for *2.4a* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-phenylbutanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
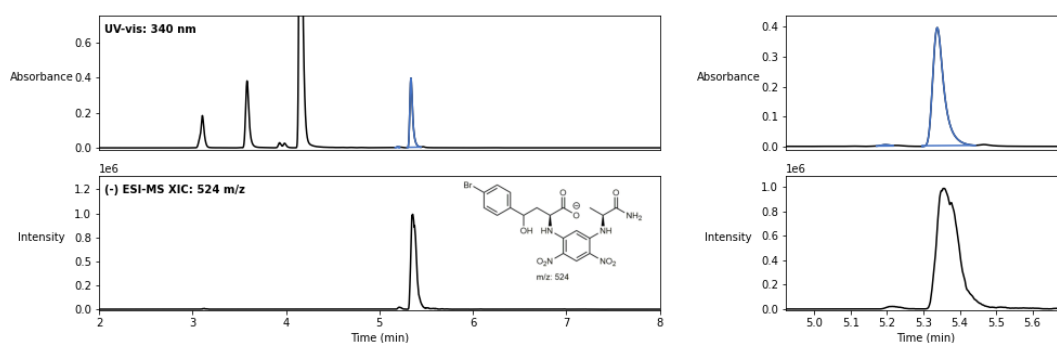
Figure 2S.8. UPLC trace for *2.4b - 2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-(4-bromophenyl)-4-hydroxybutanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
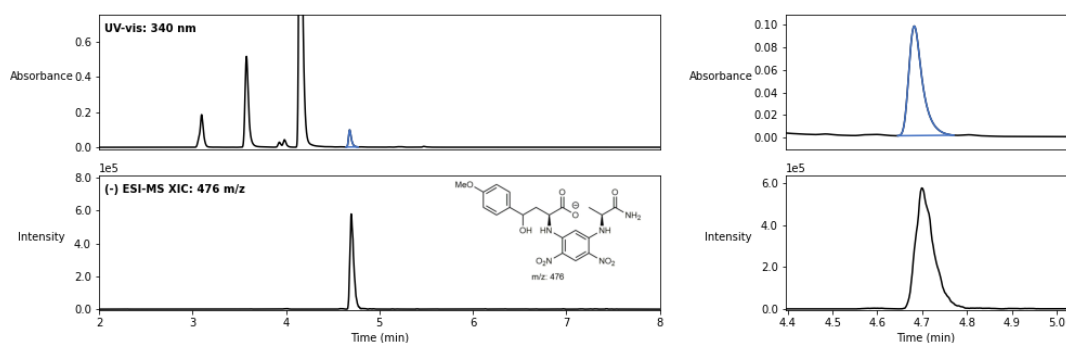
Figure 2S.9. UPLC trace for *2.4c - 2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-(4-methoxyphenyl)butanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
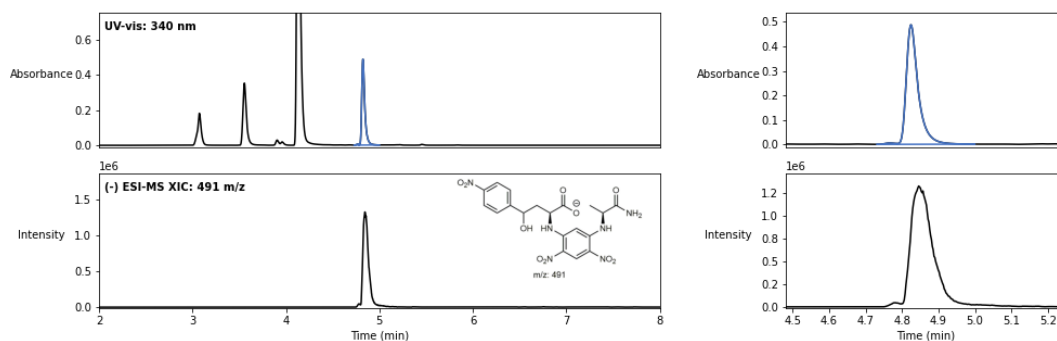
Figure 2S.10. UPLC trace for *2.4d* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-(4-nitrophenyl)butanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
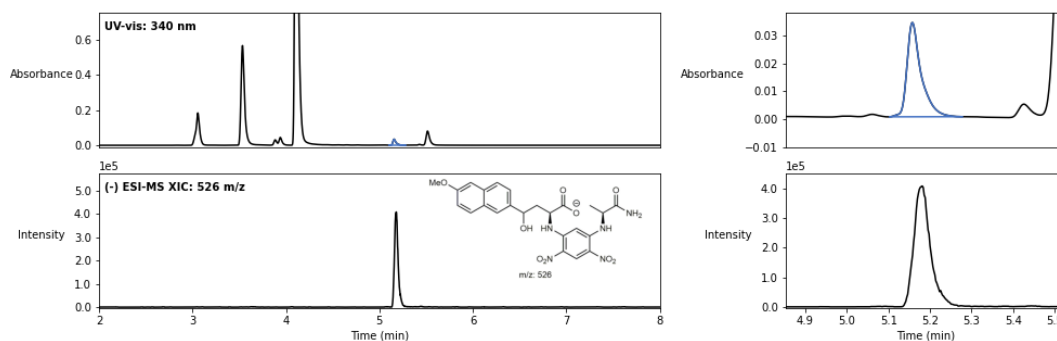
Figure 2S.11. UPLC trace for *2.4e* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-(6-methoxynaphthalen-2-yl)butanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
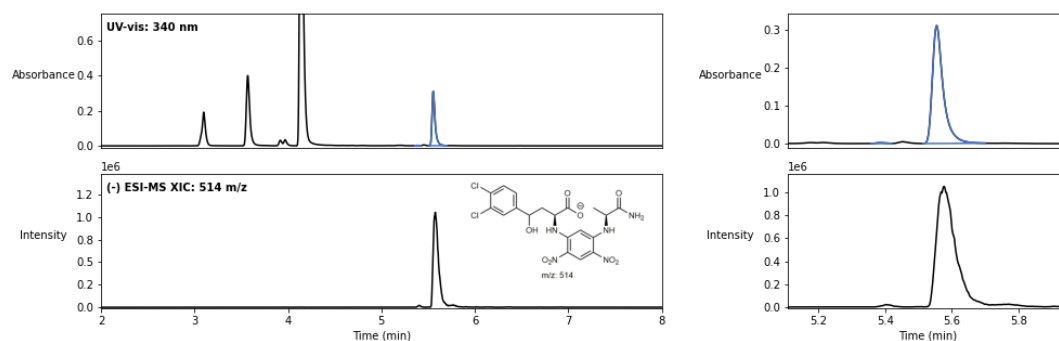
Figure 2S.12. UPLC trace for *2.4f* **-** *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-(3,4-dichlorophenyl)-4-hydroxybutanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
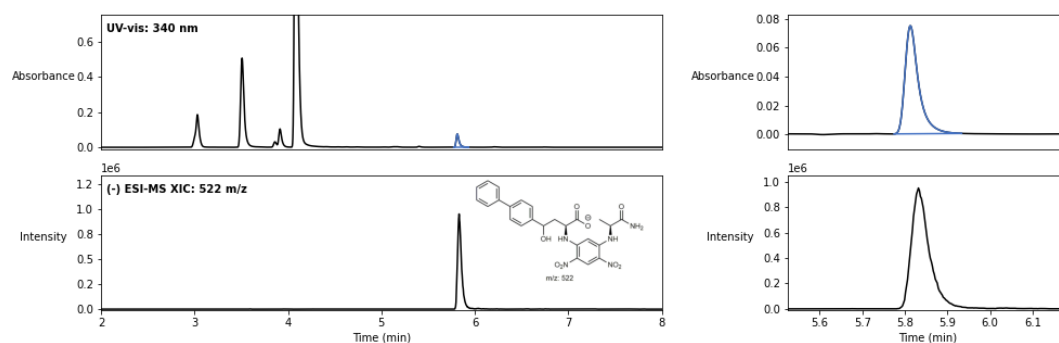
Figure 2S.13. UPLC trace for *2.4g* **-** *4-([1,1'-biphenyl]-4-yl)-2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxybutanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
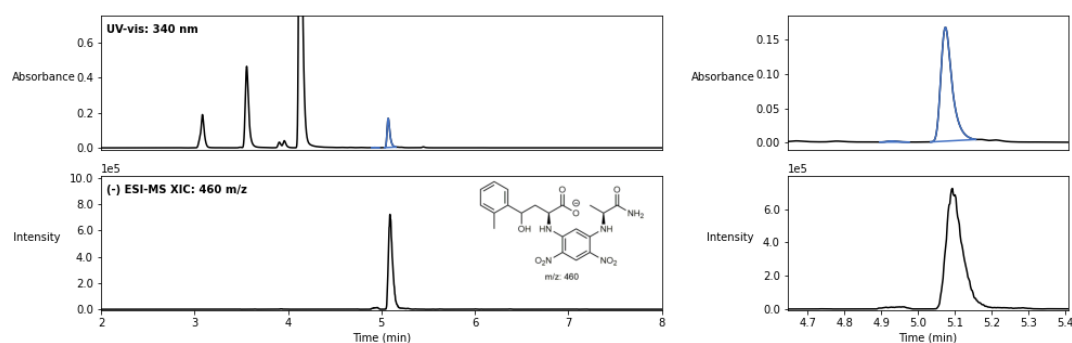
Figure 2S.14. UPLC trace for *2.4h* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-(o-tolyl)butanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
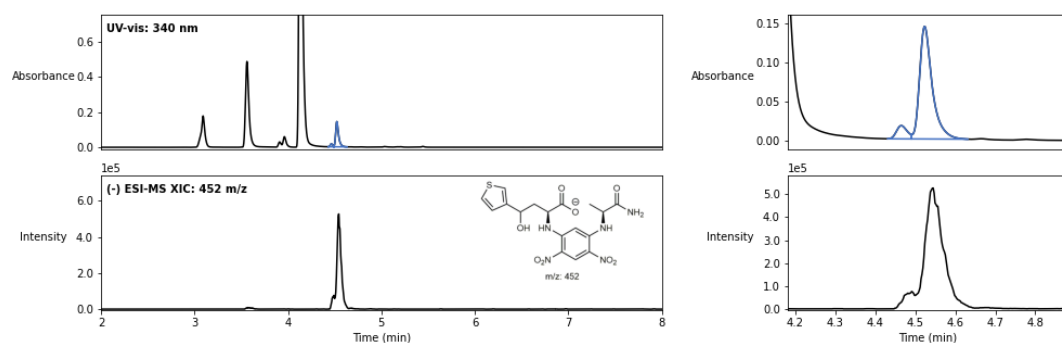
Figure 2S.15. UPLC trace for *2.4i* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-(thiophen-3-yl)butanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
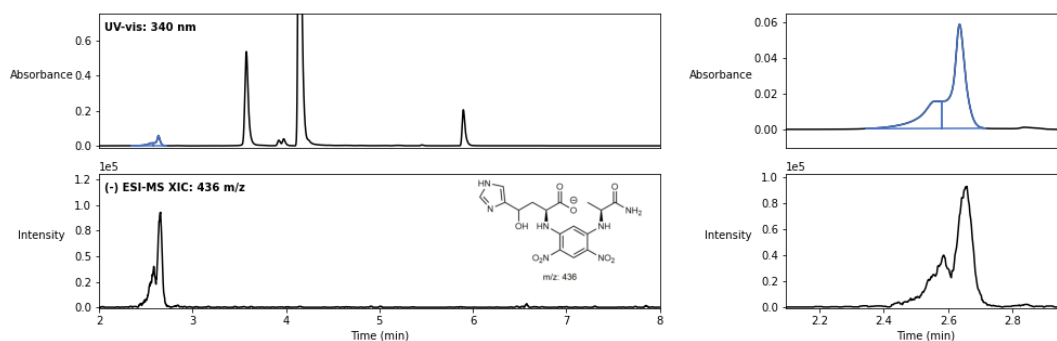
Figure 2S.16. UPLC trace for *2.4j* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-4-(1H-imidazol-4-yl)butanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
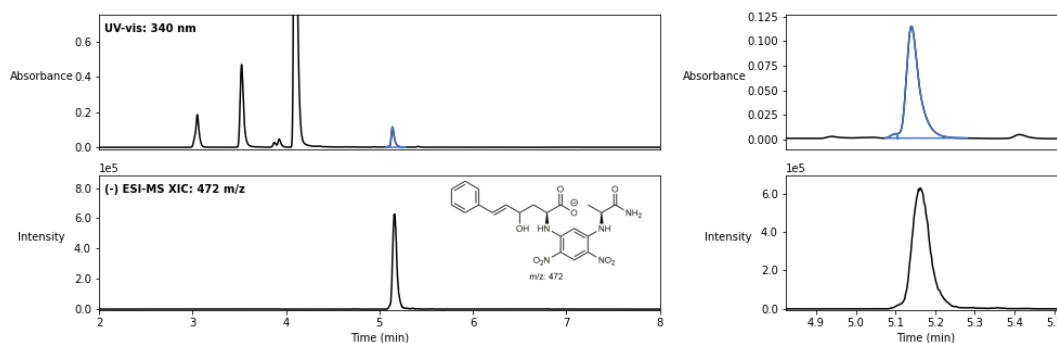
Figure 2S.17. UPLC trace for *2.4k* - *(E)-2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-6-phenylhex-5-enoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
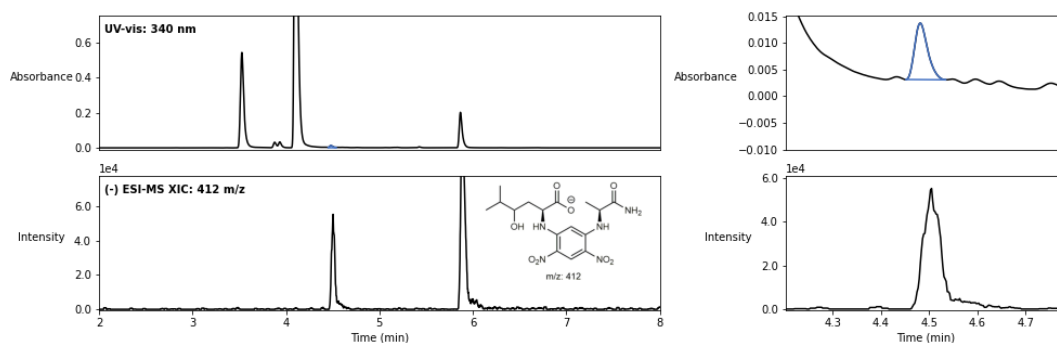
Figure 2S.18. UPLC trace for *2.4l* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxy-5-methylhexanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
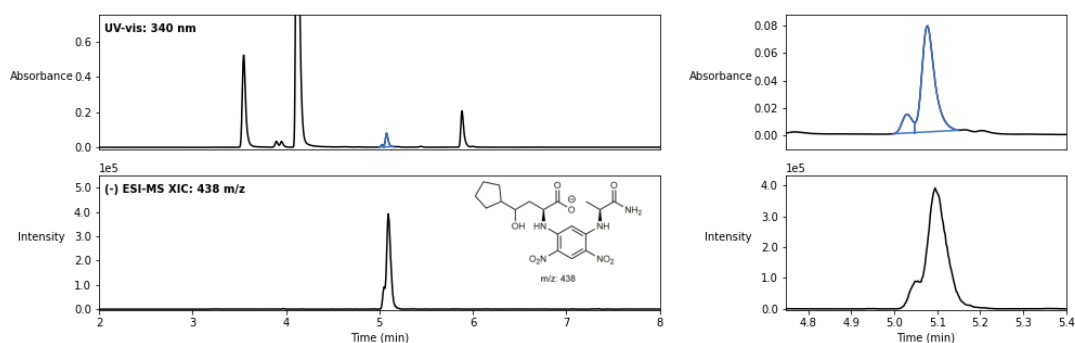
Figure 2S.19. UPLC trace for *2.4m* - *2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-cyclopentyl-4-hydroxybutanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
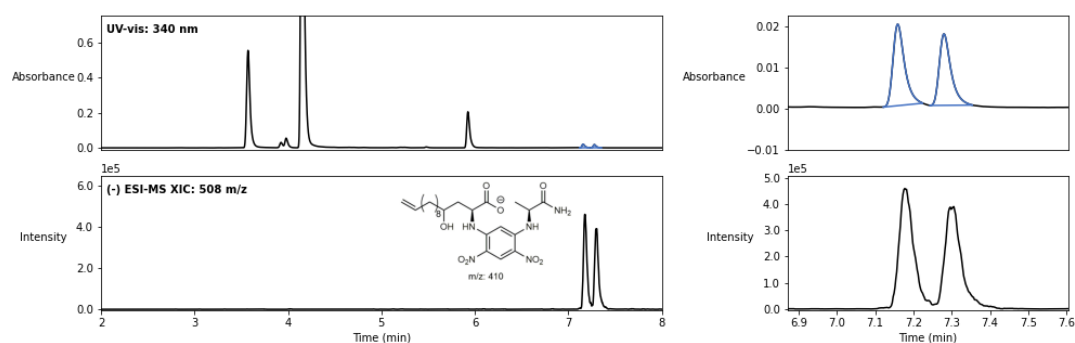
Figure 2S.20. UPLC trace for *2.4n - 2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxytetradec-13-enoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
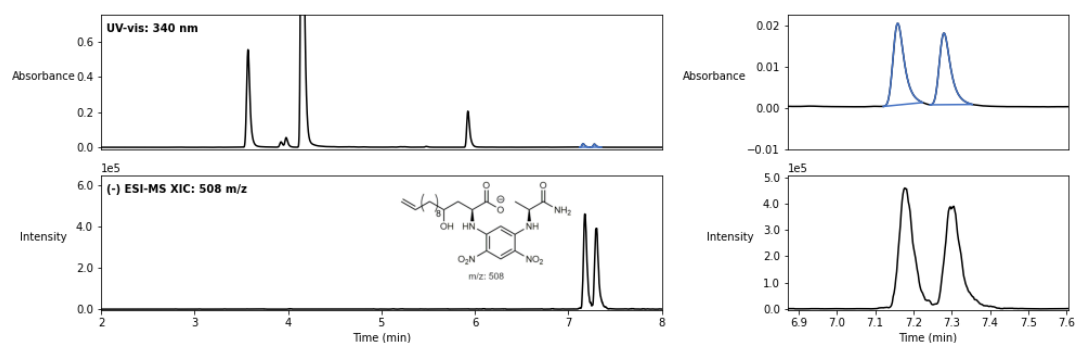
Figure 2S.21. UPLC trace for *2.4o - 2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4-hydroxypentanedioic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
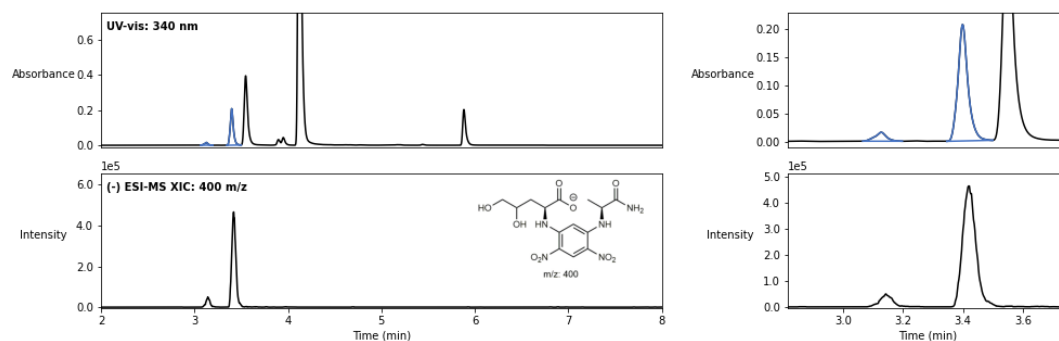
**Figure 2S.22. UPLC trace for *2.4p - 2-((5-((1-amino-1-oxopropan-2-yl)amino)-2,4-dinitrophenyl)amino)-4,5-dihydroxypentanoic acid.* Top left: UV-vis trace at 340 nm with diastereomers shown in blue. Top right: enlarged UV-vis trace of the diastereomers. Bottom left: extracted ion chromatogram of the diastereomers depicted by the chemical structure. Bottom right: enlarged extracted ion chromatogram of the diastereomers.
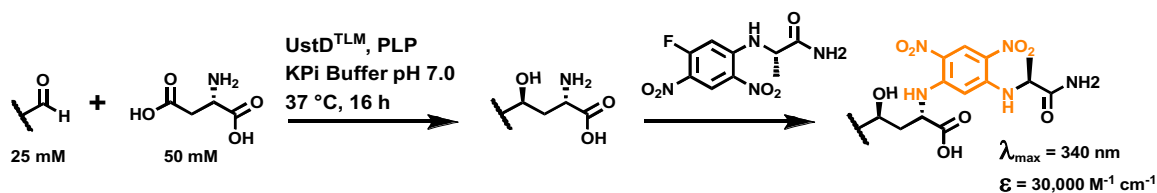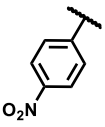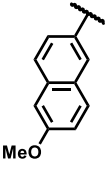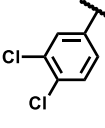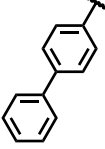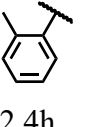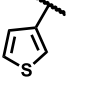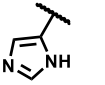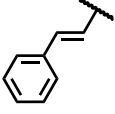
Table 2S.1. Performance analysis of UstD

	wt-UstD		UstD^{FVF}		UstD^{TLM}	
Product	Avg TTN	Avg d.r.	Avg TTN	d.r.	Avg TTN	d.r.
2.4a	3500 ± 200	93:7	4640 ± 40	96:4	8300 ± 500	96:4
2.4b	8800 ± 900	97:3	9300 ± 700	98:2	9600 ± 500	98:2
2.4c	800 ± 100	>99:1	1340 ± 70	>99:1	2300 ± 300	>99:1

Product	Avg TTN	Avg d.r.	Avg TTN	Avg d.r.	Avg TTN	Avg d.r.
2.4d (O$_2$N-aryl)	8700 ± 600	97:3	11400 ± 600	98:2	12300 ± 700	98:2
2.4e (MeO-naphthyl)	600 ± 100	>99:1	750 ± 40	>99:1	900 ± 100	>99:1
2.4f (Cl,Cl-aryl)	8000 ± 1000	98:2	6000 ± 2000	97:3 ± 1	7100 ± 800	97:3 ± 0
2.4g (biphenyl)	1500 ± 1500	>99:1	1200 ± 200	>99:1	2000 ± 400	>99:1
2.4h (methylphenyl)	190 ± 10	>99:1	900 ± 30	>99:1	4400 ± 200	95:5
	wt-UstD		UstD^{FVF}		UstD^{TLM}	
Product	Avg TTN	Avg d.r.	Avg TTN	Avg d.r.	Avg TTN	Avg d.r.
2.4i (thiophene)	430 ± 10	57:43	2300 ± 100	92:8	3900 ± 300	88:12
2.4j (imidazole)	3800 ± 200	71:29	2100 ± 200	70:30	3000 ± 100	73:27

2.4k	900 ± 300	93:7	1700 ± 100	99:1	2960 ± 70	98:2
2.4l	140 ± 30	>99:1	110 ± 20	>99:1	200 ± 30	>99:1
2.4m	700 ± 200	97:3	3400 ± 300	95:5	1900 ± 300	87:13
2.4n	830 ± 70	67:33	630 ± 60	74:26	800 ± 100	46:54
2.4o	1100 ± 100	88:12	2230 ± 50	92:8	3300 ± 300	96:4
2.4p	2800 ± 500	86:14	4700 ± 500	81:19	5600 ± 500	92:8

Average apparent TTN number (with standard deviation) and diastereomeric ratio (d.r.) are shown for each of the substrates. Measurements were conducted in at least duplicate. The reaction conditions are 25 mM aldehyde, 50 mM L-aspartate (**2.1**), 1.7 µM UstD, 17 µM PLP, 5% DMSO, buffer (100 mM KPi+ NaCl, pH = 7.0). Reactions proceeded at 37 °C overnight (~16 h), prior to derivatization with Marfey's reagent. The derivatization reaction went to >95% conversion for all cases. Product identity was further confirmed by high-resolution mass spectrometry. See SI methods for details. The maximum TON observable under these conditions was 15,000. Data in this table correspond to the bar graph in **Figure 2.6**.

Table 2S.2. X-Ray crystallographic data collection and refinement statistics

PDB ID		7MKV
Protein	*C*-His- UstD$^{v2.0}$	*C*-His- UstD$^{v2.0}$
Derivative	Au(III)	Native
Data Collection		
Space group	I422	I422
Cell dimensions (Å)	a,b,c = 162.3, 162.3, 221.4	a,b,c = 162.1, 162.1, 221.0
Cell angles	α = β = γ = 90°	α = β = γ = 90°
Wavelength (Å)	1.0394	1.0040
Beamline	21ID-D	21ID-D
Resolution (Å)	40 – 2.5	40 – 2.25
Last bin (Å)	(2.58 – 2.50)	(2.30 – 2.25)
No. observations	1,297,349	1,884,455
Completeness (%)	97.6 (81.6)	99.7 (96.5)
R$_{pim}$	0.051 (0.796)	0.035 (0.506)
CC(1/2)	0.999 (0.376)	0.999 (0.587)
I/σI	12.1 (1.0)	16.0 (1.5)
Redundancy	26.0 (18.5)	27.3 (21.9)
Refinement		
	Total no. of reflections	65625
	Total no. of atoms	9782
	TLS operators	30
	Final bin (Å)	(2.31 – 2.25)

	R$_{work}$ (%)	21.4 (31.3)
	R$_{free}$ (%)	24.6 (31.9)
	Average B factor (Å^2)	42.7
	Ramachandran plot Favored, %	98
	Allowed, %	2
	Outliers, %	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.

Table 2S.3. Crystal data and structure refinement for 2.3b

Empirical formula	$C_{10}H_{12}BrNO3 \cdot H_2O$
Formula weight	292.13
Temperature/K	100.0
Crystal system	monoclinic
Space group	P2$_1$
a/Å	7.4261(18)
b/Å	5.3296(16)
c/Å	15.004(4)
α/°	90
β/°	104.077(15)
γ/°	90
Volume/Å^3	576.0(3)
Z	2

ρ_{calc}g/cm^3	1.684
μ/mm^{-1}	3.567
F(000)	296.0
Crystal size/mm^3	0.193 × 0.173 × 0.163
Radiation	MoKα (λ = 0.71073)
2Θ range for data collection/°	2.798 to 61.048
Index ranges	-10 ≤ h ≤ 10, -7 ≤ k ≤ 7, -21 ≤ l ≤ 21
Reflections collected	16771
Independent reflections	3389 [R_{int} = 0.0206, R_{sigma} = 0.0135]
Data/restraints/parameters	3389/1/163
Goodness-of-fit on F^2	1.086
Final R indexes [I>=2σ (I)]	R_1 = 0.0143, wR_2 = 0.0388
Final R indexes [all data]	R_1 = 0.0145, wR_2 = 0.0389
Largest diff. peak/hole / e Å^{-3}	0.29/-0.27
Flack parameter	-0.006(2)

General Methods

All chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, VWR, Chem-Impex International, Alfa Aesar, Combi-blocks, Oakwood Products) at the highest quality available and used without further purification unless stated otherwise. Genes were purchased as gBlocks from Integrated DNA Technologies (IDT). *E. coli* cells were electroporated with an Eppendorf E-porator at 2500 V. New Brunswick I26R shaker incubators (Eppendorf) were used for cell growth. Cell disruption via sonication was performed with a

Sonic Dismembrator 550 (Fisher Scientific) sonicator. UV-vis spectroscopic measurements were collected on a UV-2600 Shimadzu spectrophotometer. Optical density measurements were collected using an optical density reader (Amersham Biosciences). Ultra-high pressure liquid chromatography-mass spectrometry (UPLC-MS) data were collected on an Acquity UPLC (Waters) equipped with an Acquity PDA and QDA MS detector using either a BEH C18 column (Waters) for substituted benzaldehyde reactions, or an Intrada Amino Acid column (Imtakt) for aliphatic aldehyde reactions. All UPLC-MS data were processed using Empower 3 (Waters). Preparative column separations were performed on an Isolera One Flash Purification system (Biotage). NMR data were collected on Bruker 400 or 500 MHz spectrometers equipped with BBFO and DCH cryoprobes, respectively. All NMR chemical shifts were referenced either to a residual solvent peak or TMS internal standard. Spectra recorded using DMSO-d^6 were referenced to the residual DMSO signal at 2.5 ppm for ^{1}H and 39.52 ppm for ^{13}C NMR. Spectra recorded using CDCl$_3$ were referenced to the residual CHCl$_3$ peak at 7.26 ppm for ^{1}H NMR and 77.16 for ^{13}C NMR. Spectra recorded using CD$_3$OD were referenced to the CH$_3$OD residual solvent peak at 3.31 ppm for ^{1}H and 49.00 ppm for ^{13}C NMR   Spectra recorded using D$_2$O:ACN-d^3 as the solvent were referenced to the residual H$_2$O signal at 4.79 ppm for ^{1}H and absolute referenced to the ^{1}H spectrum for ^{13}C NMR. Signal positions were recorded in ppm with the abbreviations s, d, t, q, dd, and m, denoting singlet, doublet, triplet, quartet, doublet of doublets, and multiplet respectively. All coupling constants J are measured in Hz. High resolution mass spectrometry data were collected with a Q Extractive Plus Orbitrap (NIH 1S10OD020022-1) instrument with samples ionized by ESI.

Cloning of wild-type UstD

A codon-optimized copy of the *Aspergillus flavus* UstD gene was purchased as a gBlock

from Integrated DNA Technologies. This DNA fragment was inserted into a pET-22b(+) vector

by the Gibson Assembly method[41] and transformed into electrocompetent BL21(DE3) *E. coli*

cells via electroporation. After a 30-minute recovery period in Luria-Burtani (LB) media, cells

were plated onto LB plates containing 100 µg/mL ampicillin (LB$_{amp}$) and incubated overnight. A

single colony was then used to inoculate 50 mL of Terrific Broth II media containing 100 µg/mL

ampicillin (TB$_{amp}$), which was then incubated overnight at 37 °C with shaking at 200 rpm. 500

µL of the saturated cell culture was then mixed with 500 µL of sterile 80% glycerol and snap-

frozen in liquid nitrogen to generate a glycerol stock.

Plasmid Preparations

A 5-mL overnight culture of *E. coli* harboring the plasmid of interest was grown

overnight at 37 °C with shaking at 200 rpm. The plasmid was isolated and purified using Zymo

Plasmid Miniprep kits and sequenced through Functional Biosciences.

Protein Expression

An overnight culture of *E. coli* BL21(DE3) harboring a pET-22b(+) plasmid encoding a

given UstD variant was created by inoculating 50 mL of TB$_{amp}$ media with a single colony. This

culture was shaken at 37 °C and 200 rpm for ~16 h. 10 mL of overnight culture was then used to

inoculate 1 L of TB$_{amp}$, which was shaken at 37 °C and 200 rpm for approximately 1.5 h or until

an optical density (OD) of 0.4-0.6 was reached. Cultures were removed from the incubator and

cooled on ice for 30 min, followed by induction with 100 µM IPTG. The cultures were allowed

to continue to grow for an additional ~16 h at 20 °C and shaking at 200 rpm. Cells were then

harvested by centrifugation (4 °C, 30 min, 4,000 xg), and the cell pellets were stored at -20 °C overnight.

Whole Cell Preparation of E. coli harboring UstD and variants

Following protein expression, cells were harvested by centrifugation (4 °C, 30 min, 4,000 xg). The cell pellets were then resuspended in water and centrifuged two times to remove all media. The cell pellets were transferred to 50 mL conical tubes and freeze dried by lyophilization. The dried cells were stored at -80 °C until further use.

Protein purification of UstD and variants

To purify UstD, cell pellets were thawed on ice and then resuspended in lysis buffer, comprised of enzyme storage buffer (100 mM potassium phosphate buffer, pH 7.0, 100 mM sodium chloride) containing 20 mM imidazole, 1 mg/mL Hen Egg White Lysozyme (GoldBio), 0.2 mg/mL DNase (GoldBio), 1 mM MgCl$_2$, and 150 μM pyridoxal 5'-phosphate (PLP). A ratio of 4 mL lysis buffer per gram of wet cell pellet was used. Cells lysis began by shaking for 1 h at 37 °C. The resuspended cells were subsequently sonicated (20 min, 0.8 s on, 0.2 s off, power setting 5). The resulting lysate was then spun down at 75,600 xg to pellet cellular debris. Ni/NTA beads were pre-equilibrated in storage buffer containing 20 mM imidazole. 1 mL of resin for 25 g of cells was added to the cleared lysis supernatant and incubated with nutation on ice for 1 h. The beads were then collected in a gravity column with plastic frit, and the flow-through was re-passed once to collect any remaining beads from the original vessel. The collected beads were washed with 10-20 column volumes of storage buffer containing 60 mM imidazole. Protein was eluted with 5 mL of storage buffer containing 250 mM imidazole and collecting the flow-through until the eluent was no longer yellow (color due to the enzymatically bound PLP cofactor). The eluent was then transferred to a centrifugal filter tube (Amicon®

Ultra-15, 30k MWCO) and concentrated by centrifugation (4,000 xg, 15 min). Imidazole was then removed either through dialysis or through repeated dilution (with enzyme storage buffer) and concentration steps until < 1 μM imidazole.

Generation of 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 or 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.

The PCR product was purified using a preparative agarose gel. Purified DNA fragment was inserted into a pET-22b(+) 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 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 for both the concentrations of MnCl2.

Protein engineering (library expression, screening, and validation)

Electrocompetent BL21(DE3) were transformed with mutagenized plasmid DNA and allowed to recover for 45 min in 800 μL of Terrific Broth (TB). After recovery, the cells were plated onto LB plates containing 100 μg/mL ampicillin (LB$_{amp}$) and incubated overnight. A 96-well plate containing 500 μL of TB$_{amp}$ per well was inoculated with single colonies. Each plate

included parent positive controls (from a fresh transformation), negative controls and a sterile control that was not inoculated. The plates were grown overnight at 37 °C, 200 rpm. Expression plates were prepared with 630 μL of TB_{amp} per well and inoculated with 20 μL of overnight culture. Glycerol stocks of each starter plate well were made from the remaining culture to ensure the sequence of any mutants of interest could be determined. The expression cultures were grown at 37 °C, 200 rpm for 2.5 h. Expression plates were then placed on ice for 30 min and induced with a final concentration of 0.1 mM IPTG in 50 μL of fresh TB_{amp}. The expression culture was grown overnight at 20 °C, 200 rpm. Following overnight growth, the plate was centrifuged (4,000 x*g*, 30 min, 4 °C) and all media was removed by striking plates against a paper towel on a table. Expression plates were stored at -20 °C until further use.

A lysis buffer containing 100 mM potassium phosphate buffer (pH 7.0), 100 mM sodium chloride, 1 mg/mL Hen Egg White Lysozyme (GoldBio), 0.2 mg/mL DNase (GoldBio), 1 mM $MgCl_2$, and 150 μM pyridoxal 5'-phosphate (PLP) was added to each well and the plate was subsequently lysed for 1 h at 37 °C. The lysate was pelleted at 4,000 x*g* for 30 min. Clarified lysate was added to a 96-well reaction plate where each well contained a master mix solution, such that the end reaction concentrations were 25 mM aldehyde, 25 mM L-asp, PLP, and buffer (100 mM KPi + NaCl, pH 7.0). The ratio of clarified lysate to reaction master mix was varied over the course of engineering to maintain a reasonable product measurement dynamic range. The reactions were allowed to incubate overnight at 37 °C, and were subsequently quenched with 100 μL acetonitrile and pelleted at 4,000 x*g* for 30 min. The cleared reaction mixture was transferred to a 0.2 μm centrifuge filter plate (PALL) and filtered at 1,500 x*g* for 10 min into a clean 96-well plate before being sealed prior to analysis by UPLC-MS.

The relative amount of product formed in the reactions compared to the positive control reaction was measured by absorbance at 210 nm via UPLC/MS. Given the relatively high variability in the parent signal in this assay, wells typically require an apparent 1.5-fold increase in product compared to the parent to be carried forward for validation of hits. Using the glycerol stocks from the starter culture plate, wells of interest could be streaked on to a fresh LB$_{amp}$ plate for subsequent sequencing and validation.

Every mutant of interest was validated by heterologous expression and Ni-NTA purification, accounting for changes in soluble enzyme concentration as well as changes in activity. To study how the activity profile of UstD changed over the course of engineering, each key variant in the evolutionary lineage was expressed and purified in tandem as described above (**Figure 2.3A**). Parallel triplicate 200 μL reactions containing 25 mM benzaldehyde, 50 mM L-aspartate sodium salt monohydrate, 2.5 μM PLP, and 0.25 μM UstD variant (0.001 mol% cat., 100,000 max TON) were allowed to react at 37 °C for 16 h. After, each reaction was quenched with 200 μL of ACN containing 1 mM tryptamine as an internal standard, and the reaction mixtures were analyzed by UPLC-MS. A standard curve was made using previously purified **2.3a** to facilitate total turnover number calculations. The variants were also trialed against several other aldehydes including: biphenyl-4-carboxaldehyde (20,000 max TON), *p*-anisaldehyde (20,000 max TON), and glycolaldehyde (100,000 max TON). Reactions were run using the same reaction conditions and procedure, with catalyst loading changed to match the indicated maximum turnover number. Simple fold-response measurements were used to quantify activity differences between variants (**Figure 2.5**).

UstD^{TLM} reaction condition optimization

All optimization reactions were conducted in triplicate on analytical scale (100 µL). PLP and L-aspartate stock solutions were made with 100 mM potassium phosphate buffer containing 100 mM sodium chloride (reaction buffer) at the indicated pH. Post-reaction quenching was done by adding 100 µL of 99:1 acetonitrile:ethanol with 1 mM tryptamine as an internal standard. Quenched reactions were then centrifuged at 15,000 x*g* to remove aggregated protein, and diluted with 200 µL of 1:1 water:acetonitrile. Quantification was performed by UPLC-MS analysis. Measurement of internal standard, benzaldehyde and product concentrations was done by separation on a BEH C18 column (Waters) and measurement of the corresponding 210 nm UV peak areas. Measurement of internal standard, product, L-aspartate, and L-alanine concentrations were done by separation on an Intrada amino acids column (Imtakt) 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. Optimization for each reaction condition component is listed below.

PLP concentration. A reaction master mix containing 27.5 mM L-aspartate monosodium monohydrate, 27.5 mM benzaldehyde, and 5.5% DMSO was made in 100 mM potassium phosphate buffer (pH 8.0) with 100 mM sodium chloride reaction buffer. 1.00, 0.20, 0.08, and 0.02 mM stocks of PLP were made by diluting a 20 mM PLP stock solution in reaction buffer. 0.5 dram glass vials were charged with 90.9 µL reaction master mix and 5 µL of the appropriate PLP stock (or buffer, in the case of no added PLP), and catalysis was initiated by addition of 4.1 µL of 25 µM UstD^{TLM} (Final concentrations: 25 mM L-aspartate, 2.5 µmol; 25 mM benzaldehyde, 2.5 µmol; 1 µM UstD^{TLM}, 0.004 mol% cat., 25,000 max TON; 50 µM, 10 µM, 4

UstD[TLM] reaction condition optimization

All optimization reactions were conducted in triplicate on analytical scale (100 µL). PLP and L-aspartate stock solutions were made with 100 mM potassium phosphate buffer containing 100 mM sodium chloride (reaction buffer) at the indicated pH. Post-reaction quenching was done by adding 100 µL of 99:1 acetonitrile:ethanol with 1 mM tryptamine as an internal standard. Quenched reactions were then centrifuged at 15,000 x*g* to remove aggregated protein, and diluted with 200 µL of 1:1 water:acetonitrile. Quantification was performed by UPLC-MS analysis. Measurement of internal standard, benzaldehyde and product concentrations was done by separation on a BEH C18 column (Waters) and measurement of the corresponding 210 nm UV peak areas. Measurement of internal standard, product, L-aspartate, and L-alanine concentrations were done by separation on an Intrada amino acids column (Imtakt) 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. Optimization for each reaction condition component is listed below.

PLP concentration. A reaction master mix containing 27.5 mM L-aspartate monosodium monohydrate, 27.5 mM benzaldehyde, and 5.5% DMSO was made in 100 mM potassium phosphate buffer (pH 8.0) with 100 mM sodium chloride reaction buffer. 1.00, 0.20, 0.08, and 0.02 mM stocks of PLP were made by diluting a 20 mM PLP stock solution in reaction buffer. 0.5 dram glass vials were charged with 90.9 µL reaction master mix and 5 µL of the appropriate PLP stock (or buffer, in the case of no added PLP), and catalysis was initiated by addition of 4.1 µL of 25 µM UstD[TLM] (Final concentrations: 25 mM L-aspartate, 2.5 µmol; 25 mM benzaldehyde, 2.5 µmol; 1 µM UstD[TLM], 0.004 mol% cat., 25,000 max TON; 50 µM, 10 µM, 4

µM, 1 µM, 0 µM PLP; 5% DMSO). Reactions were allowed to proceed in a 37 °C incubator for 16 h prior to quenching with 100 µL ACN and quantification (**Figure 2.4A**).

L-aspartate concentration. A reaction master mix containing 55.6 mM benzaldehyde, 111.1 µM PLP, and 11.1% DMSO was made in 100 mM potassium phosphate buffer (pH 8.0) with 100 mM sodium chloride reaction buffer. 500, 250, 100, and 50 mM stocks of L-aspartate monosodium monohydrate were made in reaction buffer. 0.5 dram glass vials were charged with 45 µL reaction master mix and 50 µL of the appropriate L-aspartate stock, and catalysis was initiated by addition of 5 µL 5 µM UstD^{TLM} (Final concentrations: 25 mM benzaldehyde, 2.5 µmol; 25, 50, 125, 250 mM L-aspartate, 2.5, 5.0, 12.5, 25.0 µmol, respectively; 0.25 µM UstD^{TLM}, 0.001 mol% cat., 100,000 max TON; 2.5 µM PLP, 10 equiv. relative to UstD^{TLM}; 5% DMSO). Reactions were allowed to proceed in a 37 °C incubator for 16 h prior to quenching with 100 µL ACN and quantification (**Figure 2.4B**).

pH. Five separate master mix solutions containing 25 mM benzaldehyde, 130 mM L-aspartate monosodium monohydrate, 1.3 mM PLP, and 5.2% DMSO were prepared in 100 mM potassium phosphate with 100 mM NaCl reaction buffer at pH 6.0, 6.5, 7.0, 7.5, and 8.0 (pH of buffer not altered after addition of reaction components). 0.5 dram glass vials were charged with 96.1 µL of the appropriate reaction master mix, and catalysis was initiated by addition of 3.9 µL 6 µM UstD^{TLM} (Final concentrations: 25 mM benzaldehyde, 2.5 µmol; 125 mM L-aspartate, 12.5 µmol; 0.25 µM UstD^{TLM}, 0.001 mol% cat., 100,000 max TON; 2.5 µM PLP, 10 equiv. relative to UstD^{TLM}; 5% DMSO). Reactions were allowed to proceed in a 37 °C incubator for 16 h prior to quenching with 100 µL ACN and quantification (**Figure 2.4C**).

mol% catalyst. A reaction master mix containing 50 mM benzaldehyde, 100 mM L-aspartate monosodium monohydrate, and 10% DMSO was made in 100 mM potassium

phosphate, pH 7.0, with 100 mM sodium chloride. UstD^{TLM} stock solutions containing 50 µM, 5.0 µM, 1.7 µM, and 1 µM were made, each containing 10 equivalents of PLP. 0.5 dram glass vials were charged with 50 µL reaction master mix, and catalysis was initiated by addition of 50 µL of the appropriate UstD stocks (Final concentrations: 25 mM benzaldehyde, 2.5 µmol; 50 mM L-aspartate, 5.0 µmol; 25 µM (0.1 mol% cat., 1,000 max TON), 2.5 µM (0.01 mol% cat., 10,000 max TON), 0.83 µM (0.003 mol% cat., 30,000 max TON), 0.25 µM (0.001 mol% cat., 100,000 max TON) UstD^{TLM}; 10 equiv. PLP relative to UstD^{TLM}; 5% DMSO). Reactions were allowed to proceed in a 37 °C incubator for 16 h prior to quenching with 100 µL ACN and quantification (**Figure 2.4D**).

UstD Performance Evaluation using Marfey's Derivatization

A 0.5-dram glass vial was charged with a master mix of L-aspartate sodium salt monohydrate (0.005 mmol, 2 equiv., 50 mM final concentration), pyridoxal-5'-phosphate (10 equiv. relative to final UstD concentration), and buffer. The master mix composition was varied to ensure a uniform concentration of each UstD variant at the completion of reaction setup. To this solution the aldehydes corresponding to compounds **2.2a-2.2p** (0.0025 mmol, 1 equiv., 25 mM final concentration) were added to the reaction mixtures. The reactions were initiated by the addition of UstD (0.007 mol% cat., 15,000 maximum turnover number). The reaction vessels were placed in a dark 37 °C incubator for 18 h and subsequently quenched with 200 µL of ACN. A Marfey's derivatization reaction was then performed in order to determine ee and dr of each enzymatic reaction. In a new flat bottom glass LC vial, 6 µL of quenched reaction mix (1 equiv., 0.5 mM final total amines from unreacted L-aspartate and formed L-alanine and γ-hydroxy amino acid product) was added to a solution of 144 µL of 10.41 mM NaHCO$_3$ (10 equiv., 5 mM final concentration) with 0.21 mM of either L-arg (0.1 mM final concentration, aldehydes **2.2a-2.2k**)

or tryptamine (0.1 mM final concentration, aldehydes **2.2l-2.2p**), followed by addition of 150 μL

5 mM L-FDAA dissolved in ACN (5 equiv., 2.5 mM final concentration) to bring the total

reaction volume to 300 μL. Each reaction vial was sealed with a pierceable LC vial cap, placed

in a dark 37 °C incubator for 18 h, then quenched with 300 μL of 1:1 ACN:60 mM HCl (15 mM

post-quench). Quenched reaction mixtures were analyzed by UPLC-MS no later than 24 h after

quenching, results are shown in **Table 2S.1** and **Figures 2S.6-22**.

Protein and DNA Sequences and Storage

Protein and DNA Sequences

Sequences with coding mutations relative to parent have the target mutation bolded.

Protein sequence of C-His-UstD: (Uniprot accession code: XP_002381324.1 for wild-type)
MKSVATSSLDDVDKDSVPLGSSINGTAQAETPLENVIDVESVRSHFPVLGGETAAFNNA
SGTVVLKEAIESTSNFMYSFPFPPGVDAKSMEAITAYTGNKGKVAAFINALPDEITFGQS
TTCLFRLLGLSLKPMLNNDCEIVCSTLCHEAAASAWIHLSRELGITIKWWSPTTTPNSPD
DPVLTTDSLKPLLSPKTRLVTCNHVSNVVGTIHPIREIADVVHTIPGCMLIVDGVACVPH
RPVDVKELDVDFYCFSWYKLFGPHLGTLYASRKAQDRYMTSINHYFVSSSSLDGKLAL
GMPSFELQLMCSPIVSYLQDTVGWDRIVRQETVLVTILLEYLLSKPSVYRVFGRRNSDPS
QRVAIVTFEVVGRSSGDVAMRVNTRNRFRITSGICLAPRPTWDVLKPKSSDGLVRVSFV
HYNTVEEVRAFCSELDEIVTRDTLEHHHHHH

DNA sequence of ustD codon optimized using IDT Codon Optimization Tool and including the flanking Gibson Assembly insertion and C-terminal 6xHis-Tag sequences.
GTTTAACTTTAAGAAGGAGATATACATATGAAGAGCGTAGCGACGAGTTCCCTTGAT
GACGTAGATAAAGATTCCGTCCCCCTGGGCAGTTCGATCAATGGCACTGCACAAGC
GGAAACTCCGCTGGAGAATGTGATCGACGTCGAATCAGTGCGCTCACATTTCCCGGT
ATTAGGGGGGGAAACGGCCGCGTTTAACAATGCATCAGGAACCGTAGTTTTGAAGG
AGGCAATTGAATCGACTTCAAATTTCATGTATAGCTTTCCTTTTCCCCCGGGTGTTGA
CGCTAAGTCAATGGAGGCTATTACCGCATATACGGGGAATAAGGGCAAGGTTGCGG
CATTTATCAATGCACTTCCTGATGAAATTACATTCGGGCAGTCCACAACTTGTCTGTT
CCGTTTATTAGGTCTGTCGCTTAAACCTATGCTGAATAACGATTGTGAAATCGTATG
CTCAACATTATGTCACGAAGCAGCAGCTTCCGCATGGATTCATTTAAGTCGCGAATT
AGGAATTACCATTAAGTGGTGGAGCCCAACTACTACACCGAATAGTCCCGATGATCC
AGTTCTGACGACTGACTCATTGAAGCCCTTGCTTAGTCCAAAAACGCGCCTTGTTAC
ATGTAATCACGTGTCGAATGTTGTAGGAACCATCCACCCTATTCGTGAGATTGCCGA
CGTGGTACATACCATTCCTGGATGCATGCTTATCGTTGACGGTGTGGCATGTGTCCC
GCATCGTCCAGTTGATGTTAAAGAATTGGATGTAGATTTTTACTGCTTTTCCTGGTAC
AAGTTGTTCGGACCGCATCTTGGAACCCTGTATGCTTCCCGCAAAGCCCAAGACCGC

TATATGACCTCAATTAACCATTACTTCGTCTCATCGTCGAGCCTTGATGGTAAGCTGG
CATTAGGCATGCCGTCCTTTGAACTGCAGTTGATGTGCTCTCCAATTGTTTCGTATTT
GCAAGATACGGTGGGCTGGGACCGTATCGTGCGCCAAGAGACTGTGCTGGTAACTA
TTTTGTTGGAGTATTTACTTAGCAAGCCATCTGTATATCGTGTGTTCGGACGTCGTAA
TTCTGATCCCAGTCAGCGTGTAGCAATCGTAACTTTTGAAGTCGTGGGACGTAGTTC
CGGGGATGTGGCAATGCGCGTAAATACGCGTAATCGCTTCCGCATTACCTCTGGAAT
TTGCCTGGCACCGCGCCCGACATGGGACGTCTTGAAACCGAAGAGTAGCGACGGAC
TTGTTCGCGTCAGCTTTGTACATTACAACACGGTTGAGGAAGTGCGTGCGTTCTGCA
GCGAGTTAGACGAGATTGTGACACGCGACACCCTCGAGCACCATCACCATCACCAT
TGAGATCCGGCTGC

DNA sequence of *ustD*^{C392L,L393M}

DNA sequence of ustD^{C392L,L393M}

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTTGTCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGTGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGATGCATG
CTTATCGTTGACGGTGTGGCATGTGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGAATT**TGCATG**GCACCGCGCCCGACATGGG
ACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTACA
ACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACGC
GACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of *ustD*^{TLM}

DNA sequence of ustD^{TLM}

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT

TACATTCGGGCAGTCCACAACTTGTCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGTGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGATGCATG
CTTATCGTTGACGGTGTGGCATGTGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGA**ACCTTAATG**GCACCGCGCCCGACATGG
GACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTAC
AACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACG
CGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^{FVF}
ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTTGTCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGTGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGATGCATG
CTTATCGTTGACGGTGTGGCATGTGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGA**TTCGTATTT**GCACCGCGCCCGACATGGG
ACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTACA
ACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACGC
GACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^{TLM-ACASC}

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACT**GCC**CTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGA**GCG**ATG
CTTATCGTTGACGGTGTGGCA**AGC**GTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGA**ACCTTAATG**GCACCGCGCCCGACATGG
GACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTAC
AACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACG
CGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^{TLM-ASCSC}

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACT**GCC**CTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGAT**TCC**GAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGC
TTCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCC
AACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCC
CTTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGG
AACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGATGCAT
GCTTATCGTTGACGGTGTGGCA**AGC**GTCCCGCATCGTCCAGTTGATGTTAAAGAATT
GGATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACC
CTGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTC
GTCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTG
CAGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGT
ATCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAG

CCATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCA
ATCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAAT
ACGCGTAATCGCTTCCGCATTACCTCTGGA**ACCTTAATG**GCACCGCGCCCGACATG
GGACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTA
CAACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACAC
GCGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD$^{TLM-ASASC}$

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACT**GCC**CTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGAT**TCC**GAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGC
TTCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCC
AACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCC
CTTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGG
AACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGA**GCG**AT
GCTTATCGTTGACGGTGTGGCA**AGC**GTCCCGCATCGTCCAGTTGATGTTAAAGAATT
GGATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACC
CTGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTC
GTCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTG
CAGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGT
ATCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAG
CCATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCA
ATCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAAT
ACGCGTAATCGCTTCCGCATTACCTCTGGA**ACCTTAATG**GCACCGCGCCCGACATG
GGACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTA
CAACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACAC
GCGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD$^{TLM-SCASC}$

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACT**TCC**CTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC

TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGA**GCC**ATG
CTTATCGTTGACGGTGTGGCA**TCT**GTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGA**ACCTTAATG**GCACCGCGCCCGACATGG
GACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTAC
AACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACG
CGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD$^{TLM-CCASC}$
ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTTGCCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGA**GCC**ATG
CTTATCGTTGACGGTGTGGCA**TCT**GTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGA**ACCTTAATG**GCACCGCGCCCGACATGG
GACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTAC
AACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACG
CGACACCCTCGAGCACCATCACCATCACCATTGA

Table 2S.4. Primer Sequences.

Protein	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
pET22b(+)-UstD	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG	GCCGGATCTCAATGGTGATGGTGATGGTGCTCGAG
C392X	AAATACGCGTAATCGCTTCCGCATTACCTCTGGAATT**X****XX***CTGGCACCGCGCCC	GATGTGGCAATGCGCGTAAATACGCGTAATCGCTTCCGCATTA
C392L-L393M	CGCTTCCGCATTACCTCTGGAATT**CTGATG**GCACCGCGCCCGACATGGG	CCAGAGGTAATGCGGAAGCGATTACGCGTATTTACG
I391[NYC] C392[KYA] L393[WTK] A394[SSA]	CGCTTCCGCATTACCTCTGG**ANYCKYAWTKSSA**CCGCGCCCGACATGGG	CCAGAGGTAATGCGGAAGCGATTACGCGTATTTACG
C122[KBC]	CGGGCAGTCCACAACT**KBC**CTGTTCCGTTTATTAG	CTAATAAACGGAACAG**GV****M**AGTTGTGGACTGCCCG
C139[TSC]	CCTATGCTGAATAACGAT**T****SC**GAAATCGTATGCTCAAC	GTTGAGCATACGATTTC**GS****A**ATCGTTATTCAGCATAGG
C227[KSC]	GGTACATACCATTCCTGGA**KSC**ATGCTTATCG	CGATAAGCAT**GSM**TCCAGGAATGGTATGTACC
C236[TSK]	CGGTGTGGCA**TSK**GTCCCGCATC	GATGCGGGAC**MSA**TGCCACACCG
C428[KSC]	GCGTGCGTTC**KSC**AGCGAGTTAG	CTAACTCGCT**GSM**GAACGCACGC
X139A	CCTATGCTGAATAACGAT**G****CG**GAAATCGTATGCTCAAC	GTTGAGCATACGATTTC**CG****C**ATCGTTATTCAGCATAGG
X236A	CGGTGTGGCA**GCG**GTCCCGCATC	GATGCGGGAC**CGC**TGCCACACCG

    * **XXX** indicates a 22-codon library made as a mixture of 3 degenerate codon primers (NDT, VHG, TGG), as described by Kille, S. et al.[42] All primers were purchased from Integrated DNA Technologies

DNA Isolation and Storage: DNA was purified via gel electrophoresis and isolated using a DNA gel extraction kit (Zymo Research). All isolated DNA was stored at -20 ºC.

Protein Characterization and Storage.

Concentration measurement: Enzyme concentration was determined by Bradford assay, using bovine serum albumin for a standard concentration curve.

Gel Electrophoresis: Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using 12% polyacrylamide gels (**Figure 2S.1**).

Storage: 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 ºC until further use. Frozen pellets were thawed at room temperature and centrifuged before use (**Figure 2S.1**).

Library Generation for Directed Evolution

Production of UstD random mutagenesis libraries:

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.

Production of UstD degenerate codon libraries:

Primers containing degenerate codons were purchased from IDT and are listed above.

Enzymatic Activity Experiments

UstD Performance Evaluation using Marfey's Derivatization

General Methods: All reactions were done in triplicate on analytical scale (200 μL). The buffer

used for all enzymatic reactions was 100 mM KPi, pH 7.0, 100 mM NaCl. PLP and L-aspartate

stock solutions were made in buffer. Aldehyde stock solutions were made in DMSO, except for

glycolaldehyde, glyoxylic acid, DL-glyceraldehyde, and imidazole-4-carboxaldehyde which were

made in buffer. All UstD variant stock solutions concentrations were quantified by Bradford

assay prior to setting up reactions. All samples were analyzed following Marfey's derivatization

by Waters Acquity UPLC-MS using a BEH C18 column (Waters). To correct for small

deviations in injection volume, an internal derivatization standard was included (0.1 mM L-

arginine for reactions containing aromatic aldehydes, 0.1 mM tryptamine for reactions

containing aliphatic aldehydes). Derivatized amino acid product quantitation was performed by

integrating chromatograph peaks at 340 nm and corrected by dividing by the internal standard

peak area. To calculate product concentrations, a standard curve was generated by subjecting

stock solutions of L-phenylalanine (25 mM – 0.25 mM) in buffer to the identical procedure used

to process and derivatize enzymatic reaction solutions, in tripliciate, with both internal standards.

These curves were used to calculate the concentrations of UstD products in solution, and

subsequently total turnover numbers after dilution factor correction.

Preparative Scale In vitro Biocatalytic Reactions and Protections

Procedure P1: *Preparative scale production of unprotected aromatic γ-hydroxy amino acids*

A 50-mL round bottom flask was charged with a given aldehyde (0.2 mmol, 1.0 equiv, 25 mM

final concentration), which was then dissolved in an appropriate amount of DMSO (5% v/v final

concentration). This solution was then diluted with 100 mM potassium phosphate buffer (pH 7.0)

containing 100 mM sodium chloride. L-aspartate sodium salt monohydrate (0.4 mmol, 2.0 equiv,

50 mM final concentration) and 10 molar equivalents of pyridoxal-5'-phosphate (PLP) relative to

final UstD concentration were then added, followed by addition of UstD (0.0002-0.001% mol

cat). The total reaction volume was 8 mL. The reaction flask was placed in the dark at 37 °C for

16 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 (4,000

rpm, 15 min) to remove aggregated protein. The decanted supernatant was then concentrated to

~2 mL by rotary evaporation and loaded onto a preparative reverse-phase C18 column pre-

equilibrated at 1:20 methanol: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.

Procedure P2: *Preparative scale production of Fmoc-protected aromatic γ-hydroxy amino acids*

A 100-mL round bottom flask was charged with a given aldehyde (0.2 mmol, 1.0 equiv., 25 mM

final concentration), which was then dissolved in an appropriate amount of DMSO (5% v/v final

concentration). This solution was then diluted with 100 mM potassium phosphate buffer (pH 7.0)

containing 100 mM sodium chloride. L-aspartate sodium salt monohydrate (0.4 mmol, 2.0 equiv.,

50 mM final concentration) and PLP (10 equiv. relative to final UstD concentration) were then

added. The reaction was initiated by addition of UstD (0.0002 – 0.001% mol. cat.), and the

reaction flask was placed in a dark 37 °C incubator for 16 h. Product formation was monitored by

UPLC-MS. The reaction mixture was treated with 6.0 M NaOH until pH ~10 (measured by

universal indicator paper). In parallel, Fmoc-Cl (0.6 mmol, 1.5 equiv.) was weighed out and

dissolved in acetonitrile to a final volume equal to the enzymatic reaction, typically 8 mL. The

Fmoc-Cl solution was then added to the alkaline enzymatic reaction mixture with vigorous

stirring and was allowed to react at room temperature for 4 h or until no amino acid product was

detected by UPLC/MS. The acetonitrile was then removed by rotary evaporation, and the

concentrated solution was treated with 2.0 M citric acid until pH = 3 by universal indicator

paper. The reaction was then extracted with ethyl acetate (3 x 25 mL), and the organic layers

were retained and combined. The organic layer was then washed with sat. $NaHCO_3$ (2 x 25 mL)

and sat. NaCl (2 x 25 mL) and dried over magnesium sulfate. The solution was then concentrated

by rotary evaporation and applied to a Biotage Samplet charged with silica before being placed

in a 25-gram Biotage SNAP KP-Sil cartridge pre-equilibrated with hexane. The Fmoc-amino

acid product was then purified against hexane:ethyl acetate by gradient elution and automated

fractionation (absorption monitoring at 210 and 254 nm). Fractions bearing product were

confirmed by UPLC-MS mass detection, pooled, and evaporated by rotary evaporation before

overnight drying under high vacuum in a pre-weighed vial. Product isolation and identity was

confirmed by NMR and high-resolution mass-spectrometry.

Procedure P3: *Preparative scale production of lactonized aliphatic γ-hydroxy amino acids*

A 100-mL round bottom flask was charged with a given aldehyde (0.2 mmol, 1.0 equiv., 25 mM

final concentration), L-aspartate sodium salt monohydrate (0.4, 2.0 equiv., 50 mM final

concentration), and pyridoxal 5'-phosphate (10 equiv. relative to final UstD concentration)

before dissolving everything in 100 mM potassium phosphate buffer (pH 7.0) containing 100

mM sodium chloride. The reaction was initiated by addition of UstD (0.0002-0.001% mol. cat.),

and the reaction flask was placed in a dark 37 °C incubator for 16 h. Product formation was

monitored by UPLC-MS. The reaction mixture was then treated with 6.0 M NaOH until pH ~10

(measured by universal indicator paper). A given amount of Fmoc-Cl (0.6 mmol, 1.5 equiv.) was weighed out and dissolved in an equal volume of acetonitrile (relative to the enzymatic reaction mixture volume). The Fmoc-Cl solution was then added to the alkaline enzymatic reaction mixture with vigorous stirring and was allowed to react at room temperature for 4 h or until no free amino acid product was visible by UPLC-MS. Acetonitrile was then removed from the reaction by rotary evaporation. The reaction flask was placed in an ice bath and allowed to cool before being acidified with concentrated HCl (16 mL, 8 M final concentration). This solution was allowed to react at room temperature for 24 h with vigorous stirring. The formation of the lactonized product was monitored by UPLC-MS. Upon completion of the lactonization, the reaction mixture was extracted with ethyl acetate (4 x 25 mL) without neutralization. The organic layer was washed with 10 mL portions of water until the washings were pH >5 by universal indicator paper. Subsequently, the organic layer was washed with saturated sodium chloride (2 x 10 mL), dried over magnesium sulfate, and concentrated by rotary evaporation. The lactonized product was then purified by normal phase chromatography using hexane:ethyl acetate containing 0.5% acetic acid. Fraction purity was initially verified by TLC, followed by UPLC-MS, and fractions containing purified product were pooled. Acetic acid was removed by three cycles of addition of heptane and rotary evaporation concentration before overnight drying under high vacuum in a pre-weighed vial. Product isolation and identity was confirmed by NMR and high-resolution mass-spectrometry.

Hydration State NMR Experiments

For reactions analyzed in 1:1 D_2O/ACN-d_3, 2 mL of solvent was prepared in a 1-dram glass vial. For reactions analyzed in DMSO-d_6, two 1 mL ampules of solvent were opened and combined in a 1-dram glass vial. Two 800 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, 750 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 and 75 s for DMSO-d_6 and D_2O/ ACN-d_3 samples, respectively) 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.

Linear Regression Modelling

To improve the expression properties of UstD[TLM], putative surface cysteines were identified

from a homology model generated by I-TASSER[4323,43,44] (**Figure 2.9A**). Because mutation to

Ala might have a deleterious effect, we sought to build a degenerate codon library that sampled a

variety of alternative residues that were present among UstD homologs. To analyze the

homologous sequences, a sequence similarity network was generated using the EFI-EST web

service[45–48] and was processed using Cytoscape[49] (**Figure 2.9B**). A cluster representing the

nearest 1000 homologues to UstD was used as the input for a multiple sequence alignment. From

these data, we identified five Cys residues that had a low degree of conservation as well as other

residues that were prominent at these positions. Primers were designed for simultaneous

degenerate codon mutagenesis at each site, constituting a library with 576 possible sequences (**Figure 2.9C**).

The library was constructed using two rounds of polymerase cycling assembly and validated by Sanger sequencing (Functional Biosciences). Transformation and screening of two 96-well plates was conducted as described above. Glycerol stocks of both library plates were also sent for sequencing, and 104 of the 176 non-control wells produced high quality sequencing results containing 95 unique sequences. The sequences were then one-hot encoded to represent the presences of mutations at each position (input vector size: 15) and matched with the UPLC screening fold-activity (relative to UstD^{TLM}) data to generate a modelling dataset. Linear regression modelling was performed using Python 3.8.6[50] and the scikit-learn 0.23.2 package[51]. To linearize the data, the natural log of all fold-activities was used during model fitting, and the model outputs were exponentiated before calculating model statistics or making predictions. As such, observations with a measured fold-activity less than zero were removed. An initial linear regression data against the remaining 77 measurements (69 unique sequences) showed a relatively modest global fit (Mean absolute error: 0.16, R^2: 0.69) with poor predictive power at high fold-activity (**Figure 2S.4A**).

The C428G and C428S mutations were highly deleterious and removed during the noise-filtering step. We hypothesized the dominance of experimental noise in the zero-activity measurements was reducing the predictive power of the model. Notably, a visual inspection of the sequences showed that those bearing mutations at C428 had very low, typically no measurable catalytic activity. Therefore, to mitigate the influence of noise and improve the extrapolation capability of the model, we removed all measurements with a recorded fold-activity < 0.1 (corresponding to the approximate noise level of the instrument.) This had the

effect of removing C428G and C428S from the modellable sequence space, leaving 32 sequence-activity measurements with 26 unique sequences.

We re-trained the linear regression model (input vector size: 13) on this more limited data set which only consisted of bonafide turnover measurements and not catalytically 'dead' enzymes. To our satisfaction, the newly trained model showed significant improvements. Leave-one-out cross-validation produced reasonable model statistics during cross-validation (mean absolute error: 0.21, R^2: 0.79; **Figure 2S.4B**), with the complete model showing improved fit at modest-to-high activities (mean absolute error: 0.13, R^2: 0.92; **Figure 2.10**). We then re-mapped the trained model weight matrix to the corresponding sequence and position from one-hot encoding, resulting in predictive matrix (**Figure 2S.4C**) for the UstD Cys shuffle sequence space. We queried this sequence space and the top three variants with the lowest cysteine content were chosen for production and screening along with the variants with the highest activity from each library plate.

Whole Cell Biocatalytic Reactions and Protections

Whole Cell 2.3a - A 1-L baffled flask was charged with 100 mM potassium phosphate buffer (pH 7.0) containing 100 mM sodium chloride and DMSO (5% v/v). Then, **2.2a** (5 mmol, 1 equiv., 30 mM final concentration), L-aspartate sodium salt monohydrate (10 mmol, 2.0 equiv., 60 mM final concentration), and PLP (8.4 µmol, 0.05 mM final concentration) were added to the solution. The reaction was initiated upon the addition of lyophilized *E. coli* cells harboring expressed UstD$^{2.0}$ (840 mg, 5 mg/mL). The reaction vessel was moved to a shaking 37 °C incubator and allowed to react overnight. Product formation was monitored by UPLC/MS. The reaction was quenched with an equal volume of ACN (168 mL). To remove cell debris, the reaction mixture was centrifuged at 4000 rpm for 20 min. The resulting supernatant was

concentrated to ~100 mL by rotary evaporation and purified in two batches by preparative

reverse-phase chromatography. Purification was performed via gradient elution on an Isolera

One Flash Purification system (Biotage) using a 60 g, C18 column pre-equilibrated at 1%

methanol. Fractions bearing **2.3a** (confirmed by UPLC-MS sampling of fraction tubes) were

pooled and evaporated to dryness by rotary evaporation. The resulting powder of **2.3a** was then

resuspended in a minimal quantity of water, transferred to a pre-weighed 20 mL scintillation vial,

frozen, and subjected to lyophilization. After confirming the product obtained from both Biotage

runs was equally pure, the two samples were combined for subsequent characterization.

Whole Cell 2.3d - A 1-L baffled flask was charged with 100 mM potassium phosphate buffer

(pH 7.0) containing 100 mM sodium chloride and DMSO (5% v/v). Then, **2.2d** (5 mmol, 1

equiv., 30 mM final concentration), L-aspartate sodium salt monohydrate (10 mmol, 2.0 equiv.,

60 mM final concentration), and PLP (16.8 µmol, 0.1 mM final concentration) were added to the

solution. The reaction was initiated upon the addition of lyophilized *E. coli* cells harboring

expressed UstD$^{2.0}$ (168 mg, 1 mg/mL). The reaction vessel was moved to a shaking 37 °C

incubator and allowed to react overnight (~16 h). Formation of **2.3d** was monitored by

UPLC/MS. The reaction was quenched with an equal volume of ACN (168 mL). To remove cell

debris, the reaction mixture was centrifuged at 4000 rpm for 30 min and the pellet was rinsed

with water twice to remove residual product. The ACN and more than half the water was

removed by rotary evaporation and the remaining solution was placed in the fridge overnight (If

the solution is too concentrated then a yellow-white sludge will form, to mitigate this add a small

amount of 1:1 H$_2$O:MeOH.) The precipitate was filtered and rinsed with ice cold water (3 x 5

mL). The resulting solid, **2.3d**, was then transferred to a pre-weighed 20 mL scintillation vial,

resuspended in a minimal quantity of water, frozen, and subjected to lyophilization. The filtrate

was analyzed by UPLC/MS and was found to still contain a small amount of **2.3d**. Therefore, the filtrate was concentrated to ~2 mL by rotary evaporation and loaded onto a preparative reverse-phase C18 column pre-equilibrated at 1% methanol. Purification was performed via gradient elution on a Isolera One Flash Purification system (Biotage). Fractions bearing **2.3d** (confirmed by UPLC-MS sampling of fraction tubes) were pooled and evaporated to dryness by rotary evaporation. The resulting solid was then resuspended in a minimal quantity of water, transferred to a pre-weighed 20 mL scintillation vial, frozen, and subjected to lyophilization. After confirming the **2.3d** obtained from both methods (precipitation and column chromatography) was equally pure, the two samples were combined for subsequent characterization.

To assess whether the diastereomeric ratios could change due to retro-aldol cleavage of the product, we assessed the stability of **2.3d** when re-subjected reaction conditions in the presence and absence of enzyme at high concentrations. To a 0.5-dram vial, buffer (100 mM KPi, 100 mM NaCl, pH 7.0), PLP (50 µM), and UstD$^{2.0}$ (0.1 mol % catalyst, 1000 max TON) were combined. Then, **2.3d** (5 mM in 5% DMSO) was added to the reaction mixture and kept in the dark for 18 h at 37 °C. Reactions were quenched with one reaction volume of ACN (250 µL). In a new flat bottom glass LC vial, 25 µL of quenched reaction mix (1 equiv., 0.2 mM **2.3d**) was added to a solution of 125 µL of 10.41 mM NaHCO$_3$ (20 equiv., 4.3 mM final concentration) with 0.21 mM of either L-arg (0.08 mM) followed by addition of 150 µL 5 mM L-FDAA dissolved in ACN (12.5 equiv., 2.5 mM final concentration) to bring the total reaction volume to 300 µL. Each reaction vial was sealed with a pierceable LC vial cap, placed in a dark 37 °C incubator for 18 h, then quenched with 300 µL of 1:1 ACN:60 mM HCl (15 mM post-quench). Quenched reaction mixtures were analyzed by UPLC-MS no later than 24 h after quenching. The

analysis (spectra shown below in **Figure 2S.3**) revealed a dr of 98:2 for **2.3d** with no erosion of the dr.

Protein Crystallization, Experimental Phasing, and Model Building

Crystallization

Crystallization conditions for $UstD^{2.0}$ 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 using 20 and 45 mg/mL $UstD^{2.0}$. Crystals appeared in conditions containing 10% PEG-MME 5000, 0.1 M HEPES pH 7.0, and 5% Tacsimate pH 7.0. Crystals were reproduced by mixing a 1:1 ratio of well solution (5-15% PEG-MME 5000, 0.1 M HEPES pH 7.0) and 20 mg/mL $UstD^{2.0}$ to generate 2μL of mother liquor, which was allowed to equilibrate against 1 mL well solution. Initially, yellow tetragonal crystals grew over a period of days. Initial diffraction experiments showed crystals diffracted to 2.7-3.5 Å, and our early efforts failed to acquire sufficient anomalous signal with heavy atom soaks to determine the substructure, which was compounded by ambiguities in the space group due to significant non-translational pseudosymmetry.

During the course of the COVID-19 pandemic, crystallization trays continued to sit in a dark room and, to our great joy, conditions that initially had precipitant concentrations formed beautiful yellow crystals at some point during 2020, over a year after the trays were initially laid. The best native dataset was later acquired from crystals that were soaked overnight with 1 μL of 50 mM thimerosal dissolved in well solution. Experimental phases were acquired from samples that were soaked overnight with 0.5 μL of 76 mM $AuCl_3$. Sample was looped with a Mitegen Micromount through Fomblin Y, and were flash frozen in liquid N_2 until diffraction.

Experimental Phasing

X-ray data were collected remotely at the Advanced Photon Source, Structural Biology Center Beamline 21-ID-D. Data were processed with XDS and Aimless[52,53]. Crystals soaked with thimerosal showed no anomalous signal, but from these conditions we acquired a native data set with good statistics. We observed a modest anomalous signal with the Au-derivatized protein. We determined the structure of UstD[2.0] by the SIRAS method, as implemented in Crank2.0[54], which used SHELX for substructure detection[55], BP3 for refinement[56]. Density modification was performed with SOLOMON[57].

Model Building

Automated model building with Buccaneer generated a model that was ~90% complete[58]. The asymmetric unit contained three subunits; two protomers in a biological dimer and a third protomer related through non-translation pseudosymmetry and formed a dimer with itself through crystallographic symmetry. Iterative cycles of manual model building with Coot[59] and refinement with REFMAC5[60]. After all atoms were placed, 10 TLS operators were added per chain using the TLS motion server and refined prior to deposition[60]. X-ray data collection and model refinement statistics are listed in **Table 2S.2**.

Characterization of γ-hydroxy amino acid products

2.3a - **Synthesis of (2S,4S)-2-amino-4-hydroxy-4-phenylbutanoic acid**
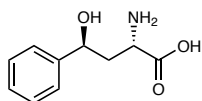
Prepared from **2.2a** using procedure **P1**. **Isolated yield:** 47% **^{1}H NMR** (400 MHz, Deuterium Oxide: Acetonitrile-d^3) δ 7.90 (d, J = 4.4 Hz, 4H), 7.87 – 7.77 (m, 1H), 5.36 (dd, J = 10.0, 3.2 Hz, 1H), 4.17 (dd, J = 7.5, 4.2 Hz, 1H), 2.66 (ddd, J = 14.3, 9.9, 4.1 Hz, 1H), 2.48 (dd, J = 7.5, 3.2 Hz, 2H). **^{13}C NMR** (126 MHz, D$_2$O:Acetonitrile-d^3) δ 176.41, 144.59, 129.14, 128.16, 126.26, 71.54, 53.78, 40.26. **Hydration State:** C$_{10}$H$_{13}$NO$_3$ · 6 H$_2$O **HRMS (ESI):** [M-H]⁻ calcd. for C$_{10}$H$_{13}$NO$_3$, 194.0823; found, 194.0823.

2.3a - **Synthesis of (2S,4S)-2-amino-4-hydroxy-4-phenylbutanoic acid (wt-UstD)**
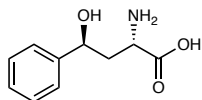
Prepared from **2.2a** using wt-UstD. **Isolated yield:** 43% **^{1}H NMR** (500 MHz, Deuterium Oxide) δ 7.47 – 7.41 (m, 4H), 7.40 – 7.35 (m, 1H), 4.90 (dd, J = 9.5, 3.9 Hz, 1H), 3.76 (dd, J = 7.3, 4.4 Hz, 1H), 2.29 (ddd, J = 14.8, 9.5, 4.3 Hz, 1H), 2.15 (ddd, J = 14.9, 7.3, 3.9 Hz, 1H). **^{13}C NMR** (126 MHz, D$_2$O) δ 176.13, 143.15, 128.86, 128.11, 125.86, 71.07, 53.10, 39.03.

2.5b - **Synthesis of (2S,4S)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-(4-bromophenyl)-4-hydroxybutanoic acid**
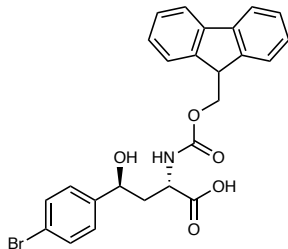
Prepared from **2.2b** using procedure **P2**. **Isolated yield:** 56% ^{1}**H NMR** (500 MHz, Chloroform-

d) δ 7.76 (d, *J* = 7.5 Hz, 2H), 7.58 (d, *J* = 7.4 Hz, 2H), 7.54 (d, *J* = 8.4 Hz, 2H), 7.40 (t, *J* = 7.6

Hz, 2H), 7.36 – 7.28 (m, 2H), 7.24 (d, 2H), 5.49 – 5.28 (m, 2H), 4.58 (s, 1H), 4.45 (d, *J* = 6.9 Hz,

2H), 4.23 (t, *J* = 6.7 Hz, 1H), 3.13 (s, 1H), 2.16 (q, *J* = 11.5, 11.0 Hz, 1H). ^{13}**C NMR** (126 MHz,

Chloroform-*d*) δ 173.73, 155.93, 143.60, 141.36, 136.75, 132.08, 127.83, 127.44, 127.13,

125.00, 123.04, 120.05, 77.98, 67.42, 52.17, 47.06, 38.83.

HRMS (ESI): 297.9722

Exact Mass: 297.97
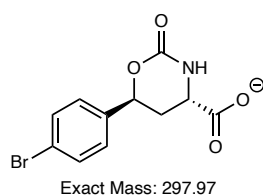
2.5c - Synthesis of (2*S*,4*S*)-2-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-hydroxy-4-(4-methoxyphenyl)butanoic acid
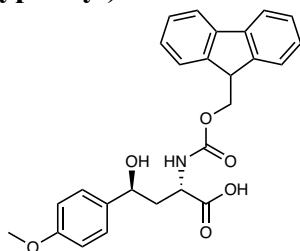
Prepared from **2.2c** using procedure **P2**. **Isolated yield:** 20% ^{1}**H NMR** (500 MHz, Chloroform-

d) δ 7.77 (d, *J* = 7.6 Hz, 2H), 7.59 (d, *J* = 7.5 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.36 – 7.28 (m,

4H), 6.93 (d, 2H), 5.39 (m, 2H), 4.59 (s, 1H), 4.45 (d, *J* = 7.0 Hz, 2H), 4.24 (t, *J* = 6.9 Hz, 1H),

3.82 (s, 3H), 3.09 (s, 1H), 2.21 (dd, *J* = 25.5, 13.3 Hz, 1H). ^{13}**C NMR** (126 MHz, Chloroform-*d*)

δ 174.12, 160.23, 156.02, 143.68, 141.35, 129.35, 127.81, 127.66, 127.13, 125.05, 120.03,

114.23, 78.86, 67.37, 55.37, 52.39, 47.09, 38.82. **HRMS (ESI):** [M+NH$_4$]$^+$ calcd. for

C$_{26}$H$_{23}$NO$_5$, 447.1915; found, 447.1912

2.5d - Synthesis of (2*S*,4*S*)-2-(((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-hydroxy-4-(4-nitrophenyl)butanoic acid
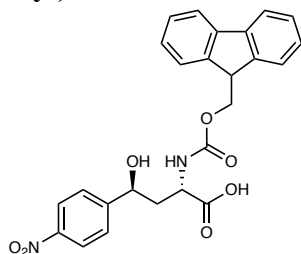
Prepared from **2.2d** using procedure **P2**. **Isolated yield:** 64% **¹H NMR** (500 MHz, Chloroform-*d*) δ 8.27 (d, *J* = 8.8 Hz, 2H), 7.76 (d, *J* = 7.6 Hz, 2H), 7.62 – 7.51 (m, 4H), 7.40 (t, *J* = 7.5 Hz, 2H), 7.30 (td, *J* = 7.5, 1.1 Hz, 2H), 5.50 (dd, *J* = 10.6, 5.3 Hz, 1H), 5.43 (d, 1H), 4.59 (s, 1H), 4.45 (d, *J* = 7.0 Hz, 2H), 4.22 (t, *J* = 6.8 Hz, 1H), 3.26 – 3.12 (m, 1H), 2.19 (q, *J* = 12.0 Hz, 1H).

¹³C NMR (126 MHz, Chloroform-*d*) δ 173.28, 155.86, 148.17, 144.97, 143.53, 141.35, 127.85, 127.13, 126.40, 124.96, 124.20, 120.07, 77.09, 67.47, 51.99, 47.04, 38.51. **HRMS (ESI):** [M-H]⁻ calcd. for $C_{25}H_{22}N_2O_7$, 461.1354; found 461.1363

2.3f - Synthesis of (2*S*,4*S*)-2-amino-4-(3,4-dichlorophenyl)-4-hydroxybutanoic acid
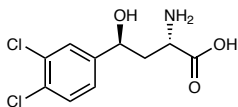
Prepared from **2.2f** using procedure **P1**. **Isolated yield:** 39% **¹H NMR** (400 MHz, DMSO-*d₆*) δ 7.54 (dd, *J* = 4.9, 3.4 Hz, 2H), 7.30 (dd, *J* = 8.4, 2.0 Hz, 1H), 6.21 (s, 2H), 4.73 (t, *J* = 6.2 Hz, 1H), 3.30 (s, 3H), 1.81 – 1.69 (m, 2H).**¹³C NMR** (126 MHz, DMSO) δ 148.34, 131.11, 130.69, 129.24, 127.97, 126.38, 70.47, 53.74, 42.66, 40.51, 40.34, 40.18, 40.01, 39.84, 39.68, 39.51.

Hydration State: $C_{10}H_{11}Cl_2NO_3$ **HRMS (ESI):** [M-H]⁻ calcd. for $C_{10}H_{11}Cl_2NO_3$, 262.0043; found, 262.0045

2.3h - Synthesis of (2*S*,4*S*)-2-amino-4-hydroxy-4-(*o*-tolyl)butanoic acid
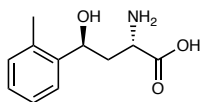
Prepared from **2.2h** using procedure **P1**. **Isolated yield:** 25% **^{1}H NMR** (500 MHz, DMSO-d_6) δ 7.45 (d, J = 7.6 Hz, 1H), 7.15 – 7.09 (m, 1H), 7.05 (d, J = 4.1 Hz, 2H), 4.88 – 4.83 (m, 1H), 3.17 (s, 1H), 2.25 (s, 3H), 1.60 (s, 2H). **^{13}C NMR** (126 MHz, DMSO) δ 145.33, 135.17, 133.42, 129.63, 125.75, 125.43, 125.37, 69.20, 44.02, 40.02, 39.85, 39.69, 39.52, 39.35, 39.19, 39.02, 18.60. **Hydration State:** $C_{11}H_{15}NO_3 \cdot$ 5 H_2O **HRMS (ESI):** [M-H]$^-$ calcd. for $C_{11}H_{15}NO_3$, 208.0979; found, 208.0976

2.5j **-Synthesis of (2S,4S)-2-(((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-hydroxy-4-(1H-imidazol-4-yl)butanoic acid**
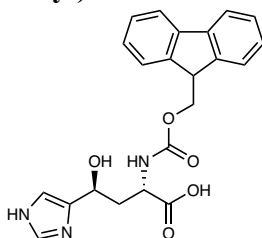
Prepared from **2.2j** using **P2**. **Isolated Yield:** 17% **^{1}H NMR** (500 MHz, Methanol-d_4) δ 7.98 (s, 1H), 7.79 (d, J = 7.5 Hz, 2H), 7.68 (dd, J = 10.5, 7.4 Hz, 2H), 7.38 (td, 2H), 7.31 (tt, J = 7.4, 1.3 Hz, 2H), 7.12 (s, 1H), 4.83 (d, J = 3.7 Hz, 1H), 4.43 – 4.29 (m, 3H), 4.24 (t, J = 7.0 Hz, 1H), 2.33 – 2.22 (m, 1H), 2.08 (ddd, J = 13.5, 9.1, 3.6 Hz, 1H). **^{13}C NMR** (500 MHz, MeOD) δ 177.74, 158.66, 145.44, 142.57, 135.77, 128.76, 128.16, 126.31, 126.26, 120.89, 117.09, 67.99, 65.08, 54.20, 44.92, 41.15. **HRMS (ESI):** [M+H]$^+$ calcd. for $C_{22}H_{21}N_3O_5$, 408.1554; found, 408.1551

***2.6o* -Synthesis of 4-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-5-oxotetrahydrofuran-2-carboxylic acid**
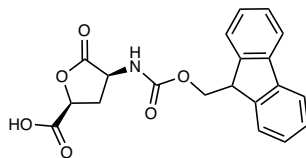
Prepared from **2.2o** using **P3**. **Isolated Yield:** 49% **¹H NMR** (500 MHz, Acetonitrile-d_3) δ 7.84 (d, J = 7.5 Hz, 2H), 7.65 (d, J = 7.5 Hz, 2H), 7.42 (t, J = 7.4 Hz, 2H), 7.34 (t, J = 7.5 Hz, 2H), 4.86 (dd, J = 10.4, 7.0 Hz, 1H), 4.46 (t, J = 10.0 Hz, 1H), 4.42 – 4.33 (m, 2H), 4.25 (t, J = 6.9 Hz, 1H), 2.88 – 2.79 (m, 1H), 2.24 (q, J = 11.4 Hz, 1H). **¹³C NMR** (126 MHz, CD$_3$CN) δ 174.55, 170.53, 156.69, 144.97, 142.13, 128.71, 128.11, 126.08, 120.96, 73.47, 67.52, 51.05, 47.92, 32.26. **HRMS (ESI):** [M-H]⁻ calcd. for C$_{20}$H$_{17}$NO$_6$, 366.09831; found, 366.0986

Large scale 2.3a **- Synthesis of (2*S*,4*S*)-2-amino-4-hydroxy-4-phenylbutanoic acid**
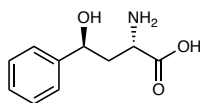
Prepared from **2.2a** using the whole cell **2.3a** procedure above. **Isolated yield:** 47% **¹H NMR** (500 MHz, DMSO-d_6) δ 7.48 – 7.04 (m, 8H), 4.75 (dd, J = 9.9, 2.7 Hz, 1H), 3.44 (t, J = 6.2 Hz, 1H), 1.95 – 1.77 (m, 2H). **¹³C NMR** (126 MHz, DMSO) δ 171.99, 146.32, 127.85, 126.44, 125.60, 125.43, 70.74, 53.01, 41.82, 40.05, 39.89, 39.72, 39.55, 39.38, 39.20, 39.05. **Hydration State:** C$_{10}$H$_{13}$NO$_3$ · 0.5 H$_2$O

Large scale 2.3d **- Synthesis of (2*S*,4*S*)-2-amino-4-hydroxy-4-(4-nitrophenyl)butanoic acid**
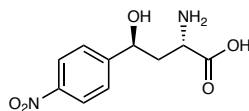
Prepared from **2.2d** using the whole cell **2.3d** procedure above. **Isolated Yield:** 98% **¹H NMR** (400 MHz, Deuterium Oxide:Acetonitrile-d^3) δ 8.72 (d, J = 8.8 Hz, 2H), 8.12 (d, J = 8.8 Hz, 2H), 5.50 (t, J = 6.6 Hz, 1H), 4.34 (t, J = 5.5 Hz, 1H), 2.67 (dd, J = 7.2, 5.0 Hz, 2H). **¹³C NMR**

(101 MHz, D$_2$O:Acetonitrile-d^3) δ 174.65, 152.60, 148.06, 127.67, 124.63, 119.45, 70.96, 53.83, 39.15, 1.32. **Hydration State:** C$_{10}$H$_{12}$N$_2$O$_5$ · 2 H$_2$O **HRMS (ESI):** [M-H]$^-$ calcd. for C$_{10}$H$_{12}$N$_2$O$_5$ 239.0673; found 239.0675

(2R,4S)-2-amino-5,5,5-trifluoro-4-hydroxy-4-methylpentanoic acid
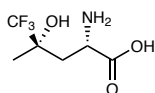
Prepared from 1,1,1-trifluoroacetone using Marfey's derivatization procedure. **HRMS (ESI):** [M+H]$^+$ calcd. for C$_6$H$_{10}$F$_3$NO$_3$, 202.0686; found, 202.0689

NMR Spectra

2.3a

^{1}H NMR spectrum in D$_2$O:ACN-d^3 at 400 MHz

^{13}C NMR spectrum in D$_2$O:ACN-d_3 at 126 MHz

2.3a- wt-UstD

^{1}H NMR spectrum in D$_2$O at 500 MHz

^{13}C NMR spectrum in D$_2$O at 126 MHz

2.5b

^{1}H NMR spectrum in CDCl$_3$ at 500 MHz

^{13}C APT NMR spectrum in CDCl$_3$ at 126 MHz

2.5c

^{1}H NMR spectrum in CDCl$_3$ at 500 MHz

^{13}C NMR spectrum in CDCl$_3$ at 126 MHz

2.5d

^{1}H NMR spectrum in CDCl$_3$ at 500 MHz

^{13}C APT NMR spectrum in CDCl$_3$ at 126 MHz

2.3f
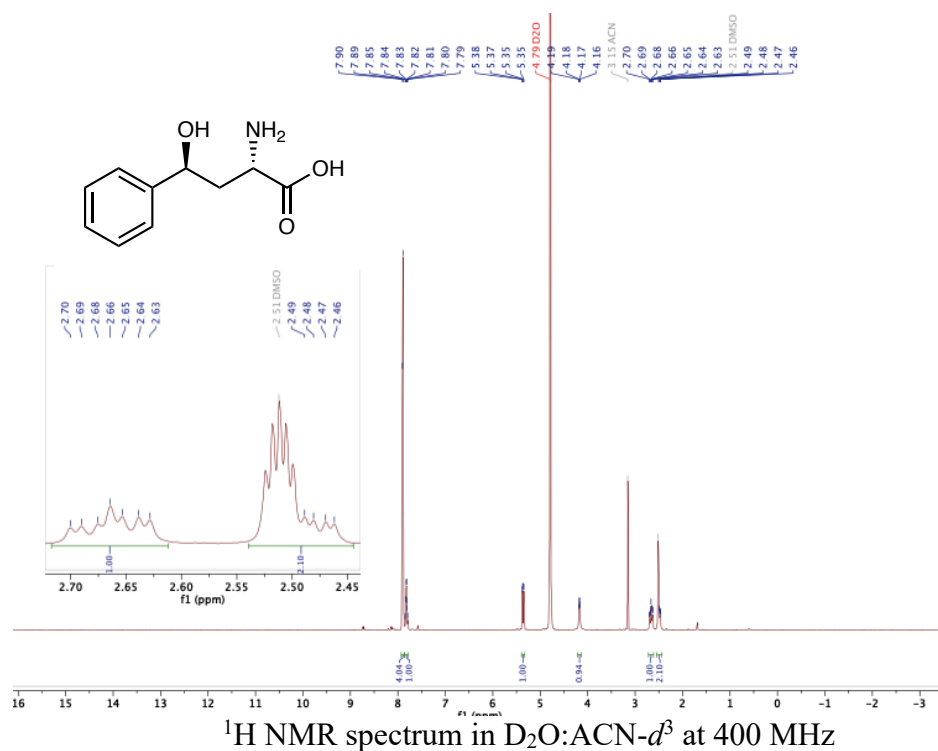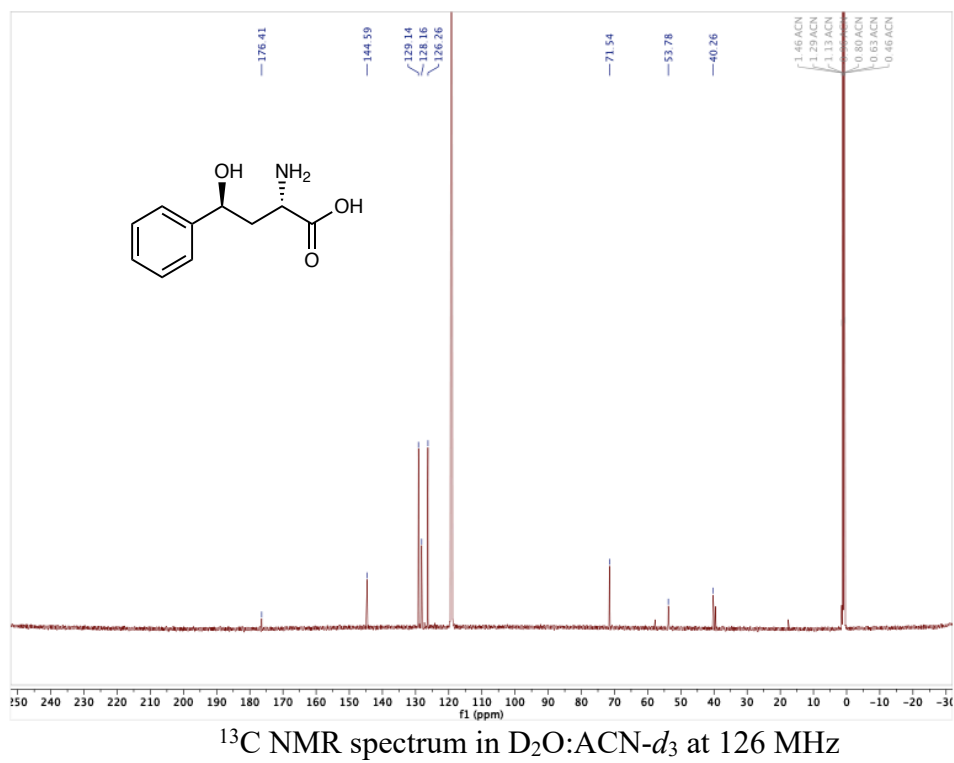
^{1}H NMR spectrum in DMSO-d_6 at 400 MHz
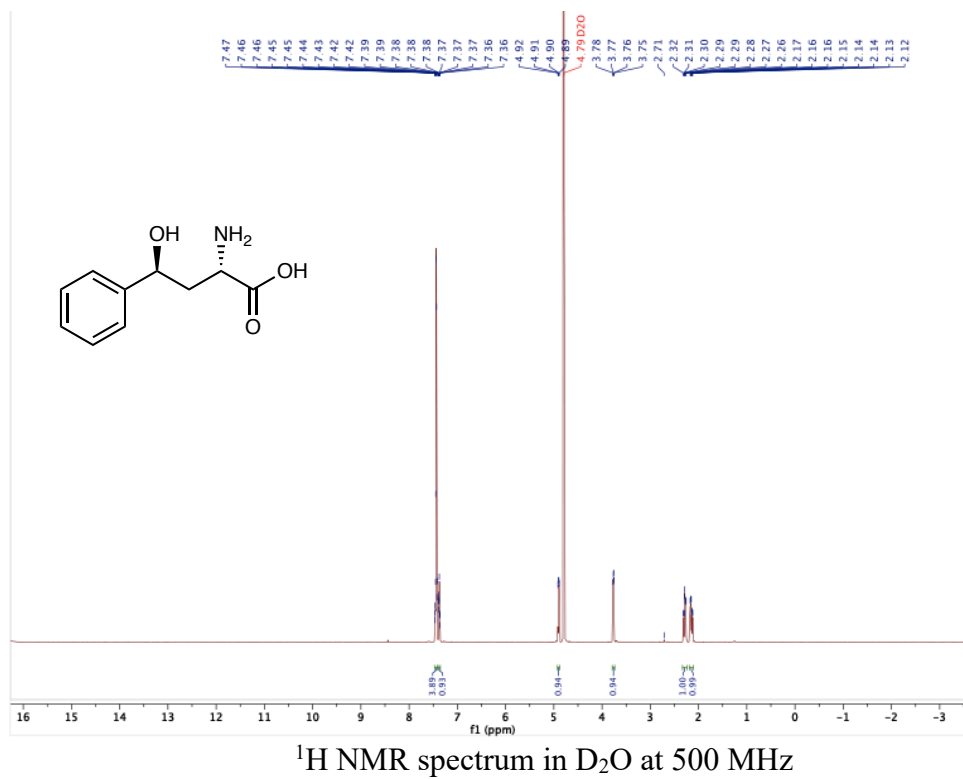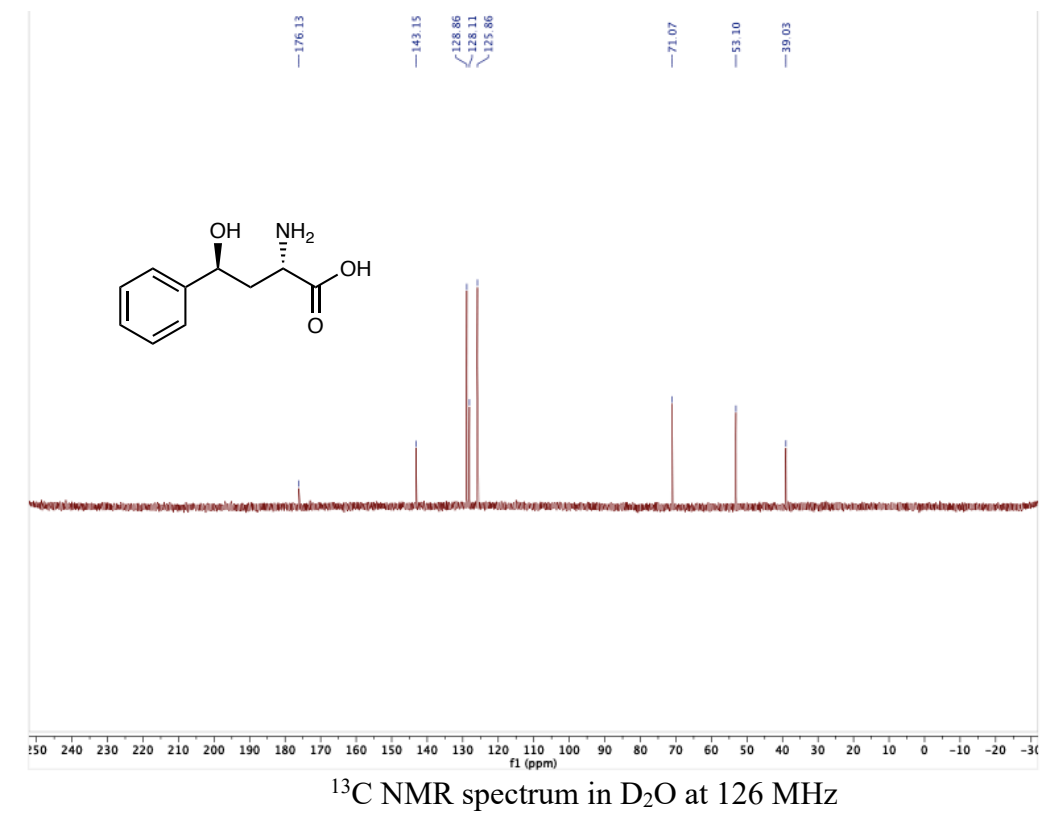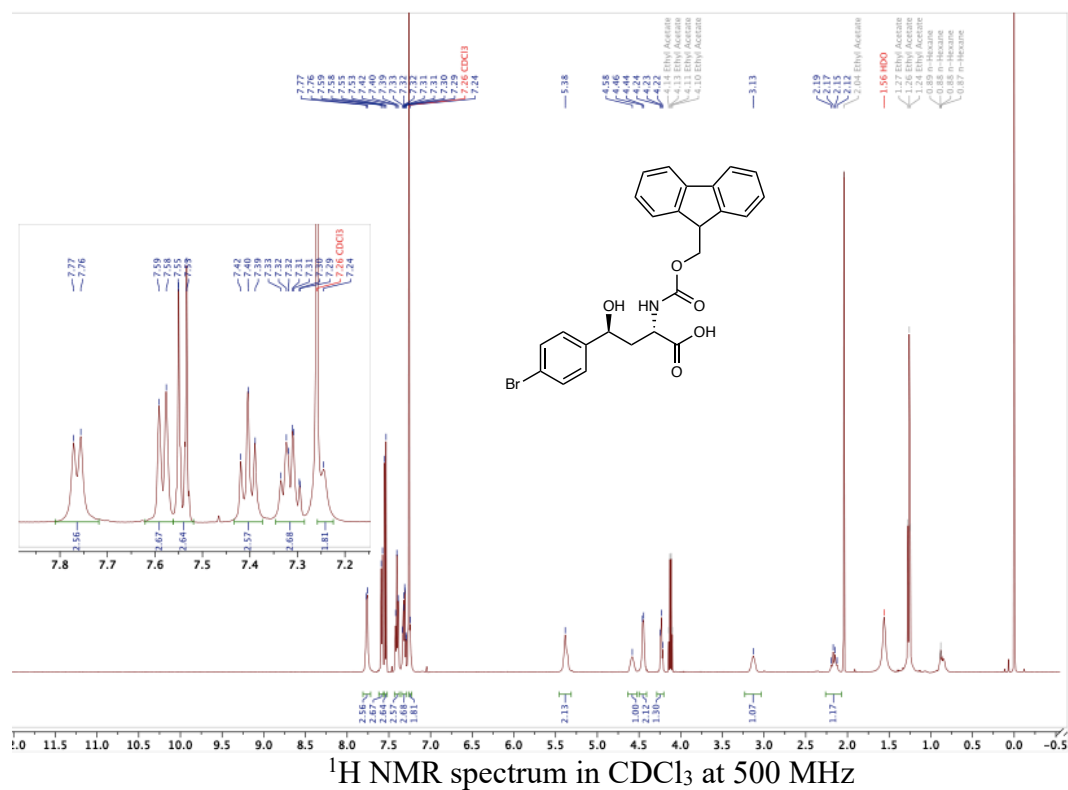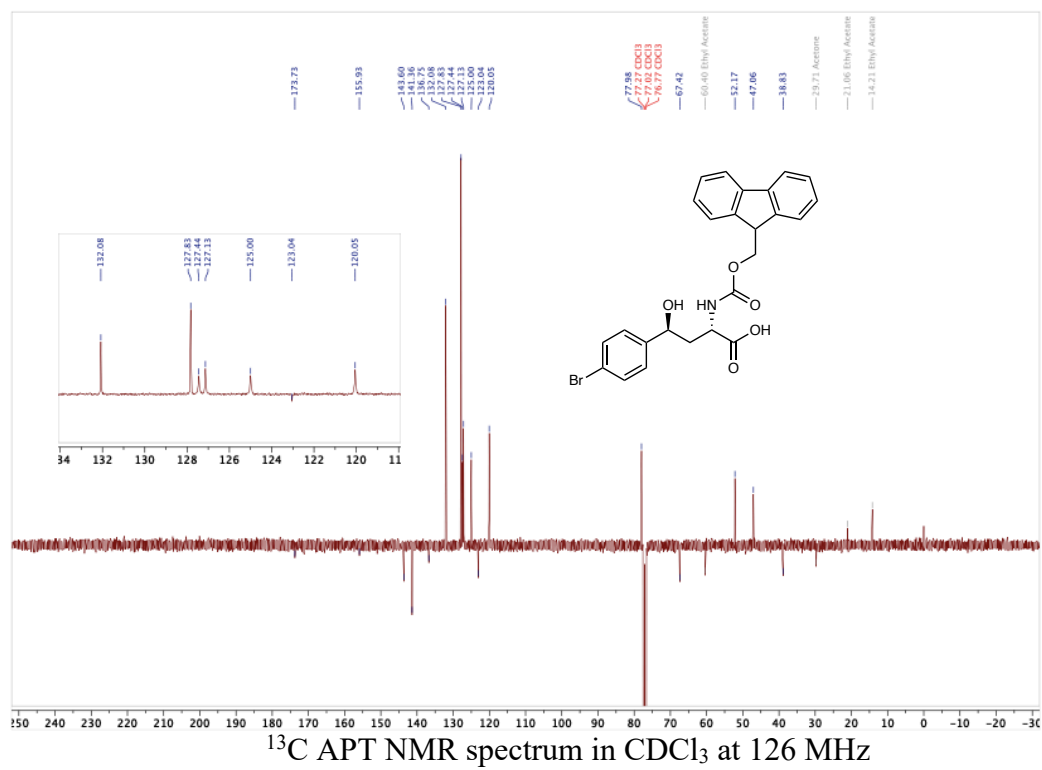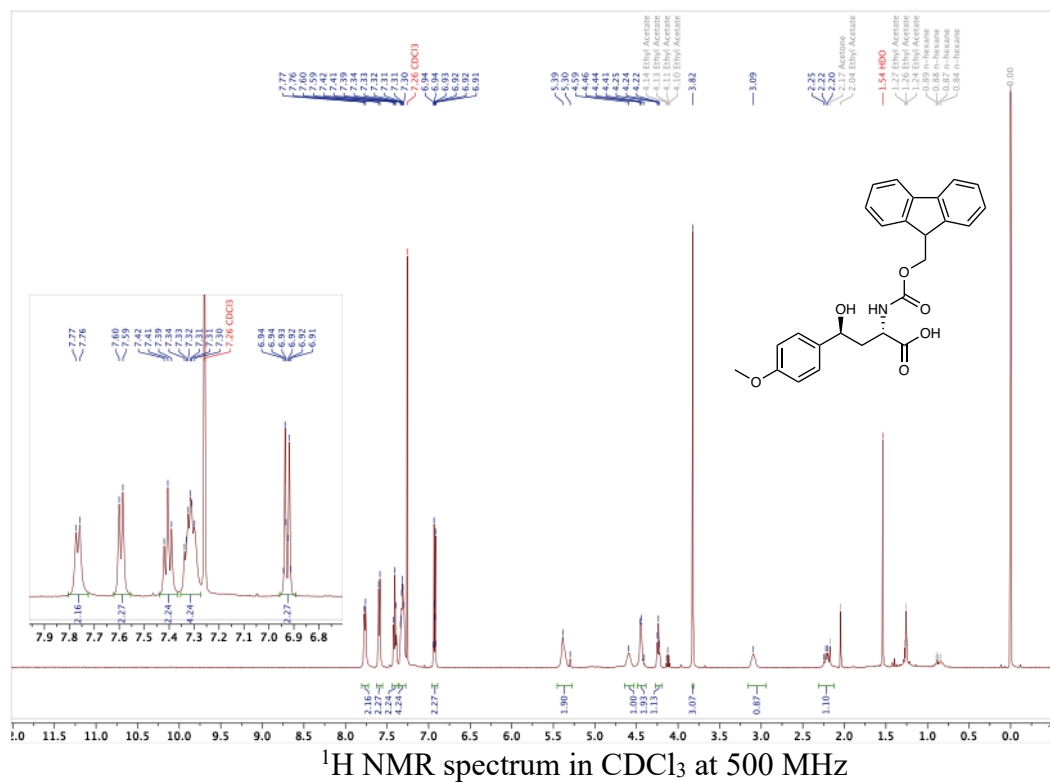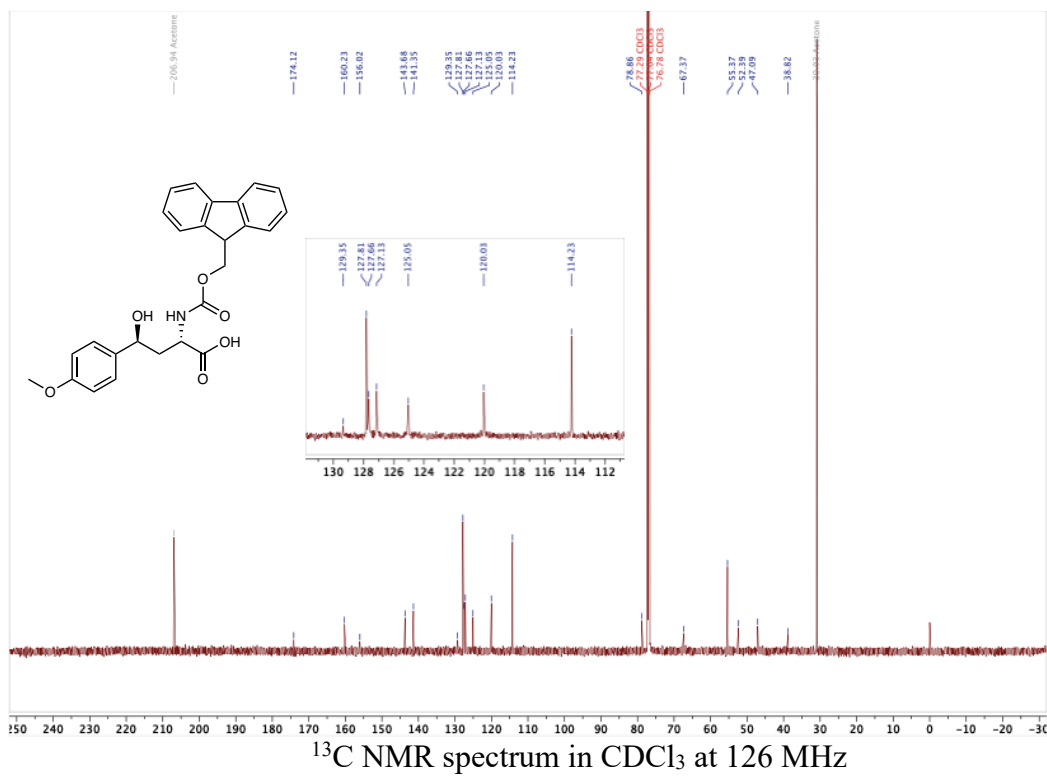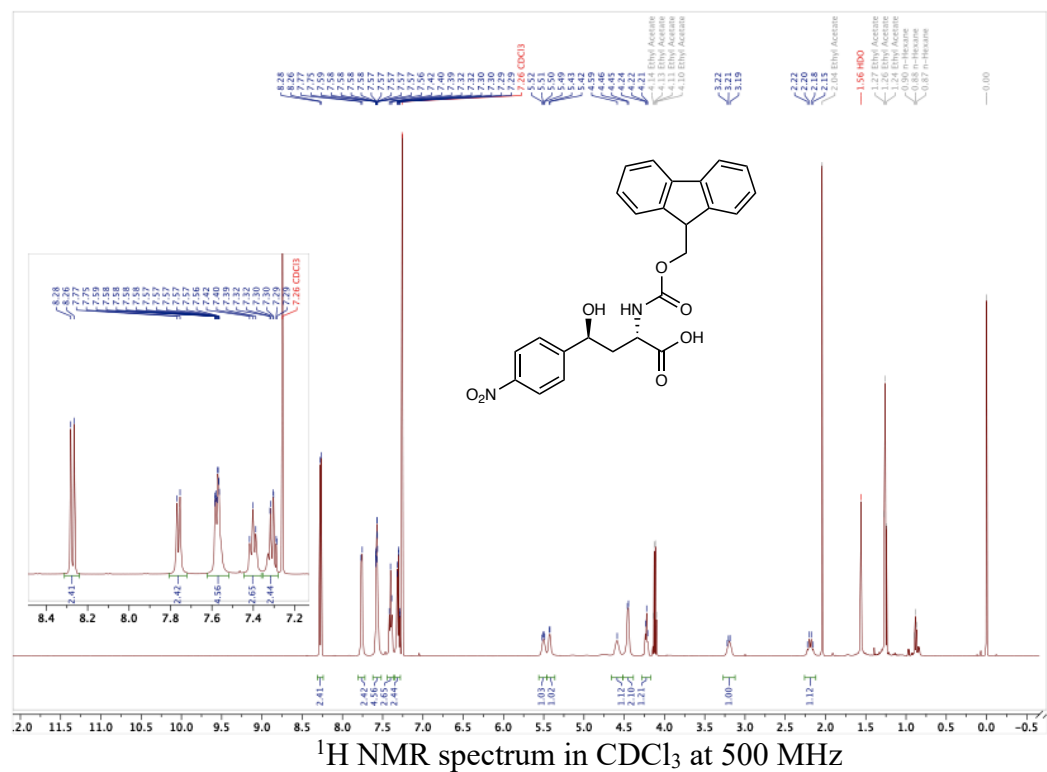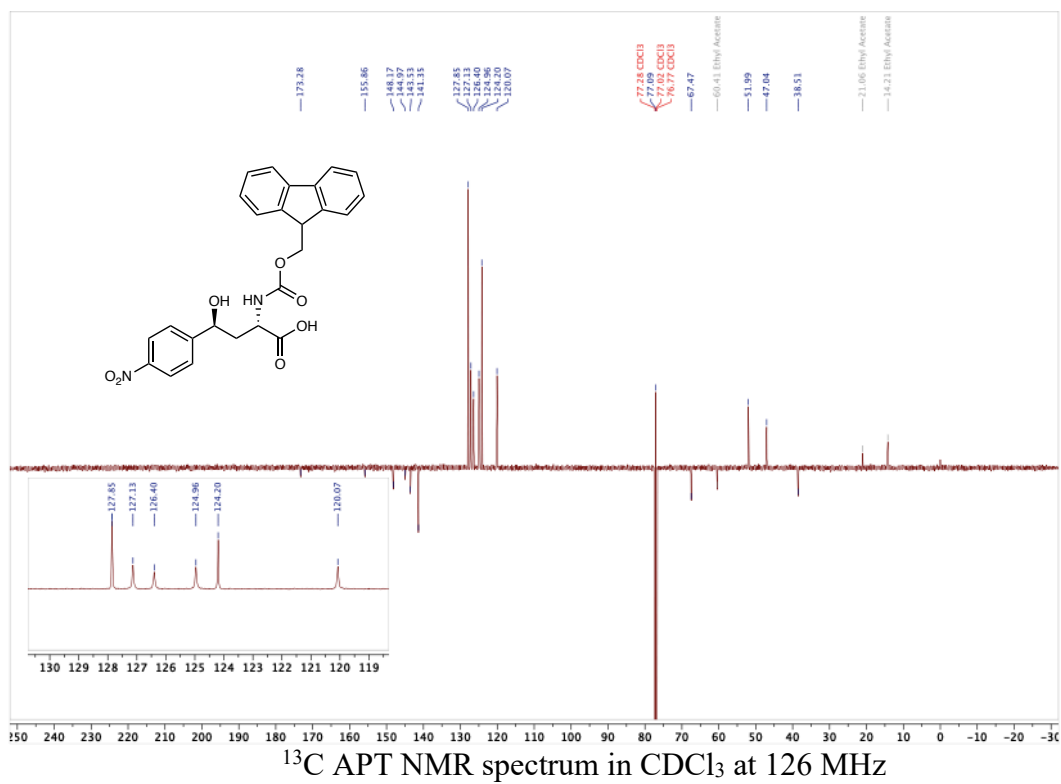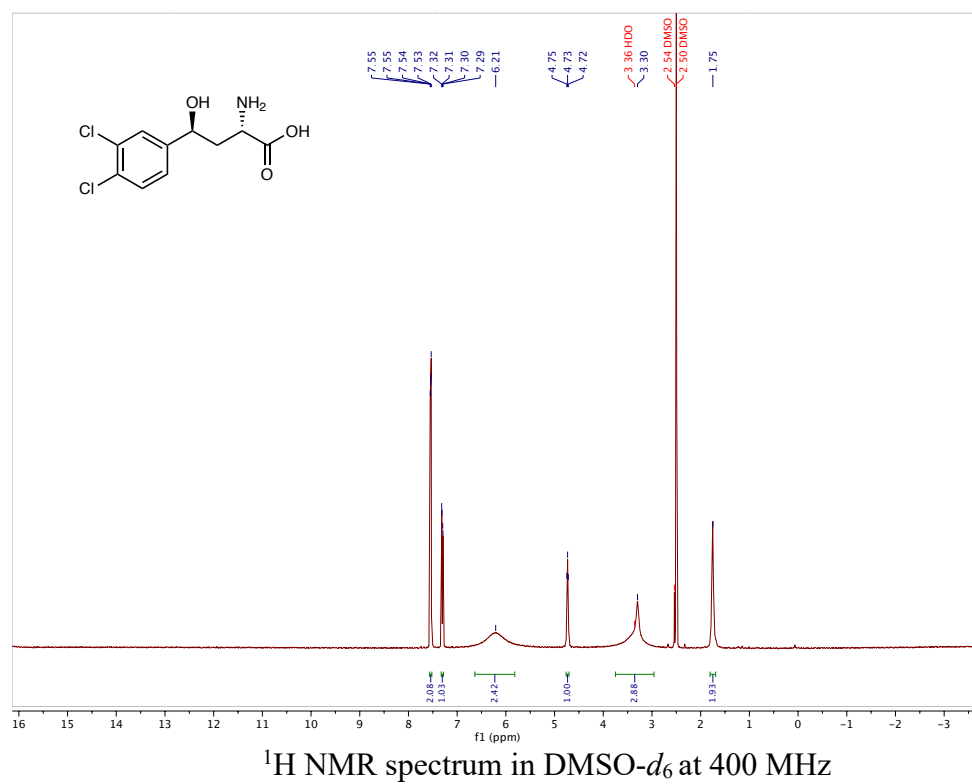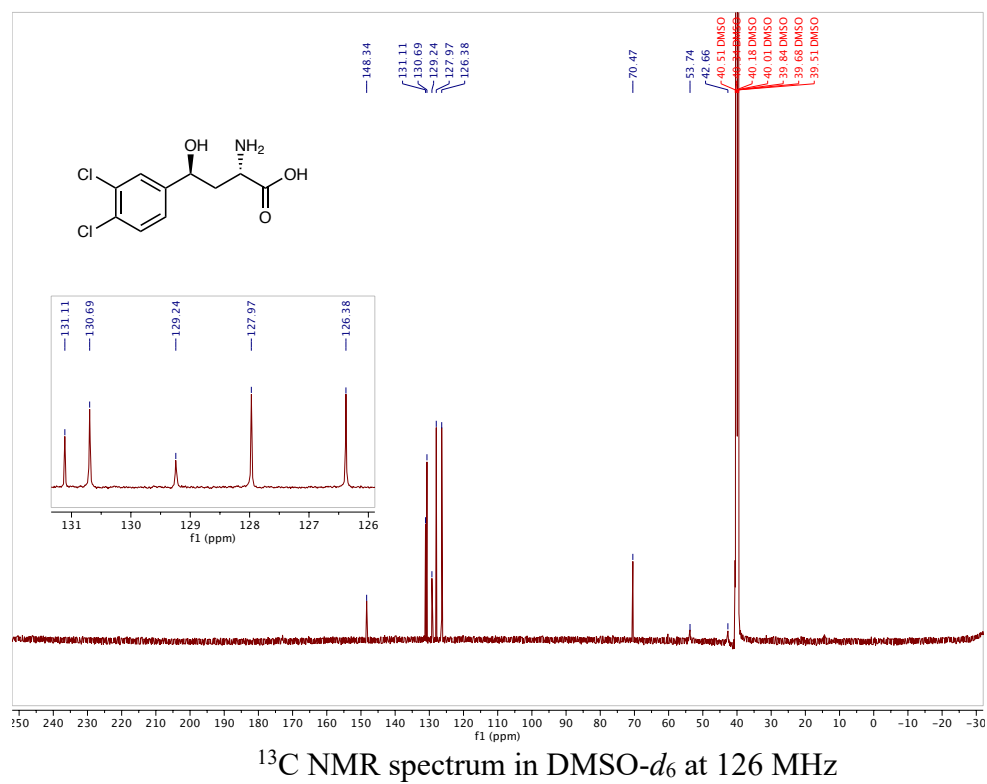
^{13}C NMR spectrum in DMSO-d_6 at 126 MHz

2.3h

^{1}H NMR spectrum in DMSO-d_6 at 500 MHz

^{13}C NMR spectrum in DMSO-d_6 at 126 MHz

2.5j

^{1}H NMR spectrum in MeOD at 500 MHz

^{13}C NMR spectrum in MeOD at 126 MHz

2.6o

^{1}H NMR spectrum in ACN-d^3 at 500 MHz

^{13}C NMR spectrum in ACN-d^3 at 126 MHz

Large scale-2.3a

^{1}H NMR spectrum in DMSO-d_6 at 500 MHz

^{13}C APT NMR spectrum in DMSO-d_6 at 126 MHz

Large scale-2.3d

^{1}H NMR spectrum in D$_2$O:ACN-d_3 at 400 MHz

^{13}C APT NMR spectrum in D$_2$O:ACN-d_3 at 101 MHz
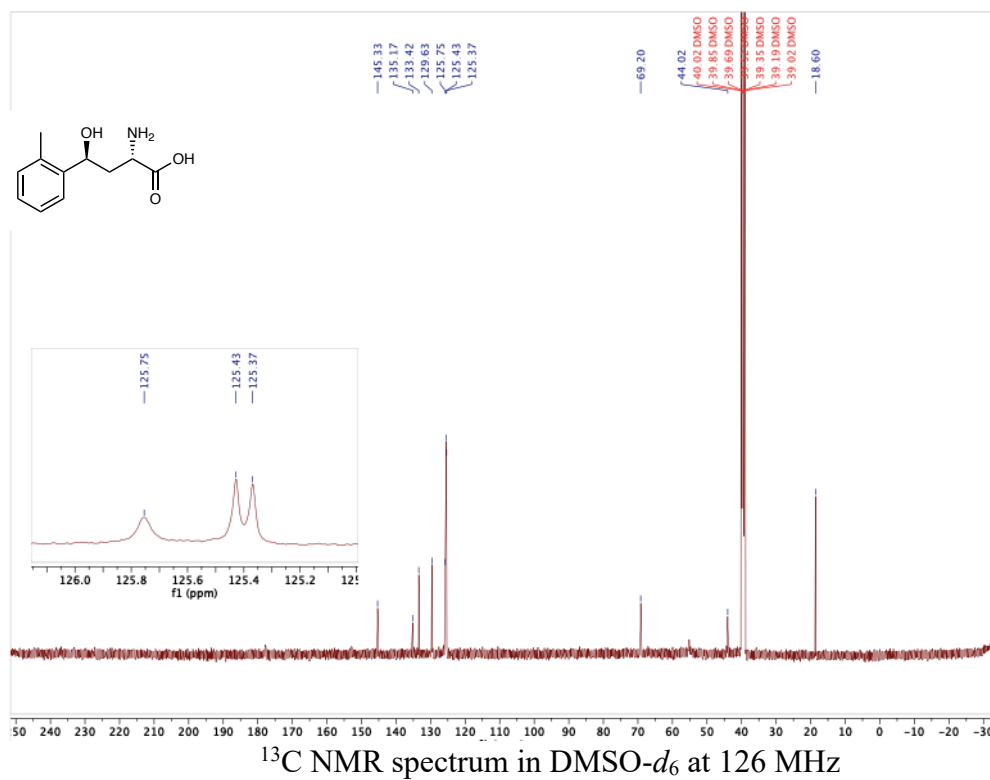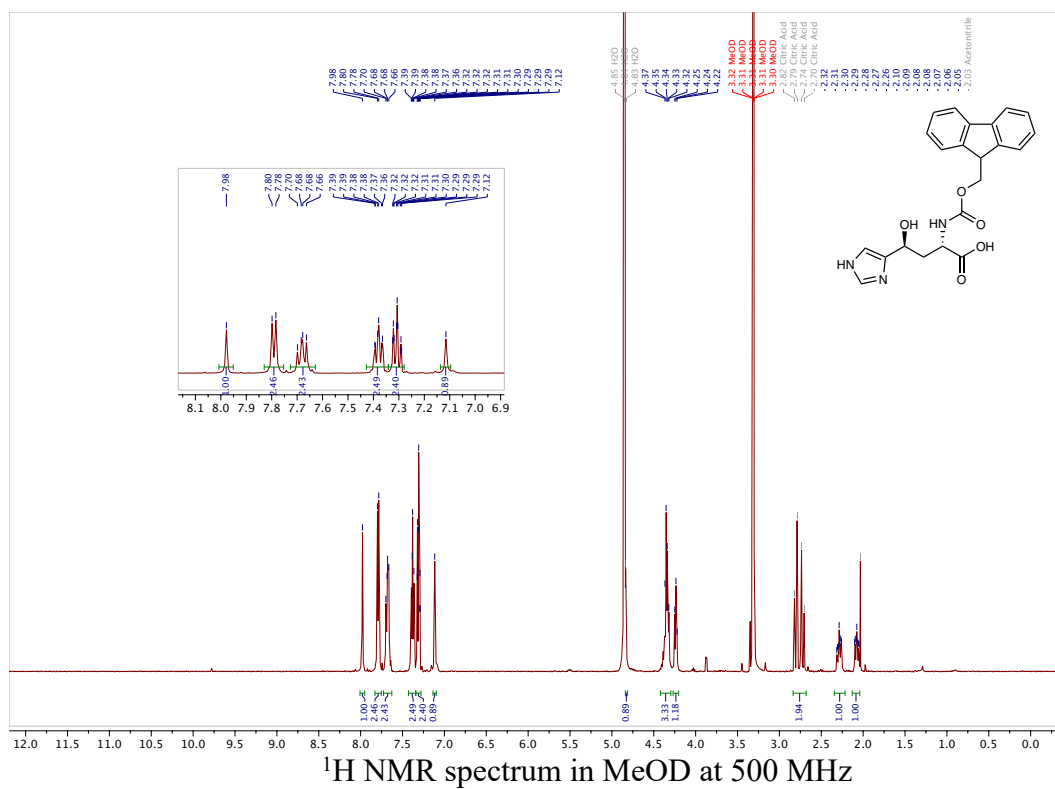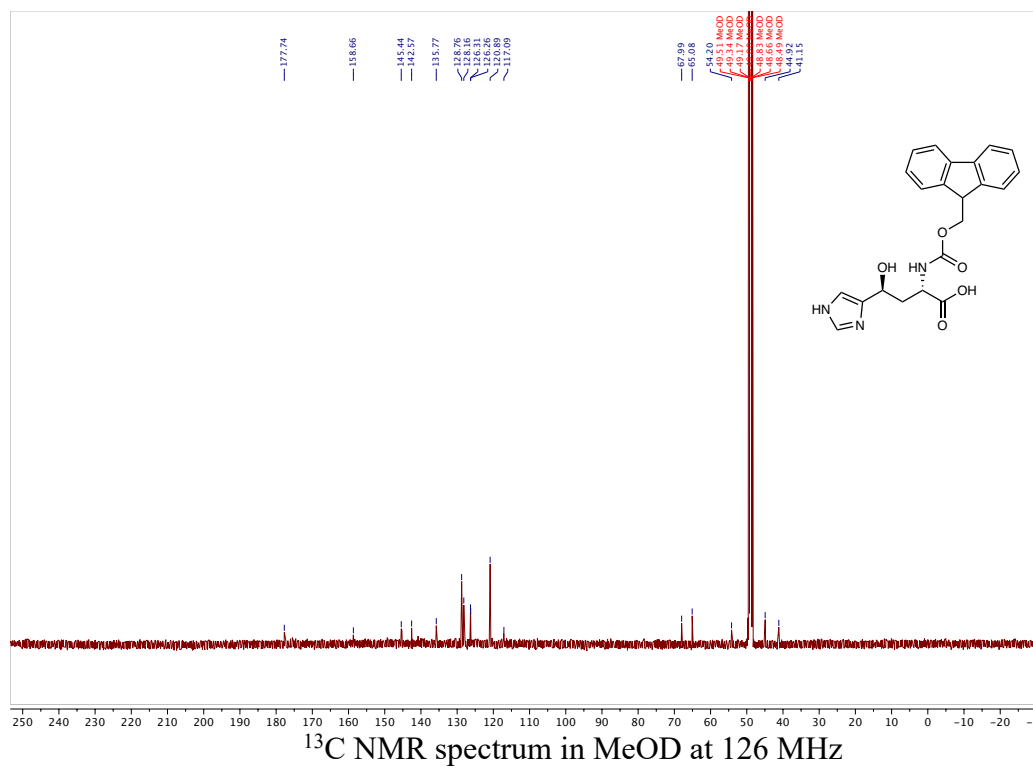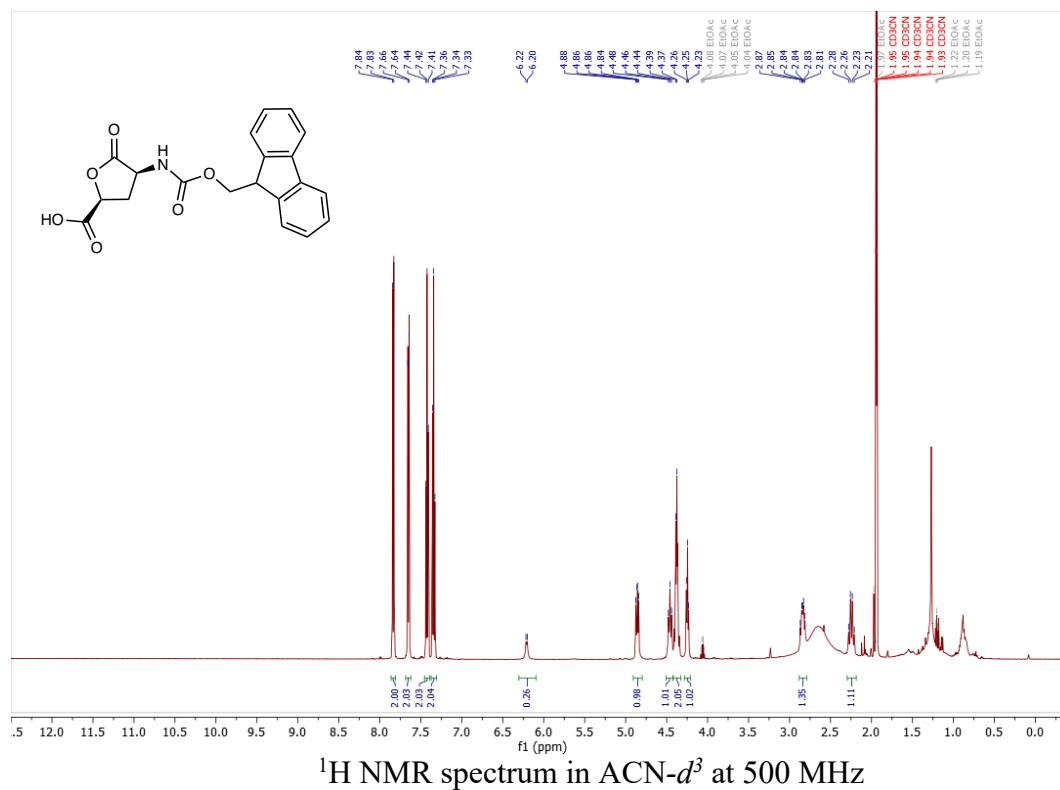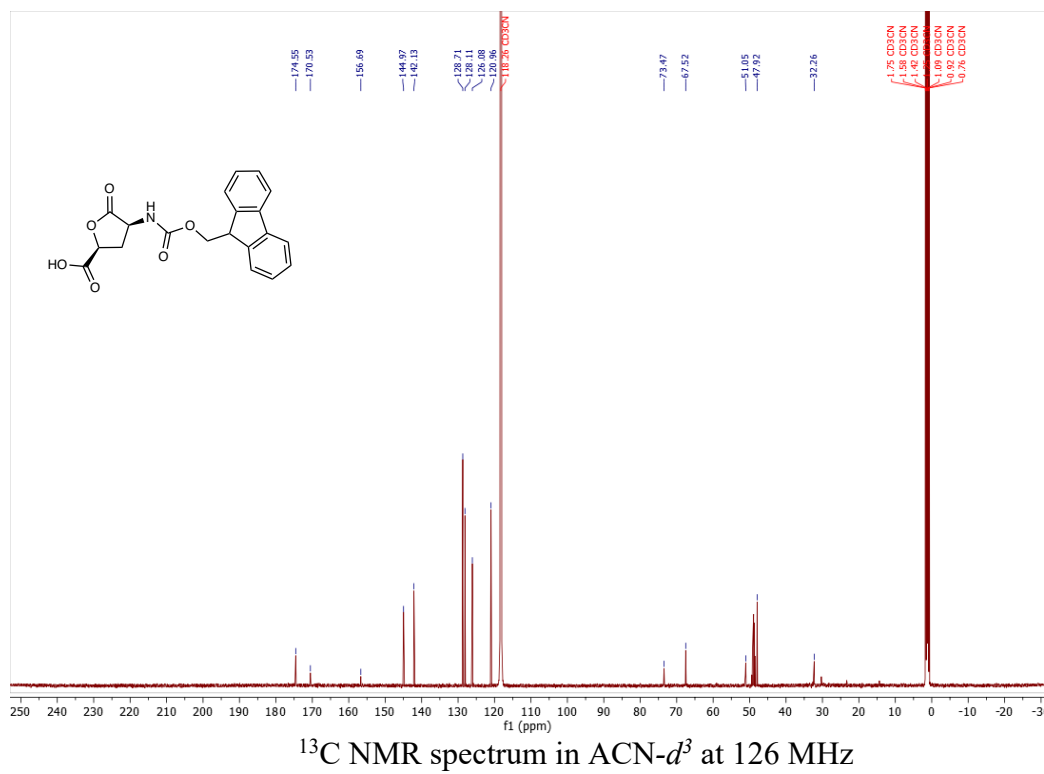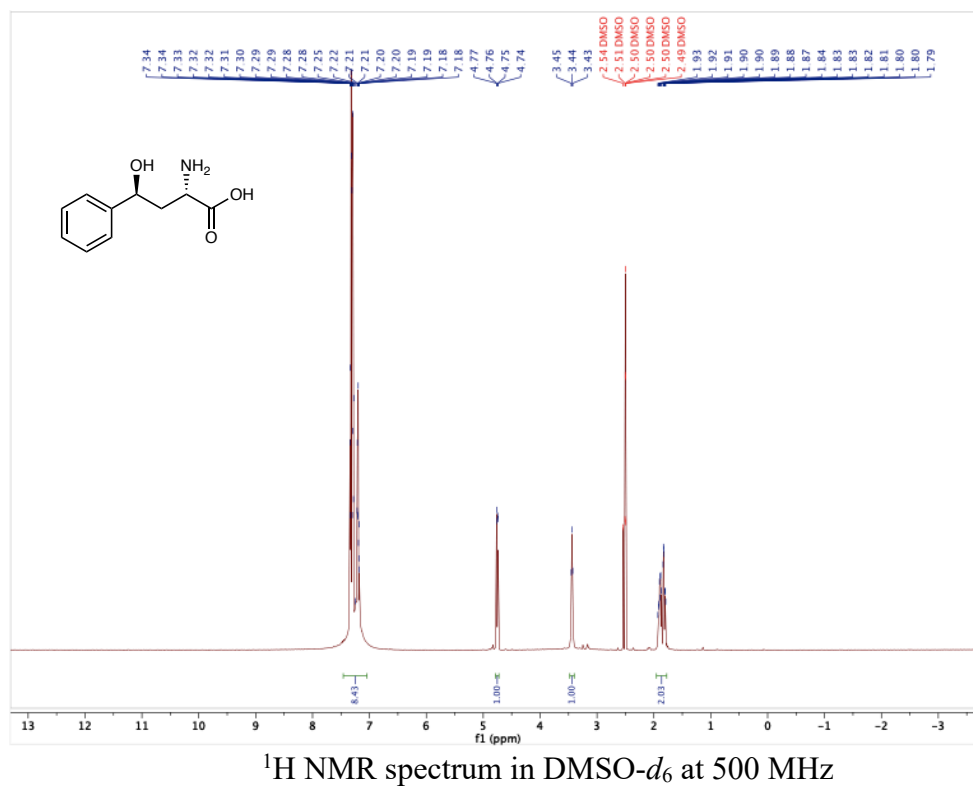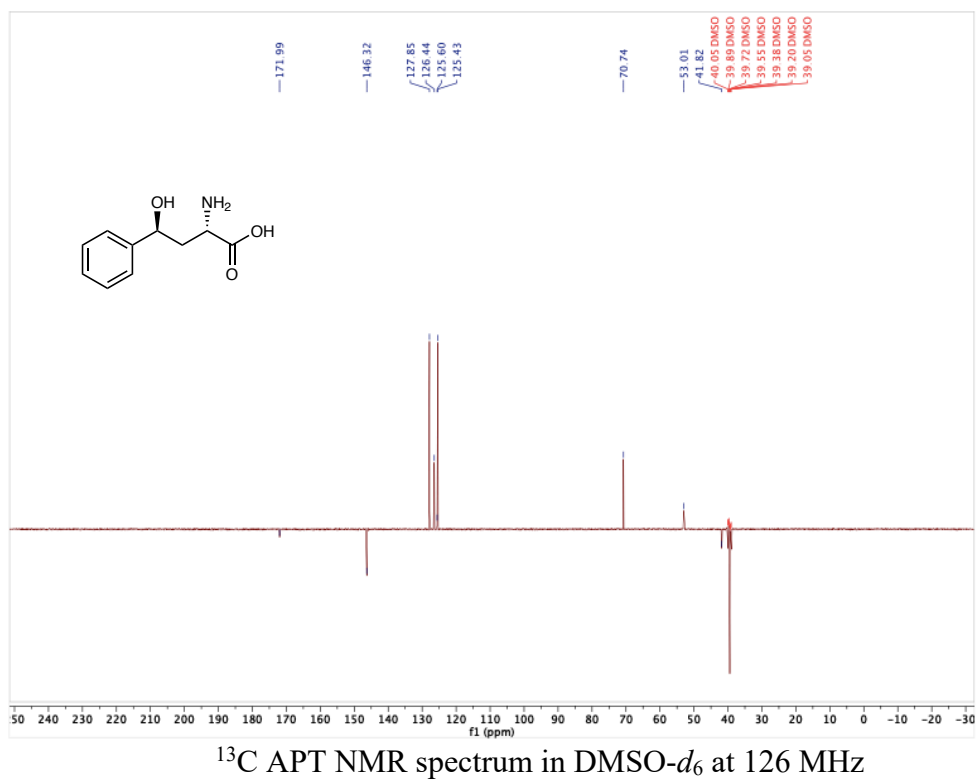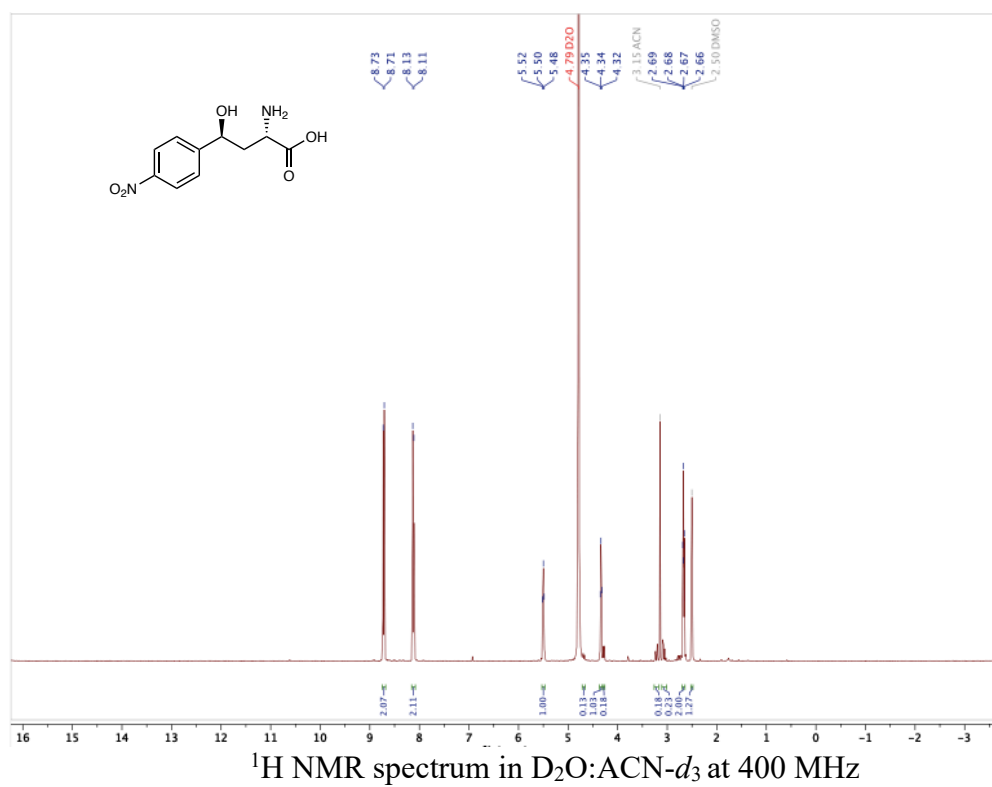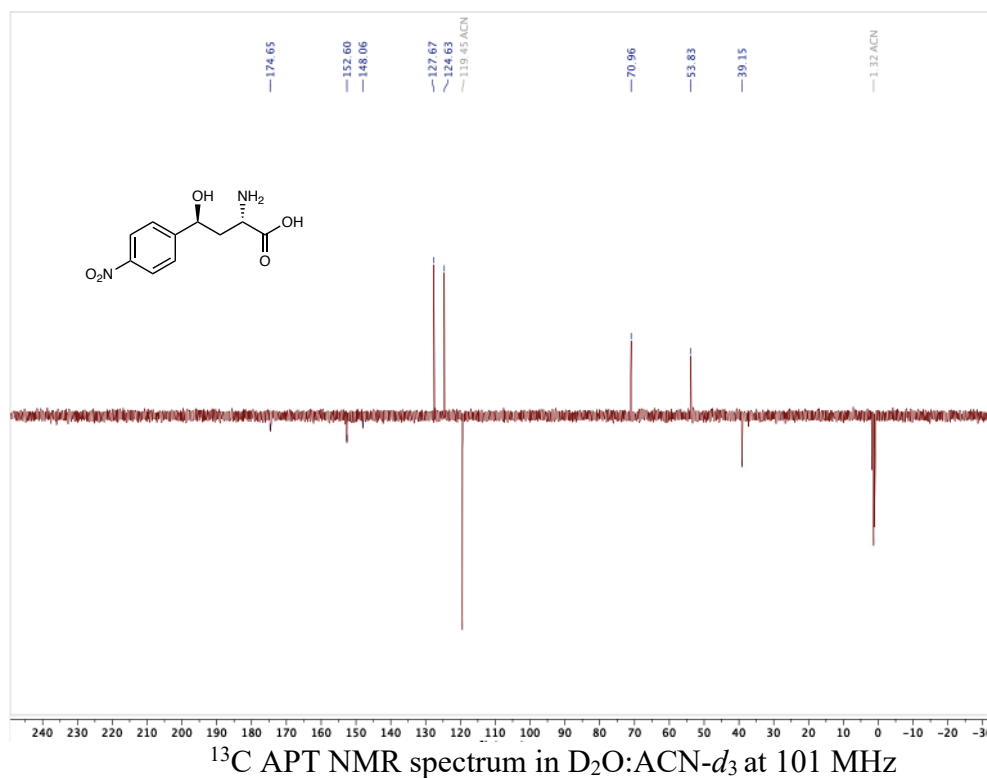
2.5 References

1.   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).

2.   Zhang, X. *et al.* Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science.* **369**, 799–806 (2020).

3.   Brown, D. G. & Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **59**, 4443–4458 (2016).

4.   Fesko, K. & Gruber-Khadjawi, M. Biocatalytic Methods for C-C Bond Formation. *ChemCatChem* **5**, 1248–1272 (2013).

5.   Schmidt, N. G., Eger, E. & Kroutil, W. Building Bridges: Biocatalytic C-C-Bond

Formation toward Multifunctional Products. *ACS Catalysis* **6**, 4286–4311 (2016).

6.   Fujii, I. Heterologous expression systems for polyketide synthases. *Nat. Prod. Rep.* **26**, 155–169 (2009).

7.   Heine, A. *et al.* Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science.* **294**, 369–374 (2001).

8.   Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochromeP450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chemie - Int. Ed.* **53**, 6810–6813 (2014).

9.   Berkeš, D., Kolarovič, A., Manduch, R., Baran, P. & Považanec, F. Crystallization-induced asymmetric transformations (CIAT): Stereoconvergent acid-catalyzed lactonization of substituted 2-amino-4-aryl-4-hydroxybutanoic acids. *Tetrahedron Asymmetry* **16**, 1927–1934 (2005).

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

11.   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).

12.   Zetzsche, L. E. & Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nat. Rev. Chem.* **4**, 334–346 (2020).

13.   Ye, Y. *et al.* Unveiling the Biosynthetic Pathway of the Ribosomally Synthesized and Post-translationally Modified Peptide Ustiloxin B in Filamentous Fungi. *Angew. Chemie Int. Ed.* **55**, 8072–8075 (2016).

14.   Prier, C. K. & Arnold, F. H. Chemomimetic Biocatalysis: Exploiting the Synthetic
        Potential of Cofactor-Dependent Enzymes to Create New Catalysts. *J. Am. Chem. Soc.*
        137, 13992–14006 (2015).

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

16.   Marsden, S. R., Gjonaj, L., Eustace, S. J. & Hanefeld, U. Separating Thermodynamics
        from Kinetics—A New Understanding of the Transketolase Reaction. *ChemCatChem* **9**,
        1808–1814 (2017).

17.   Ariza, J., Font, J. & Ortuño, R. M. An efficient and concise entry to (-)-4,5-dihydroxy-d-
        threo-l-norvaline. Formal synthesis of clavalanine. *Tetrahedron Lett.* **32**, 1979–1982
        (1991).

18.   Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *J. Med. Chem.* **59**,
        10807–10836 (2016).

19.   Moreno, C. J. *et al.* Synthesis of γ-Hydroxy-α-amino Acid Derivatives by Enzymatic
        Tandem Aldol Addition–Transamination Reactions. *ACS Catal.* **11**, 4660–4669 (2021).

20.   Hernandez, K. *et al.* Combining Aldolases and Transaminases for the Synthesis of 2-
        Amino-4-hydroxybutanoic Acid. (2017). doi:10.1021/acscatal.6b03181

21.   Vargas-Rodriguez, O., Sevostyanova, A., Söll, D. & Crnković, A. Upgrading aminoacyl-
        tRNA synthetases for genetic code expansion. *Current Opinion in Chemical Biology* **46**,
        115–122 (2018).

22.   Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid
        biosynthesis. *Nature* **567**, 420–424 (2019).

23.   Yang, J. *et al.* The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2014).

24.   Ho, T. H. *et al.* Catalytic Intermediate Crystal Structures of Cysteine Desulfurase from the Archaeon Thermococcus onnurineus NA1. *Archaea* **2017**, 1–11 (2017).

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

26.   Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y. & Bocola, M. Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: Rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **132**, 9144–9152 (2010).

27.   Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

28.   Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chemie Int. Ed.* **44**, 4192–4196 (2005).

29.   Romney, D. K., Sarai, N. S. & Arnold, F. H. Nitroalkanes as Versatile Nucleophiles for Enzymatic Synthesis of Noncanonical Amino Acids. *ACS Catal.* **9**, 8726–8730 (2019).

30.   Marfey, P. DETERMINATION OF D-AMINO ACIDS . II . USE OF A BIFUNCTIONAL REAGENT, 1,5-DIFLUORO-2,4-DINITROBENZENE. *Carlsb. Res. Commun.* **49**, 591–596 (1984).

31.   Wu, G. *et al.* Proline and hydroxyproline metabolism: Implications for animal and human nutrition. *Amino Acids* **40**, 1053–1063 (2011).

32.   MÜLLER, J.-C., TOOME, V., PRUESS, D. L., BLOUNT, J. F. & WEIGELE, M. Ro 22-

5417, a new clavam antibiotic from Streptomyces clavuligerus. III. Absolute stereochemistry. *J. Antibiot. (Tokyo).* **36**, 217–225 (1983).

33. Wahab, R. A., Elias, N., Abdullah, F. & Ghoshal, S. K. On the taught new tricks of enzymes immobilization: An all-inclusive overview. *Reactive and Functional Polymers* **152**, 104613 (2020).

34. 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).

35. Al-Ayyoubi, M., Gettins, P. G. W. & Volz, K. Crystal structure of human maspin, a serpin with antitumor properties: Reactive center loop of maspin is exposed but constrained. *J. Biol. Chem.* **279**, 55540–55544 (2004).

36. Fox, R. Directed molecular evolution by machine learning and the influence of nonlinear interactions. *J. Theor. Biol.* **234**, 187–199 (2005).

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

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

39. Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E. & Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).

40. Boville, C. E. *et al.* Engineered Biosynthesis of β-Alkyl Tryptophan Analogues. *Angew. Chemie - Int. Ed.* **57**, 14764–14768 (2018).

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

42.   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).

43.   Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).

44.   Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).

45.   Gerlt, J. A. Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **56**, 4293–4308 (2017).

46.   Zallot, R., Oberg, N. O. & Gerlt, J. A. 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Op. Chem. Biol.* **47**, 77–85 (2018).

47.   Zallot, R., Oberg, N. & Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169–4182 (2019).

48.   Gerlt, J. A. *et al.* Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **1854**, 1019–1037 (2015).

49.   Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

50.   Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual*. (CreateSpace, 2009).

51. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in {P}ython. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

52. Kabsch, W. & IUCr. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

53. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).

54. Pannu, N. S. *et al.* Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 331–337 (2011).

55. Sheldrick, G. M. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography* **64**, 112–122 (2008).

56. Pannu, N. S. & Read, R. J. The application of multivariate statistical techniques improves single-wavelength anomalous diffraction phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 22–27 (2004).

57. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1ATPase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52**, 30–42 (1996).

58. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011 (2006).

59. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

60. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 122–133 (2001).

CHAPTER THREE: PROMISCUITY-GUIDED ENGINEERING OF A

DECARBOXYLATIVE ALDOLASE

3.1 Introduction

3.1.1 Limitations of traditional protein engineering for identifying generalists

Directed evolution has found widespread use for developing biocatalysts that are applied in a myriad of industrial settings, including pharmaceuticals, fine chemicals, and bioremediation.[1–4] The power of directed evolution lies in the iterative cycles of genetic diversification and selection, allowing researchers to identify mutations that enhance enzyme activity, either by increasing catalyst concentration, stability, or activity under researcher-defined conditions.[5,6] The rapid identification of stabilizing mutations that do not compromise function still requires trial and error, but recent advances in computational design and machine learning are making rapid strides in this area.[7–9] In contrast, strategies to identify mutations that alter catalytic activity are more limited.[10,11] Mutagenesis of active site residues is a reliable strategy, but direct identification of distal sites that influence reactivity is a long-standing challenge in protein engineering because it requires in depth structural and mechanistic information.[12,13]

In standard directed evolution workflows, a single model substrate is typically chosen for screening even when synthetic utility is the goal for enzyme evolution. While this approach is often quite successful, there are also many cases where directed evolution yields a catalyst that has high activity for the substrate under selection but struggles to react with substrate analogs (low substrate promiscuity).[14–16] In some cases, intermediates along a directed evolution lineage are more promiscuous and evolution inadvertently limited the scope of the transformation.[14–16] More broadly, the inability to efficiently track substrate promiscuity during the initial screening phase hinders the ability to engineer enzymes toward generality.

3.1.2 Substrate multiplexed screening (SUMS) as an alternative engineering approach

Recently, we advanced substrate multiplexed screening (SUMS) as an alternative screening strategy for directed evolution to aid in tracking changes to enzyme promiscuity while minimizing researcher intervention and screening time.[17] By screening substrates in competition, total activity and the distribution of the products, representing promiscuity, are measured simultaneously (**Fig. 3.1A**).
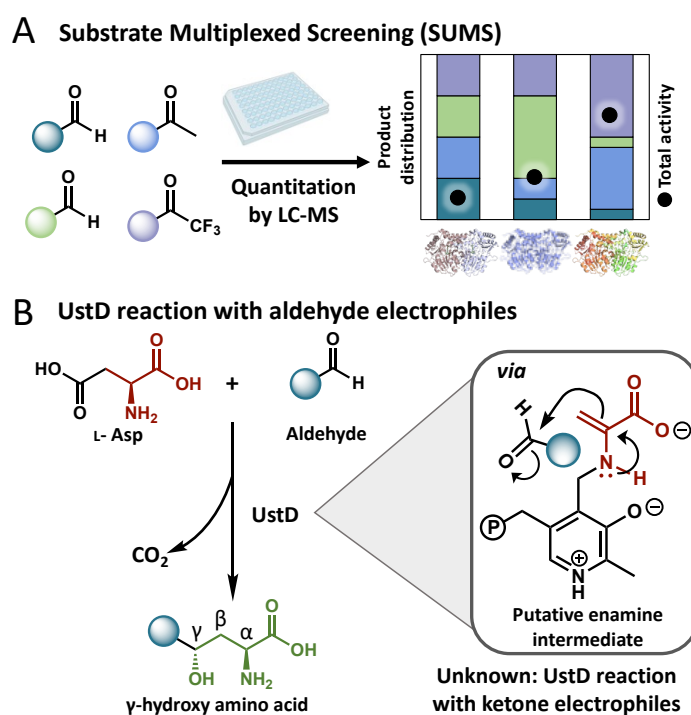
Figure 3.1 Application of SUMS. A. General SUMS procedure. **B.** UstD reaction scheme depicting the native reactivity with aldehyde electrophiles.

This method has been used previously in a few cases to effectively discover large boosts in activity with defined subclasses of substrates using small active site libraries.[18–20] In another case, the Kreis lab used SUMS for multi-generational, iterative active site mutagenesis efforts on adenylate forming enzymes to maintain previously existing promiscuity and to ensure that mutations that might turn on undesired activity are not accumulated during evolution.[21]

Researchers in the Buller lab have reported a single test of the utility of SUMS with globally random mutagenesis.[17] It was observed that mutation at a residue flanking the active site of the fold type II PLP dependent enzyme decreased activity but altered promiscuity. Subsequent site-saturation mutagenesis at this position then lead to a variant with a general boost in activity, including with substrates outside the substrate mixture.[17] This intriguing result prompted us to explore how simultaneous promiscuity and activity measurements could be leveraged to engineer for generality beyond simplifying screening efforts.

Many uncertainties must be addressed to develop SUMS into a broadly applicable protein engineering tool to effectively traverse a fitness landscape. In a complex mixture of substrates, there is no longer a single metric to identify desirable variants. We must determine how best to interpret the rich datasets afforded by SUMS. More significantly, the specific engineering goals with which these methods are assessed should be relevant to the goals of biocatalysis community. Previous implementations of SUMS used model enzymes and reactions that began with exceptional properties for biocatalysis, such as favorable thermodynamics, high speed, expression, and thermal stability.[17] It is common to use such platforms for method development. However, the history of protein engineering is replete with advances in methodology that only operate under contrived settings. We posit that the strengths and limitations of a method would become more apparent when embarking on challenging engineering goals, such as the evolution of a stereoselective C–C bond formation.

3.1.3 Proposed application of SUMS to engineer a biocatalytic aldol addition into ketone electrophiles

Here, we chose to explore the effectiveness of multi-generational, promiscuity-guided engineering in the context of a historically challenging reaction in protic solvents, aldol addition

into ketones. In the previous chapter, we discussed the classical evolution of PLP-dependent decarboxylative aldolase, UstD, to react with benzaldehyde.[7] This is a convergent C–C bond forming reaction yielding a γ-hydroxy non-canonical amino acid product (**Fig 3.1B**). The evolution, serendipitously, resulted in a 'generalist' enzyme, UstD[2.0], capable of reacting with diverse aldehyde electrophiles to generate chiral secondary alcohols. Reactions with ketone electrophiles would yield chiral tertiary alcohols, a highly sought-after motif in medicinal chemistry.[22–24] However, aldol-type reactions with ketones are deceptively challenging compared to aldehydes, as ketones are thermodynamically more stable and kinetically slower to react than aldehydes for both steric and electronic reasons.[25,26] The challenging reactivity in combination with the synthetic relevance of the products makes the evolution of UstD[2.0] for aldol addition into ketones an excellent testing ground for exploring the SUMS method.

3.2 Practical considerations for implementing SUMS

To successfully implement SUMS, we considered the effect of several parameters, notably the composition of the substrate mixture and reaction time. We hypothesized that, by screening on a substrate mixture, changes in the relative rates of reactivity would aid in the identification of residues outside the active site that are influencing catalysis. We considered an initial substrate mixture of both aldehydes and ketones to provide UstD[2.0] with distinct steric and electronic challenges to reactivity. In direct competition, robust UPLC-MS signals for the secondary alcohol products were observed and indicated each aldehyde substrate reacted with similar efficiency. Therefore, the selection of the substrate mixture was devoid of substrates acting as strong competitive inhibitors or skewing the promiscuity profile due to high reactivity. However, the signal arising from ketone reactivity was low under these conditions and a five-fold excess of ketone to aldehyde was deployed.

In a typical protein engineering assay where high yield is the goal and catalyst activity is limiting, reactions are quantified at the reaction endpoint because the greatest difference in enzyme activity would be visible at the end of a single substrate reaction. In contrast, the greatest differences in promiscuity are observed at early timepoints. If the promiscuity were measured under initial velocity conditions that would allow us to determine the specificity of each variant. However, maintaining initial velocity conditions would have required tedious optimization and monitoring of the assay. Therefore, we elected to only measure promiscuity for technical simplicity. Therefore, we assayed the variants after a one-hour reaction time, which provided a small but reproducible signal for the tertiary alcohol product. With resolved screening parameters, engineering towards ketone activity began using globally random mutagenesis.

3.3 Identification of distal 'hotspots' from global random mutagenesis

We screened a small set of global random mutagenesis libraries prepared through error-prone PCR, comprising 880 random clones (see SI for details). From this screen we identified no variants with general boosts in activity on all substrates. Some variants, however, appeared to have changes in promiscuity (**Fig. 3.2A**). Diverse strategies to quantitate promiscuity have been developed.[27,28] However, applications of these metrics to our data demonstrated they have limited utility in this context because activity on the ketone substrate is very low. Most variants had reduced activity relative to parent and the noise associated with low-intensity measurements was indistinguishable from a change in promiscuity. Nevertheless, visual inspection clearly showed that a range of mutational effects were observed with some variants displaying apparent shifts in promiscuity. We selected 7 such variants for sequencing from which we identified 12 mutations (**Fig 3.2A**). Analysis of the UstD$^{2.0}$ crystal structure showed that no mutations were in the active site and the alpha carbon (Cα) of the mutations were an average 18 Å away from the

cofactor (**Fig 3.2B**). Consequently, we hypothesized that these mutations may be allosterically altering activity. Although these specific mutations are deleterious, some other mutation at the same site may be beneficial for catalysis, which we tested with site saturation mutagenesis (SSM).
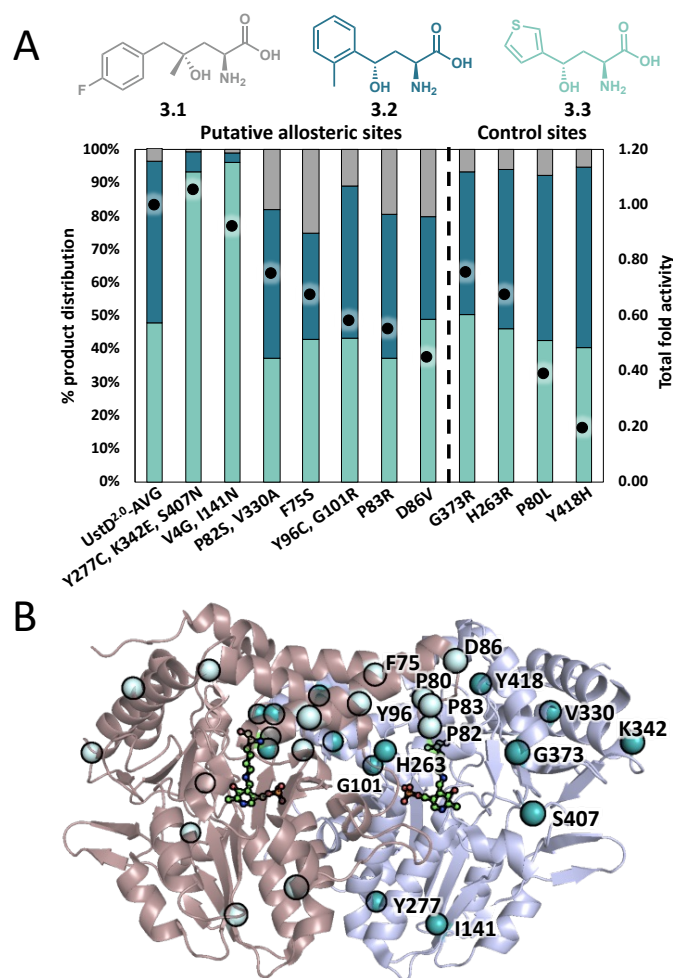
Figure 3.2 Distal putative allosteric sites identified by globally random mutagenesis. A. Combined results showing all promiscuity-shifting variants of interest from global random mutagenesis. Relative total activity for all products is represented by the dots. Product distribution is represented by the bars. The mutations found in each variant are displayed on the x-axis. Conditions: 50 mM L-asp, 5 mM thiophene-3-carboxaldehyde, 5 mM *o*-tolualdehyde, 40 mM (4-fluorophenyl)acetone, 50 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, *E. coli* whole cells over-expressing UstD[2.0] variants, 37 °C, 200 rpm, 1 h reaction time. **B.** UstD[2.0] structure showing the location of the distal residues in each monomer. The teal spheres are located on chain B (purple) while the light cyan spheres are located on chain C (pink). The PLP cofactor (green) is shown in the internal aldimine form.

3.3.1 Mutation of putative allosteric sites reveals activating mutations

We selected P82 and F75 as initial sites to test our allosteric site hypothesis. Residue P82 is 10.7 Å from the catalytic Lys, and the crystal structure of UstD$^{2.0}$ shows that it formed a cis-peptide bond. Mutation at P82 therefore has the potential to introduce more pervasive structural changes. Residue F75, is located 18 Å from the catalytic Lys and mutation at this residue during globally random mutagenesis displayed the largest apparent shift towards reactivity with a ketone. We also considered two sites from random mutagenesis that had a comparable decrease in total activity to the putative allosteric sites, but no significant change in promiscuity, G373 and P80. The initial round of global random mutagenesis gave no suggestion that these mutations might impact reaction promiscuity. They therefore acted as control sites to the relative efficiency of SSM at a putative allosteric site compared to some other site in the protein. To our surprise, the SUMS profiles of the top variants from each SSM library suggested that the control site, G373, contained generally activating mutations. In contrast, mutation at P82 had similar total activity relative to parent, but with an altered reactivity profile suggestive of a more generalist enzyme (**Fig 3.3**).
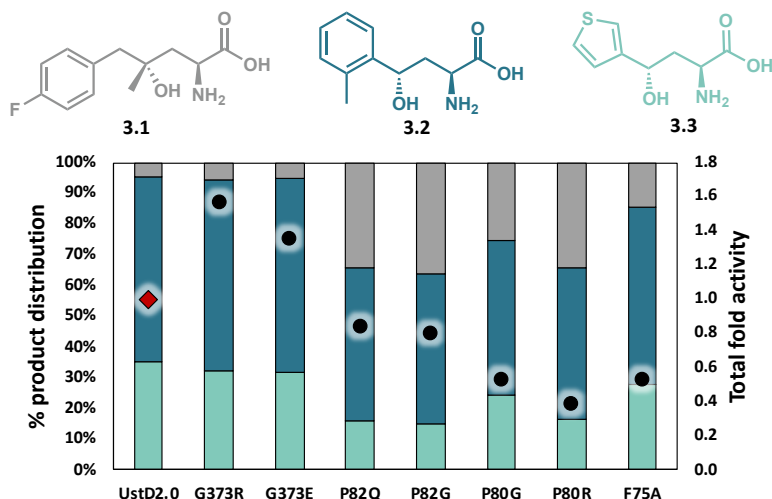
Figure 3.3 Combined results showing all promiscuity-shifting variants of interest from site saturation mutagenesis. Relative total activity for all products is represented by the dots. Product distribution is represented by the bars. The mutations found in each variant are displayed on the x-axis. **Conditions:** 50 mM L-asp, 5 mM thiophene-3-carboxaldehyde, 5 mM *o*-tolualdehyde, 40 mM (4-fluorophenyl)acetone, 50 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, *E. coli* whole cells over-expressing UstD$^{2.0}$ variants, 37 °C, 200 rpm, 1 h reaction time.

These variants were re-screened outside of a competition setting on just the desired ketone substrate, but we chose to mimic the screening conditions closely using whole-cell catalyst and only allowing each reaction to run for one hour (**Fig 3.4**). Under the whole-cell conditions, none of the P80 variants displayed increased activity for **3.1**, confirming our hypothesis that the site would not lead to activated variants. Mutation at one of the other three sites displayed increased activity for **3.1**. These results confirmed our allosteric site hypothesis as mutation at P82 and F75 significantly increased enzyme activity on a ketone substrate, as well as G373E that was discovered serendipitously.
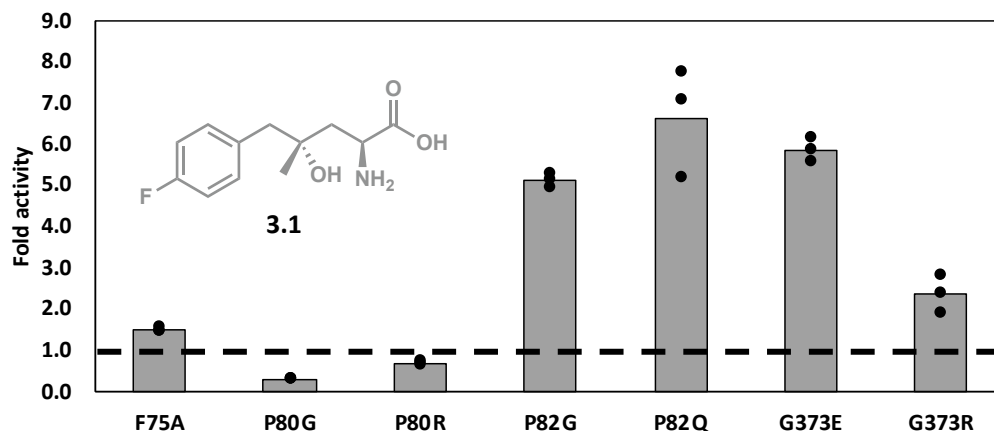
Figure 3.4 Single substrate reactions for F75X, P80X, P82X, and G373X at a one-hour reaction time. Reactions were performed in duplicate or triplicate, and activity was compared to UstD$^{2.0}$ reactions that were run simultaneously. The bars represent the average fold change of the replicates while the dots represent each individual measurement. The dotted line represents UstD$^{2.0}$ activity which is set to one. **Conditions:** 50 mM L-asp, 50 mM (4-fluorophenyl)acetone, 50 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, 20 mg/mL *E. coli* whole cells over-expressing UstD$^{2.0}$ variants, 37 °C, 200 rpm, 1 h reaction time.

We subjected the top two variants (P82Q and G373E) to more stringent assay conditions which challenge catalyst stability and turnover numbers (dilute catalyst, 16 h reaction time) to help differentiate between the variants. Under these more challenging conditions, G373E displayed a 1.2-fold boost in activity for **3.1** while P82Q had 9.6-fold higher activity for **3.1** compared to parent. (**Fig 3.5**). To probe potential additive or cooperative effects between these distal sites, we generated a new variant combining the mutations P82Q and G373E to make a new variant, QE, with 11.7-fold improvement in **3.1** (**Fig 3.5**). As the mutations are distal to the active site, it is difficult to even speculate on the molecular mechanisms through which these mutations operate. Nevertheless, this increase in activity is significant because activity with the desired ketone electrophile has crossed the threshold from stoichiometric turnover with UstD$^{2.0}$, to a modest 17 turnovers with QE.
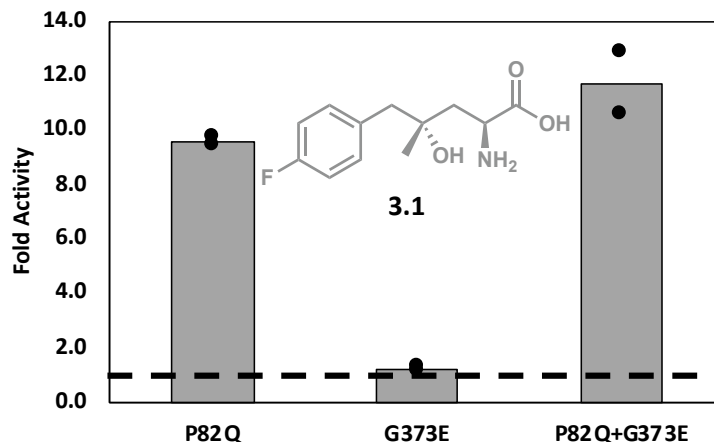
Figure 3.5 Single substrate reactions of P82Q and G373E SSM libraries, and QE double mutant under challenging conditions. Reactions were performed in duplicate, and activity was compared to UstD$^{2.0}$ reactions that were run simultaneously. The bars represent the average fold change of the replicates while the dots represent each individual measurement. The dotted line represents UstD$^{2.0}$ activity which is set to one. **Conditions:** 50 mM L-asp, 50 mM (4-fluorophenyl)acetone, 5 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, 0.01 mol% cat (10,000 max turnovers), 37 °C, 16 h reaction time.

3.3.2 Substrate space redesign to increase sensitivity to changes in reactivity

Our initial substrate mixture was chosen to minimize data complexity. Spurred by the success of the promiscuity-guided evolution above, we considered whether more information might be accessible from SUMS when using a substrate mixture with more diverse electrophiles. To maintain some continuity between the global random mutagenesis and the SSM screens both (4-fluorophenyl)acetone and *o*-tolualdehyde were retained in the substrate mixture, whereas the thiophene aldehyde was omitted. We introduced two new ketones, 1,1,1-trifluoro-3-phenyl-2-propanone (TFMK) to diversify ketone electrophilicity and 4'-nitroacetophenone (4NO2AP) to diversify structures in the substrate mixture. In simple mixtures with equimolar concentrations, the activated electrophiles dominated the overall reactivity and the signal for the tertiary alcohol products was again on the order of experimental noise (**Fig. 3S.1**). We therefore altered the substrate concentrations to ensure reproducible and robust signals for all products via UPLC-MS. The resulting mixture contained a 9:1 ratio of ketone to aldehyde substrates. In this

straightforward way, substrate mixtures for multi-generational evolutionary campaigns can be re-tuned as reactivity is expanded.

We used QE as the parent enzyme for additional SSM at distal residues identified from globally random mutagenesis. SSM at eight out of twelve of the putative allosteric sites displayed at least 1.5-fold increases in total activity during screening (**Fig 3.6**). We observed that mutations D86V and V330A both produced a 2-fold increase in activity. We also found, serendipitously, that synonymous codon changes at I141 and S371 led to boosts in whole-cell catalyst activity, presumably by increasing the soluble enzyme expression. We did not observe increases in activity from SSM at K342, a site that was previously identified in the context of a triple mutant (**Fig 3.2A**). We again considered the possibility that SSM at distal sites might lead to improvements in total activity independent of any a priori knowledge about changes in promiscuity. We therefore screened a SSM library at another control site, Y418, where mutation to His was identified by global random mutagenesis as one that decreased activity without a significant shift in promiscuity. Unlike SSM at G373, this control site did not lead to increases in activity (**Fig. 3S.2**).

We selected the top variants from five sites (16 variants) to re-screen outside of competition with longer reaction times (**Fig 3.7**). Of the 16 variants four had reproducible increases in activity on at least one ketone substrate after long reaction times (16 h). While only 25% of the identified variants were activated in this context, three of the five sites identified had at least one reproducible hit. Hence, 60% of the sites identified with SUMS promiscuity profiles resulted in bonafide activated variants for ketone substrates. Notably, D86V displays 2.5 and 1.5-fold boosts in activity with **3.4** and **3.5** respectively, which matches well with the competition promiscuity profile.
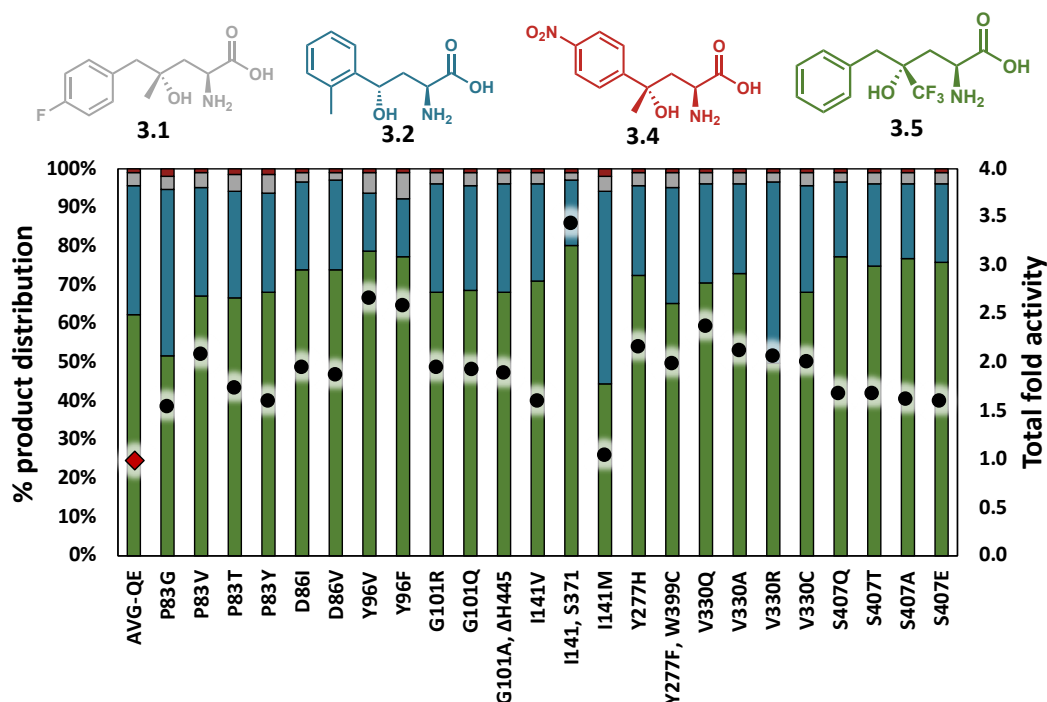
Figure 3.6 Combined SSM results for libraries using QE as parent. All sequenced variants are displayed on the x-axis. The stacked bars depict the product distribution with each color representing the corresponding chiral alcohol product. The black dots show the total fold change for all products. The total fold change for QE is represented by the red diamond and is set to one. The QE product distribution is the average QE product distribution across all library plates. **Conditions:** 50 mM L-asp 10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 50 µM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, *E. coli* whole cells over-expressing QE variants, 37 °C, 200 rpm, 1 h reaction time.

We note that not all the promiscuity profiles matched well with the screening data. The

differences between the competition promiscuity profiles (**Fig 3.6**) and the single substrate

activity (**Fig 3.7**) could be due to differences in reaction time. The libraries were screened at

early timepoints (1 h) to maximize differences in promiscuity, which has the potential to select

for enzymes that are initially activating for ketones but are unstable and become inactive quickly.

In contrast, validations were conducted overnight (16 h) to challenge the stability of the variants

to the reaction conditions and identify variants that persist in solution. Taken together, these data

demonstrate how SUMS can aid in the identification of variants with distal, cooperative

interactions influencing reactivity. By further mutagenizing these sites, this two-step process allowed identification of activating mutations with no prior information about protein structure, dynamics, or evolution. While future studies may untangle the basis of these activating effects, the present investigation is focused on the development of a practical engineering approach that incorporates promiscuity information, which we continued by deploying another common step in enzyme evolution: recombination of activating mutations.
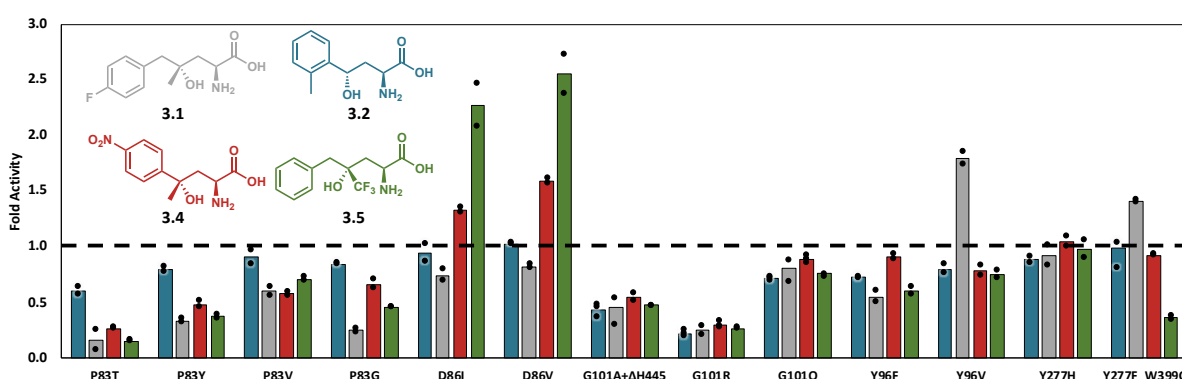
Figure 3.7 Subset of variants from SSM libraries re-screened in lysate against single substrates. Reactions were performed in at least duplicate, and activity was compared to QE reactions that were run simultaneously to determine fold change (y-axis). The bars represent the average fold change of the replicates while the dots represent each individual measurement. The dotted line represents QE activity which is set to one. The color of the bars corresponds to the product structure displayed on the chart. **Conditions:** 50 mM L-asp, 50 mM electrophile, 50 µM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, 40-50 mg/mL lysate, 37 °C, 16 h reaction time, QE parent enzyme.

3.4 Identification of cooperative mutational effects in a recombination library

There are many successful strategies for designing and screening recombination libraries.[29–32] We considered a sequence space comprising five different positions (F75, D86, I141, V330, S407) that are distributed across the protein structure. We used screening data to identify degenerate codons that limit the inclusion of mutations that are deleterious when introduced independently. The resulting library consisted of ~2,800 possible variants (See **Table 3S.1**). With this space, we again acknowledge a tradeoff between library size, fitness, and

screening intensity. For this research, it is not necessary to exhaustively screen all possible combinations, but rather to use a carefully crafted library to efficiently traverse the largest recombination space without over-sampling inactive catalysts.

At this stage of evolution, we maintained the substrate mixture from the previous round and increased reaction time from one hour to eight hours to reduce the likelihood of selecting destabilizing mutations or those that accelerate decomposition of product through retro-aldol cleavage. We hypothesized that increasing the reaction time would also lead to greater consistency between competition activity and single substrate activity. We screened 704 clones, representing ~24% of the theoretical recombination space, and found many generally activated variants (**Fig 3.8A**) The majority of the activated variants displayed a modest ~2-fold increase in total activity compared to QE. We validated a subset of the hits using purified enzyme in triplicate using single substrates (**Fig 3.8B**). As a control, D86V was included in the validation to ensure that the chosen recombination variants had higher ketone activity than a single point mutant. All hits displayed increased activity for multiple ketone electrophiles when validated. This result confirms our hypothesis that increasing reaction time increases consistency between competition and single substrate activity. Of the proteins, three variants dubbed AVILE, AVIQE, and AIIRQ, showed a ~2-fold increase in activity with all three representative ketone electrophiles while slightly decreasing activity with *o*-tolualdehyde. Of these enzymes, the quadruple mutant AIIRQ had higher soluble expression (~80 mg protein/ L culture) compared to QE and the other variants (~50 mg protein/ L culture). Therefore, AIIRQ was chosen as the new parent enzyme for subsequent evolution.

Figure 3.8 A. All sequenced variants are displayed on the x-axis. The stacked bars depict the product distribution with each color representing the corresponding chiral alcohol product. The black dots show the total fold change for all products. The total fold change for QE is represented by the red diamond and is set to one. The QE product distribution is the average QE product distribution across all library plates. **Conditions:** 50 mM L-asp, 10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 50 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, E. coli whole cells over-expressing QE variants, 37 °C, 200 rpm, 8 h reaction time. **B.** Subset of variants from recombination libraries re-screened against single substrates with purified protein. Reactions were performed in triplicate, and activity was compared to QE reactions that were run simultaneously to determine fold change (y-axis). The bars represent the average fold change of the replicates while the dots represent each individual measurement. The dotted line represents QE activity which is set to one. The color of the bars corresponds to the product structure displayed on the chart. **Conditions:** 50 mM L-asp, 50 mM electrophile, 5 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, 0.01 mol% catalyst (10,000 Max turnovers), 37 °C, 16 h reaction time.

3.5 Targeted mutagenesis in the active site reveals two mutants with distinct promiscuity profiles

In addition to distal hotspots, the active site provides a trove of sites capable of shifting activity. Previously, we identified a loop region in the UstD active site encompassing residues

391-393 which was key to unlocking high activity with challenging aldehydes.[7] We

hypothesized retargeting the 391-393 residues would reveal high activity ketone variants. Two

additional residues, M299 and T388 were also considered because both appear to have side

chains that protrude into the active site (**Fig 3.9**).
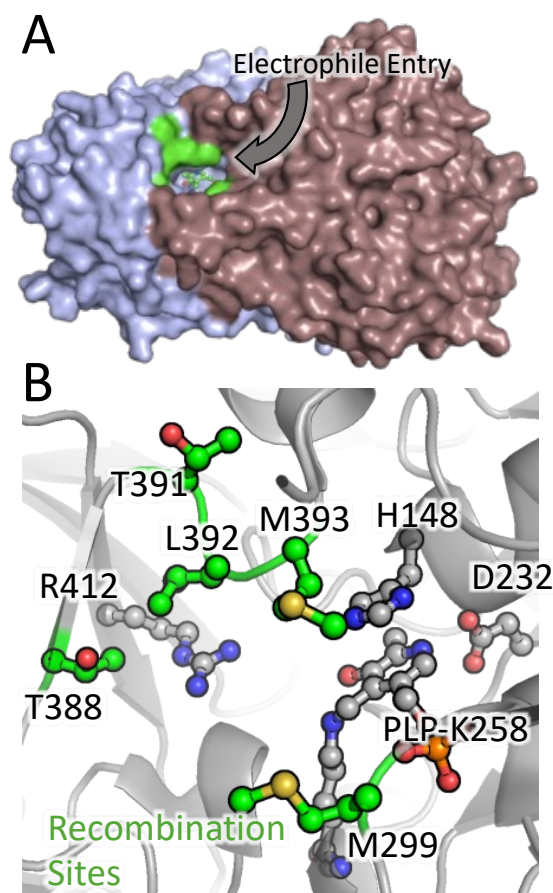
Figure 3.9 UstD active site targeted for mutagenesis. A. Space filling view of overall UstD structure. Individual monomers are colored purple (chain B) and pink (chain C) while the active site and PLP are shown in green. **B**. Close up view of UstD active site. The PLP complex and anchoring residues are shown in grey. The residues targeted for mutagenesis are shown in green.

Simple modelling of potential mutations at T388 indicated that smaller residues would

minimize the likelihood of blocking the active site. Residue M299 appears to be positioned in the

substrate channel where the electrophile substrate approaches the nucleophilic enamine

intermediate. We had previously screened a SSM library at this site with QE as the parent, which showed conservative mutations to hydrophobic residues were generally neutral to activating (**Fig 3S.3**). With this information, we designed a focused recombination library of active site variants comprising ~2,300 variants (**Table 3S.2**). Because cooperative effects are particularly common in enzyme active sites, there is a higher propensity for multiple active site mutations to be deleterious, raising the specter of laborious screening of predominately inactive sequence space. We therefore biased our search to lower mutagenesis rates by constructing the library such that the parent codon is re-sampled at greater frequency (see methods for details). This strategy decreased the expected mutational load from 3.9 to 3.3 mutations per variant.

We screened 968 clones within this focused library, comprising 42% of the theoretical sequence space. These results showed much greater diversity in the promiscuity profile from SUMS, consistent with the role of the active site in determining substrate specificity (**Fig. 3.10**).

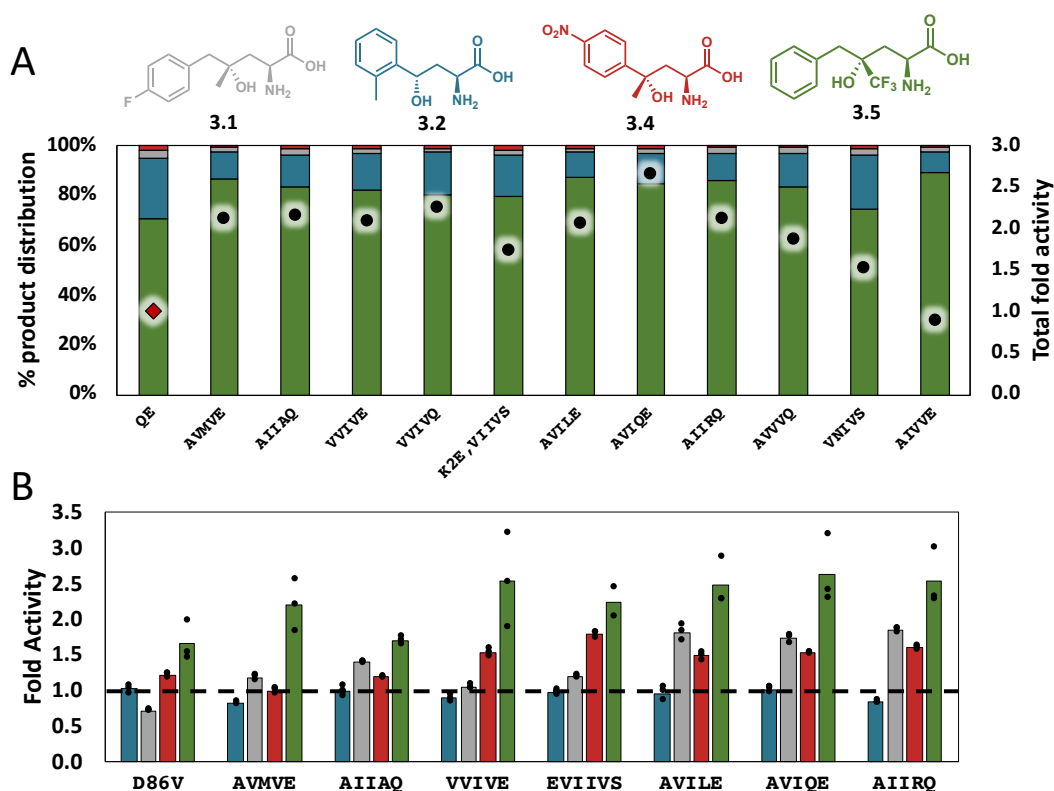**Figure 3.10 Combined library data of all sequenced active site recombination variants.** All sequenced variants are displayed on the x-axis. The stacked bars depict the product distribution with each color representing the corresponding chiral alcohol product. The distribution is shown starting at 40% to make visualization of the promiscuity shifts clearer for less reactive substrates. The black dots show the total fold change for all products. The total fold change for AIIRQ is represented by the red diamond and is set to one. The AIIRQ product distribution is the average product distribution across all library plates. Conditions: 50 mM L-asp, 10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 50 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, *E. coli* whole cells over-expressing QE variants, 37 °C, 200 rpm, 6 h reaction time.

From these variants, we identified seven with distinct promiscuity profiles, each of which are activating with the ketone electrophiles, which was consistent in the single substrate validation (**Fig. 3.11**). Notably, these activated enzymes contained an average of 3.1 mutations, in line with our hypothesis that higher mutational rates would enrich the recombination space in inactive variants. Each of these seven active site variants have distinct sequence and activity profiles and may be stimulating sources of future inquiry. Nevertheless, our focus is on developing practical catalysts with high activity and the broadest possible substrate scope. For

this reason, we selected two enzymes with distinct reactivity profiles, 7G11 (M299V, T391S, M393W) and 7B05 (T391S, M393F) for further exploration.

3.6 Lineage analysis of enzyme generality during promiscuity-guided evolution

Characterizing key variants in an evolutionary lineage is often used to show how successive rounds of mutagenesis affect activity. Such retrospective analysis is typically limited to the single transformation that was under selective pressure, with the notable exception of evolutions deploying a substrate walking strategy.[33] We wanted to perform the same type of analysis but focus on characterizing how promiscuity changed through evolution. To accomplish this goal, activity of several different substrates was measured to understand how the promiscuity-guided evolutionary process led to the activity observed by the final variants.

In order to compare activity under optimized conditions, we performed a brief survey of reaction conditions using the final variant 7G11. Increasing the concentration of L-Asp from 50 to 250 mM increased yields with both unactivated and trifluoromethyl ketones (see SI for details, **Fig. 3S.4**). While such high concentrations are not ideal, the amino acid is cheap and commercially available. We further observed that additional equivalents of PLP relative to enzyme (between 10-50-fold excess) were beneficial for all substrates.
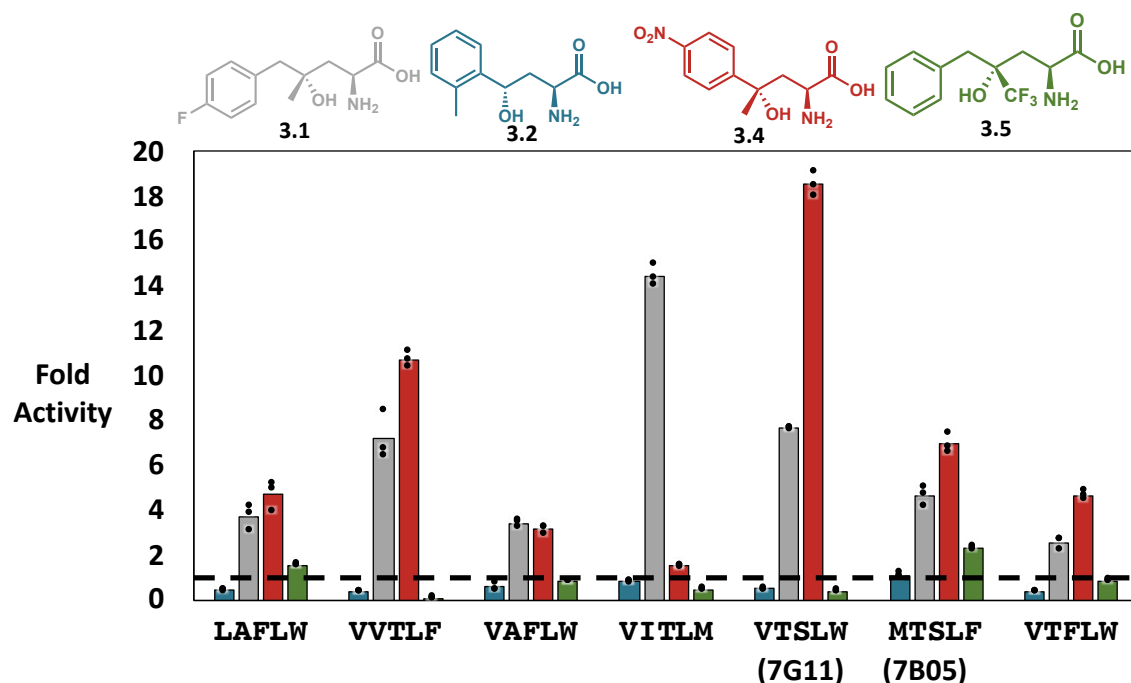
Figure 3.11 Subset of variants from active-site recombination libraries re-screened against single substrates with purified protein. Reactions were performed in triplicate, and activity was compared to AIIRQ reactions that were run simultaneously to determine fold change (y-axis). The bars represent the average fold activity of the replicates while the dots represent each individual measurement. The dotted line represents AIIRQ activity which is set to one. The color of the bars corresponds to the product structure displayed on the chart. **Conditions:** 50 mM L-asp, 50 mM electrophile, 5 µM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, 0.01 mol% catalyst (10,000 Max turnovers), 37 °C, 16 h reaction time.

Using these optimized conditions, we assayed the lineage variants against several electrophiles in single substrate reactions. We found that there were significant changes in activity for substrates that were under direct selective pressure (4-fluorophenylacetone and TFMK) and on those that were not (benzaldehyde and furylacetone). Cumulatively, activity with ketone substrates increased throughout evolution, leading to the two ketone generalists 7B05 and 7G11 (**Fig. 3.12**). The increases in activity with ketones correlated with decreases in formation of the shunt product, L-ala (**Fig. 3.12**). This could indicate that increases in activity with ketone substrates are partially due to suppression of the shunt reaction. However, further mechanistic studies would be required to support this hypothesis which are beyond the scope of this work.
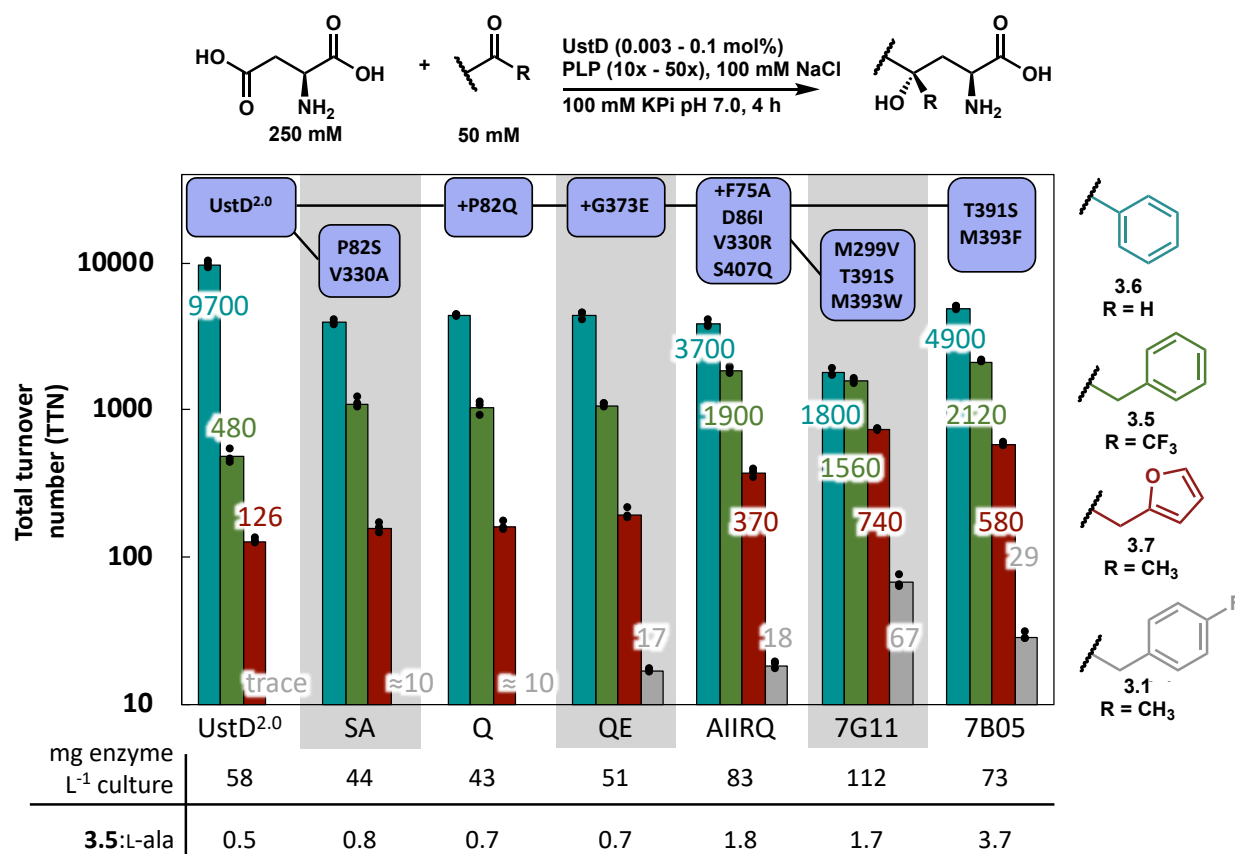
Figure 3.12 Lineage analysis of enzyme generality. Reactions were performed with single substrates in triplicate. The substrates are displayed to the right of the chart. The total turnover (TTN) number is displayed on the y-axis on a logarithmic scale. The bar represents the average of the technical replicates, and the dots represent the TTN measurement of each individual replicate. The color of each bar corresponds to its product. The mutations associated with each variant are displayed at the top of the graph with the connecting lines representing the relationship between them. Below the chart, the yield of protein per liter of culture is displayed for each variant. The ratio of **3.5** to the shunt pathway product, L-ala, is displayed below the chart to demonstrate that the ratio changes over the lineage.

Benzaldehyde was too reactive with the enzyme for inclusion in the multiplexing mixtures, meaning it was not under direct selective pressure during evolution. UstD[2.0], which was evolved using benzaldehyde as a model substrate, performs 9,700 turnovers with this aldehyde substrate. The observed activity is in good agreement with previously reported turnover numbers.[7] However, activity on benzaldehyde decreased through the evolution with 7G11 performing 1,800 turnovers and 7B05 performing 4,900 turnovers. While data are plotted on a log scale, we emphasize that the changes in activity on benzaldehyde are statistically significant.

Overall, through evolution, 7G11 and 7B05 display a ~5-fold and ~2-fold decrease in activity with benzaldehyde.

In contrast, activity with the different ketones steadily increased throughout the lineage. Of particular note, a ~6-fold increase in activity was observed with furylacetone even though it was never utilized as a substrate during evolution (**Fig. 3.12**). The top TFMK variant, 7B05, performed 2,120 turnovers, representing a ~4-fold increase in activity with this activated ketone. Parent activity with (4-fluorophenyl)acetone was particularly abysmal with only trace reactivity detected. Through our evolution, we enhanced the reaction with (4-fluorophenyl)acetone from stoichiometric to catalytic representing a 67-fold increase in activity using the best unactivated ketone variant, 7G11. Most gratifyingly, these data show concretely how promiscuity-guided evolution is an effective strategy for developing generalists.

3.7 Conclusions

To the best of our knowledge, this work represents the first example of a multigenerational promiscuity-guided protein engineering campaign leveraging mutations distal to the active site. We posit the residues we identified are acting as allosteric sites to tune reactivity. While additional studies, both experimental and computational, are likely needed to understand the molecular mechanism for how these mutations influence the active site, our results demonstrate that cooperativity exists. Using SUMS, these distal sites were discovered and leveraged without the need for detailed structural or mechanistic information, a task that remains difficult using traditional engineering methods.[13,34] While we have demonstrated that SUMS engineering provides benefits traditional engineering cannot, it also has several drawbacks. There is modestly more reaction optimization prior to screening and the data analysis is more complex. The promiscuity profiles allow researchers to directly observe reaction scope, a boon

when developing generalists. However, it becomes more difficult to define the "best" variant, as activated variants can encompass changes in promiscuity, activity, or both. Despite these additional complexities, our method was successful in generating variants with surprisingly different promiscuities throughout the evolutionary lineage. This work provides a detailed road map of how promiscuity-guided engineering can be employed to traverse protein fitness landscapes more effectively without structural or mechanistic information and shows promise for engineering challenging reactions.

3.8 Supplementary Information, Figures and Tables
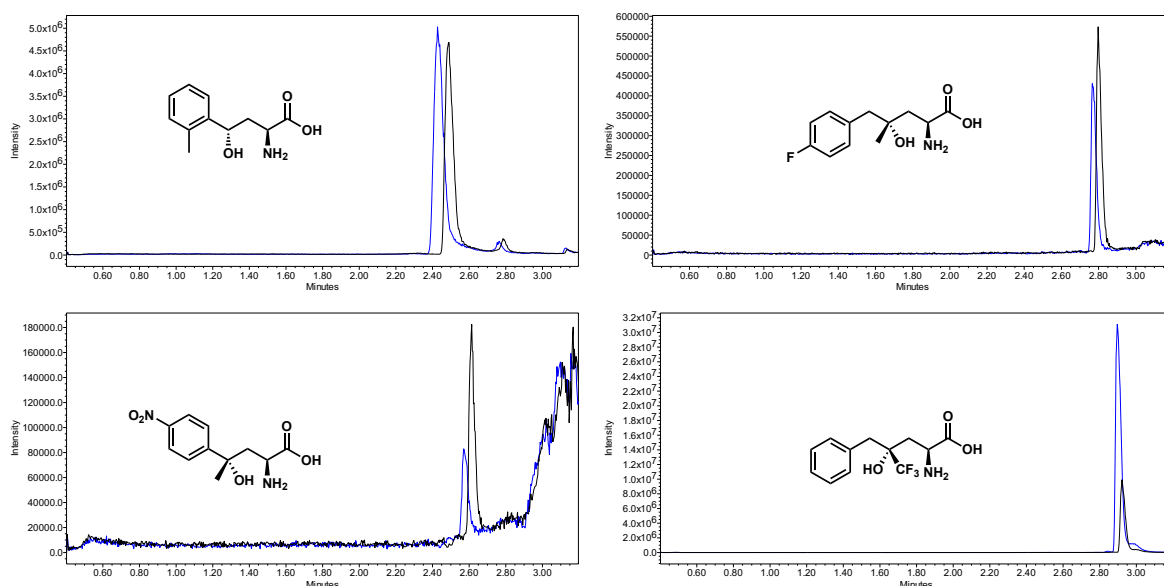
Figure 3S.1 Overlayed product mass traces, indicated by the structure, for the different electrophile mixtures. In blue, substrates are added in equal amounts (12.5 mM each). In black, the substrates are added in differing amounts according to their electrophilicity (5 mM *o*-tolualdehyde, 10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 2.5 mM 1,1,1-trifluoror-3-phenyl-2-propanone). **Conditions:** 50 mM L-asp, 50 mM total electrophiles, 5 µM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, QE (0.01 mol% catalyst, 10,000 Max TON), 37 °C, 1 h reaction time.
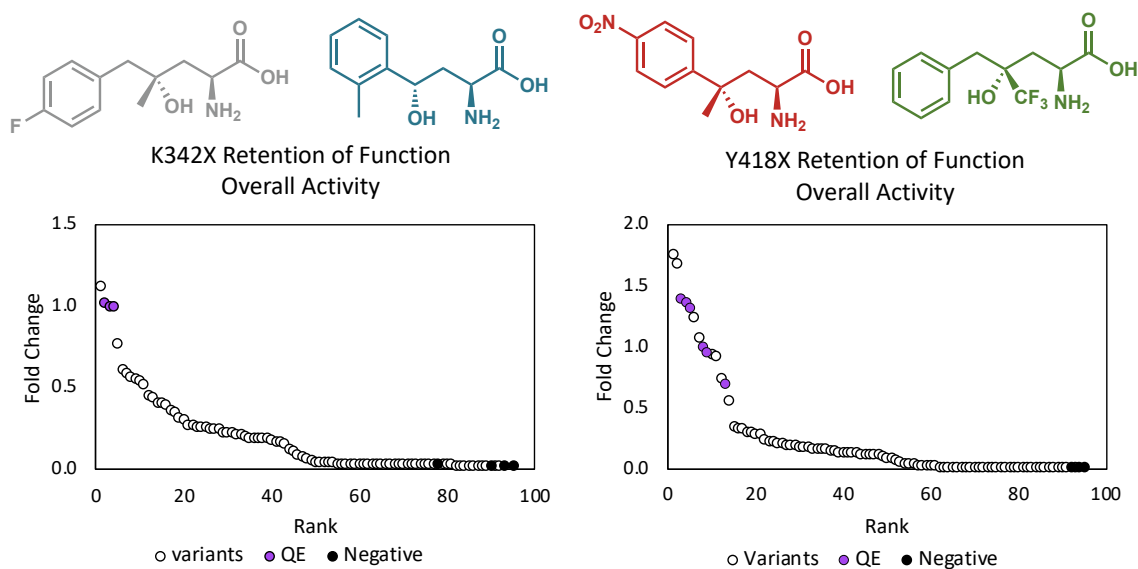
Figure 3S.2. Retention of function curves for total activity at sites K342 and Y418 for their respective site saturation libraries. Neither library shows any activated variants (white dots) compared to the QE parent enzyme (purple dots). Negative controls and the sterile well are shown in black. **Conditions:** 50 mM L-asp, 10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM o-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 50 μM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, E. coli whole cells over-expressing QE variants, 37 °C, 200 rpm, 1 h reaction time.

Table 3S.1. Distal recombination library amino acid residues possible at each site.

site	F75	D86	I141	V330	S407
mutation	F	D	I	Q	S
	A	I	V	A	T
	V	V	M	R	Q
		N		E	A
				P	E
				L	
				G	
				V	
size	3	4	3	8	5
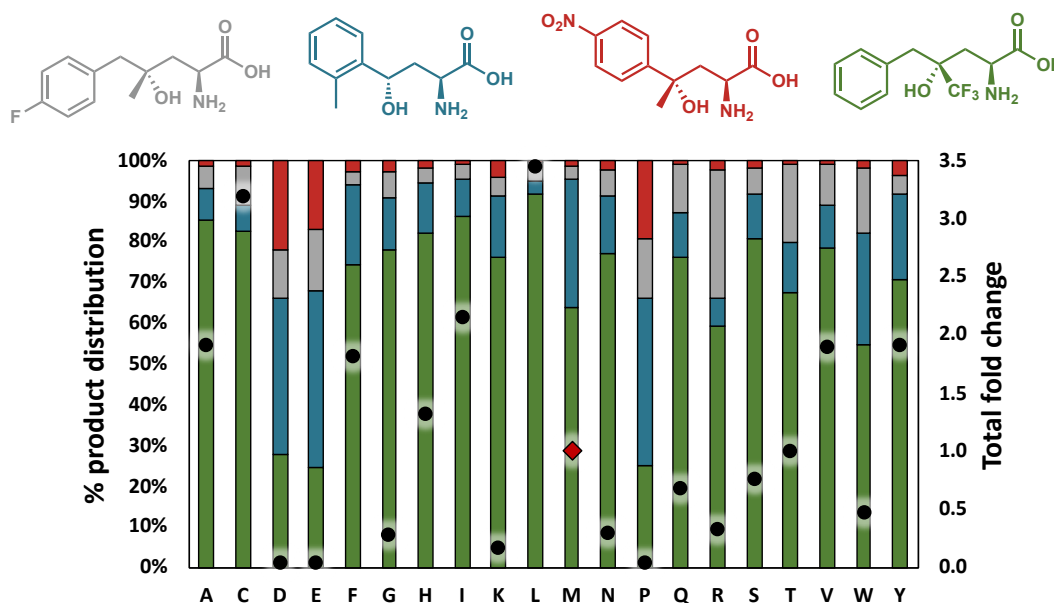
Figure 3S.3. M299X SSM library data. All sequenced variants are displayed on the x-axis. The stacked bars depict the product distribution with each color representing the corresponding chiral alcohol product. The black dots show the total fold change for all products. The total fold change for QE is represented by the red diamond and is set to one. **Conditions:** 50 mM L-asp, 10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 50 µM PLP, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, *E. coli* whole cells over-expressing QE variants, 37 °C, 200 rpm, 1 h reaction time.

Table 3S.2. Active site recombination library amino acid residues possible at each site.

site	M299	T388	T391	L392	M393
mutation	M	T	T	L	M
	L	V	F	V	W
	V	I	S	A	F
		A	L	S	L
			P		R
			I		C
					I
					S
size	3	4	6	4	8
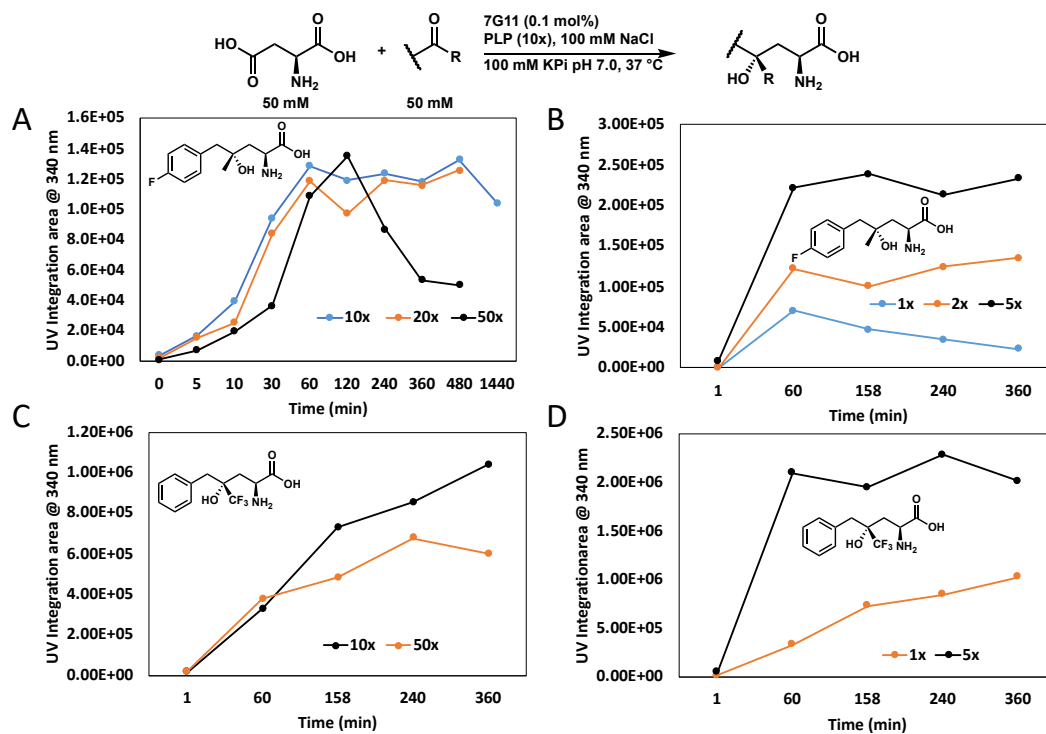
Figure 3S.4. Reaction condition optimization of 7G11. Initial reaction conditions are depicted in the top scheme. Conditions were optimized sequentially: **A.** 50 equivalents PLP relative to catalyst for (4-fluorophenyl)acetone, **B.** 5 equivalents of L-aspartate relative to (4-fluorophenyl)acetone, **C.** L-asp **D.** 5 equivalents of L-aspartate relative to TFMK.
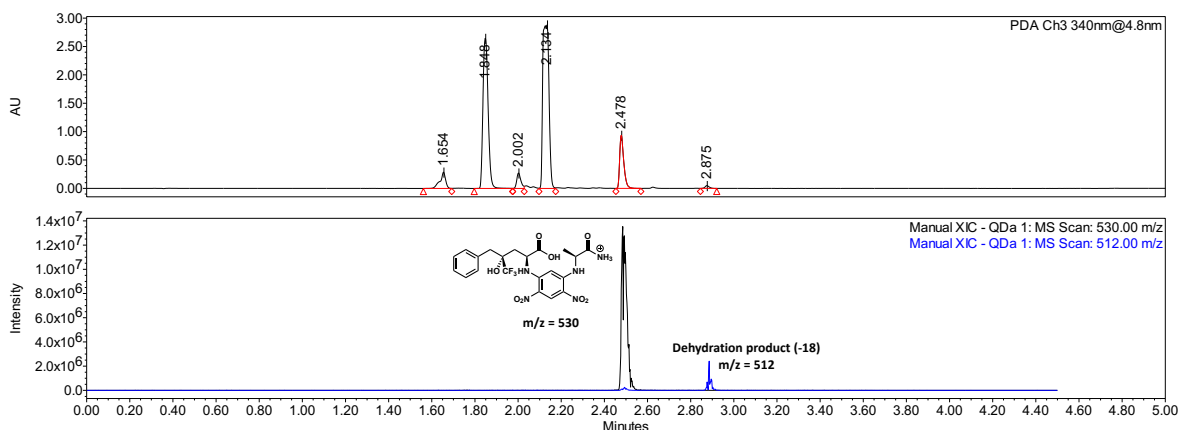
Figure 3S.5 Representative UPLC trace of TFMK for lineage analysis following Marfey's derivatization. The top trace displays the absorbance at 340 nm while the bottom trace displays the relevant product masses. Note that Marfey's reactions are quenched with 60 mM HCl which causes a dehydration product of the amino acid to form resulting in two product peaks with masses differing by 18 mass units. The molecular ion is shown in black.
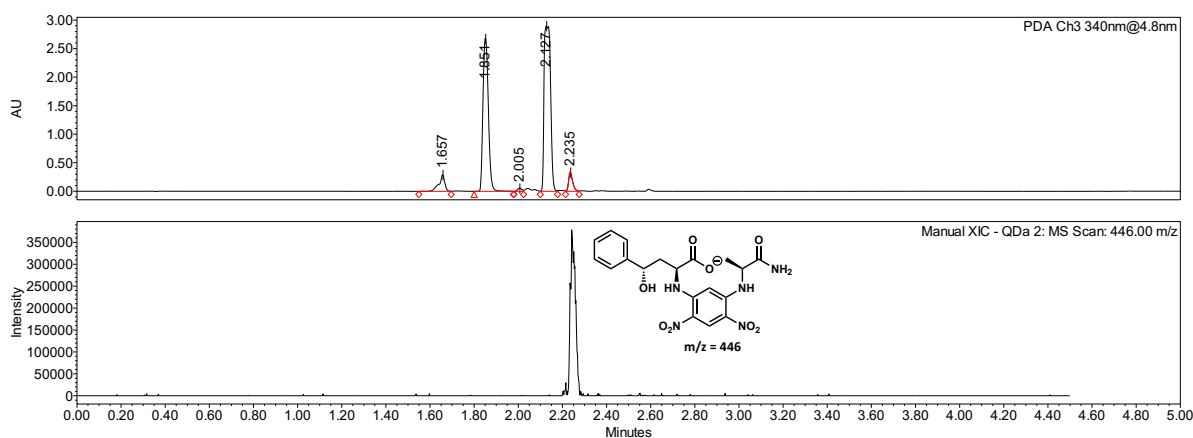
Figure 3S.6 Representative UPLC trace for benzaldehyde following Marfey's derivatization. The top trace displays the absorbance at 340 nm while the bottom trace displays the relevant product masses and structures. The molecular ion is shown in black.
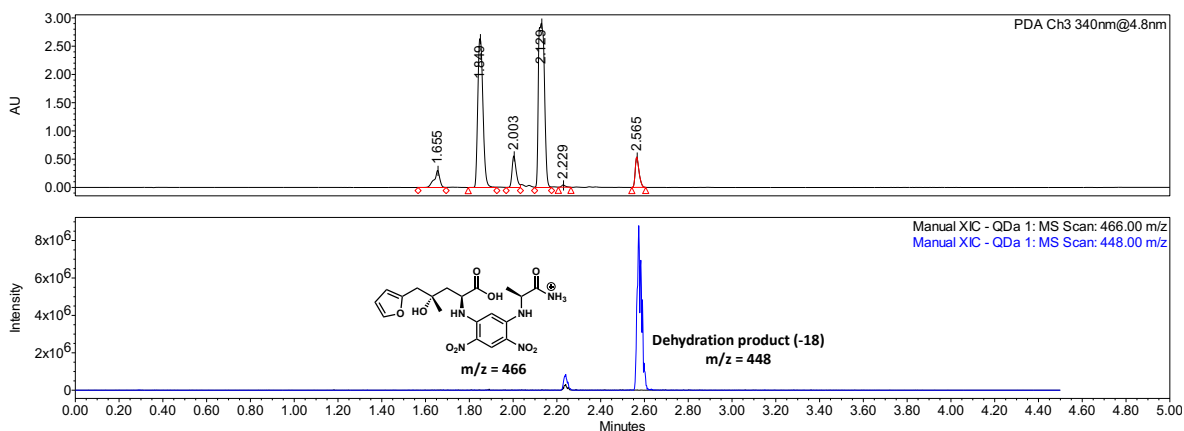
Figure 3S.7 Representative UPLC trace of furylacetone for lineage analysis following Marfey's derivatization. The top trace displays the absorbance at 340 nm while the bottom trace displays the relevant product masses and structures. Note that Marfey's reactions are quenched with 60 mM HCl which causes a dehydration product of the amino acid to form resulting in two product peaks with masses differing by 18 mass units. The molecular ion is shown in black.
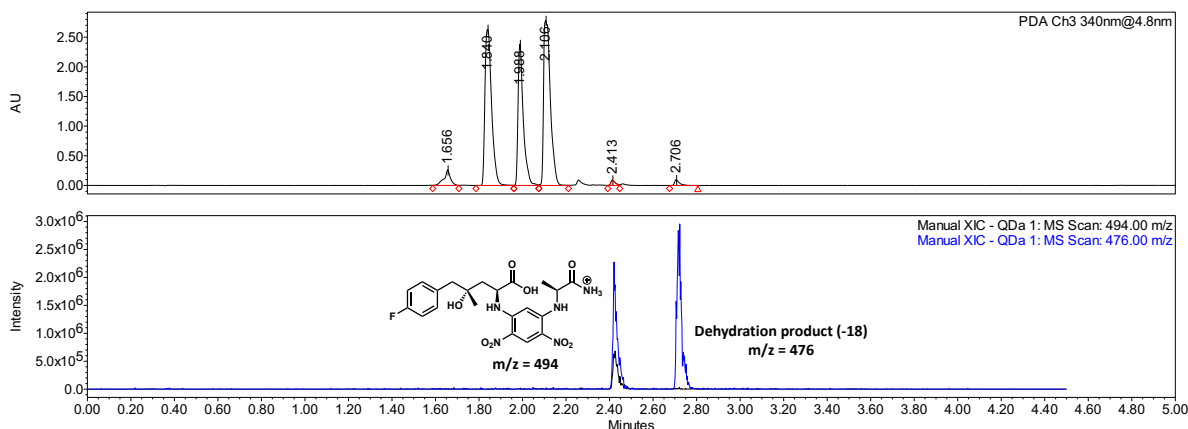
Figure 3S.8 Representative UPLC trace of (4-fluorophenyl)acetone for lineage analysis following Marfey's derivatization. The top trace displays the absorbance at 340 nm while the bottom trace displays the relevant product masses and structures. Note that Marfey's reactions are quenched with 60 mM HCl which causes a dehydration product of the amino acid to form resulting in two product peaks with masses differing by 18 mass units. The molecular ion is shown in black.

General materials and methods

All chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, VWR, Chem-Impex International, Alfa Aesar, Combi-blocks, Oakwood Products) at the highest quality available and used without further purification unless stated otherwise. Genes were purchased as gBlocks from Integrated DNA Technologies (IDT). *E. coli* cells were electroporated with an Eppendorf E-porator at 2500 V. New Brunswick I26R shaker incubators (Eppendorf) were used for cell growth. Cell disruption via sonication was performed with a Sonic Dismembrator 550 (Fisher Scientific) sonicator. UV-vis spectroscopic measurements were collected on a UV-2600 Shimadzu spectrophotometer. Optical density measurements were collected using an optical density reader (Amersham Biosciences). Ultra-high pressure liquid chromatography-mass spectrometry (UPLC-MS) data were collected on an Acquity UPLC (Waters) equipped with an Acquity PDA and QDA MS detector using either a BEH C18 column (Waters) for aromatic substrates, or an Intrada Amino Acid column (Imtakt) for aliphatic substrates.

Plasmid Preparations

A 5-mL overnight culture of *E. coli* harboring the plasmid of interest was grown overnight at 37 °C with shaking at 200 rpm. The plasmid was isolated and purified using Zymo Plasmid Miniprep or Macherey-Nagel kits and sequenced through Functional Biosciences.

Protein and DNA sequences

Sequences with coding and silent mutations relative to UstD$^{2.0}$ have the target mutation bolded.

DNA sequence of ustD$^{2.0}$
ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA

CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTTCCCCCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCGATG
CTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGAACCTTAATGGCACCGCGCCCGACATGGG
ACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTACA
ACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACGC
GACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^{SA}
ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTT**TCC**CCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCCATG
CTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACT**GCG**CTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGAACCTTAATGGCACCGCGCCCGACATGGG
ACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTACA

ACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACGC
GACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^Q

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTT**CAG**CCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCCATG
CTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC
AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCCGGGGATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGAACCTTAATGGCACCGCGCCCGACATGGG
ACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTACA
ACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACGC
GACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^{QE}

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAATTTCAT
GTATAGCTTTCCTTTT**CAG**CCGGGTGTTGACGCTAAGTCAATGGAGGCTATTACCGC
ATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAAAT
TACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAACCT
ATGCTGAATAACGATTGCGAAATCGTATGCTCAACATTATGTCACGAAGCAGCAGCT
TCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCCCA
ACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGCCC
TTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAGGA
ACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCCATG
CTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAATTG
GATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAACCC
TGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTTCG
TCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACTGC

AGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCGTA
TCGTGCGCCAAGAGACTGTGCTGGTAACTATTTTGTTGGAGTATTTACTTAGCAAGC
CATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGCAA
TCGTAACTTTTGAAGTCGTGGGACGTAGTTCC**GAG**GATGTGGCAATGCGCGTAAATA
CGCGTAATCGCTTCCGCATTACCTCTGGAACCTTAATGGCACCGCGCCCGACATGGG
ACGTCTTGAAACCGAAGAGTAGCGACGGACTTGTTCGCGTCAGCTTTGTACATTACA
ACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACACGC
GACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^AIIRQ
ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAAT**GCA**A
TGTATAGCTTTCCTTTT**CAG**CCGGGTGTT**ATC**GCTAAGTCAATGGAGGCTATTACCG
CATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAA
ATTACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAAC
CTATGCTGAATAACGATTGCGAA**ATT**GTATGCTCAACATTATGTCACGAAGCAGCAG
CTTCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCC
CAACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGC
CCTTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAG
GAACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCCA
TGCTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAAT
TGGATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAAC
CCTGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTT
CGTCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACT
GCAGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCG
TATCGTGCGCCAAGAGACT**CGA**CTGGTAACTATTTTGTTGGAGTATTTACTTAGCAA
GCCATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGC
AATCGTAACTTTTGAAGTCGTGGGACGTAGTTCC**GAG**GATGTGGCAATGCGCGTAA
ATACGCGTAATCGCTTCCGCATTACCTCTGGAACCTTAATGGCACCGCGCCCGACAT
GGGACGTCTTGAAACCGAAGAGT**CAA**GACGGACTTGTTCGCGTCAGCTTTGTACATT
ACAACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACA
CGCGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^7B05
ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAAT**GCA**A
TGTATAGCTTTCCTTTT**CAG**CCGGGTGTT**ATC**GCTAAGTCAATGGAGGCTATTACCG
CATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAA
ATTACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAAC
CTATGCTGAATAACGATTGCGAA**ATT**GTATGCTCAACATTATGTCACGAAGCAGCAG
CTTCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCC
CAACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGC

CCTTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAG
GAACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCCA
TGCTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAAT
TGGATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAAC
CCTGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTT
CGTCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGCATGCCGTCCTTTGAACT
GCAGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCG
TATCGTGCGCCAAGAGACT**CGA**CTGGTAACTATTTTGTTGGAGTATTTACTTAGCAA
GCCATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGC
AATCGTAACTTTTGAAGTCGTGGGACGTAGTTCC**GAG**GATGTGGCAATGCGCGTAA
ATACGCGTAATCGCTTCCGCATTACCTCTGGA**TCC**TTA**TTT**GCACCGCGCCCGACAT
GGGACGTCTTGAAACCGAAGAGT**CAA**GACGGACTTGTTCGCGTCAGCTTTGTACATT
ACAACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACA
CGCGACACCCTCGAGCACCATCACCATCACCATTGA

DNA sequence of ustD^{7G11}

ATGAAGAGCGTAGCGACGAGTTCCCTTGATGACGTAGATAAAGATTCCGTCCCCCTG
GGCAGTTCGATCAATGGCACTGCACAAGCGGAAACTCCGCTGGAGAATGTGATCGA
CGTCGAATCAGTGCGCTCACATTTCCCGGTATTAGGGGGGGAAACGGCCGCGTTTAA
CAATGCATCAGGAACCGTAGTTTTGAAGGAGGCAATTGAATCGACTTCAAAT**GCA**A
TGTATAGCTTTCCTTTT**CAG**CCGGGTGTT**ATC**GCTAAGTCAATGGAGGCTATTACCG
CATATACGGGGAATAAGGGCAAGGTTGCGGCATTTATCAATGCACTTCCTGATGAA
ATTACATTCGGGCAGTCCACAACTGCCCTGTTCCGTTTATTAGGTCTGTCGCTTAAAC
CTATGCTGAATAACGATTGCGAA**ATT**GTATGCTCAACATTATGTCACGAAGCAGCAG
CTTCCGCATGGATTCATTTAAGTCGCGAATTAGGAATTACCATTAAGTGGTGGAGCC
CAACTACTACACCGAATAGTCCCGATGATCCAGTTCTGACGACTGACTCATTGAAGC
CCTTGCTTAGTCCAAAAACGCGCCTTGTTACATGTAATCACGTGTCGAATGTTGTAG
GAACCATCCACCCTATTCGTGAGATTGCCGACGTGGTACATACCATTCCTGGAGCCA
TGCTTATCGTTGACGGTGTGGCAAGCGTCCCGCATCGTCCAGTTGATGTTAAAGAAT
TGGATGTAGATTTTTACTGCTTTTCCTGGTACAAGTTGTTCGGACCGCATCTTGGAAC
CCTGTATGCTTCCCGCAAAGCCCAAGACCGCTATATGACCTCAATTAACCATTACTT
CGTCTCATCGTCGAGCCTTGATGGTAAGCTGGCATTAGGC**GTG**CCGTCCTTTGAACT
GCAGTTGATGTGCTCTCCAATTGTTTCGTATTTGCAAGATACGGTGGGCTGGGACCG
TATCGTGCGCCAAGAGACT**CGA**CTGGTAACTATTTTGTTGGAGTATTTACTTAGCAA
GCCATCTGTATATCGTGTGTTCGGACGTCGTAATTCTGATCCCAGTCAGCGTGTAGC
AATCGTAACTTTTGAAGTCGTGGGACGTAGTTCC**GAG**GATGTGGCAATGCGCGTAA
ATACGCGTAATCGCTTCCGCATTACCTCTGGA**TCC**TTA**TGG**GCACCGCGCCCGACAT
GGGACGTCTTGAAACCGAAGAGT**CAA**GACGGACTTGTTCGCGTCAGCTTTGTACATT
ACAACACGGTTGAGGAAGTGCGTGCGTTCTGCAGCGAGTTAGACGAGATTGTGACA
CGCGACACCCTCGAGCACCATCACCATCACCATTGA

Table 3S.3 Primer Sequences. All primers were purchased from Integrated DNA Technologies

Protein	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
pET22b(+)-UstD$^{2.0}$	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG	GCCGGATCTCAATGGTGATGGTGATGGTGCTCGAG
D86X	TTTCCTTTTCAGCCGGGTGTT**XXX**GCTAAGTCAATGGAGGC	AACACCCGGCTGAAAAGGAA
F75X	GGCAATTGAATCGACTTCAAAT**XXX**ATGTATAGCTTTCCTTTTCCCCC	ATTTGAAGTCGATTCAATTGCCTCCTTC
G101X	ACCGCATATACGGGGAATAAG**XXX**AAGGTTGCGGCATTTATCAATGC	CTTATTCCCCGTATATGCGGTAATAGCCTC
G373X	GAAGTCGTGGGACGTAGTTCC**XXX**GATGTGGCAATGCGCGTAAATACG	GGAACTACGTCCCACGACTTCAAAAGTTA
I141X	CTATGCTGAATAACGATTGCGAA**XXX**GTATGCTCAACATTATGTCACGAAG	TTCGCAATCGTTATTCAGCATAGGTTTAA
K342X	GTTGGAGTATTTACTTAGC**XXX**CCATCTGTATATCGTGTGTTCGG	GCTAAGTAAATACTCCAACAAAAATAGTTACC
M299X	GATGGTAAGCTGGCATTAGGC**XXX**CCGTCCTTTGAACTGCAGTTGA	GCCTAATGCCAGCTTACCATCAAGGC
P80X	CAAATTTCATGTATAGCTTT**XXX**TTTCCCCCGGGTGTTGAC	AAAGCTATACATGAAATTTGAAGTCGATTC
P82X	CATGTATAGCTTTCCTTTT**XXX**CCGGGTGTTGACGCTAAGTC	AAAAGGAAAGCTATACATGAAATTTGAAGTCG

P83X	TTTCATGTATAGCTTTCCTT TTCAG**XXX**GGTGTTGACGC TAAGTCAATGGAGGC	CTGAAAAGGAAAGCTATAC ATGAAATTTGAAGTCG
S407X	CGTCTTGAAACCGAAGAGT **XXX**GACGGACTTGTTCGCG TCAG	ACTCTTCGGTTTCAAGACG TCC
V330X	GTATCGTGCGCCAAGAGAC T**XXX**CTGGTAACTATTTTGT TGGAG	AGTCTCTTGGCGCACGATA CGGTCCC
Y277X	CCCGCAAAGCCCAAGACCG C**XXX**ATGACCTCAATTAAC CATTACTTCG	GCGGTCTTGGGCTTTGCGG G
Y418X	GTTCGCGTCAGCTTTGTAC AT**XXX**AACACGGTTGAGGA AGTGCG	ATGTACAAAGCTGACGCGA ACAAG
Y96X	GTCAATGGAGGCTATTACC GCA**XXX**ACGGGGAATAAG GGCAAGGT	TGCGGTAATAGCCTCCATT G
F75[GYA]	GGCAATTGAATCGACTTCA AAT**GYA**ATGTATAGCTTTC CTTTTCAGCC	ATTTGAAGTCGATTCAATT GCCTCCTTC
D86[RWC]	GCTTTCCTTTTCAGCCGGGT GTT**RWC**GCTAAGTCAATGG AGGCTATTAC	AACACCCGGCTGAAAAGGA A
I141[RTK]	CCTATGCTGAATAACGATT GCGAA**RTK**GTATGCTCAAC ATTATGTCACGAAGC	TTCGCAATCGTTATTCAGC ATAGGTTTAA
V330[SNA]	CCGTATCGTGCGCCAAGAG ACT**SNA**CTGGTAACTATTTT GTTGGAGTATTTAC	AGTCTCTTGGCGCACGATA CGGTCCC
S407[SAA]	CGTCTTGAAACCGAAGAGT **SAA**GACGGACTTGTTCGCG TCAGC	ACTCTTCGGTTTCAAGACG TCC
S407[DCA]	CGTCTTGAAACCGAAGAGT **DCA**GACGGACTTGTTCGCG TCAGC	ACTCTTCGGTTTCAAGACG TCC

TTLM_4 mutation	CGCGTAATCGCTTCCGCAT**TRYA**TCTGGA**HYCDYAWKK**GCACCGCGCCCGACATGGGACG	AATGCGGAAGCGATTACGCGTATTTACGC
TTLM_3 mutation_T388	CGCGTAATCGCTTCCGCATTACCTCTGGA**HYCDYAWKK**GCACCGCGCCCGACATGGGACG	AATGCGGAAGCGATTACGCGTATTTACGC
TTLM_3 mutation_T391	CGCGTAATCGCTTCCGCAT**TRYA**TCTGGAACC**DYAWKK**GCACCGCGCCCGACATGGGACG	AATGCGGAAGCGATTACGCGTATTTACGC
TTLM_3 mutation_L392	CGCGTAATCGCTTCCGCAT**TRYA**TCTGGA**HYC**TTA**WKK**GCACCGCGCCCGACATGGGACG	AATGCGGAAGCGATTACGCGTATTTACGC
TTLM_3 mutation_M393	CGCGTAATCGCTTCCGCAT**TRYA**TCTGGA**HYCDYA**ATGGCACCGCGCCCGACATGGGACG	AATGCGGAAGCGATTACGCGTATTTACGC
M299 [DTG]	GATGGTAAGCTGGCATTAGGC**DTG**CCGTCCTTTGAACTGCAGTTG	GCCTAATGCCAGCTTACCATC
H263X	CCTGGTACAAGTTGTTCGGACCG**XXX**CTTGGAACCCTGTATGCTTCCCGC	CGGTCCGAACAACTTGTACCAGGAAAAGC
H283X	CCAAGACCGCTATATGACCTCAATTAAC**XXX**TACTTCGTCTCATCGTCGAGCCTTG	
F285X	GCTATATGACCTCAATTAACCATTAC**XXX**GTCTCATCGTCGAGCCTTGATGGTAAGCT	GTAATGGTTAATTGAGGTCATATAGCGG
P300X-7B05	GATGGTAAGCTGGCATTAGGCATG**XXX**TCCTTTGAACTGCAGTTGATGTGCTC	CATGCCTAATGCCAGCTTACCATCA

P300X-7G11	GATGGTAAGCTGGCATTAG GCGTG**XXX**TCCTTTGAACT GCAGTTGATGTGCTC	CACGCCTAATGCCAGCTTA CCATCA
S389X-7B05	TACGCGTAATCGCTTCCGC ATTACC**XXX**GGATCCTTAT TTGCACCGCGCCCG	GGTAATGCGGAAGCGATTA CGCGTA
S389X-7G11	TACGCGTAATCGCTTCCGC ATTACC**XXX**GGATCCTTAT GGGCACCGCGCCCG	GGTAATGCGGAAGCGATTA CGCGTA

* **XXX** indicates a 22-codon library made as a mixture of 3 degenerate codon primers (NDT, VHG, TGG), as described by Kille, S. et al.[35]

DNA Isolation and Storage

 DNA was purified via gel electrophoresis and isolated using a DNA gel extraction kit (Zymo Research or Macherey-Nagel). All isolated DNA was stored at -20 ºC.

Protein Expression & Purification

Optimized Expression of UstD$^{2.0}$ and variants: An overnight culture of *E. coli* BL21(DE3) harboring a pET-22b(+) plasmid encoding a given UstD$^{2.0}$ variant was created by inoculating 10 mL of TB$_{amp}$ media with a single colony. This culture was shaken at 37 ºC and 200 rpm for ~16 h. 10 mL of overnight culture was then used to inoculate 1 L of TB$_{amp}$, which was shaken at 37 ºC and 200 rpm for approximately 1.5 h or until an optical density (OD) of 0.4-0.6 was reached. Cultures were removed from the incubator and cooled on ice for 30 min, followed by induction with 100 μM IPTG. The cultures were allowed to continue to grow for an additional ~16 h at 20 ºC and shaking at 200 rpm. Cells were then harvested by centrifugation (4 ºC, 30 min, 4,000 x*g*), and the cell pellets were stored at -20 ºC overnight.

Protein Purification: To purify UstD, cell pellets were thawed at room temperature and then resuspended in lysis buffer, comprised of enzyme storage buffer (100 mM potassium phosphate buffer pH 7.0, 100 mM sodium chloride) containing 20 mM imidazole, 1 mg/mL Hen Egg White

Lysozyme (GoldBio), 0.2 mg/mL DNase (GoldBio), 1 mM $MgCl_2$, and 0.5 mg/g cell pellet pyridoxal 5'-phosphate (PLP). A ratio of 4 mL lysis buffer per gram of wet cell pellet was used. Cells lysis began by shaking for 1 h at 37 °C. The resuspended cells were subsequently sonicated (30 s per g cell pellet, 2 s on, 2 s off, 40% amplitude). The resulting lysate was then spun down at 48,384 x*g* to pellet cellular debris. Ni/NTA beads were pre-equilibrated in storage buffer containing 20 mM imidazole. 1-2 mL of resin were used per 50 mL of lysis. The flow-through was re-passed once to collect any remaining beads from the original vessel. The collected beads were washed with 5 column volumes each of storage buffer containing 20 mM, 40 mM, and 60 mM imidazole. Protein was eluted with 3 column volumes of storage buffer containing 250 mM imidazole and collecting the flow-through until the eluent was no longer yellow (color due to the enzymatically bound PLP cofactor). Imidazole was then removed using a PD10 salt exchange column.

Protein Characterization and Storage

Concentration measurement: Enzyme concentration was determined by Bradford assay, using bovine serum albumin for a standard concentration curve.

Gel Electrophoresis: Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using 12% polyacrylamide gels.

Storage: 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.

<u>Library Generation for Directed Evolution</u>

Production of UstD 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 µM MgCl$_2$, 7.5 µL 100 µM MnCl$_2$, 1 µL DMSO, 0.5 µL Taq polymerase (New England Biolabs) and diluted to a total volume of 50 µL with milliQ H$_2$O. Reactions were carried out in a thermocycle according to the following scheme:

<u>Thermocycler program</u>

Step 1: 95 °C 2 min 30 s

Step 2: 95 °C 15 s

Step 3: 54 °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-22b(+) 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 Terrific Broth (TB) media at 37 °C, 200 rpm, cells were plated onto LB plates with 100 µg/mL Ampicillin (amp) and incubated overnight. Single colonies were used to inoculate 5 mL TB + 100 µg/mL amp (TB$_{amp}$), which were grown overnight at 37 °C, 200 rpm. Colonies were sequenced and there was an average of 2 coding mutations.

Production of UstD degenerate codon libraries: Primers containing degenerate codons were purchased from IDT and are listed above. Mutagenesis was carried out via overlap-extension PCR. Reactions were setup by adding the following to a PCR tube: 10 μL 5x HF buffer (New England Biolabs), 1 μL 10 mM dNTP mix, 1 μL 10 μM forward primer, 1 μL 10 μM reverse primer, 1 μL ~100 ng/μL parent plasmid, 1 μL DMSO, 1 μL Phusion polymerase (New England Biolabs) and diluted to a total volume of 50 μL with milliQ H_2O. Reactions were carried out in a thermocycle according to the following scheme:

Thermocycler program

Step 1: 98 °C 1 min

Step 2: 98 °C 15 s

Step 3: 54 °C 20 s

Step 4: 72 °C 1 min

Step 5: 72 °C 5 min

Extension steps 2 – 4 were performed for at least 30 cycles.

The PCR product was purified using a preparative agarose gel. Purified DNA fragment was inserted into a pET-22b(+) 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 Terrific Broth (TB) media at 37 °C, 200 rpm, cells were plated onto LB plates with 100 μg/mL Ampicillin (amp) and incubated overnight. Single colonies were used to inoculate 5 mL TB + 100 μg/mL amp (TB$_{amp}$), which were grown overnight at 37 °C, 200 rpm. Colonies were sequenced to confirm correct incorporation of desired mutations.

Enzymatic Activity Experiments

General procedure for library generation and screening: Mutagenized plasmid DNA was

generated Electrocompetent BL21(DE3) were transformed with mutagenized plasmid DNA and

allowed to recover for 45 min in 800 μL of Terrific Broth (TB). After recovery, the cells were

plated onto LB plates containing 100 μg/mL ampicillin (LB$_{amp}$) and incubated overnight. A 96-

well plate containing 600 μL of TB$_{amp}$ per well was inoculated with single colonies. Each plate

included parent positive controls (from a fresh transformation), negative controls and a sterile

control that was not inoculated. The plates were grown overnight at 37 °C, 200 rpm. Expression

plates were prepared with 600-610 μL of TB$_{amp}$ per well and inoculated with 6-20 μL of

overnight culture. Glycerol stocks of each starter plate well were made using 150 μL of the

remaining culture and 100 μL of 60% sterile glycerol to ensure the sequence of any mutants of

interest could be determined. The expression plates were grown at 37 °C, 200 rpm for 3 h.

Expression plates were then placed on ice for 30 min. Cultures were induced with a final

concentration of 0.1 mM IPTG in 70 μL of fresh TB$_{amp}$ and, if necessary, diluted to a final

volume of 700 μL with fresh TB$_{amp}$. The expression culture was grown overnight at 20 °C, 200

rpm. Following overnight growth, the plate was centrifuged (4,000 x*g,* 30 min, 4 °C) and all

media was removed by striking plates against a paper towel on a table. Expression plates were

stored at -20 °C until further use.

A reaction master mix containing a final concentration of 50 mM L-asp, 50 μM PLP, and

buffer (100 mM KPi + NaCl, pH 7.0) was added to the thawed expression pellets using an

Opentrons OT-2 liquid handling robot. The pellets were resuspended by vortexing. Then, 50 mM

final concentration of electrophile mix (substrates varied throughout evolution), was added to the

reaction mixture by Opentrons OT-2 robot and reactions were allowed to incubate at 37 °C, 200

rpm for the desired reaction time (1-8 h). Subsequently, reactions were quenched with 300 µL (1 reaction volume) acetonitrile using Opentrons OT-2 robot and clarified at 4,000 x*g* for 30 min. The supernatant was transferred to a 0.2 µm centrifuge filter plate (PALL) and filtered at 1,500 rpm for 10 min into a clean Waters 96-well UPLC plate before being sealed prior to analysis by UPLC-MS.

Specific library generation and screening conditions for global random mutagenesis and P80X, F75X, G373X, and P82X site saturation libraries: Library generation and screening for these libraries follows the general procedure laid out above. The parent enzyme was **UstD$^{2.0}$** for these libraries. Expression plates were prepared with **600 µL** of TB$_{amp}$ per well and inoculated with **6 µL** of overnight culture. After induction, all expression plate wells had a final volume of **700 uL**. The enzymatic reaction time was **1 h** at 37 °C, 200 rpm with a 50 mM final concentration electrophile mix consisting of **4.2 mM thiophene-3-carboxyaldehyde, 4.2 mM *o*-tolualdehyde, and 41.7 mM (4-fluorophenyl)acetone**. The ketone:aldehyde ratio is **5:1** for these libraries.

Specific library generation and screening conditions for M299X, P83X, Y96X, Y277X, G101X, D86X, K342X, S407X, V330X, I141X, Y418X site saturation libraries: Library generation and screening for these libraries follows the general procedure laid out above. The parent enzyme was **UstD^{QE}** for these libraries. Expression plates were prepared with **600 µL** of TB$_{amp}$ per well and inoculated with **6 µL** of overnight culture. After induction, all expression plate wells had a final volume of **700 uL**. The enzymatic reaction time was **1 h** at 37 °C, 200 rpm with a 50 mM final concentration electrophile mix consisting of **10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone**. The ketone:aldehyde ratio is **9:1** for these libraries.

Specific library generation and screening conditions for distal recombination library: Library generation and screening for these libraries follows the general procedure laid out above. The parent enzyme was **UstD^{QE}** for these libraries. Expression plates were prepared with **600 µL** of TB$_{amp}$ per well and inoculated with **6 µL** of overnight culture. After induction, all expression plate wells had a final volume of **700 uL**. The enzymatic reaction time was **8 h** at 37 °C, 200 rpm with a 50 mM final concentration electrophile mix consisting of **10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone**. The ketone:aldehyde ratio is **9:1** for these libraries.

Specific library generation and screening conditions for active site mutagenesis libraries: Library generation and screening for these libraries follows the general procedure laid out above. For DNA generation, an equimolar primer mix of TTLM_4 mutation, TTLM_3 mutation_T388, TTLM_3 mutation_T391, TTLM_3 mutation_L392, TTLM_3 mutation_M393 was used as a doping strategy to lower the mutagenesis rate to 3.3 mutations per variant. All other primers were used in the standard fashion associated with overlap extension PCR. The parent enzyme was **AIIRQ** for these libraries. Expression plates were prepared with **610 µL** of TB$_{amp}$ per well and inoculated with **20 µL** of overnight culture. After induction, all expression plate wells had a final volume of **700 uL**. The enzymatic reaction time was **6 h** at 37 °C, 200 rpm with a 50 mM final concentration electrophile mix consisting of **10 mM (4-fluorophenyl)acetone, 32.5 mM 4'-nitroacetophenone, 5 mM *o*-tolualdehyde, 2.5 mM 1,1,1-trifluoro-3-phenyl-2-propanone**. The ketone:aldehyde ratio is **9:1** for these libraries.

Analysis of mutagenesis libraries

The relative amount of product formed in the reactions compared to the positive control reaction was measured by single ion retention mass analysis via UPLC/MS. Given the relatively

high variability in the parent signal in this assay, wells typically require an apparent 1.5-fold increase in product compared to the parent to be carried forward for validation. Using the glycerol stocks from the starter culture plate (described above), wells of interest could be streaked onto a fresh LB$_{amp}$ plate for subsequent sequencing and validation.

Validations of mutants of interest

Every validated mutant of interest was validated by heterologous expression in duplicate or triplicate reactions using whole cell, cell lysate or Ni-NTA purified enzyme reactions.

Validations of F75A: A reaction master mix (225 μL) containing a final concentration of 50 mM L-asp, 50 μM PLP, and buffer (100 mM KPi + NaCl, pH 7.0) was added to epi tubes. (4-fluorophenyl)acetone (15 μL, 50 mM final concentration, 5% DMSO in reaction mixture) was added to the epi tubes. Frozen cell stocks of *E. coli* cells expressing UstD$^{2.0}$+F75A were thawed to room temperature. The reaction was initiated upon addition of 60 μL of cell suspension (20 mg/mL final concentration). The reactions were allowed to incubate at 37 °C, 200 rpm for 1 h. Subsequently, reactions were quenched with 300 μL (1 reaction volume) acetonitrile and clarified at 16160 x*g* for 10 min. The supernatant was transferred to a UPLC vial and analyzed by UPLC-MS. The reactions were done in triplicate technical replicates at the same time as parent to determine the fold change most accurately.

Validations of G373E, P82Q, QE: A reaction master mix containing a final concentration of 50 mM L-asp, 5 μM PLP, and buffer (100 mM KPi + NaCl, pH 7.0) was added to epi tubes. The electrophile (4-fluorophenyl)acetone (30 μL, 50 mM final concentration, 5% DMSO in reaction mixture) was added to the epi tubes. The enzyme catalyst was thawed to room temperature and clarified at 16160 x*g* for 3 min. The reaction was initiated upon addition of enzyme (0.01 mol% catalyst, 10,000 Max TON). The reactions were allowed to incubate at 37 °C overnight.

Subsequently, reactions were quenched with 300 µL (1 reaction volume) acetonitrile and clarified at 16160 x*g* for 10 min. The supernatant was transferred to a UPLC plate and analyzed by UPLC-MS. The reactions were done in duplicate technical replicates at the same time as parent to determine the fold change most accurately.

Validations of P83T, P83Y, P83V, P83G, D86I, D86V, Y96V, Y96F, G101R, G101Q, G101A+ΔH445, Y277H, Y277F+W399C: A reaction master mix containing a final concentration of 50 mM L-asp, 50 µM PLP, and buffer (100 mM KPi + NaCl, pH 7.0) was added to epi tubes. The electrophile (4-fluorophenyl)acetone (15 µL, 50 mM final concentration, 5% DMSO in reaction mixture) was added to the epi tubes. The reaction was initiated upon addition of lysate (60 µL, 40 - 50 mg/mL final concentration). The reactions were allowed to incubate at 37 °C overnight. Subsequently, reactions were quenched with 300 µL (1 reaction volume) acetonitrile and clarified at 4300 rpm for 15 min. The supernatant (250 µL) was transferred to a filter plate (0.2 µm filter) and centrifuged at 1000 rpm into a UPLC plate and analyzed by UPLC-MS. The reactions were done in duplicate or triplicate technical replicates at the same time as parent to determine the fold change most accurately.

Validations of distal recombination and active site variants: A reaction master mix containing a final concentration of 50 mM L-asp, 5 µM PLP, and buffer (100 mM KPi, pH 7.0, 100 mM NaCl) was added to epi tubes. The electrophile (10 µL, 50 mM final concentration, 5% DMSO in reaction mixture) was added to the epi tubes. The enzyme catalyst was thawed to room temperature and clarified at 16160 x*g* for 3 min. The reaction was initiated upon addition of enzyme (0.01 mol% catalyst, 10,000 Max TON). The reactions were allowed to incubate at 37 °C for 16 h. Subsequently, reactions were quenched with 100 µL (1 reaction volume) acetonitrile and clarified at 4000 x*g* for 15 min. The supernatant was diluted and filtered through a 0.2 µm

PALL filter into a UPLC plate and analyzed by UPLC-MS. The reactions were done in triplicate technical replicates at the same time as parent to determine the fold change most accurately. Only a single electrophile was use per reaction, not a mixture. The electrophiles tested were *o*-tolualdehyde, (4-fluorophenyl)acetone, 4'-nitroacetophenone, and 1,1,1-trifluoro-3-phenyl-2-propanone. To analyze each reaction by UPLC, the enzymatic reactions were diluted different amounts. The *o*-tolualdehyde reactions were diluted 50x. (4-fluorophenyl)acetone reactions were diluted 3.3x. 4'-nitroacetophenone reactions were diluted 3.3x. 1,1,1-trifluoro-3-phenyl-2-propanone reactions were diluted 100x.

Discussion of Directed Evolution Strategy

Table 3S.4. Directed evolution summary with the sequences of variants with altered promiscuity or activity.

Round	Description	Clones Screened	Variants with altered promiscuity and/or activity in screening	Best Variant(s)
1	Random mutagenesis of the entire gene	880	G373R, D86V, F75S, Y96C+G101R, P82S+V330A, P83R, Y277C+K342E+S407N, P80L, H263R, Y418H	UstD$^{v2.0}$ All variants were generally deactivated but changed promiscuity. No mutations were fixed this round.
2	Site saturation of 'hotspots' with UstD$^{v2.0}$: F75, P80, P82, G373	352	F75: A, C, H, I, K, L, M, N, Q, R, S, T, W, Y P80: G, R P82: G, Q G373: E, R	P82Q G373E F75A
3	Double mutant P82Q+G373E	1	P82Q+G373E	QE = P82Q+G373E
4	Site saturation of 'hotspots' with QE as parent: P83, D86, Y96,	880	P83: G, V, T, Y D86: I, V G101: F, Q, A+ΔH445	None. Used mutational information from D86, I141, V330, and S407 in subsequent library

	G101, I141, Y277, V330, K342, S407, Y418,		I141: V, M, Y277: H, F+W399C V330: Q, A, R, C K342: none S407: Q, T, A, E Y418: none * silent mutations were fixed through primer design I141 (ATC→ ATT)	
5	Recombination at sites F75, D86, I141, V330, S407	704	F75A+D86I+I141V+S407E F75A+D86I+V330R+S407Q F75V+D86N+V330V F75A+D86V+I141V+S407Q F75A+D86V+I141M+S407E F75V+D86V+S407Q F75A+D86V+V330L+S407E F75A+D86I+V330A+S407Q F75V+D86V+S407E F75A+D86I+V330A+S407Q K2E+F75V+D86I+V330V F75A+D86V+V330Q+S407E *synonymous mutations not included	AIIRQ= QE+ F75A+ D86I+ I141I(ATC→ATT)+ V330R+ S407Q
6	Recombination at active site residues M299, T388, T391, L392, M393	968	M299L+T388V+T391F M299V+T388V M299V+T388V+T391S+L440P M299V+T388I+T391F+M393C M299L+T388I+T391S+M393C	7G11 = AIIRQ+ M299V+ T391S+ M393W 7B05 = AIIRQ+ T391S+ M393F

			M299V+T388I	
			T388V+T391F+M393S	
			T388I+T391S	
			T391S+M393W	
			C201C+M299V+T391S+M393F	
			M299L+T388A+T391F	
			M299V+M393W	
			M299V+T388V+M393W	
			M299V+T388V+M393W	
			T391S+M393F	
			M299V+T388A+T391F+M393W	
			M299V+T388I	
			M299V+T391S+M393W	
			M299V+T388V+T391F+M393F	
			M299L+T391S+M393W	
			M299V+T388I+T391S+M393F	
			M299V+T388I+T391F	
			M299V+T391S+L392A+M393W	
			T388I	
			V63I	
			M299V+T391F+M393W	
			M299V+T388I+M393F	
			M299V+T388I+M393L	
			M299V+T388A+T391F+M393W	
			M299L	
			T391F+M393W	
			T391F+L392I+M393W	
			T388A+T391F	

7	Site saturation of various residues both in and out of active site: Y257, H263, F285, P300, S389, H283. Each library was generated with 7G11 and 7B05 as the parent enzyme.	440	none	No variants surpassed 7G11 or 7B05 activity, so no further engineering was pursued.

Reaction Condition Optimization

Before testing additional substrates, we first attempted to optimize reaction conditions by observing the progress curve of a (4-fluorophenyl)acetone reaction with 7G11 in varying conditions. It was known to us that PLP is degraded over the course of the reaction via off-pathway deamination of alanine which is formed as a shunt product via protonation of the nucleophilic enamine intermediate. We hypothesized that increasing the amount of PLP in the reaction would allow for higher yield of the product. Therefore, we observed the reaction progress at varying amounts of PLP (10x, 20x, 50x). While each reaction appeared to reach the same yield, the 50x PLP reaction achieved the max yield and then the product was degraded in the reaction time. This indicated to us that there was enough PLP in the reaction to build-up and degrade the desired product indicating a sufficient excess for enzyme catalysis. Therefore, we elected to run reactions at 50x PLP for unactivated ketones. Next, we varied the concentration of L-asp (50 mM, 100 mM, 250 mM) in the reaction mixture to attempt to competitively inhibit product reentry and degradation. Any amount of additional L-aspartate increased the yield of the amino acid product and kept degradation from occurring.

When we repeated the same experiments using 1,1,1-trifluoro-3-phenyl-2-propanone as the substrate, we observed a worse yield with 50x PLP than 10x PLP. Therefore, we chose to use 10x PLP for highly activated ketone substrates. However, the reaction reached higher yield upon addition of 250 mM L-asp. Therefore, we determined our standard conditions to be 250 mM L-asp, 50 mM ketone, 10x PLP for activated substrates, 50x PLP for unactivated substrates, 4 hour reaction time, 37 °C, 5% cosolvent, 100 mM KPi pH 7.0, 100 mM NaCl.

<u>Ketone lineage analysis</u>

General Procedure: All reactions were done in triplicate on analytical scale (200 µL). The buffer used for all enzymatic reactions was 100 mM KPi, pH 7.0, 100 mM NaCl. PLP and L-aspartate stock solutions were made in water. Electrophile stock solutions were made in DMSO and each reaction only has a single electrophile. All UstD variant stock solutions concentrations were quantified by Bradford assay prior to setting up reactions. The difference enzyme concentration between variants was corrected for prior to addition into the reactions. All samples were analyzed following Marfey's derivatization by Waters Acquity UPLC-MS using a BEH C18 column (Waters). To correct for small deviations in injection volume, an internal derivatization standard was included (0.098 mM L-arginine). Derivatized amino acid product quantitation was performed by integrating chromatograph peaks at 340 nm and corrected by dividing by the internal standard peak area. To calculate product concentrations, a standard curve was generated by subjecting stock solutions of L-phenylalanine (50 mM – 0.4 mM) in water to the identical procedure used to process and derivatize enzymatic reaction solutions, in triplicate, with internal standard. These curves were used to calculate the concentrations of UstD products in solution, and subsequently total turnover numbers after dilution factor correction.

Enzymatic reactions: Each well of a 96-well plate was charged with electrophile (10 μmol, 1 equiv., 50 mM final concentration, 5% DMSO final concentration). A reaction master mixture containing L-aspartate sodium salt monohydrate (50 μmol, 5 equiv., 250 mM final concentration), pyridoxal-5'-phosphate (10 or 50 molar equivalents to final enzyme concentration, see **Table 3S.5**) and buffer was aliquoted into each of the wells. The plate was vortexed gently to mix. Reactions were initiated by addition of UstD (0.1-0.003 mol% catalyst, 1,000-30,000 max turnover number, see **Table 3S.5**) to bring the total reaction volume to 200 μL. The 96-well plate was sealed with a silicon lid and placed at 37 °C for 4 h. Reactions were quenched with 200 μL of acetonitrile (1 reaction volume) and diluted with 200 μL of 1:1 ACN:DI H_2O to homogenize reaction solutions. Denatured enzyme was removed by centrifugation at 2000 rpm for 10 min. The resulting supernatant was passed through a 0.2 μm PALL filter plate to remove any remaining particulates of enzyme prior to derivatization. Marfey's derivatization of the clarified enzymatic reactions were performed to quantify amino acid yield, results shown in **Table 3S.6**.

Marfey's derivatization procedure: To a fresh 96-well plate, 10 μL of quenched reaction mixture (2.7 mM total amines in reaction, 1 equiv., ~0.8 μmol total amines in reaction) 140 μL of 10.41 mM $NaHCO_3$ with 0.21 mM of L-arginine as an internal standard (9.7 mM $NaHCO_3$ final concentration, 3.5 equiv. base, 2.9 μmol $NaHCO_3$), and 10 mM L-FDAA (5 mM final concentration, 1.8 equiv., 1.5 μmol) were added in each well. The derivatization reaction was allowed to proceed at 37 °C for 18 h. The reactions were quenched with 300 μL of 60 mM HCl in acetonitrile (1 reaction volume) and analyzed via UPLC-MS no later than 24 h after quenching. Note that the amino acid products are susceptible to dehydration upon addition of the acid required to quench the reactions resulting in two product peaks with masses differing by 18

mass units. The dehydration product could result from elimination of the alcohol to form an

alkene or intramolecular lactonization. Turnover numbers were calculated based on the total

integration of linear and lactonized amino acid product peaks at 340 nm.

Table 3S.5 Lineage reaction conditions by electrophile.

Electrophile	Molar equivalents PLP	Max TON	mol% catalyst
benzaldehyde	10	30000	0.003
1,1,1-trifluoro-3-phenyl-2-propanone	10	5000	0.02
2-furylacetone	50	2500	0.04
(4-fluorophenyl)acetone	50	1000	0.1
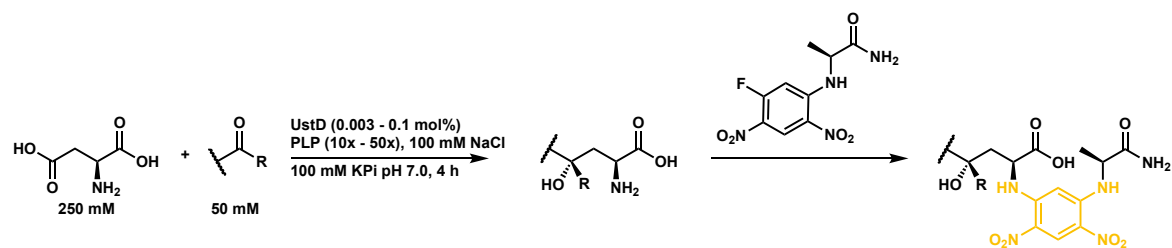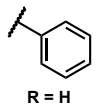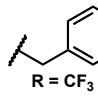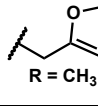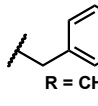
Table 3S.6. Turnover numbers for ketone lineage.

$\lambda_{max} = 340$ nm

Variant	UstD[2.0]	SA	Q	QE	AIIRQ	7G11	7B05
Product	Avg TTN	Avg TTN	Avg TTN	Avg TTN	Avg TTN	Avg TTN	Avg TTN
R = H	9700 ± 500	3900 ± 200	4300 ± 100	4300 ± 300	3800 ± 200	1800 ± 100	4900 ± 200
R = CF₃	480 ± 50	1110 ± 90	1000 ± 100	1070 ± 40	1900 ± 100	1560 ± 50	2120 ± 40
R = CH₃	126 ± 6	160 ± 10	160 ± 10	200 ± 20	370 ± 30	740 ± 10	580 ± 10
R = CH₃	trace	~ 10	~ 10	17 ± 1	18 ± 1	67 ± 6	29 ± 2

3.9 References

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

2. France, S. P., Lewis, R. D. & Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au* **3**, 715–735 (2023).

3. Radley, E. *et al.* Engineering Enzymes for Environmental Sustainability. *Angew. Chemie Int. Ed.* **62**, e202309305 (2023).

4. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chemie Int. Ed.* **60**, 88–119 (2021).

5. Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angew. Chemie Int. Ed.* **50**, 138–174 (2011).

6. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5618–22 (1993).

7. Ellis, J. M. *et al.* Biocatalytic synthesis of non-standard amino acids by a decarboxylative aldol reaction. *Nat. Catal. 2022 52* **5**, 136–143 (2022).

8. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science.* **378**, 49–56 (2022).

9. Dantas, G. *et al.* High-resolution Structural and Thermodynamic Analysis of Extreme Stabilization of Human Procarboxypeptidase by Computational Protein Design. *J. Mol. Biol.* **366**, 1209–1221 (2007).

10. Honda Malca, S. *et al.* Effective engineering of a ketoreductase for the biocatalytic

synthesis of an ipatasertib precursor. *Commun. Chem. 2024 71* **7**, 1–11 (2024).

11. Thorpe, T. W., Marshall, J. R. & Turner, N. J. Multifunctional Biocatalysts for Organic Synthesis. *J. Am. Chem. Soc.* (2024). doi:10.1021/JACS.3C09542

12. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chemie Int. Ed.* **44**, 4192–4196 (2005).

13. Gu, J., Xu, Y. & Nie, Y. Role of distal sites in enzyme engineering. *Biotechnol. Adv.* **63**, 108094 (2023).

14. Romney, D. K., Sarai, N. S. & Arnold, F. H. Nitroalkanes as Versatile Nucleophiles for Enzymatic Synthesis of Noncanonical Amino Acids. *ACS Catal.* **9**, 8726–8730 (2019).

15. Blikstad, C., Dahlström, K. M., Salminen, T. A. & Widersten, M. Substrate scope and selectivity in offspring to an enzyme subjected to directed evolution. *FEBS J.* **281**, 2387–2398 (2014).

16. Matsumura, I. & Ellington, A. D. In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds Through Non-specific Intermediates. *J. Mol. Biol.* **305**, 331–339 (2001).

17. Mcdonald, A. D., Higgins, P. M. & Buller, A. R. Substrate multiplexed protein engineering facilitates promiscuous biocatalytic synthesis. doi:10.1038/s41467-022-32789-w

18. Jakoblinnert, A. *et al.* Reengineered carbonyl reductase for reducing methyl-substituted cyclohexanones. *Protein Eng. Des. Sel.* **26**, 291–298 (2013).

19. Junker, S., Roldan, R., Joosten, H. J., Clapés, P. & Fessner, W. D. Complete Switch of Reaction Specificity of an Aldolase by Directed Evolution In Vitro: Synthesis of Generic

Aliphatic Aldol Products. *Angew. Chemie Int. Ed.* **57**, 10153–10157 (2018).

20. Knorrscheidt, A. *et al.* Simultaneous screening of multiple substrates with an unspecific peroxygenase enabled modified alkane and alkene oxyfunctionalisations. *Catal. Sci. Technol.* **11**, 6058–6064 (2021).

21. Stanišić, A., Hüsken, A. & Kries, H. HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes. *Chem. Sci.* **10**, 10395–10399 (2019).

22. Fujimori, D. G. *et al.* Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16498–16503 (2007).

23. Morokuma, K. *et al.* Total synthesis of lycoperdic acid and stereoisomers (part 2): Completion of the synthesis and biological evaluation. *Tetrahedron* **145**, 133623 (2023).

24. Maharaj, V. J., Moodley, N. & Vahrmeijer, H. Characterization of natural monatin isomers, a high intensity sweetener from the plant Sclerochiton ilicifolius from South Africa. *South African J. Bot.* **115**, 37–43 (2018).

25. Li, Z. *et al.* Kinetics and mechanism of oxirane formation by darzens condensation of ketones: Quantification of the electrophilicities of ketones. *J. Am. Chem. Soc.* **140**, 5500–5515 (2018).

26. Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry*. (AIP Publishing, 2006).

27. Weeks, A. M. & Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol. 2017 141* **14**, 50–57 (2017).

28. Nath, A. & Atkins, W. M. A quantitative index of substrate promiscuity. *Biochemistry* **47**, 157–166 (2008).

29. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the Range of

Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chemie Int. Ed.* **44**, 4192–4196 (2005).

30. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc. 2007 24* **2**, 891–903 (2007).

31. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nat. 1994 3706488* **370**, 389–391 (1994).

32. Voigt, C. A., Martinez, C., Wang, Z. G., Mayo, S. L. & Arnold, F. H. Protein building blocks preserved by recombination. *Nat. Struct. Biol. 2002 97* **9**, 553–558 (2002).

33. Savile, C. K. *et al.* Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. (2010). doi:10.1126/science.1188934

34. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol. 2009 1012* **10**, 866–876 (2009).

35. 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).

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

CHAPTER FOUR: BIOCATALYTIC SYNTHESIS OF CHIRAL TERTIARY ALCOHOL

NON-STANDARD AMINO ACIDS

4.1 Introduction

Chiral tertiary alcohols represent a class of industrially relevant compounds.[1,2] Non-standard amino acids (nsAAs) bearing a chiral tertiary alcohol have been shown to have a variety of industrial and biologically relevant properties (**Fig. 4.1**).[1–3] For example, kutzneride 1 has antifungal properties and monatin is used as a low-cal sweetner.[4,5] However, direct methods to generate this motif remain limited.
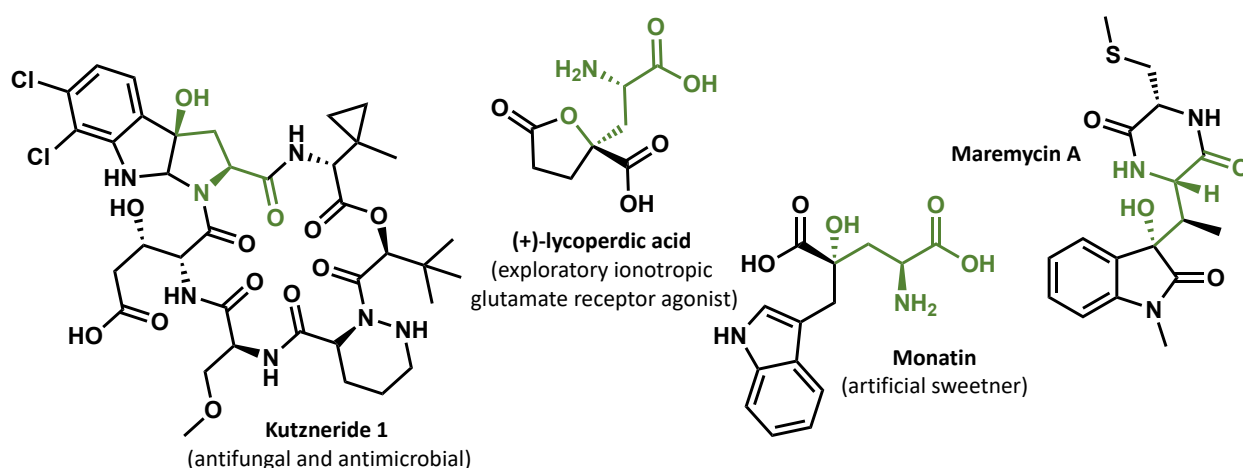
Figure 4.1 Chiral tertiary alcohol containing amino acids.

Generation of nsAAs bearing a chiral tertiary alcohol are challenging for two reasons. First, installation of both an amine and an alcohol in a stereoselective fashion for small molecules remains synthetically challenging.[6–9] While there are biocatalytic methods to generate amino acids, these often require enzyme cascades to sequentially incorporate both functional groups reducing their synthetic utility.[10,11] Second, chiral tertiary alcohols are typically generated using aldol addition into a ketone electrophile which is thermodynamically and kinetically more challenging compared to aldol addition into an aldehyde electrophile.[12–16] Direct, stereoselective methods of performing these aldol reactions suffer from limitations whether traditional synthetic or biocatalytic approaches are pursued.

The three main synthetic approaches for direct, catalytic aldol addition into ketones are Mukaiyama-aldol, decarboxylative aldol, and organocatalyst mediated aldol (**Fig. 4.2A**). The Mukaiyama-aldol reaction requires pre-formation of the enolate. This method has many iterations, but only a handful of methods demonstrated utility with unactivated ketones as electrophiles.[17–19] Unfortunately, the synthetic utility of the reactions were hindered by long reaction times, harsh conditions, low stereoselectivity, or small scopes.[17–19]
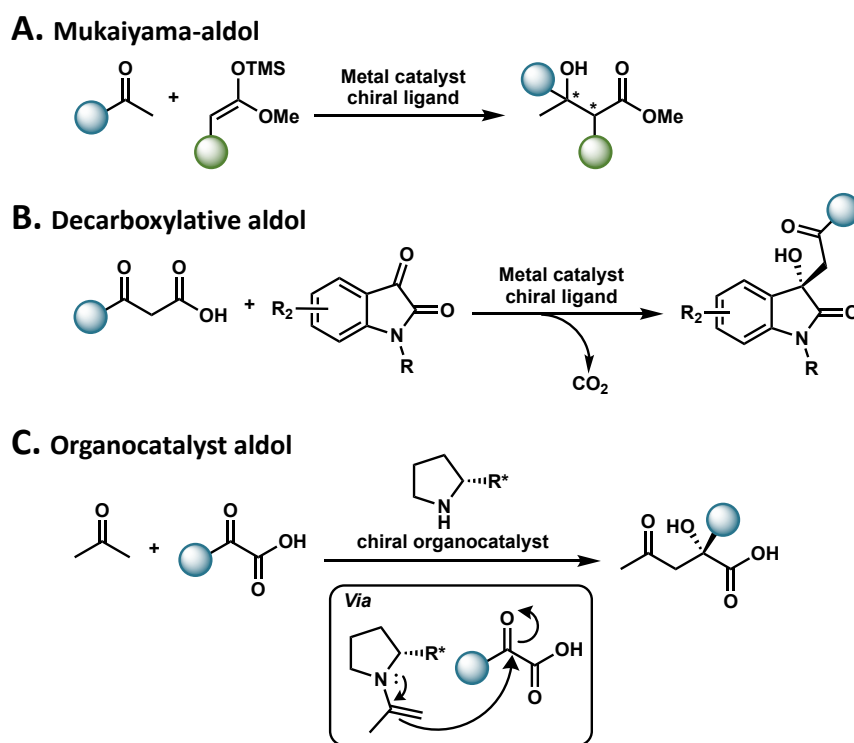
Figure 4.2 Aldol reaction types. A. Mukaiyama aldol using pre-formed enolates. **B.** Decarboxylative aldol where the nucleophile is formed *in situ* upon decarboxylation of the pro-nucleophile. **C.** Organocatalyst mediated aldol reactions where the nucleophile is generated through covalent attachment to the catalyst.

Decarboxylative aldol strategies generate the nucleophile *in situ* using metal catalyzed addition into activated diketones, such as isatin or α-ketoacid derivatives (**Fig. 4.2B**).[20–23] Again, the reactions require high catalyst loadings and long reaction times usually over 24 h. Lastly, enamine catalysis using organocatalysts can be used as biomimetic methods to generate chiral

tertiary alcohols (**Fig. 4.2C**).[24–26] However, the organocatalyst methods have limited reaction scope, low enantio-enrichment (*ee*), long reaction times, and require activated ketones to work.[24–26] Despite several decades of research, development of aldol reactions for the synthesis of chiral tertiary alcohols are still challenging and exploring alternative methods may provide solutions.[16]

Enzymes are an attractive alternative to the synthetic methods as they work in mild conditions and have exquisite regio- and stereoselectivity.[27,28] Similar to synthetic methods, biocatalytic strategies for generating chiral tertiary alcohols include aldol addition of a carbon nucleophile into a ketone electrophile. Direct addition of carbon nucleophiles can be mediated using aldolases and thiamin pyrophosphate (TPP) dependent enzymes. Classical aldolases have historically struggled to produce synthetically relevant quantities of desired tertiary alcohol product due to the inherent reversibility of the reaction.[29] Only a few examples exist in the literature demonstrating addition into ketones with these enzymes is possible, but scope and scalability are limited.[29–33] For example, dihydroxyacetone phosphate (DHAP)-dependent rhamnulose aldolases from *Bacteroides thetaiotaomicron* has been shown to catalyze addition into a small number of ketones bearing α-hydroxyl groups in high yield and good diastereoselectivity (**Fig. 4.3A**).[32] TPP dependent enzymes such as YerE and acetylacetoin synthase (AAS) have demonstrated addition of a carbon nucleophile into ketone substrates (**Fig. 4.3B**).[34–36] In both cases the scope was limited with moderate product yields and low *ee*. [34–36] A proteinase from *Aspergillus melleus* (AMP) catalyzes an asymmetric aldol reaction between β,γ-unsaturated α-keto esters and ketones affording tertiary chiral alcohols in moderate to good yield with moderate to poor stereoselectivity (**Fig. 4.3C**).[37] In this case, both the electrophile and nucleophile could be modified, but only minor changes were tolerated limiting synthetic utility.[37] In addition to the

limited scope, all of the aforementioned methods only install the chiral tertiary alcohol functional

group. Therefore, generation of the amino group to generate the desired amino acid motif would

require additional steps to install.

Figure 4.3 Biocatalytic methods for generating chiral tertiary alcohols.

Given the aforementioned limitations, direct methods to generate nsAAs bearing a chiral

tertiary alcohol remain sparse. Any method capable of providing this functionality in a

streamlined fashion would represent a significant synthetic advance. While some C–H activation

strategies exist using *cis*-4-proline hydroxylases (P4H) and isoleucine dioxygenase (IDO), these

enzymes are not building the carbon framework and are better suited to late-stage

functionalization.[38,39] A recent report demonstrated a wild-type UstD enzyme from *Aspergillus*

pseudonomius, a homolog of the UstD used in this thesis, was capable of catalyzing aldol

addition into ketones, but it was limited to activated di-ketone substrates only (**Fig. 4.3D**).[2]

Given the many challenges associated with making chiral tertiary alcohol nsAAs, our extensive

engineering of UstD$^{2.0}$ using SUMS to increase activity on ketone electrophiles may provide a

route to amino acids that were not previously accessible. To this end, the two ketone generalists

from our engineering, 7B05 and 7G11 were investigated to determine their utility for synthesis

of chiral tertiary alcohol nsAAs. The work in this chapter is ongoing and represents the progress

made in this area so far.

4.2 Exploration of biocatalytic synthesis of chiral tertiary alcohols

4.2.1 Survey of substrate scope on analytical scale

To test synthetic utility of the enzymes, analytical reactions were conducted to probe the

reaction scope under the optimized conditions reported in chapter 3 (**Fig. 4.4**). The electrophile

substrate loading was 50 mM while an excess of L-asp (5 eq.) was added to decrease retro-aldol

cleavage. Enzyme loading in all the reactions was set to 0.1 mol% catalyst loading (1000 Max

TON). PLP was added at 10-molar equivalents relative to enzyme concentration for substrates

bearing a trifluoromethyl ketone and 50-molar equivalents for all other substrates. Yield for each

reaction was quantified by Marfey's derivatization as described in the SI.

In general, both enzymes, 7B05 and 7G11 show broad activity with aromatic and

aliphatic substrates. The diastereoselectivity varies substrate to substrate but is higher when the

substituents connected to the carbonyl are sterically distinct. Presumably, the more pronounced

the steric differences in the carbonyl substituents the greater the discrimination between binding

poses leading to higher diastereoselectivity for those substrates.[14] This hypothesis is most cleanly

supported by the trend in diastereoselectivity for the alkyl trifluoromethyl ketones bearing hexyl-

butyl- and methyl- substituents. Both the hexyl and butyl substrates have high

diastereoselectivity, but trifluoroacetone has no selectivity with 7G11 and a modest 5:2 d.r. with

7B05.
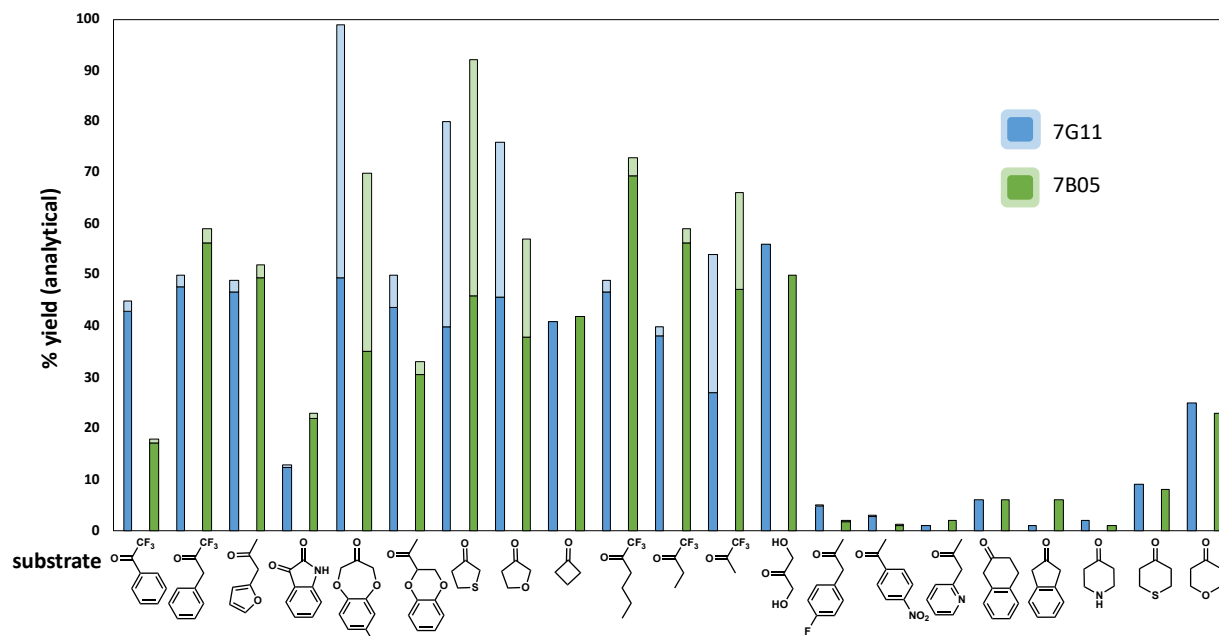
Figure 4.4 Analytical evaluation of substrate scope. All yields were quantified by Marfey's derivatization against a standard curve as described in the SI. The lighter colored bars represent the amount of syn diastereomer formed with the assumption the anti-diastereomer is the major isomer. Each entry represents only a single reaction. The blue bars are reactions with 7G11 and the green bars are reactions with 7B05. **Conditions:** 250 mM L-asp, 50 mM electrophile, 10x or 50x PLP compared to enzyme, 5% DMSO, 100 mM NaCl, 100 mM potassium phosphate pH 7.0, 7B05 or 7G11 (0.1 mol% cat, 1,000 Max TON, 37 °C, 4 h reaction time.

7B05 can convert substrates bearing a trifluoromethyl on the carbonyl with more efficiency than 7G11, albeit the difference is modest. Meanwhile, 7G11 can convert unactivated ketones more efficiently. Structurally, the four and five-membered rings are more readily accepted by the enzyme than the six-membered rings. Neither enzyme tolerates a 2-substituted pyridine ring. Despite extensive engineering, the two unactivated ketones, (4-fluorophenyl)acetone and 4-nitroacetophenone, have not reached synthetically useful levels of product formation. Nevertheless, we wanted to demonstrate synthetic utility of the enzymes. Therefore, we chose to isolate and characterize a subset of nsAAs on preparative scale. Of the

listed entries, substrates were selected based on chemical diversity to demonstrate the widest breadth of amino acids that can be synthesized.

4.2.2 Preparative scale biocatalytic synthesis of chiral tertiary alcohols

Efficient isolation of aromatic containing amino acids was possible using reverse-phase flash chromatography (**Fig. 4.5,** see SI for further details). Beginning with **4.1**, a good yield of 55% and excellent diastereoselectivity (>20:1) was observed. The enzyme converted TFMK, one of the substrates under selection during evolution, to **4.2** in 46% yield with excellent d.r. (>20:1). Excitingly, the hetereocyclic amino acid, **4.3** was isolated at 39% yield and excellent d.r. (>20:1). Product **4.4** was isolated at 31% yield and high d.r. (18:1). The high diastereoselectivity was particularly exciting in this case as there are three stereocenters present in this product. A crystal structure of the isolated product confirmed the absolute configuration of the major isomer to be *S,S,R*, with the anti-configuration of the α-amine to the γ-hydroxy groups conserved (**Fig. 4.S1,2)**. A similar ring motif to **4.4** is found in **4.5**, which was isolated in excellent yield (96%) as a 1:1 mixture of diastereomers. The ketone substrate for **4.5** is a synthetic fragrance molecule known as calone or "watermelon ketone" and is used commercially to convey a sea breeze with floral overtones.[40] Unfortunately, the fragrance properties of calone are abolished during the reaction leaving the amino acid odorless.

The aliphatic nsAA, **4.6** was sufficiently hydrophobic to be isolated using reverse phase chromatography in good yield (59%) as a 1:1 mixture of diastereomers. While both diastereomers were structurally characterized by small molecule crystallography, only the anti-isomer is displayed here for clarity (**Fig. 4.S3, Table 4.S2**). The more hydrophilic and aliphatic amino acids **4.7** and **4.8** were unable to be isolated by reverse phase chromatography. Instead, the enzymatic reaction was followed by a standard Fmoc protection reaction in a telescoped

fashion. The protected amino acids were subsequently isolated using normal-phase chromatography. In this way, the amino acid, **4.7**, was isolated as an Fmoc protected product in 62% yield. The sugar-like amino acid **4.8**, derived from reaction with 1,3-dihydroxyacetone, was isolated in 71% yield as the Fmoc protected product. For **4.8**, the enzymatic reaction was run at a higher catalyst loading (250 Max TON, 0.4 mol% catalyst) to demonstrate that increasing catalyst loading can improve yield.
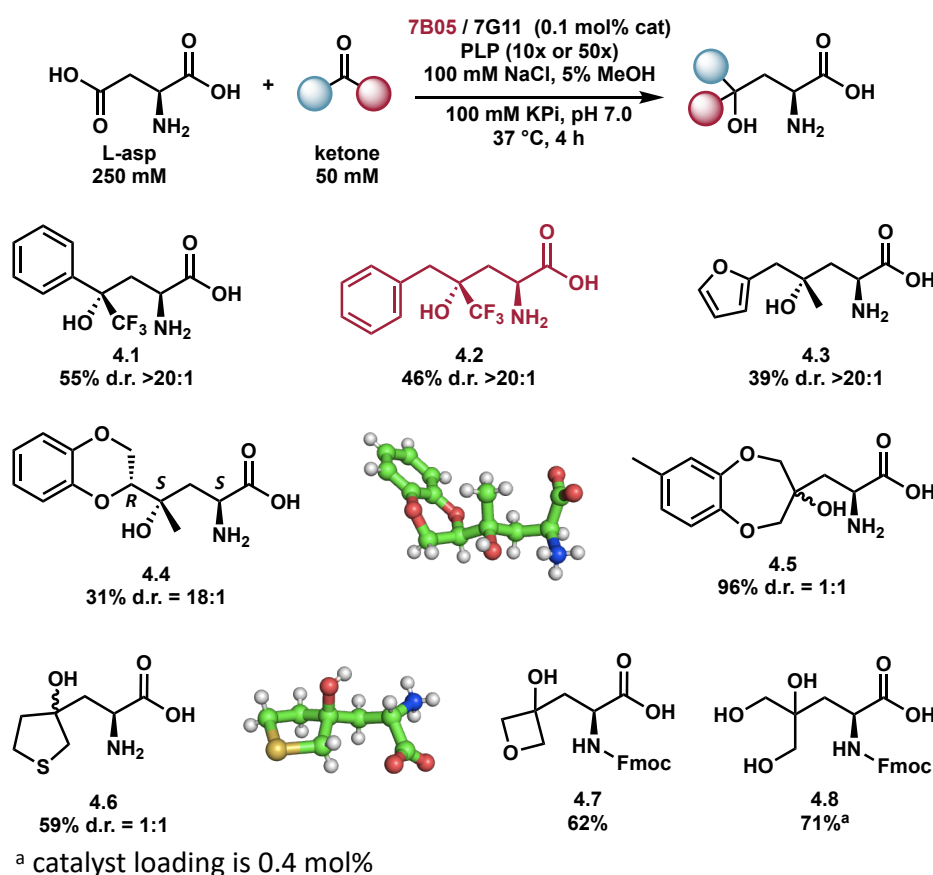
Figure 4.5 Preparative scale biocatalytic reactions.

4.3 Conclusions

Here, we have explored the scope of 7B05 and 7G11 for aldol addition into ketones. We isolated a small set of chemically diverse amino acid products in moderate to good yield using a

single enzyme system. While there are examples in the literature of biocatalytic C–C bond formation into ketone electrophiles, many of these enzymes have limited scopes making practical synthesis difficult. Here, the engineered enzymes 7G11 and 7B05 are capable of catalyzing aldol addition into many ketones, providing access to previously inaccessible amino acids. Additionally, the enzymes react with a many of these ketones in a stereoselective fashion affording many of the products as single diastereomers. This streamlined single enzyme system represents a significant advance towards developing biocatalytic platforms for C–C bond formation. Future work will encompass mechanistic investigations into the specific molecular determinants of promiscuity for each enzyme to aid in future development of new synthetic applications of UstD and its engineered variants.

4.4 Supplementary figures, tables, and information
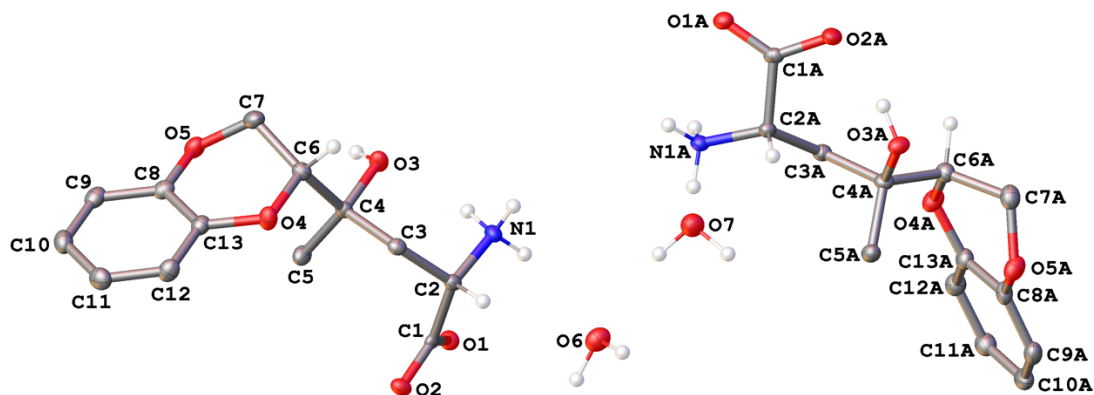
Figure 4.S1. A molecular drawing of the asymmetric unit in 4.4 shown with 50% probability ellipsoids. All H atoms (except those bound to N/O atoms or chiral centers) are omitted. The absolute configuration of the chiral atoms is C2–S, C4–S, C6–R, C2A–S, C4A–S, and C6A–R.
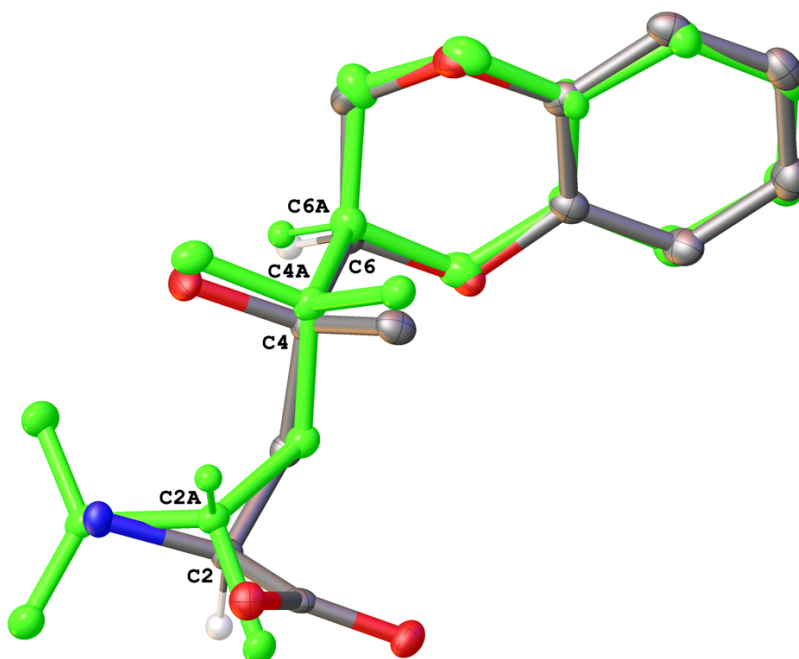
Figure 4.S2. A molecular drawing overlaying the two symmetry independent molecules in 4.4 shown with 50% probability ellipsoids. All H atoms (except those bound to chiral centers) and solvent molecules are omitted. The second symmetry independent molecule is shown in green. The absolute configuration of the chiral atoms is C2–S, C4–S, C6–R, C2A–S, C4A–S, and C6A–R.

Table 4.S1. Crystal data and structure refinement for 4.4

Empirical formula	$C_{13}H_{17}NO_5 \cdot H_2O$
Formula weight	285.29
Temperature/K	100.00
Crystal system	triclinic
Space group	$P1$
a/Å	5.7856(6)
b/Å	6.2026(6)
c/Å	20.699(2)
α/°	94.081(6)
β/°	90.513(6)

γ/°	115.481(6)
Volume/Å^3	668.19(13)
Z	2
ρ_{calc}g/cm^3	1.418
μ/mm^{-1}	0.951
F(000)	304.0
Crystal size/mm^3	0.055 × 0.041 × 0.011
Radiation	Cu Kα (λ = 1.54178)
2Θ range for data collection/°	8.574 to 144.67
Index ranges	-7 ≤ h ≤ 7, -7 ≤ k ≤ 7, -25 ≤ l ≤ 25
Reflections collected	22465
Independent reflections	5087 [R_{int} = 0.0300, R_{sigma} = 0.0235]
Data/restraints/parameters	5087/56/399
Goodness-of-fit on F^2	1.060
Final R indexes [I>=2σ (I)]	R_1 = 0.0260, wR_2 = 0.0651
Final R indexes [all data]	R_1 = 0.0266, wR_2 = 0.0656
Largest diff. peak/hole / e Å^{-3}	0.20/-0.21
Flack parameter	-0.02(6)
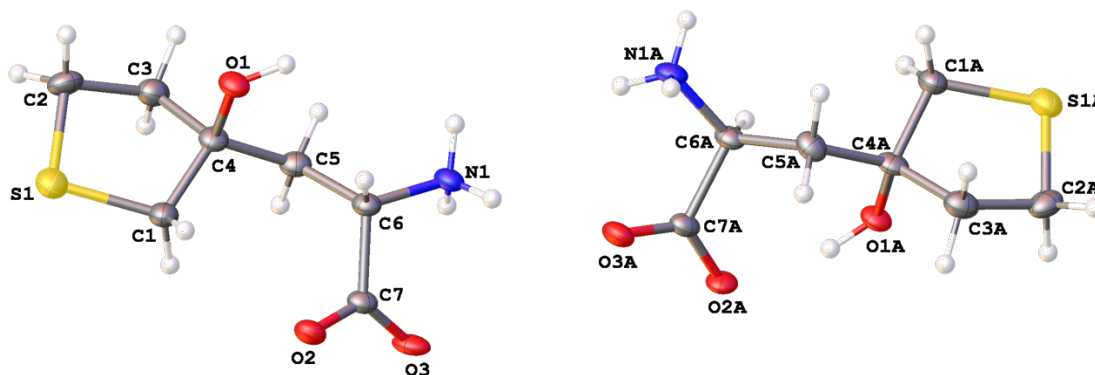
Figure 4.S3. A molecular drawing of **4.6** shown with 50% probability ellipsoids. The two molecules are diastereomers.

Table 4.S2. Crystal data and structure refinement for 4.6

Empirical formula	$C_{14}H_{26}N_2O_6S_2$
Formula weight	382.49
Temperature/K	100.00
Crystal system	triclinic
Space group	P1
a/Å	5.4808(7)
b/Å	5.6991(7)
c/Å	13.8474(14)
α/°	80.745(8)
β/°	87.828(6)
γ/°	79.666(9)
Volume/Å^3	419.96(9)
Z	1
ρ_{calc}g/cm^3	1.512
μ/mm^{-1}	3.187
F(000)	204.0

Crystal size/mm^3	0.08 × 0.02 × 0.02
Radiation	CuKα (λ = 1.54178)
2Θ range for data collection/°	6.468 to 160.754
Index ranges	-6 ≤ h ≤ 6, -7 ≤ k ≤ 7, -17 ≤ l ≤ 17
Reflections collected	15945
Independent reflections	3413 [R_{int} = 0.0636, R_{sigma} = 0.0514]
Data/restraints/parameters	3413/20/241
Goodness-of-fit on F^2	1.090
Final R indexes [I>=2σ (I)]	R_1 = 0.0423, wR_2 = 0.1111
Final R indexes [all data]	R_1 = 0.0436, wR_2 = 0.1121
Largest diff. peak/hole / e Å^{-3}	0.32/-0.38
Flack parameter	0.014(16)
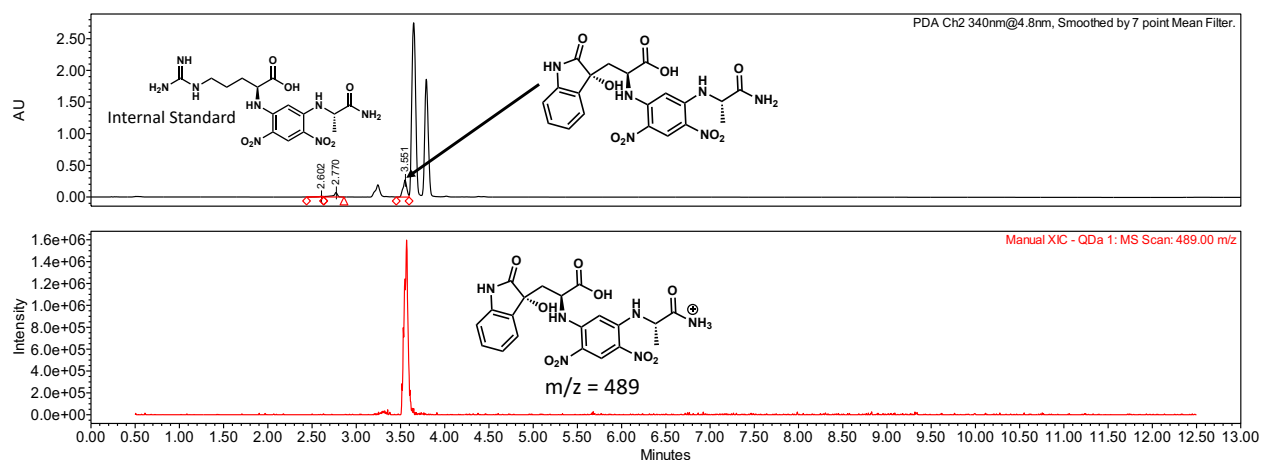
Figure 4.S4. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with isatin. **Conditions:** 50 mM isatin, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
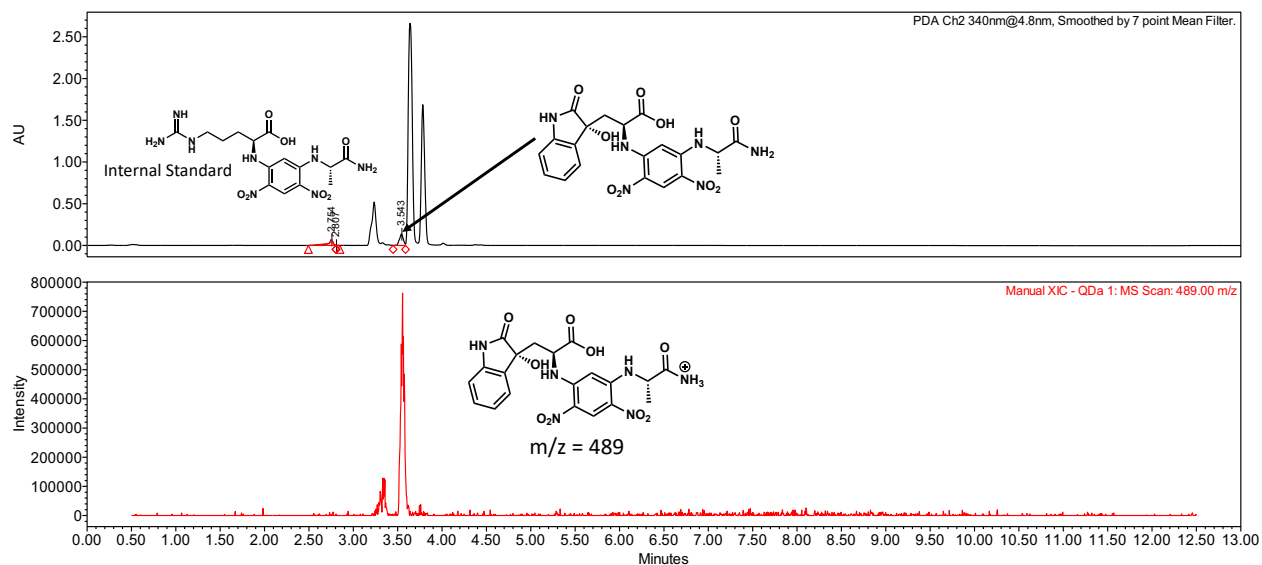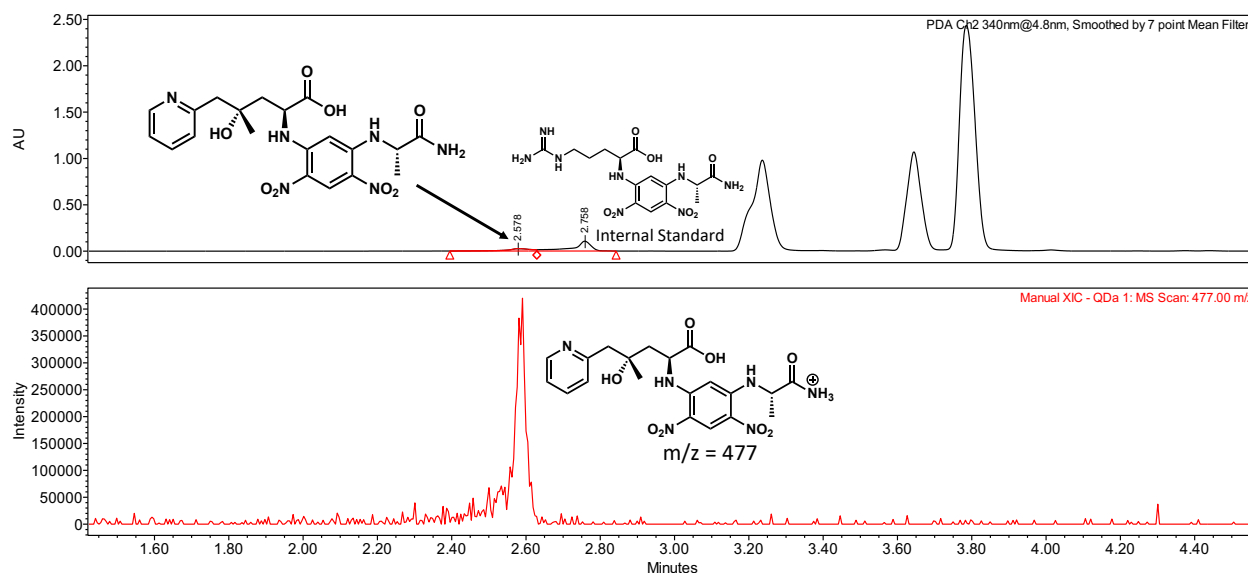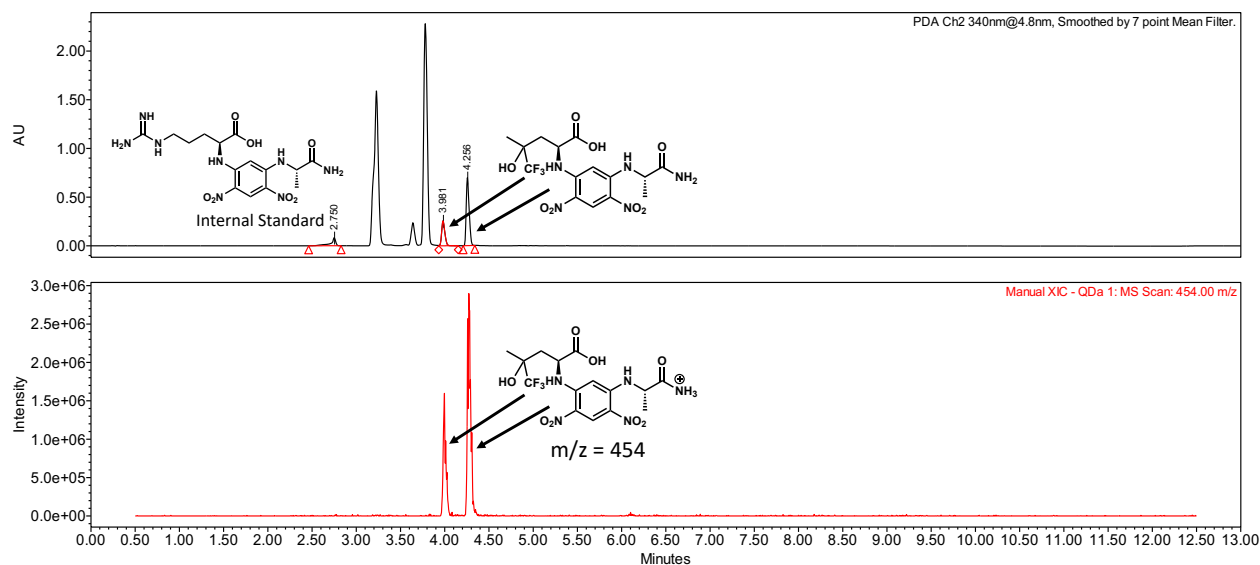
Figure 4.S5. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with isatin. **Conditions:** 50 mM isatin, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
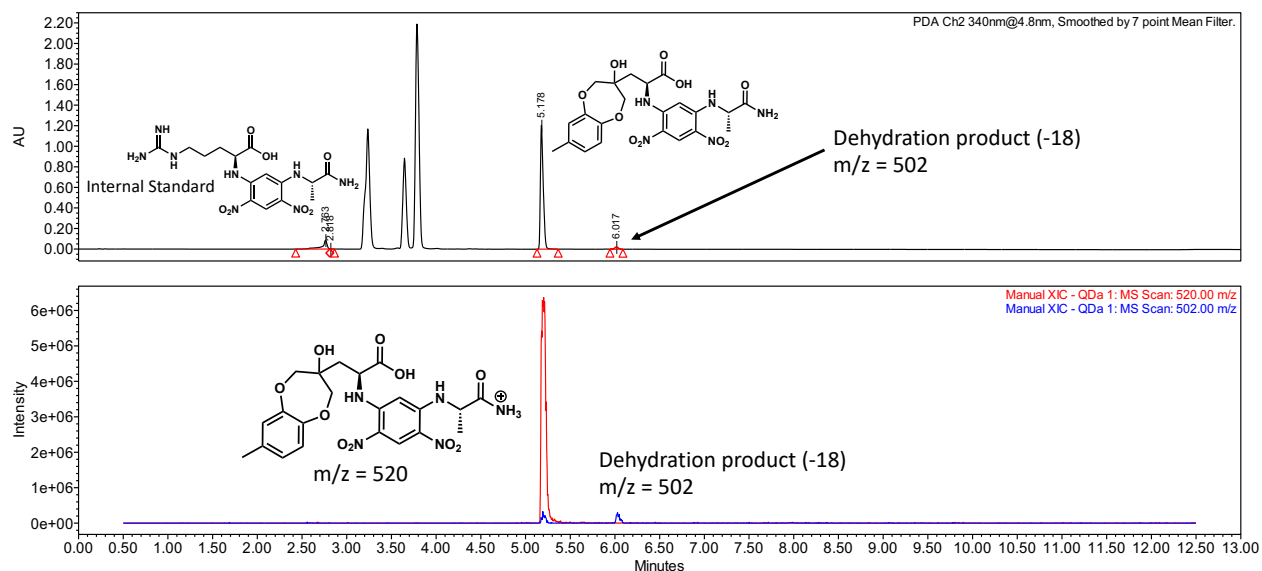
Figure 4.S6. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with calone. **Conditions:** 50 mM calone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
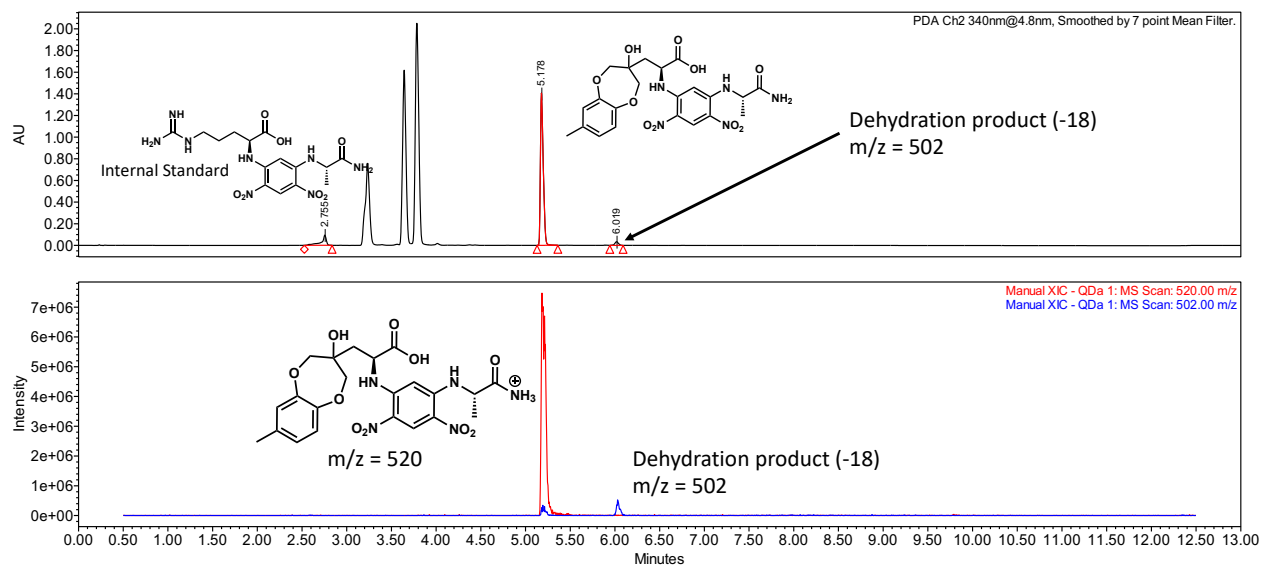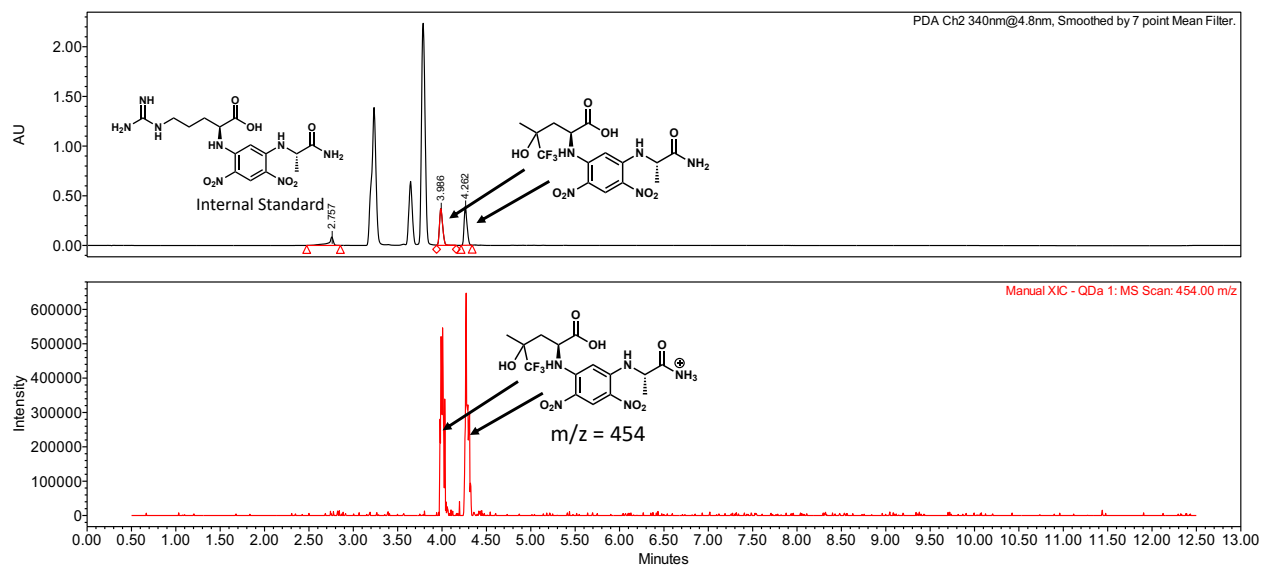
Figure 4.S7. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with calone. **Conditions:** 50 mM calone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
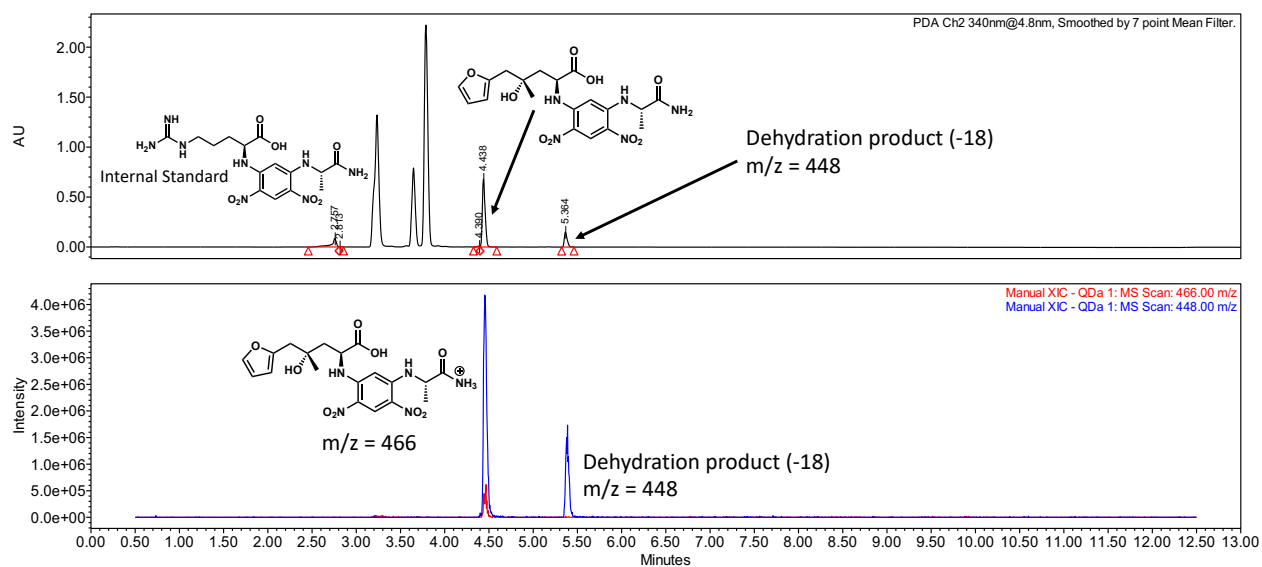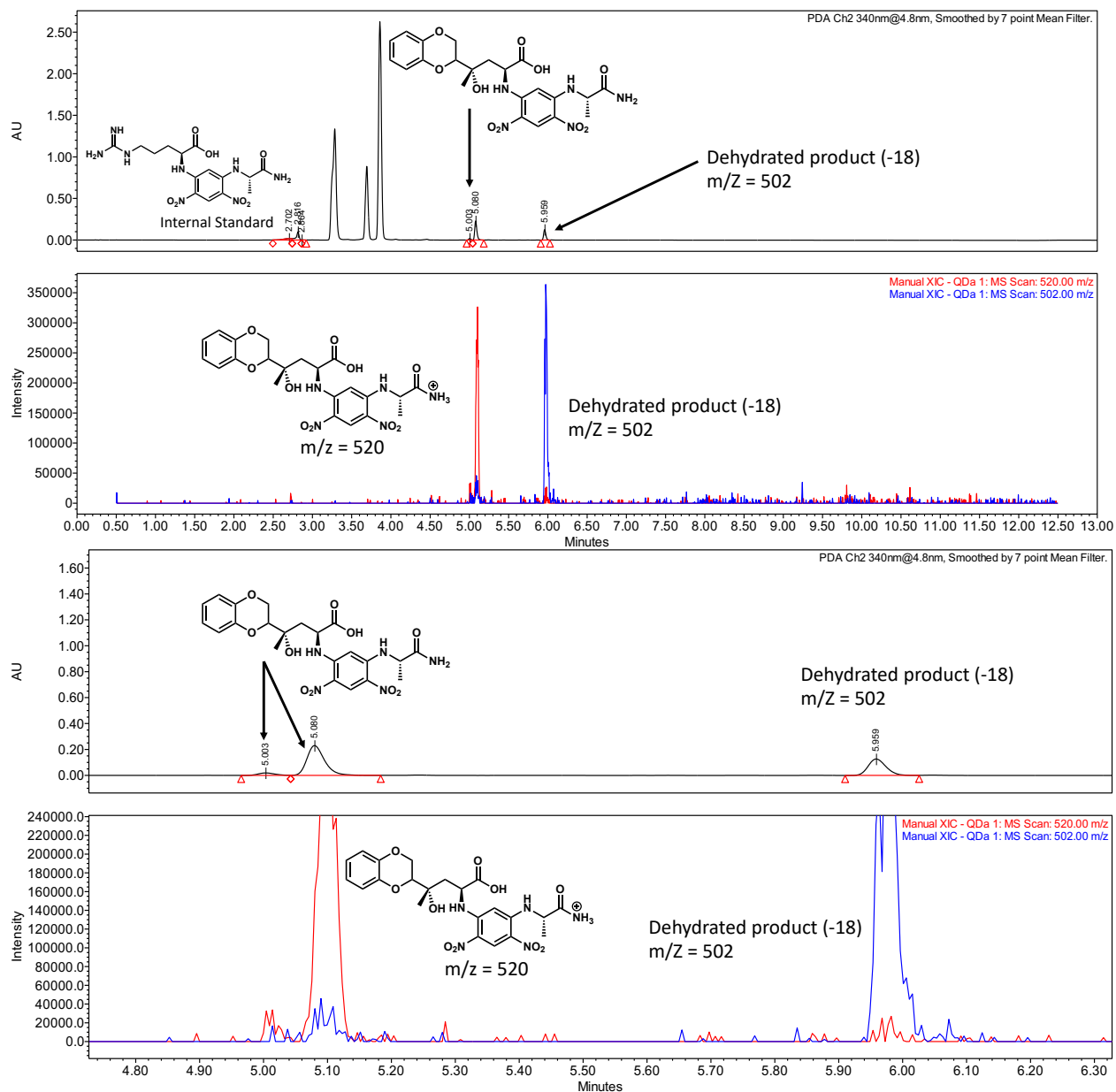

Figure 4.S8. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 2-furylacetone. **Conditions:** 50 mM 2-furylacetone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
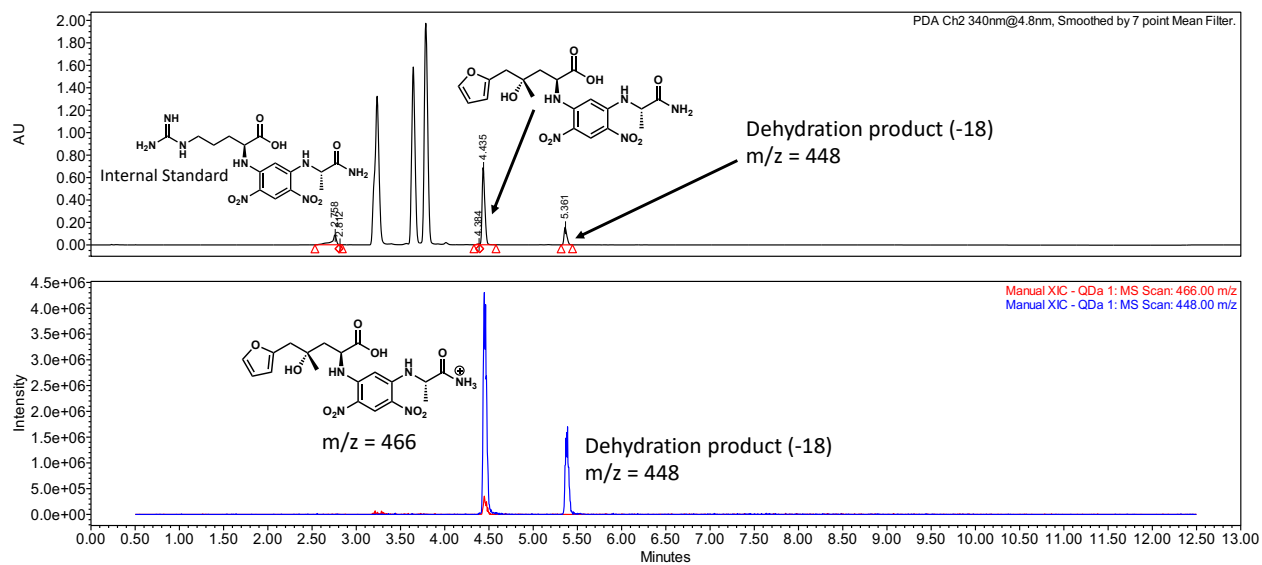
Figure 4.S9. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 2-furylacetone. **Conditions:** 50 mM 2-furylacetone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
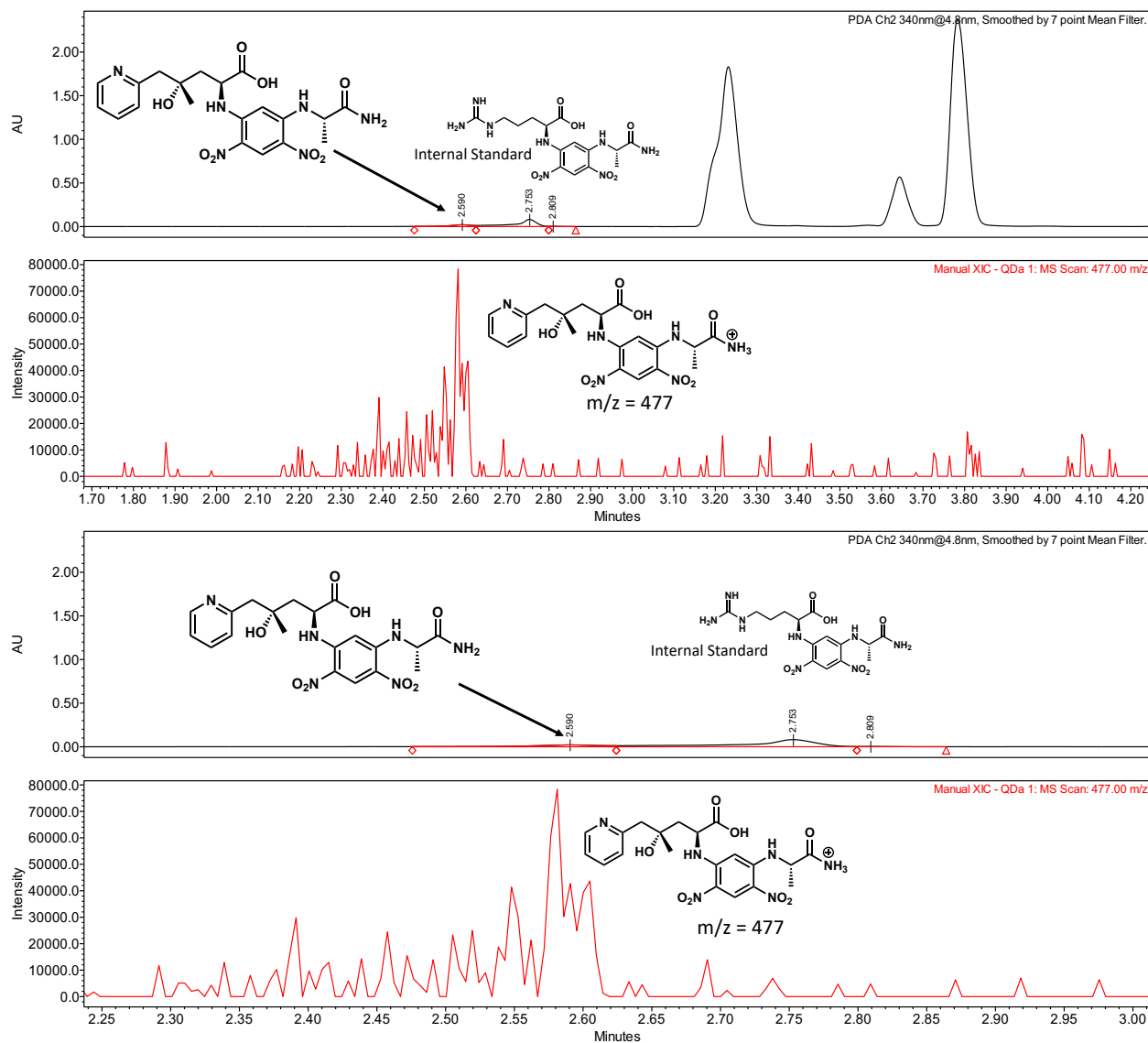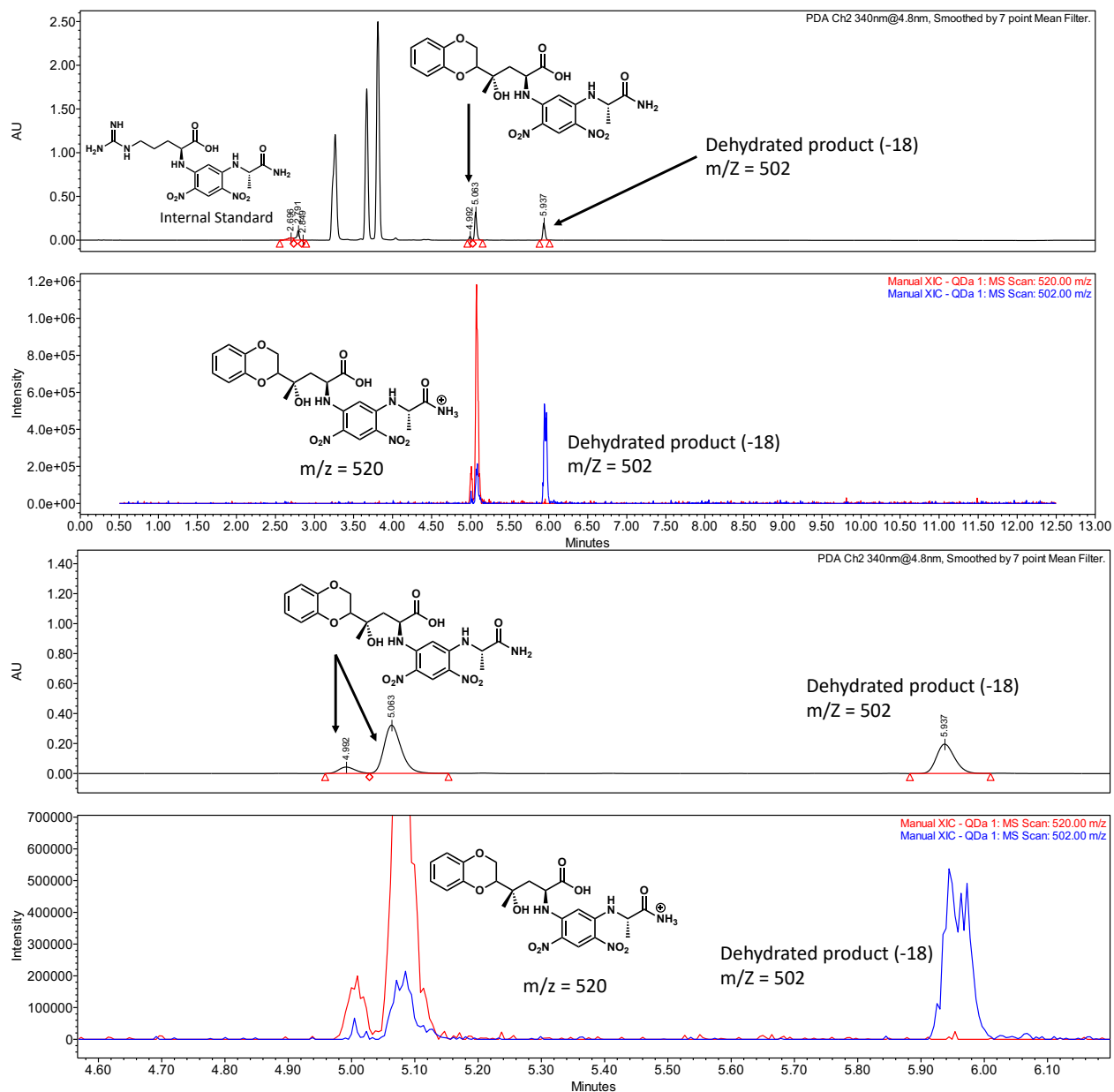
Figure 4.S10. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 1-pyridine-2-yl-propan-2-one. **Conditions:** 50 mM 1-pyridine-2-yl-propan-2-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
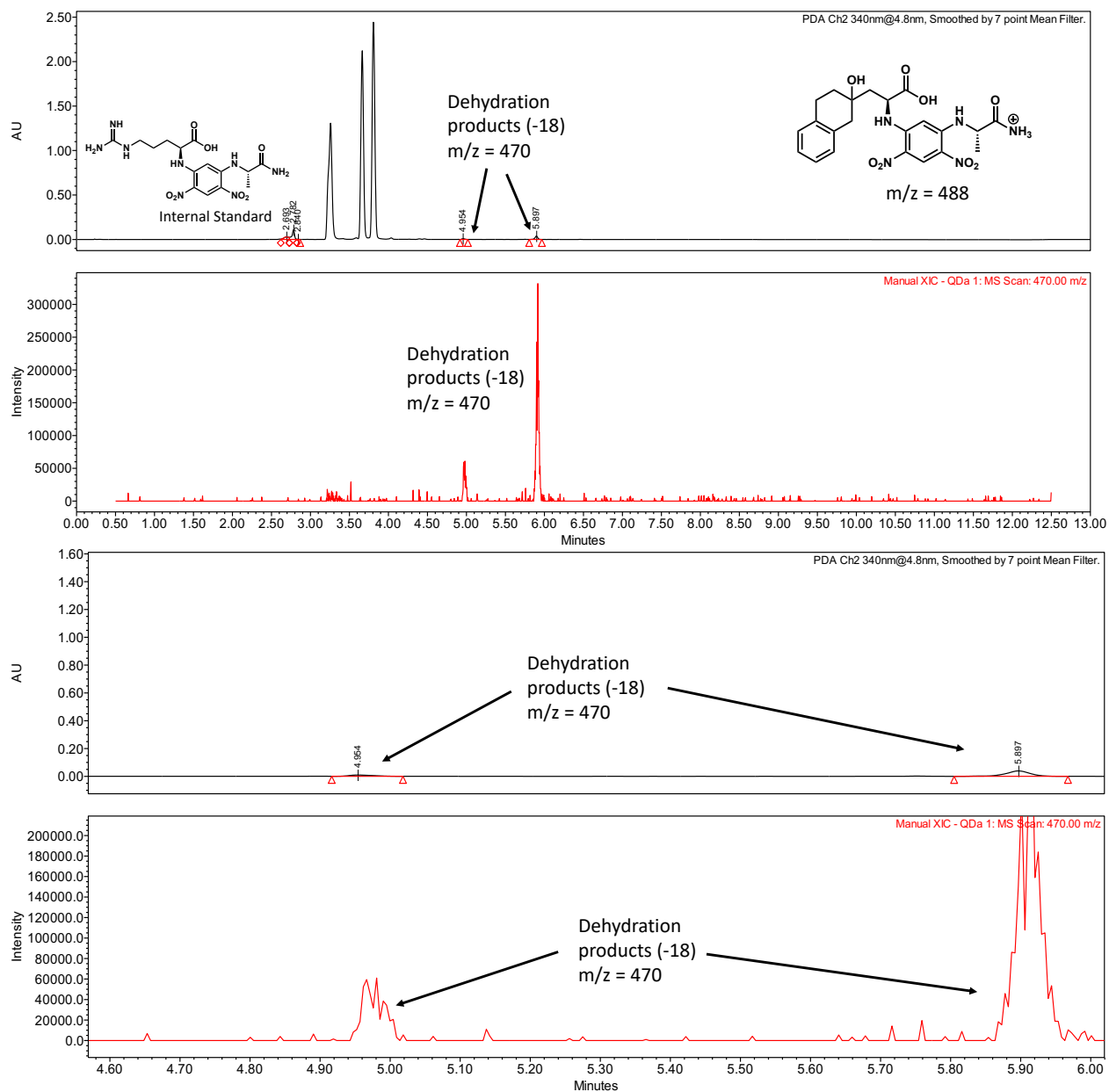
Figure 4.S11. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 1-pyridine-2-yl-propan-2-one. **Conditions:** 50 mM 1-pyridine-2-yl-propan-2-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
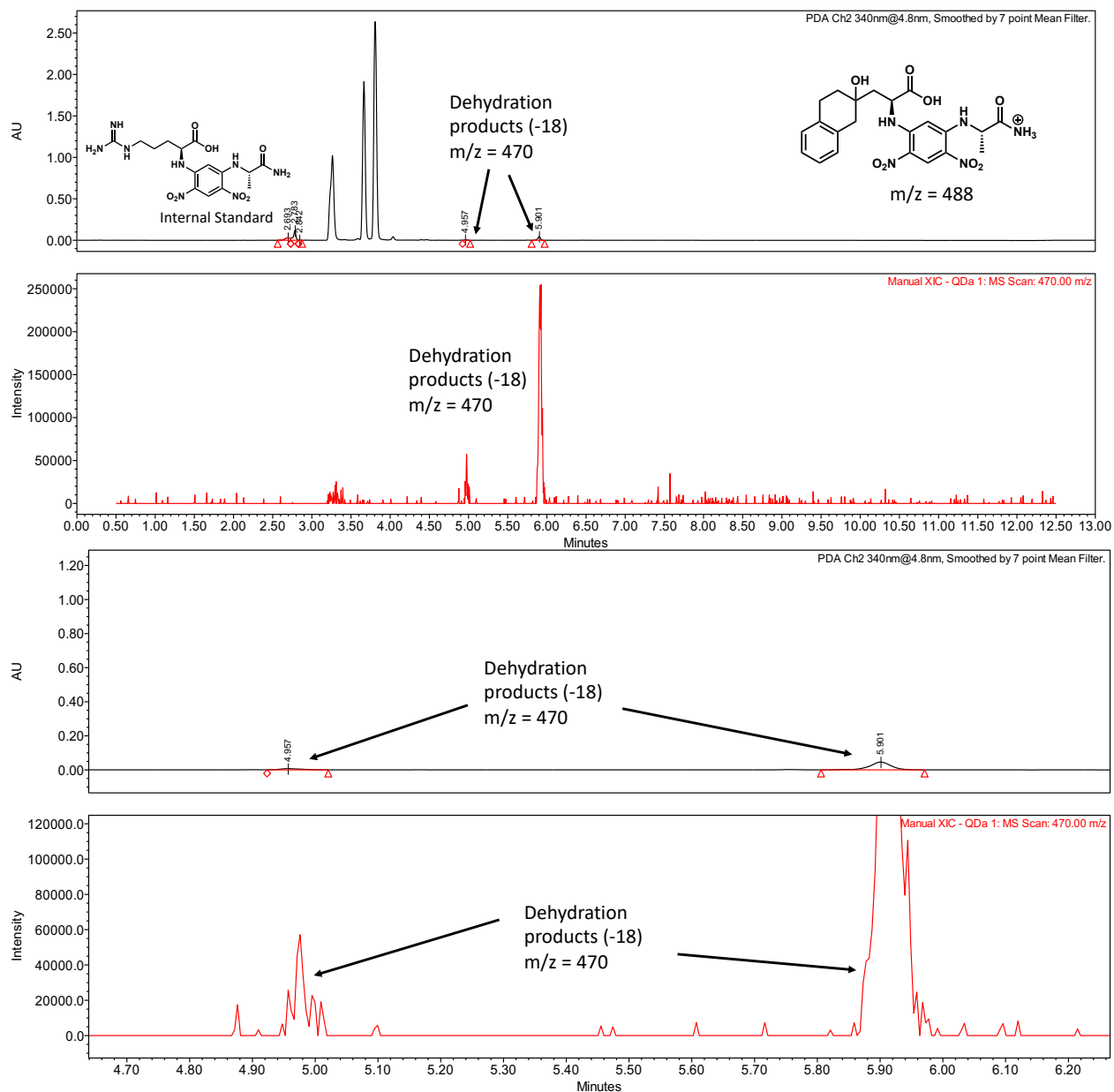

Figure 4.S12. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 1,1,1-trifluoroacetone. **Conditions:** 50 mM 1,1,1-trifluoroacetone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S13. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 1,1,1-trifluoroacetone. **Conditions:** 50 mM 1,1,1-trifluoroacetone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

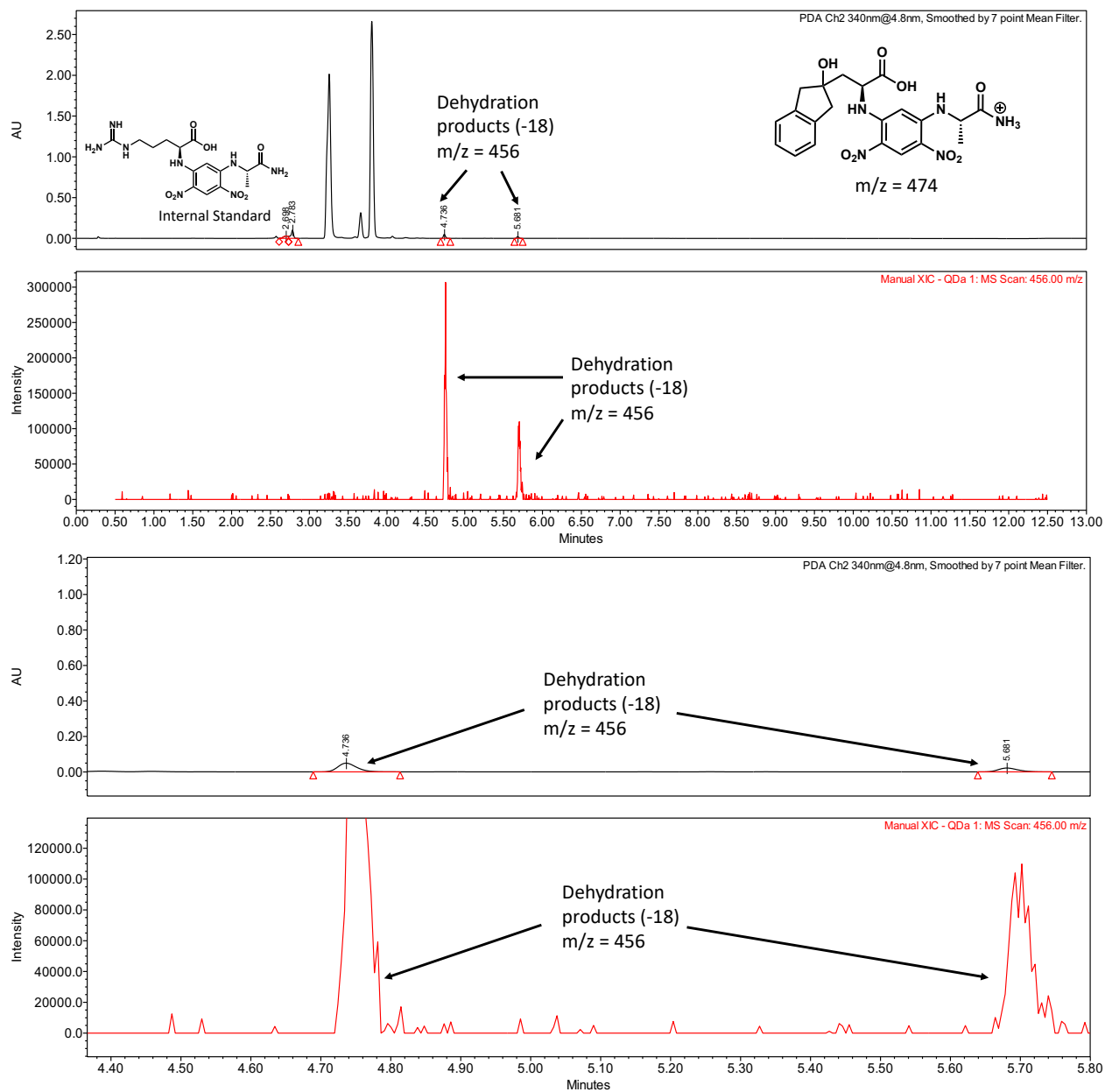**Figure 4.S14.** Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 1-(2,3-dihydro-1,4-benzodioxin-2-yl)ethanone. **Conditions**: 50 mM 1-(2,3-dihydro-1,4-benzodioxin-2-yl)ethanone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
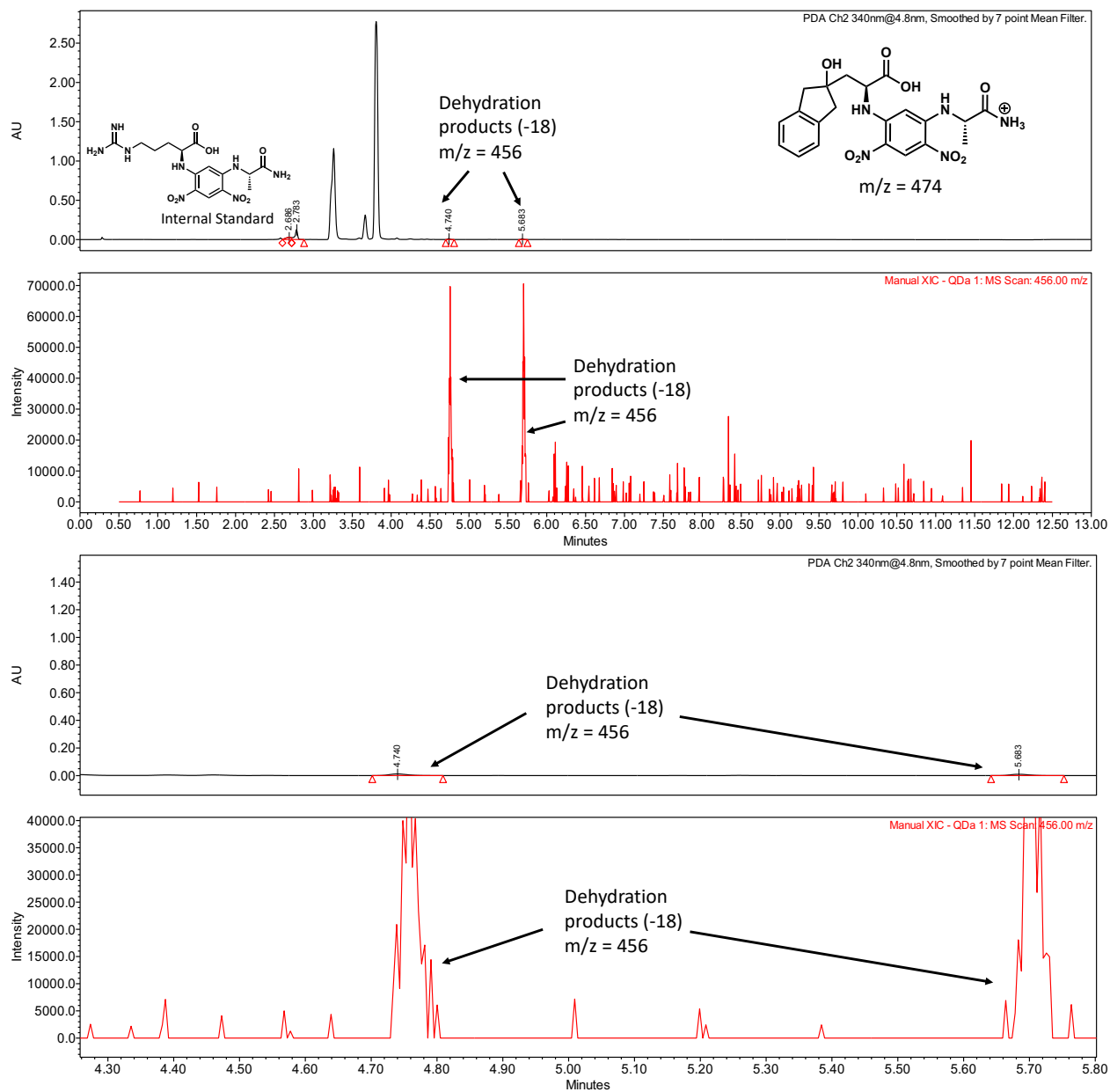
Figure 4.S15. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 1-(2,3-dihydro-1,4-benzodioxin-2-yl)ethanone. **Conditions:** 50 mM 1-(2,3-dihydro-1,4-benzodioxin-2-yl)ethanone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
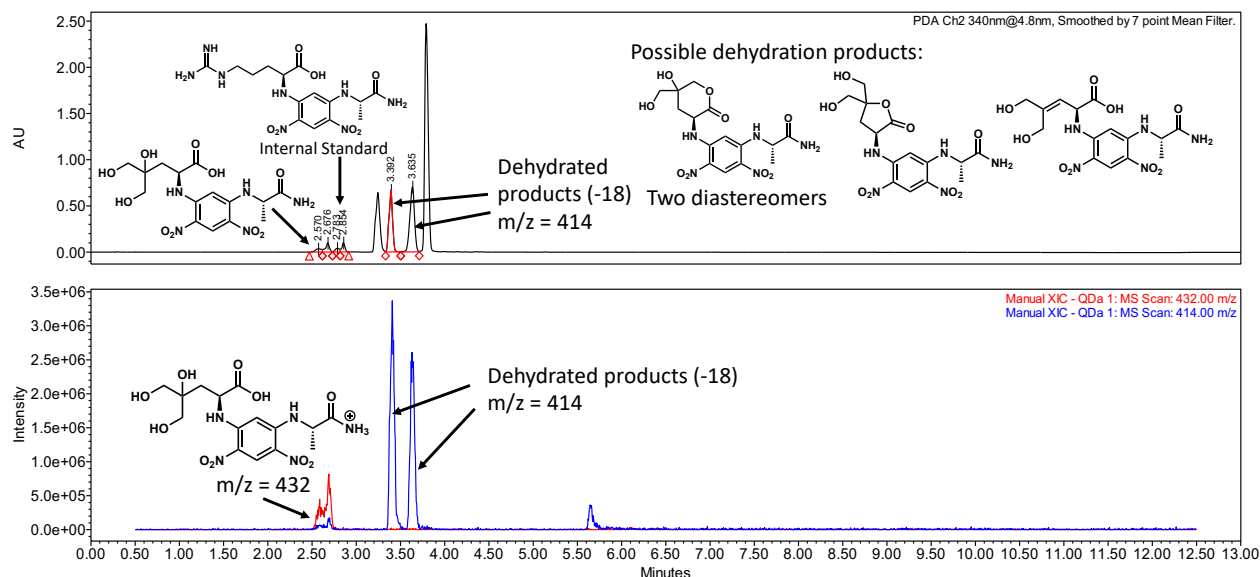
Figure 4.S16. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with β-tetralone. **Conditions:** 50 mM β-tetralone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S17. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with β-tetralone. **Conditions:** 50 mM β-tetralone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

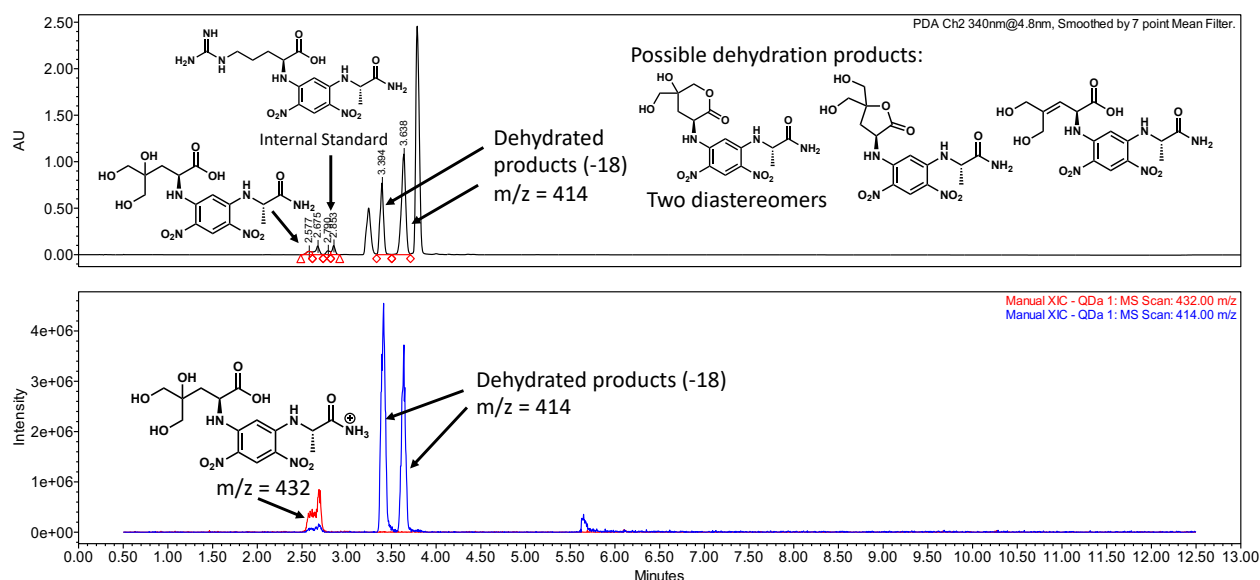**Figure 4.S18.** Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 2-indanone. **Conditions:** 50 mM 2-indanone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S19. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 2-indanone. **Conditions:** 50 mM 2-indanone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

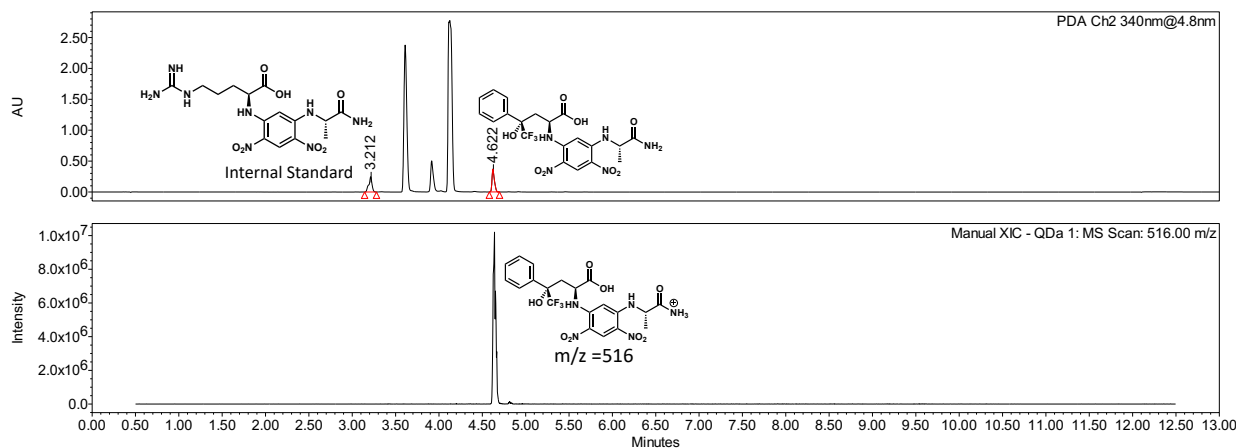**Figure 4.S20.** Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 1,3-dihydroxyacetone. **Conditions:** 50 mM 1,3-dihydroxyacetone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h. **Note:** lactonization to the six membered ring would generate two diastereomeric lactones that could be responsible for the two -18 mass peaks. Additionally, the large number of hydroxyl groups could contribute to the streakiness of the linear amino acid product peak.

Figure 4.S21. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 1,3-dihydroxyacetone. **Conditions:** 50 mM 1,3-dihydroxyacetone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h. **Note:** lactonization to the six membered ring would generate two diastereomeric lactones that could be responsible for the two -18 mass peaks. Additionally, the large number of hydroxyl groups could contribute to the streakiness of the linear amino acid product peak.

Figure 4.S22. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with 2,2,2-trifluoroacetophenone. **Conditions:** 50 mM 2,2,2-trifluoroacetophenone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
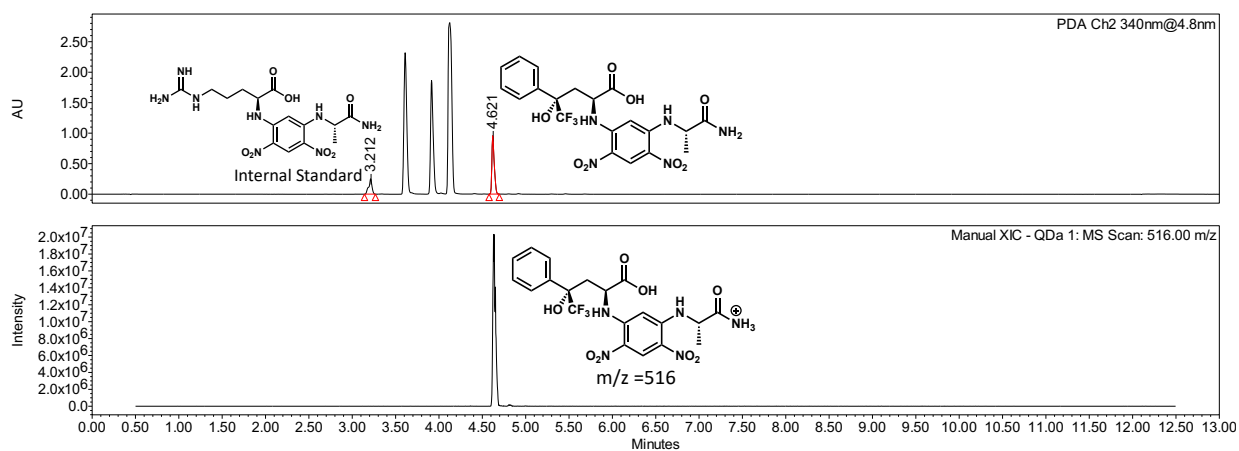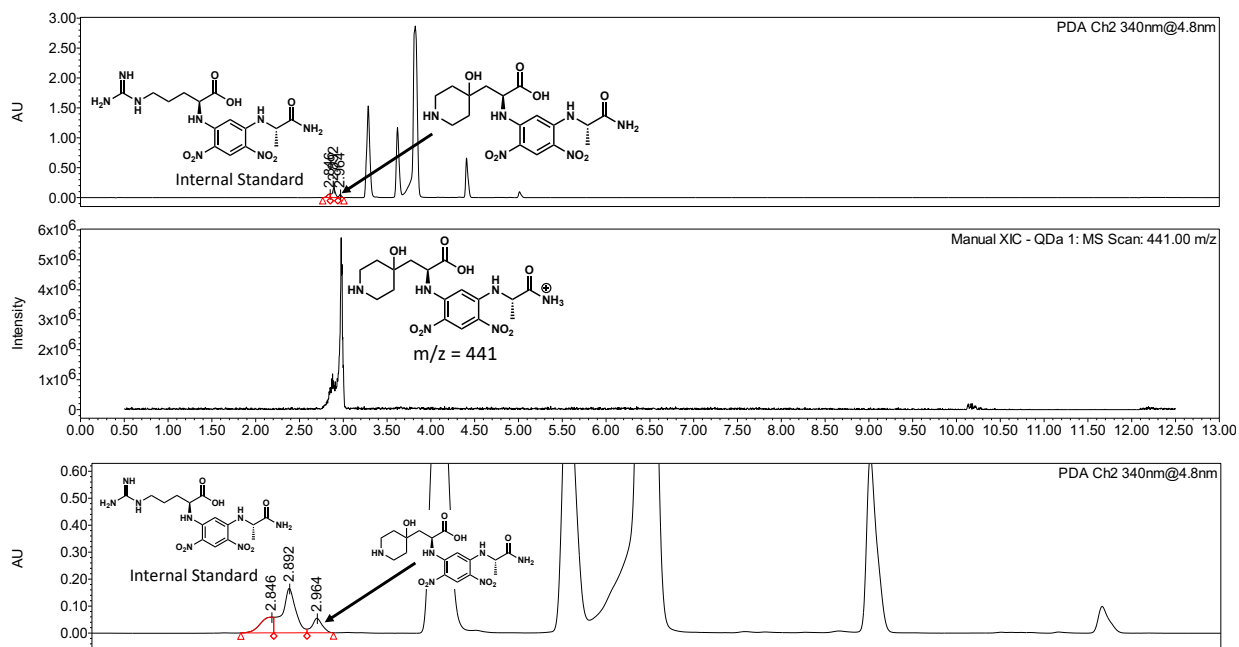

Figure 4.S23. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with 2,2,2-trifluoroacetophenone. **Conditions:** 50 mM 2,2,2-trifluoroacetophenone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
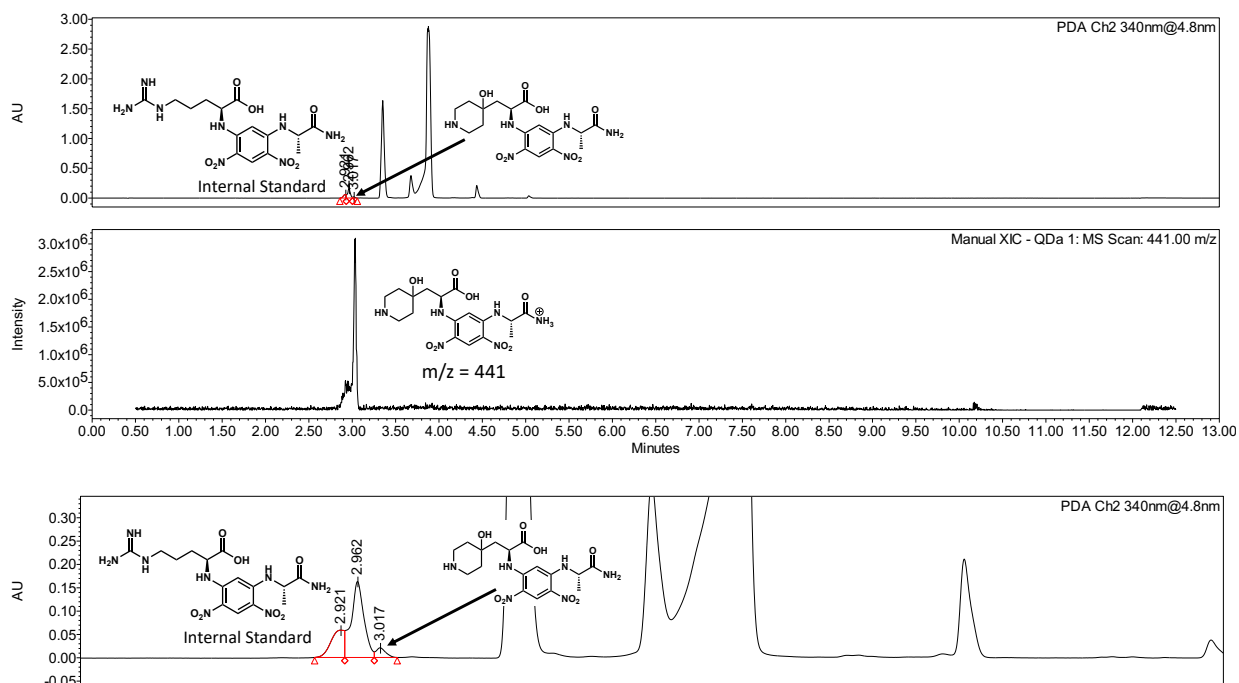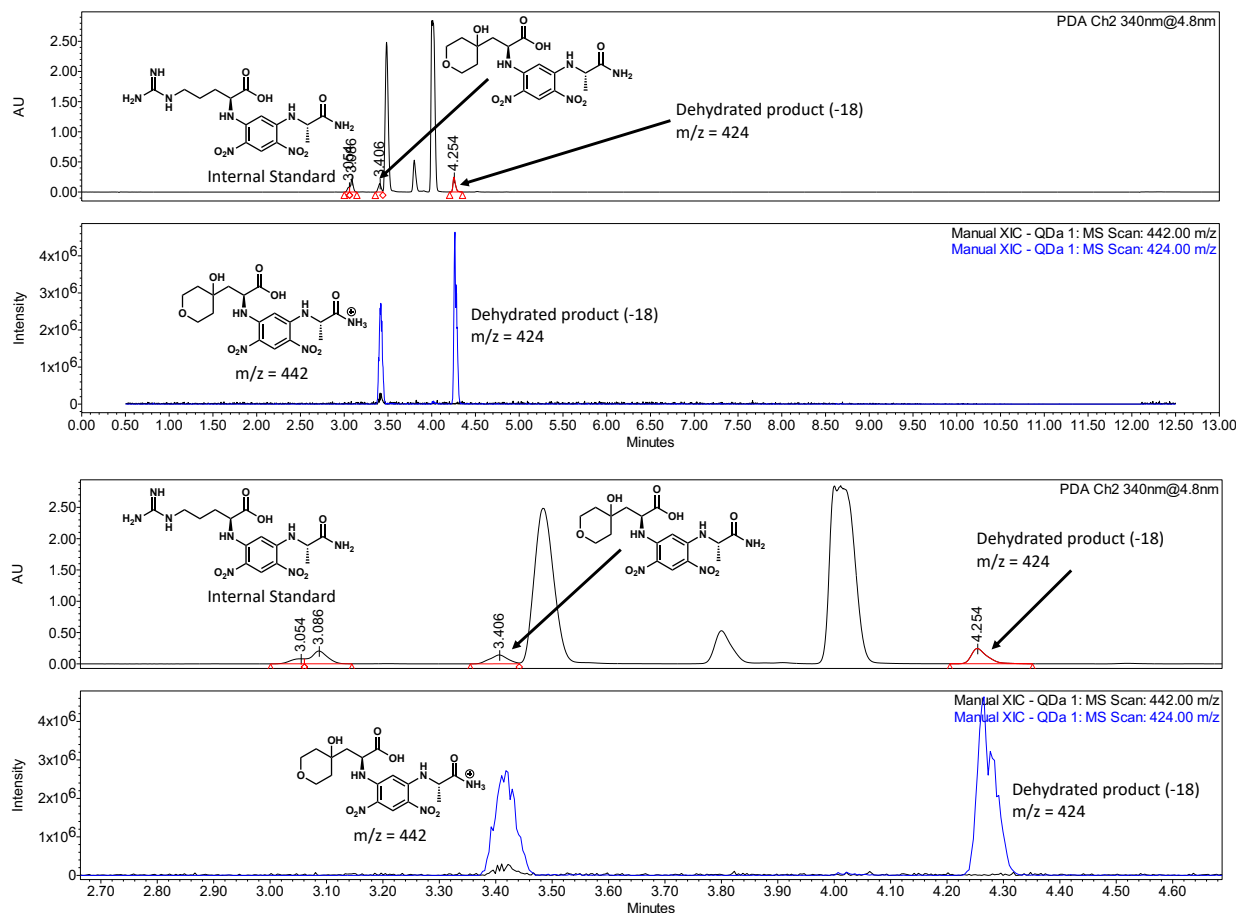
Figure 4.S24. Representative UPLC-MS trace of Marfey's derivatization for 7B05 with piperidin-4-one. **Conditions:** 50 mM piperidin-4-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
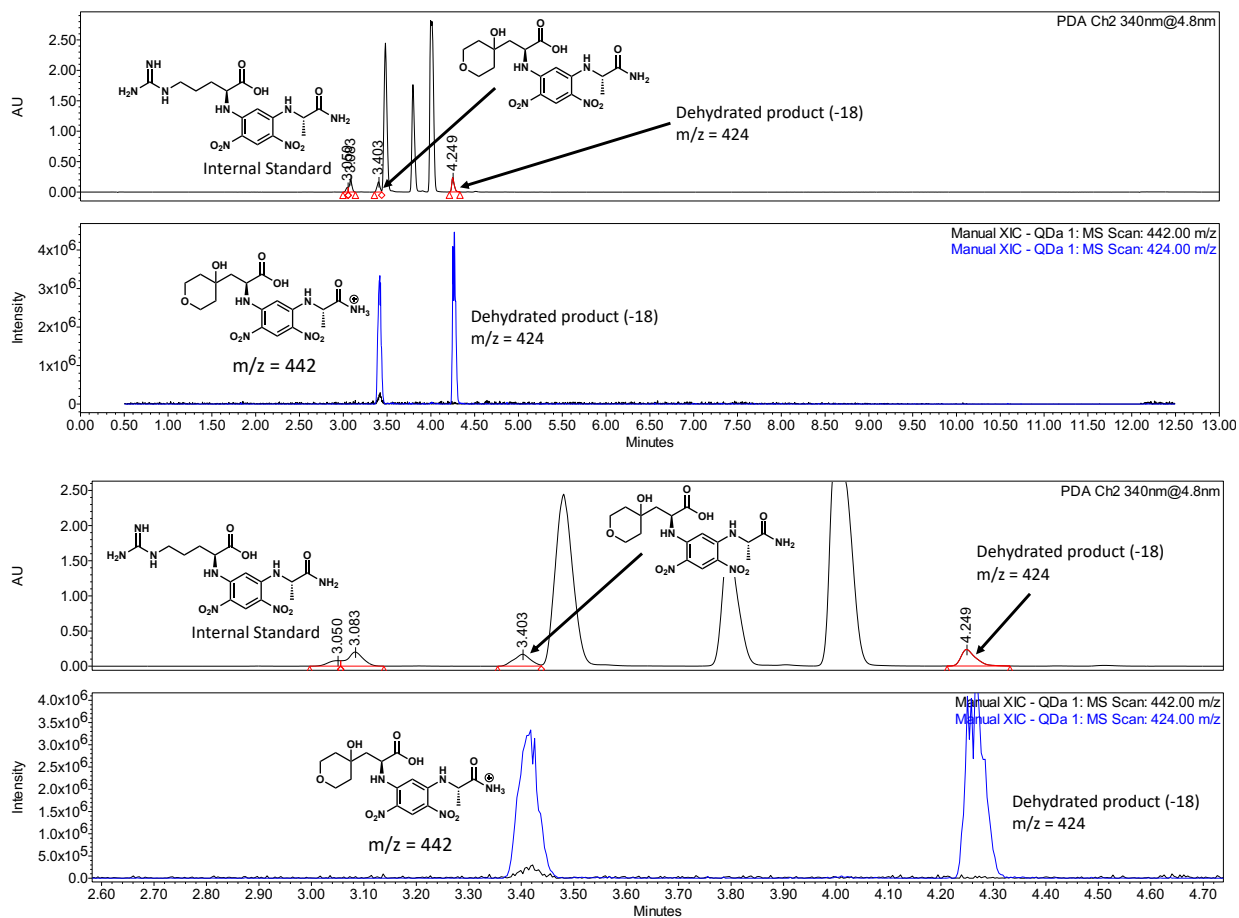

Figure 4.S25. Representative UPLC-MS trace of Marfey's derivatization for 7G11 with piperidin-4-one. **Conditions:** 50 mM piperidin-4-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S26. Representative UPLC-MS trace of Marfey's derivatization for 7B05 tetrahydro-4H-pyran-4-one. **Conditions:** 50 mM tetrahydro-4H-pyran-4-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

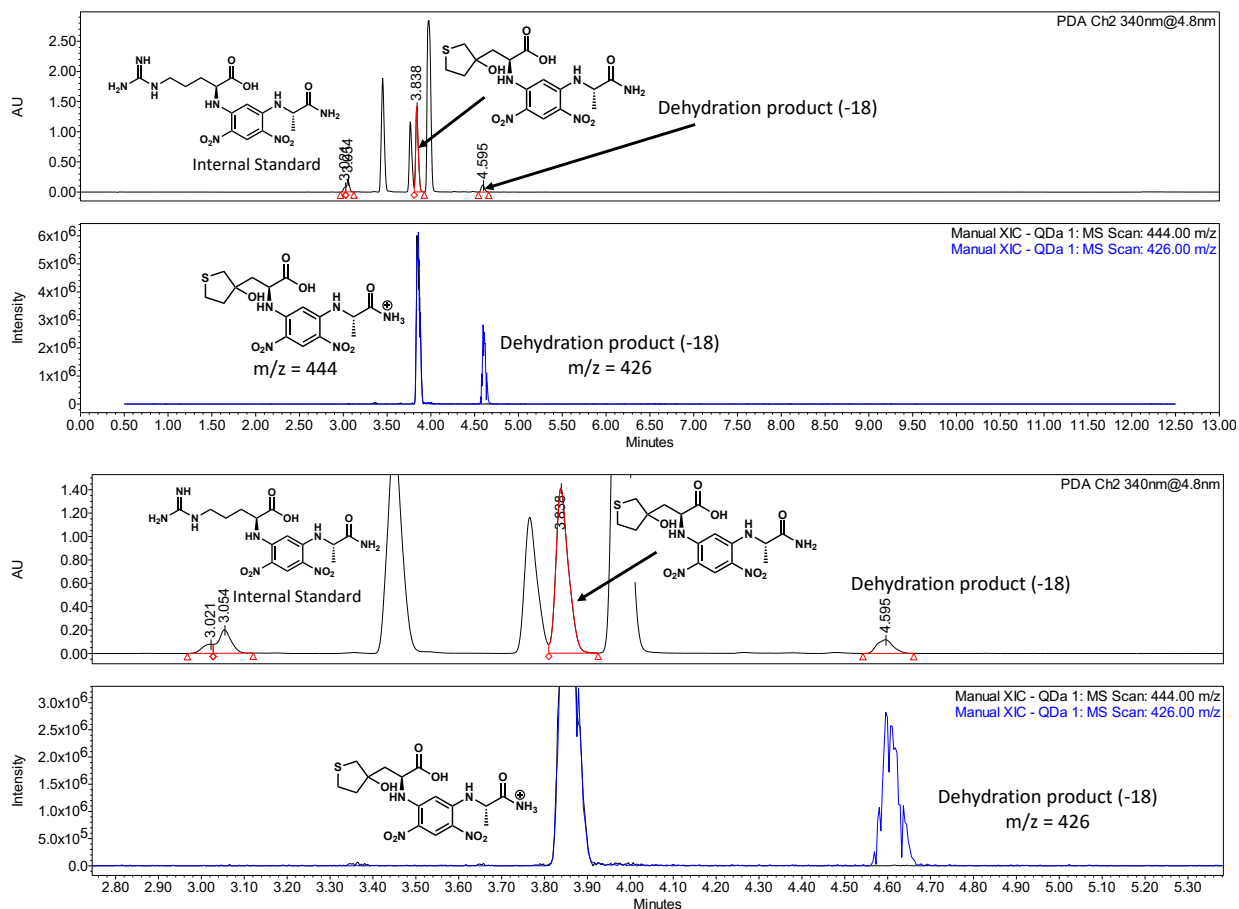**Figure 4.S27.** Representative UPLC-MS trace of Marfey's derivatization for 7G11 tetrahydro-4*H*-pyran-4-one. **Conditions:** 50 mM tetrahydro-4*H*-pyran-4-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
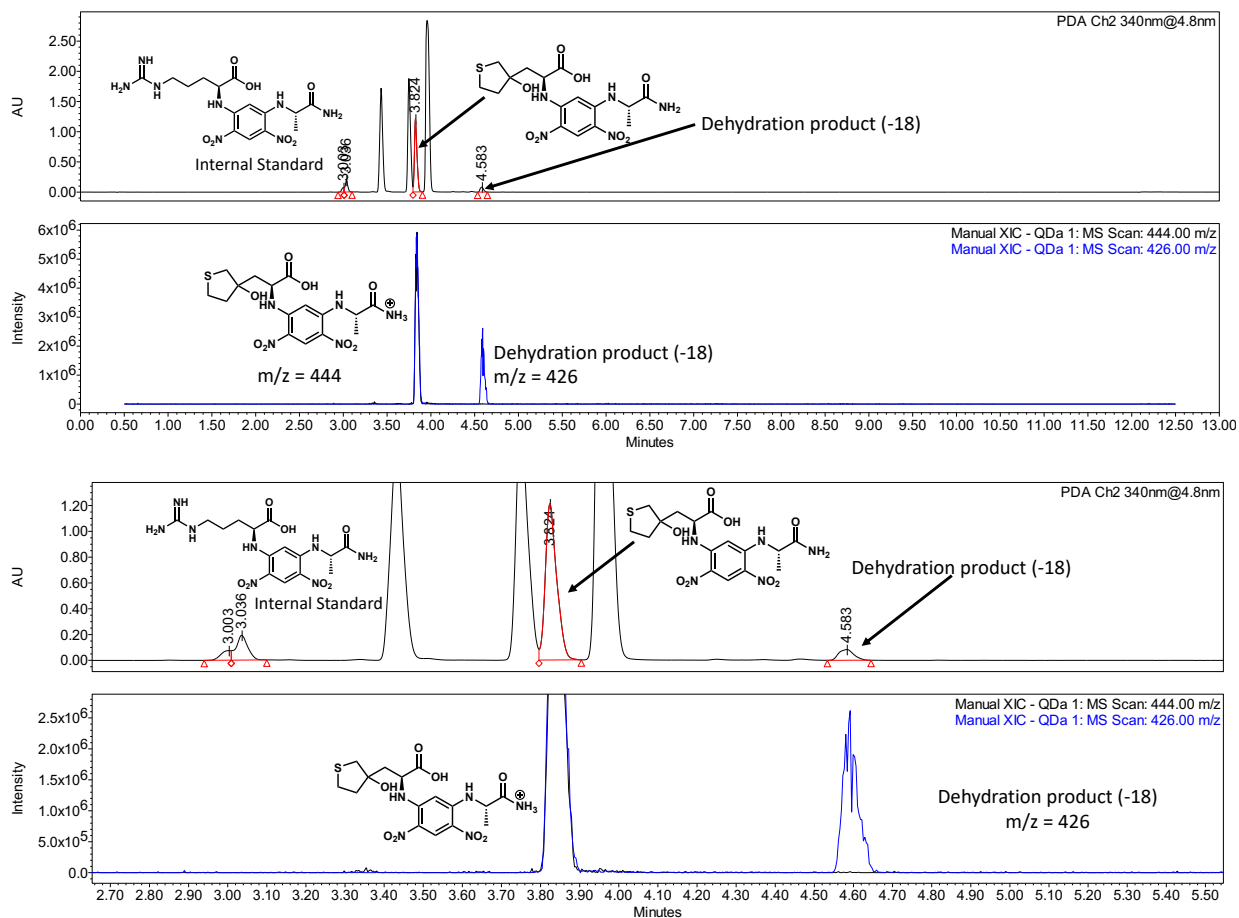
Figure 4.S28. Representative UPLC-MS trace of Marfey's derivatization for 7B05 dihydrothiophen-3(2*H*)-one. **Conditions:** 50 mM dihydrothiophen-3(2*H*)-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S29. Representative UPLC-MS trace of Marfey's derivatization for 7G11 dihydrothiophen-3(2*H*)-one. **Conditions:** 50 mM dihydrothiophen-3(2*H*)-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
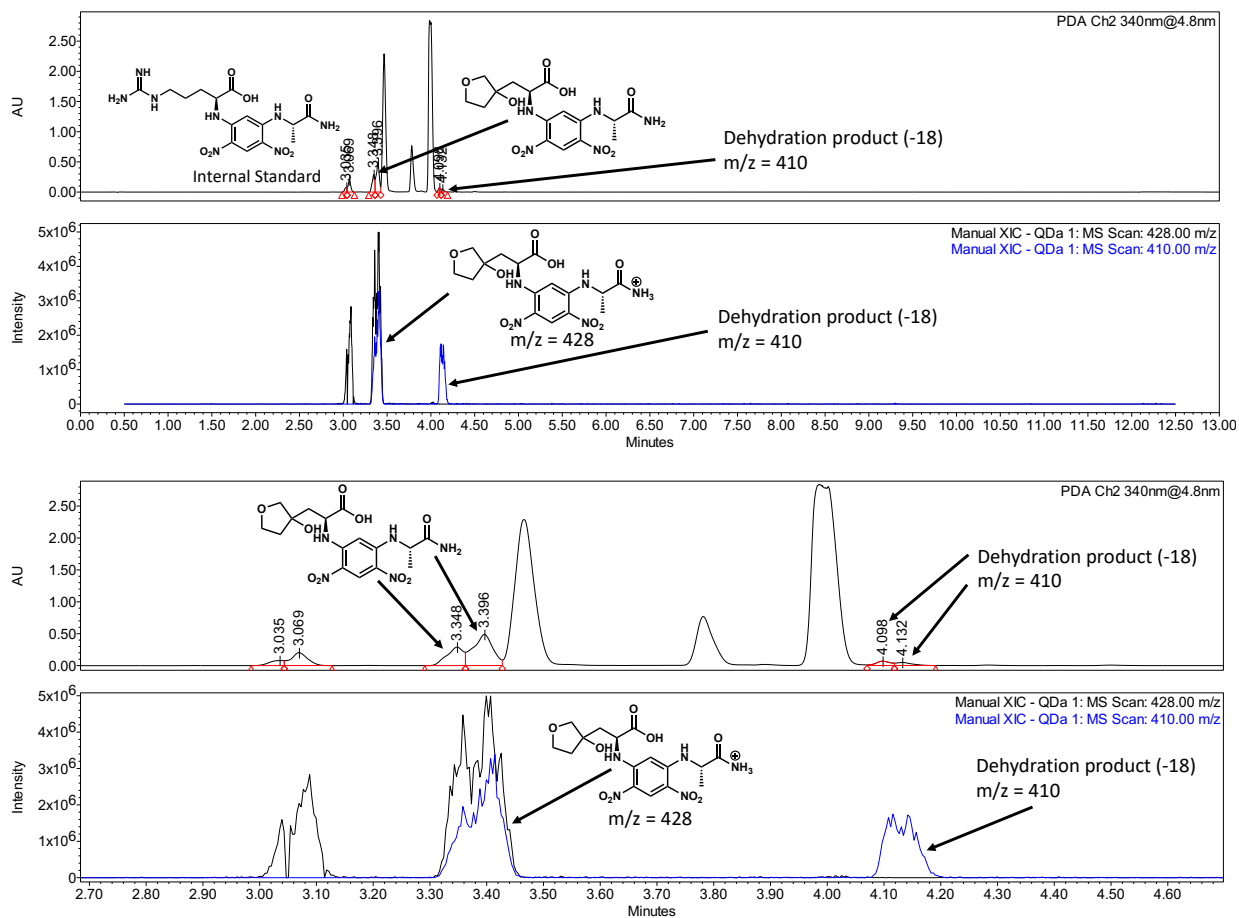
Figure 4.S30. Representative UPLC-MS trace of Marfey's derivatization for 7B05 dihydrofuran-3(2H)-one. **Conditions:** 50 mM dihydrofuran-3(2H)-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
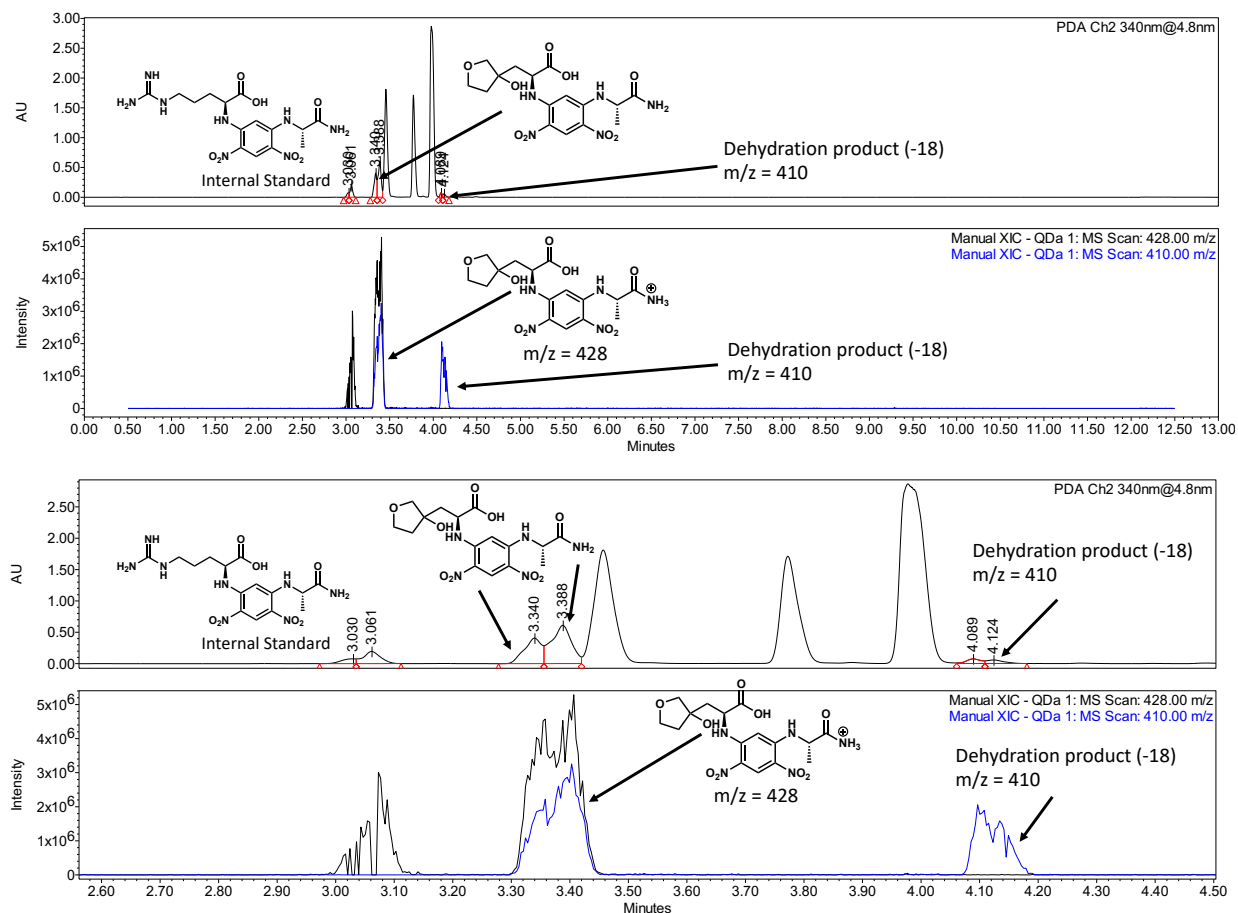
Figure 4.S31. Representative UPLC-MS trace of Marfey's derivatization for 7G11 dihydrofuran-3(2*H*)-one. **Conditions:** 50 mM dihydrofuran-3(2*H*)-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
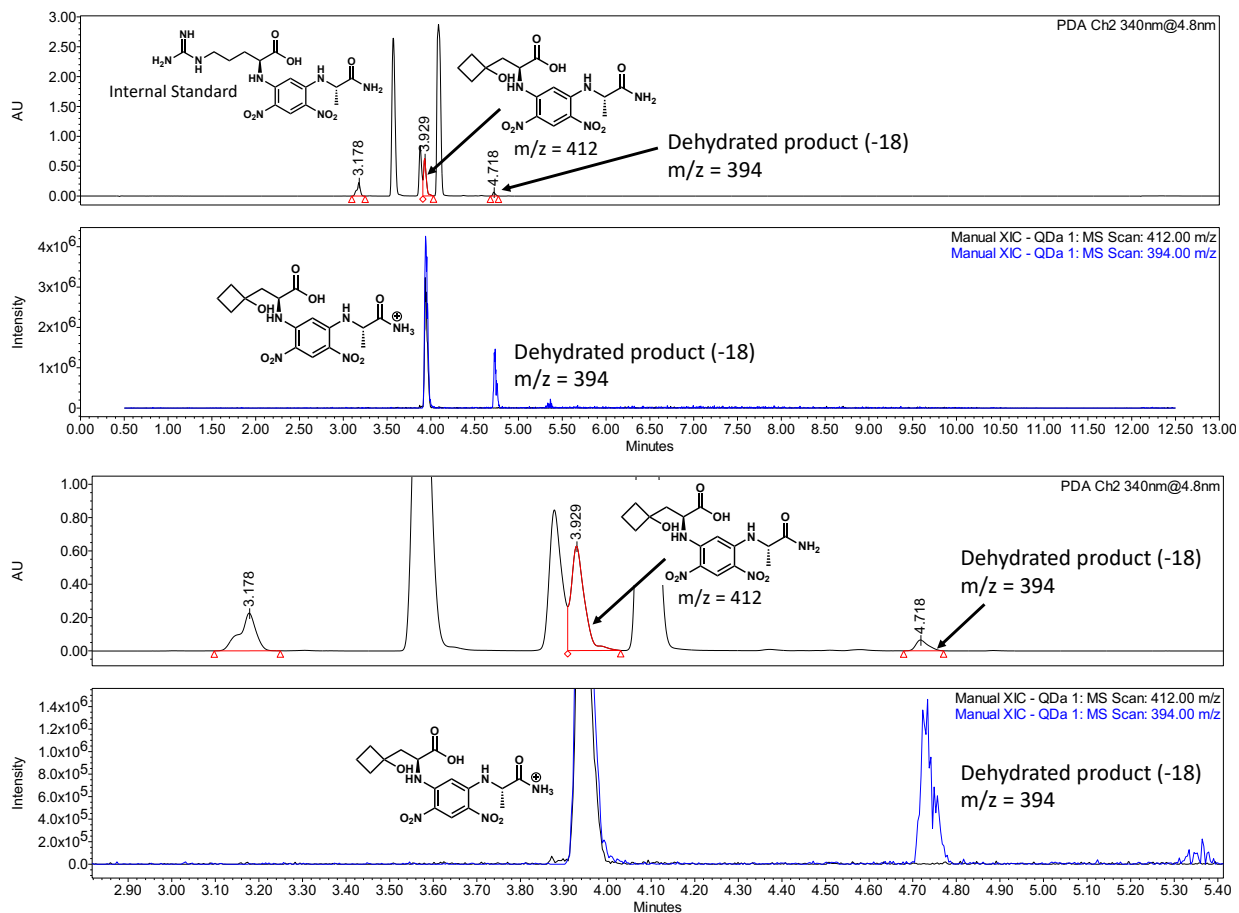
Figure 4.S32. Representative UPLC-MS trace of Marfey's derivatization for 7B05 cyclobutanone. **Conditions:** 50 mM cyclobutanone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
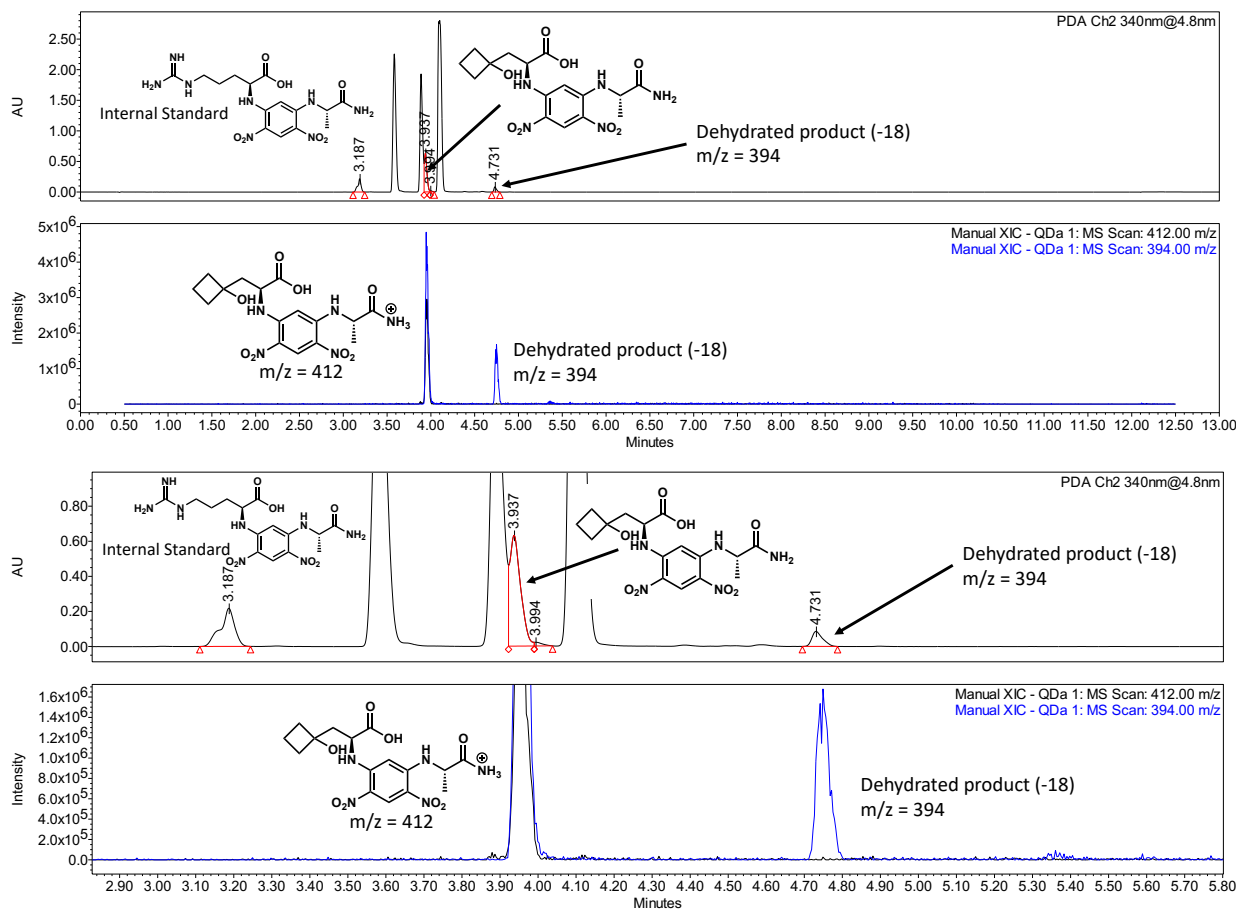
Figure 4.S33. Representative UPLC-MS trace of Marfey's derivatization for 7G11 cyclobutanone. **Conditions:** 50 mM cyclobutanone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
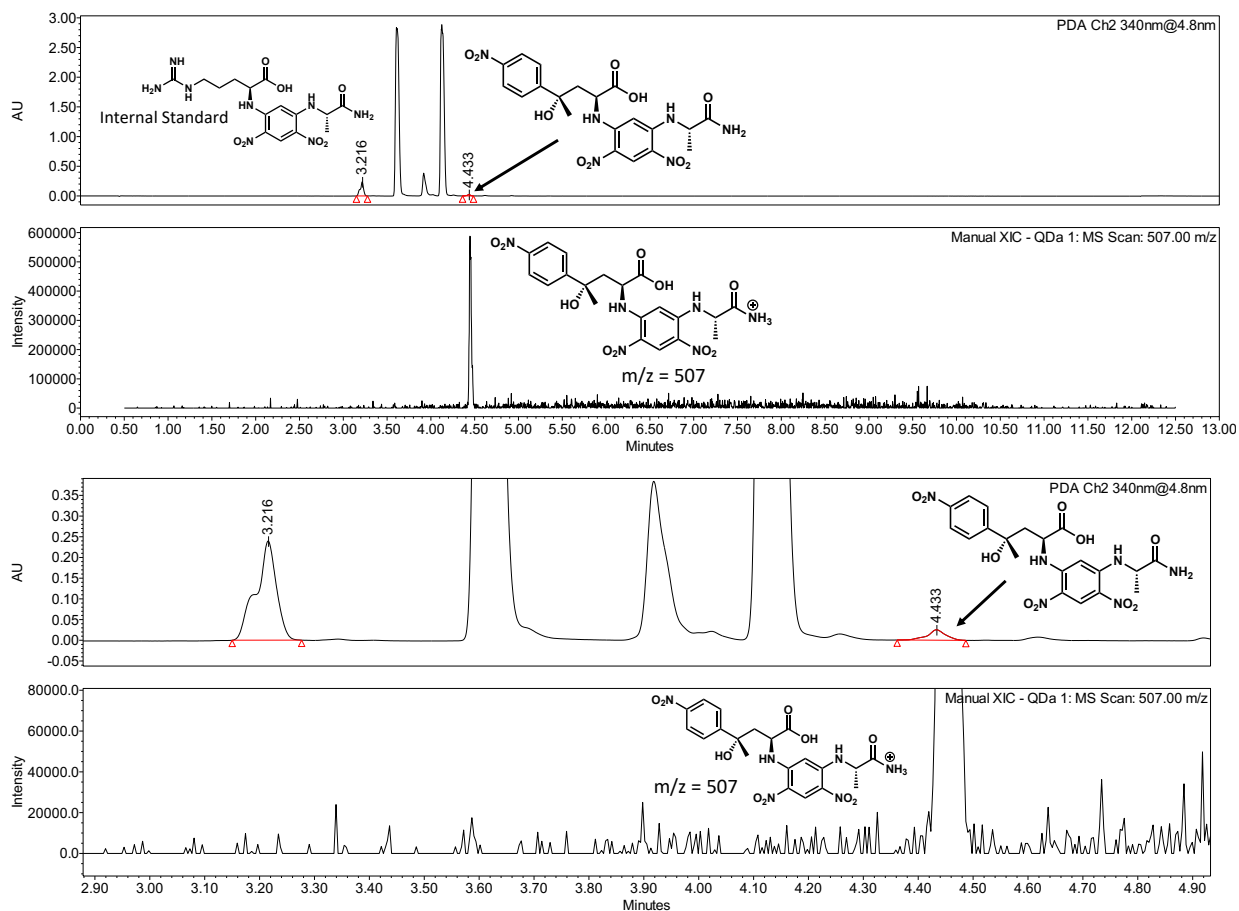
Figure 4.S34. Representative UPLC-MS trace of Marfey's derivatization for 7B05 4'-nitroacetophenone. **Conditions:** 50 mM 4'-nitroacetophenone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
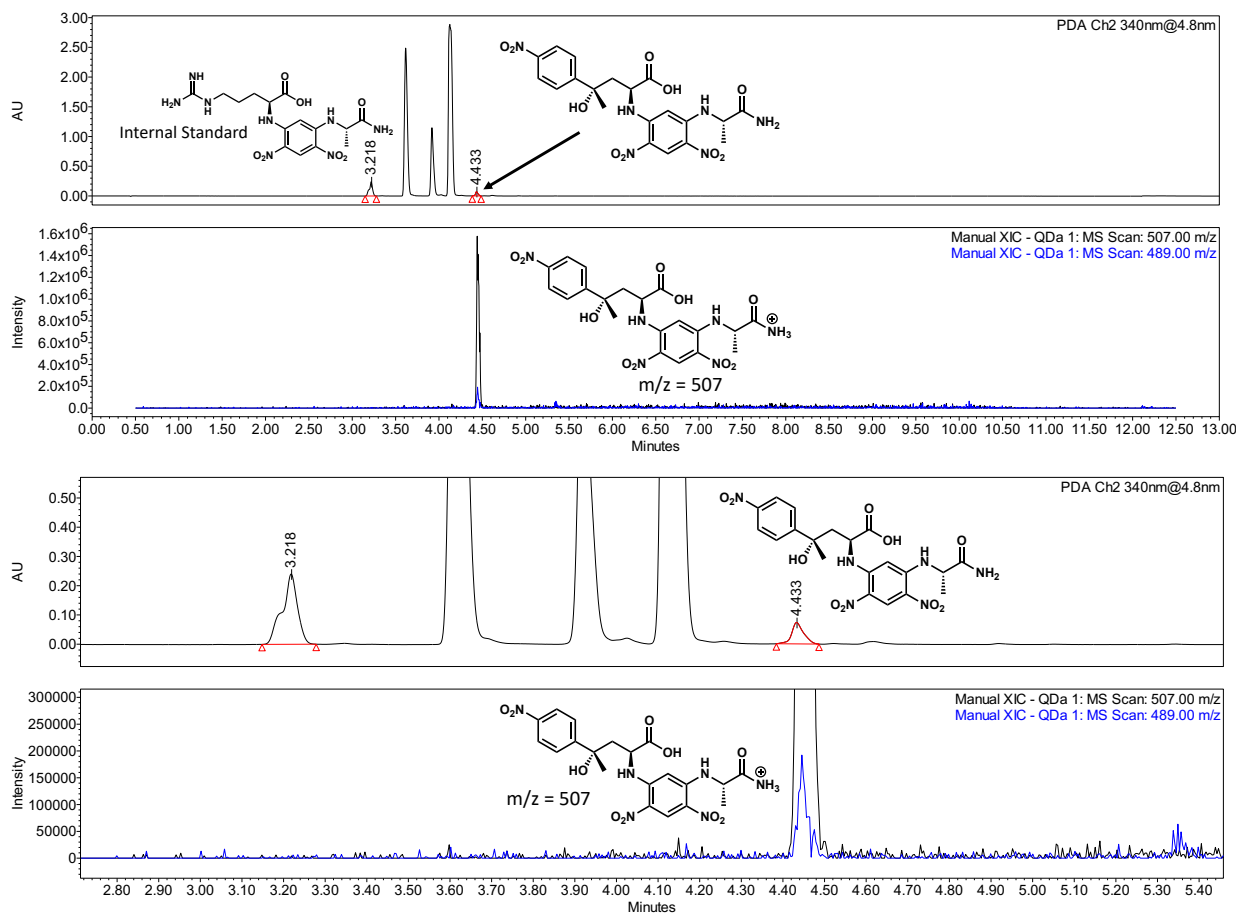
Figure 4.S35. Representative UPLC-MS trace of Marfey's derivatization for 7G11 4'-nitroacetophenone. **Conditions:** 50 mM 4'-nitroacetophenone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S36. Representative UPLC-MS trace of Marfey's derivatization for 7B05 1,1,1-trifluoro-3-phenyl-2-propanone. **Conditions:** 50 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
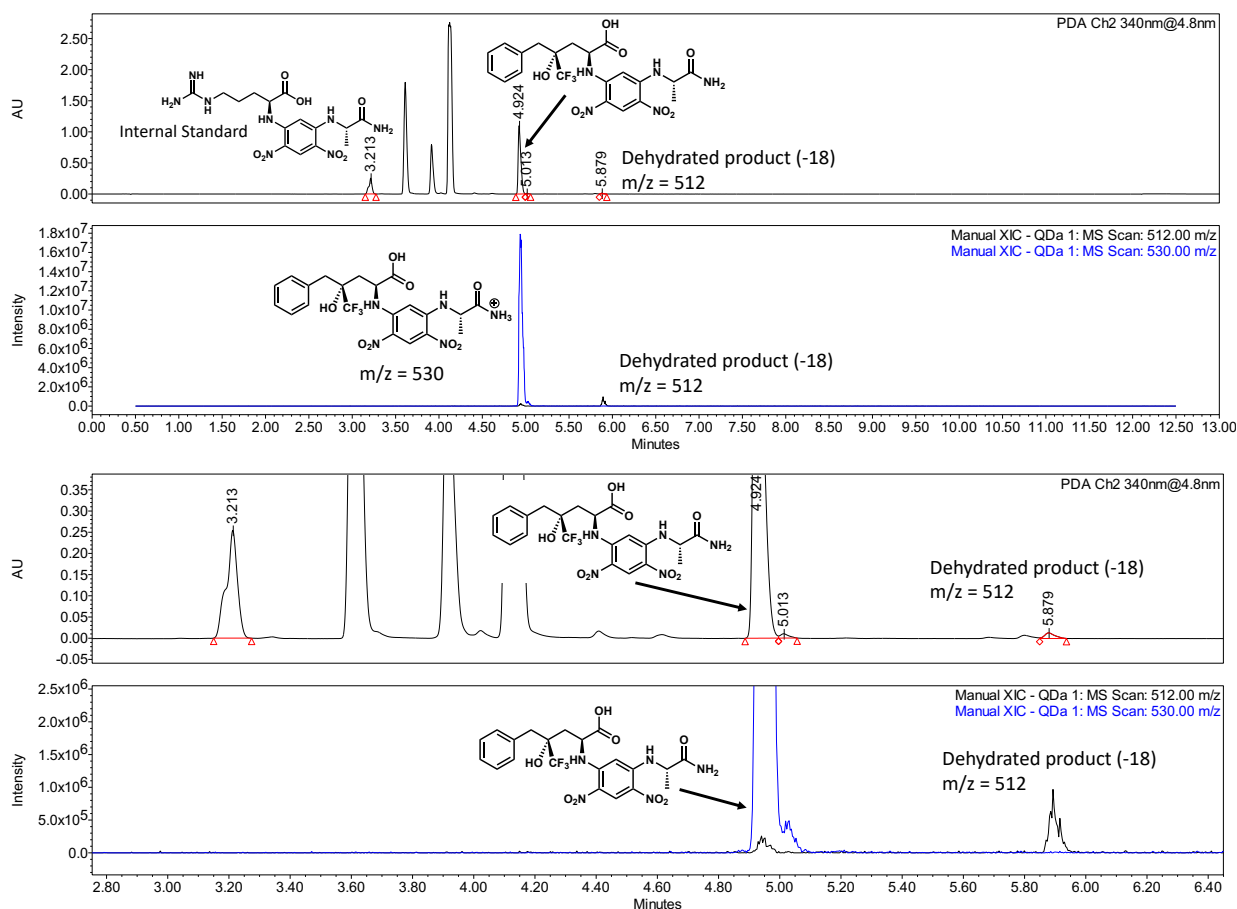
Figure 4.S37. Representative UPLC-MS trace of Marfey's derivatization for 7G11 1,1,1-trifluoro-3-phenyl-2-propanone. **Conditions:** 50 mM 1,1,1-trifluoro-3-phenyl-2-propanone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Figure 4.S38. Representative UPLC-MS trace of Marfey's derivatization for 7B05 (4-fluorophenyl)acetone. **Conditions:** 50 mM (4-fluorophenyl)acetone, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
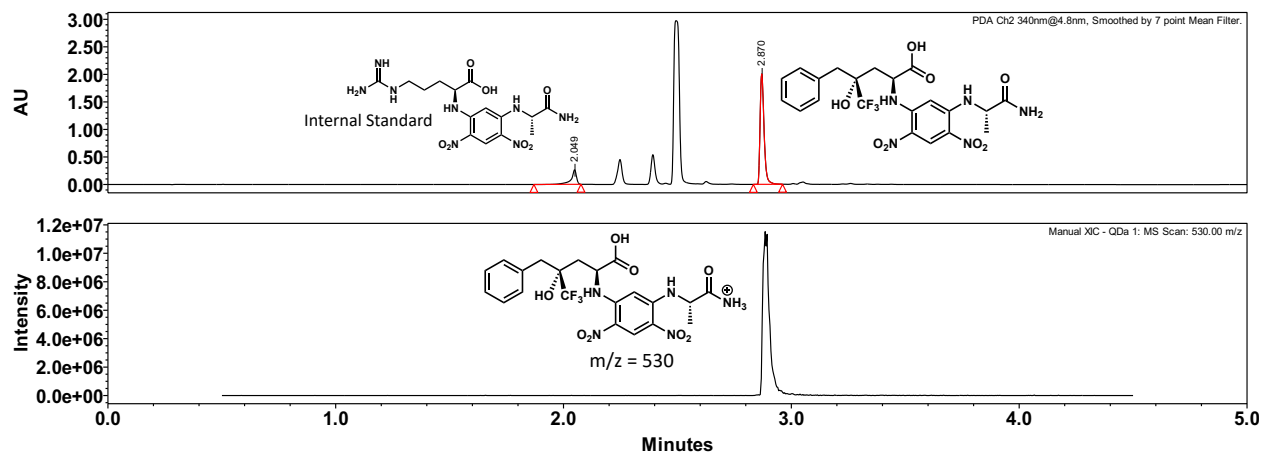
Figure 4.S39. Representative UPLC-MS trace of Marfey's derivatization for 7G11 (4-fluorophenyl)acetone. **Conditions:** 50 mM (4-fluorophenyl)acetone, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
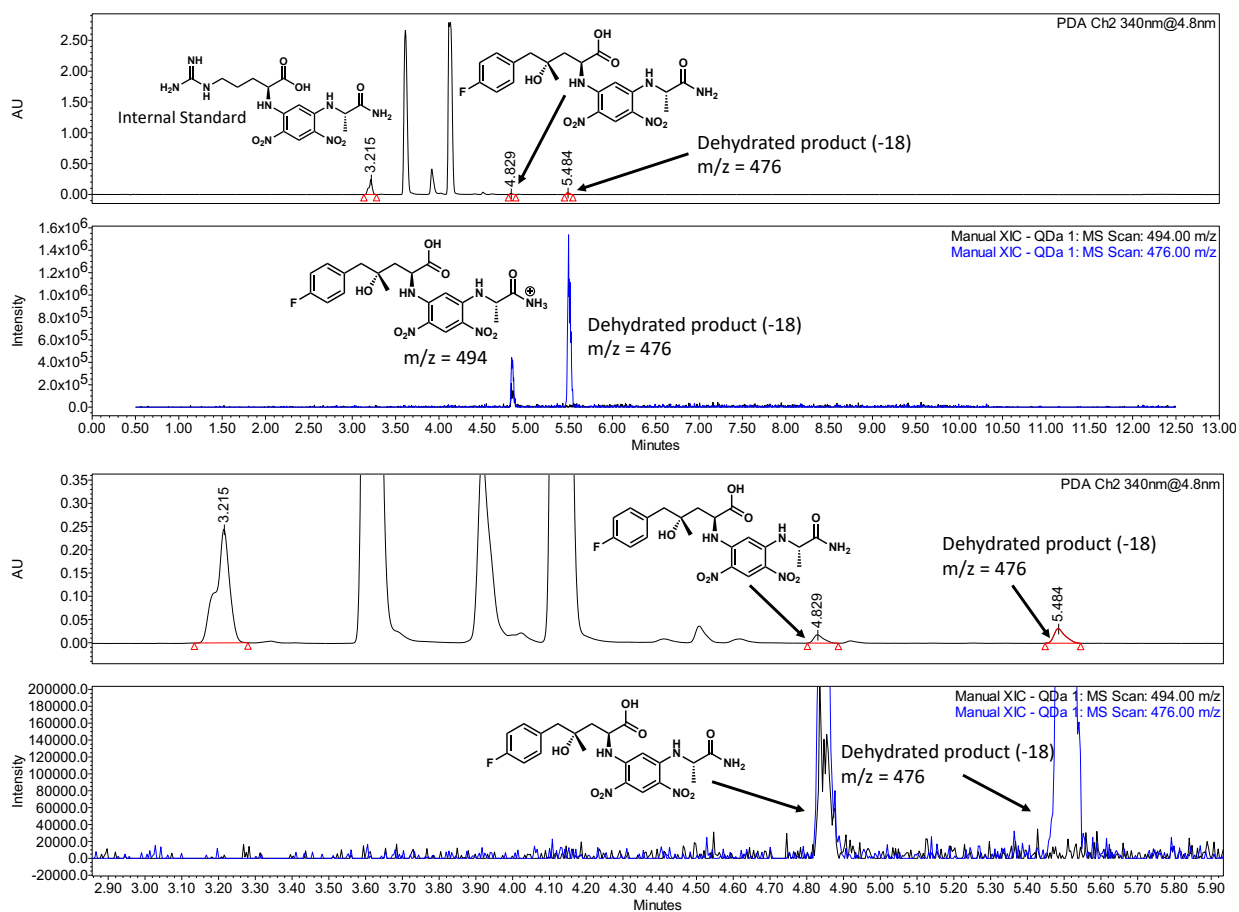
Figure 4.S40. Representative UPLC-MS trace of Marfey's derivatization for 7B05 1,1,1-trifluorohex-2-one. **Conditions:** 50 mM 1,1,1-trifluorohex-2-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
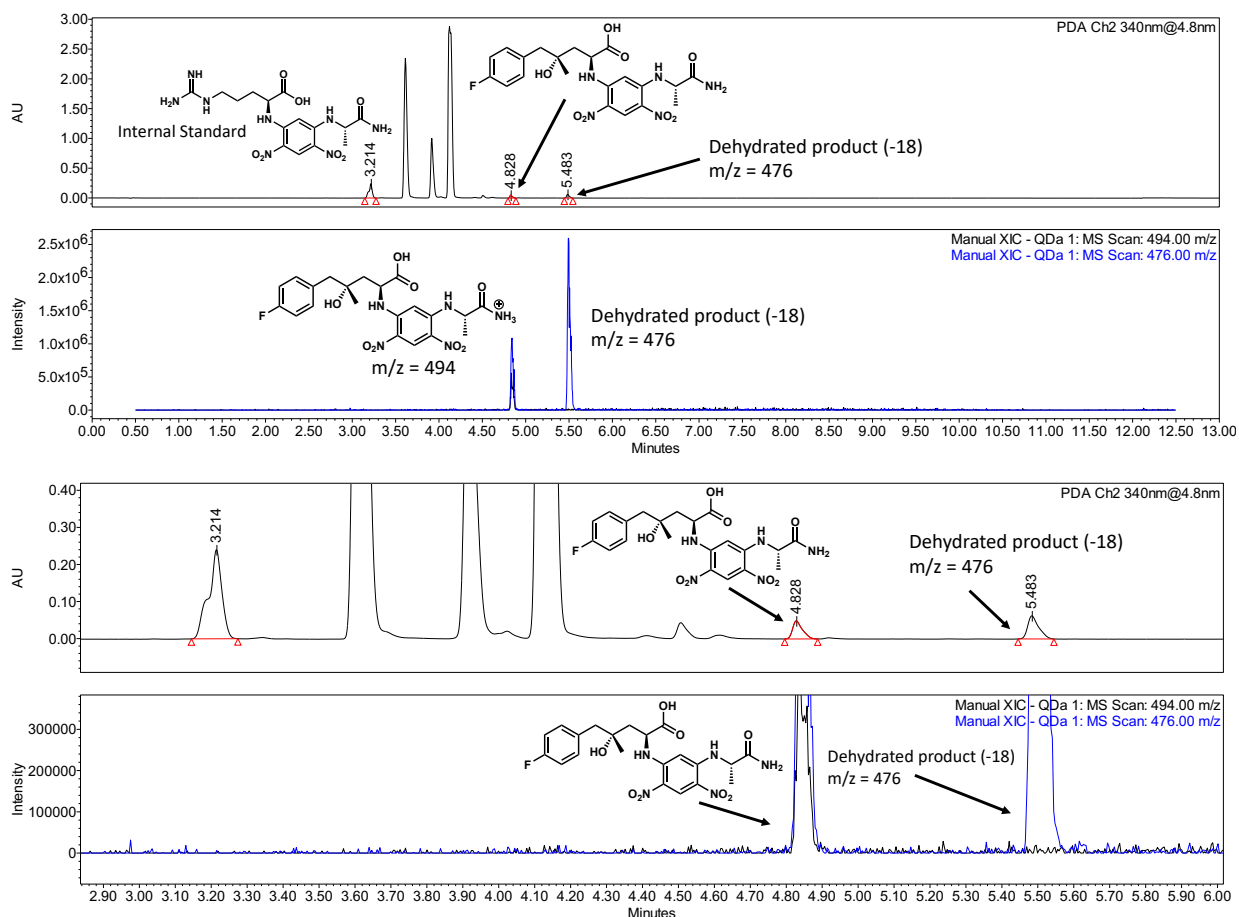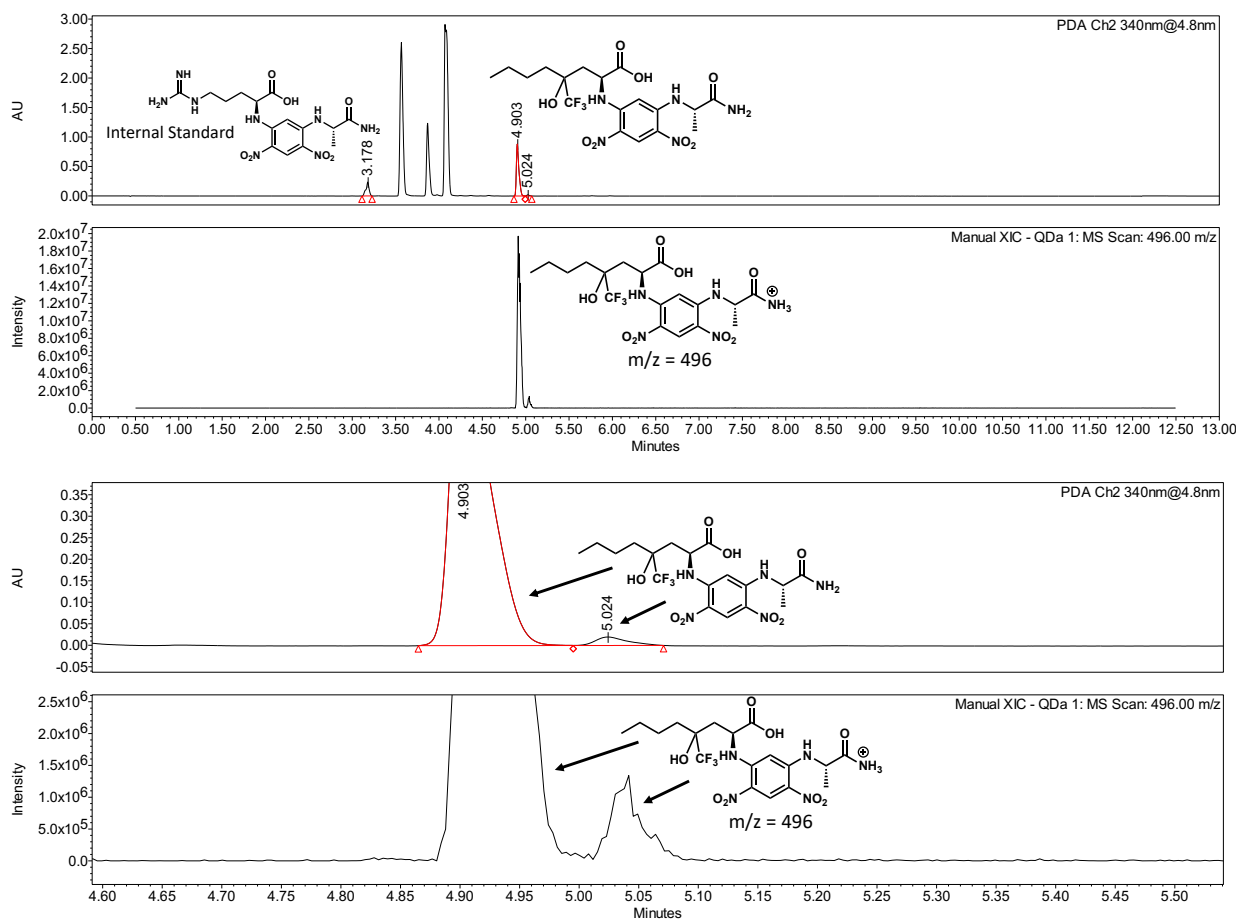
Figure 4.S41. Representative UPLC-MS trace of Marfey's derivatization for 7G11 1,1,1-trifluorohex-2-one. **Conditions:** 50 mM 1,1,1-trifluorohex-2-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
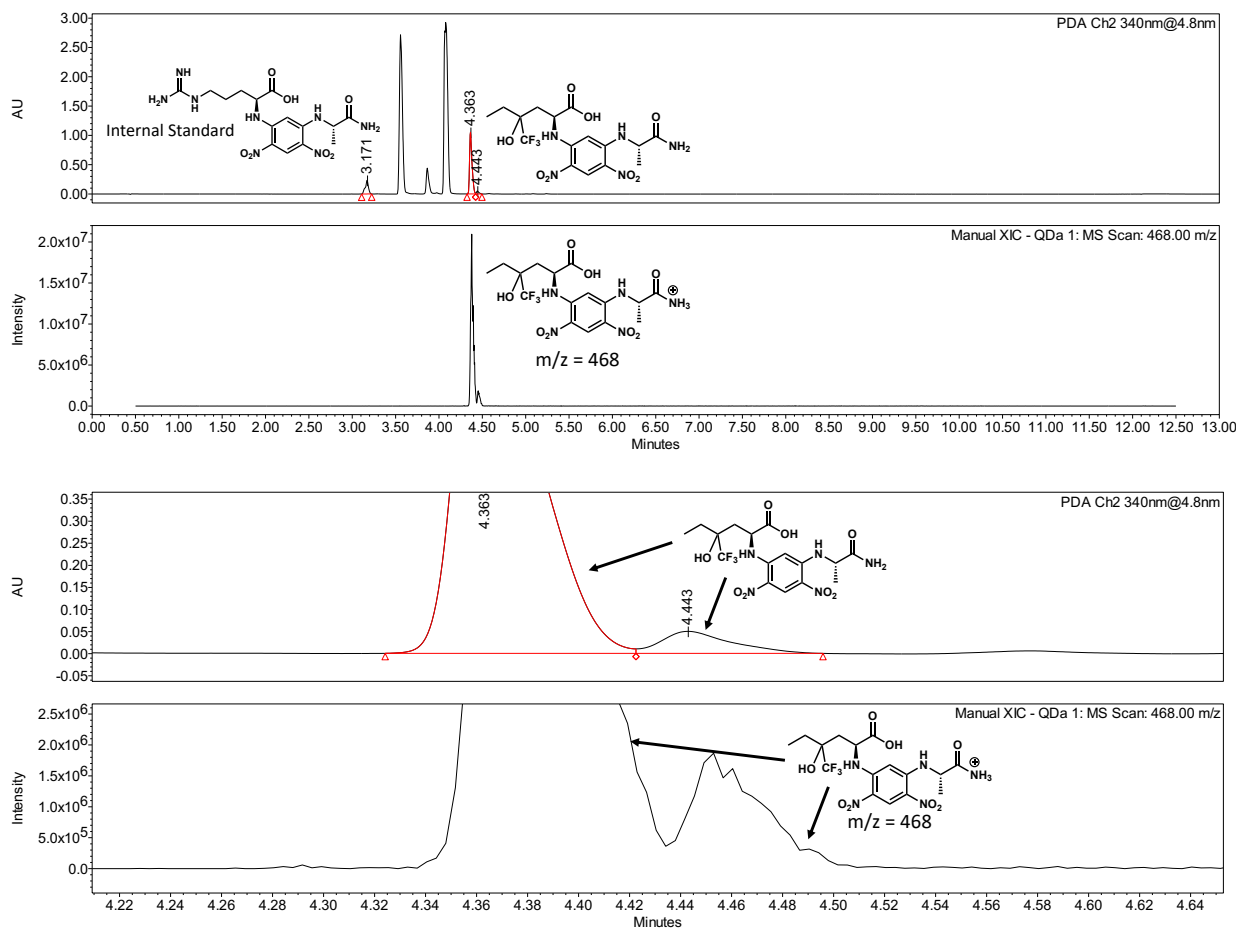
Figure 4.S42. Representative UPLC-MS trace of Marfey's derivatization for 7B05 1,1,1-trifluorobutan-2-one. **Conditions:** 50 mM 1,1,1-trifluorobutan-2-one, 250 mM L-asp, PLP (50x to catalyst), 7B05 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.
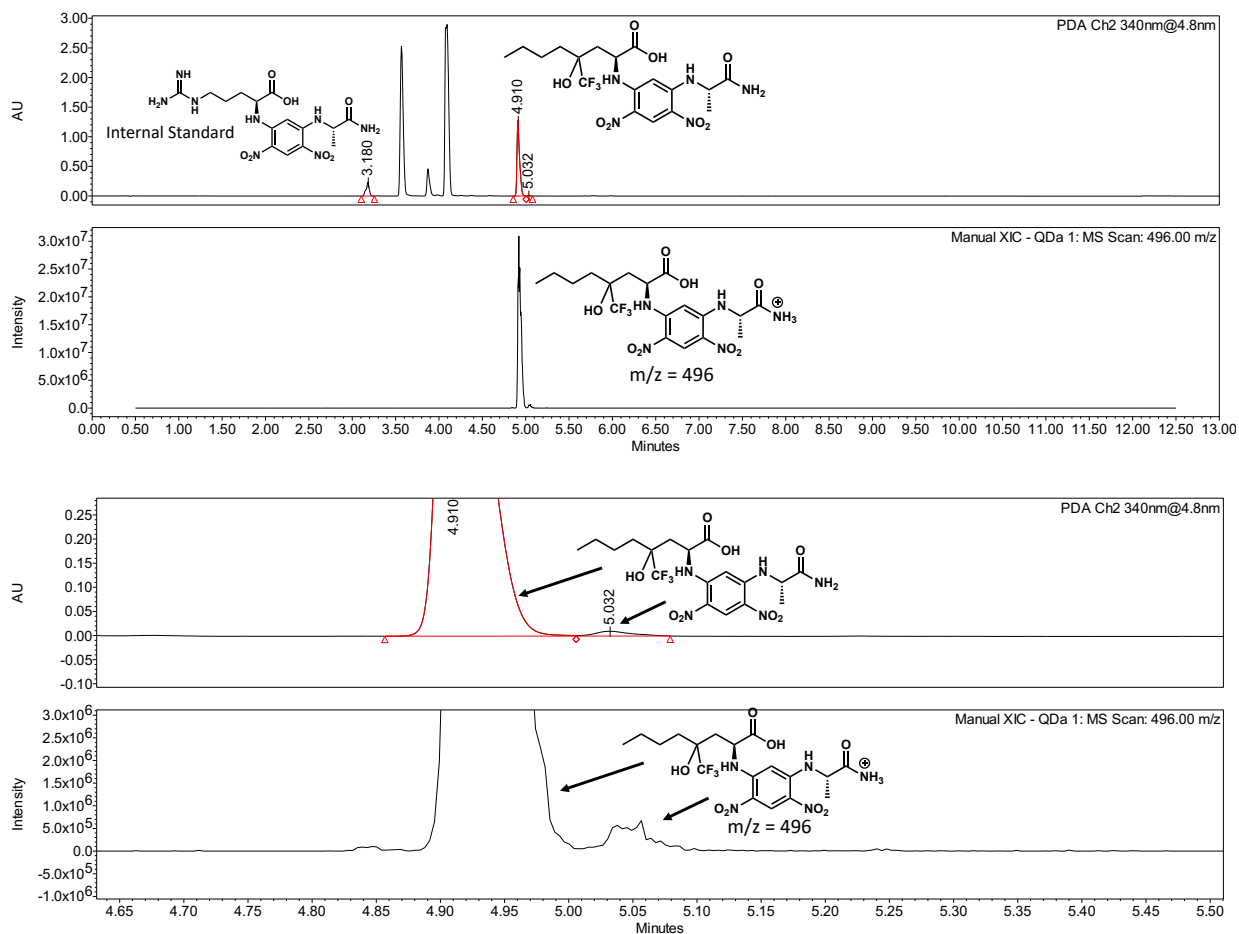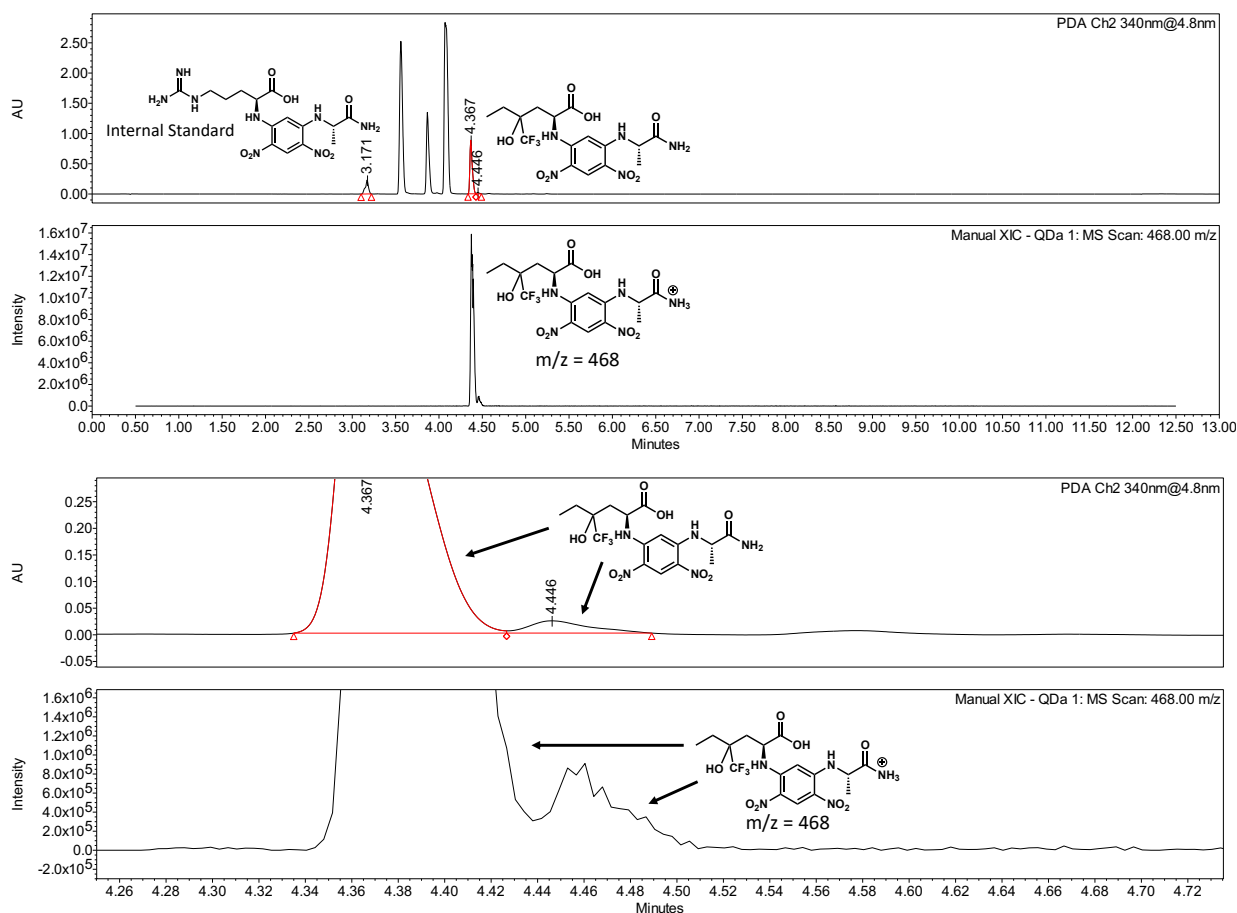
Figure 4.S43. Representative UPLC-MS trace of Marfey's derivatization for 7G11 1,1,1-trifluorobutan-2-one. **Conditions:** 50 mM 1,1,1-trifluorobutan-2-one, 250 mM L-asp, PLP (50x to catalyst), 7G11 (0.1 mol% cat, 1000 Max TON), 5% DMSO, 100 mM potassium phosphate buffer, pH 7.0, 100 mM NaCl, 37 °C, 4 h.

Analytical substrate scope

General Procedure

All reactions were done only once on analytical scale (200 μL). The buffer used for all enzymatic reactions was 100 mM KPi, pH 7.0, 100 mM NaCl. PLP and L-aspartate stock solutions were made in water. Electrophile stock solutions were made in DMSO and each reaction only has a single electrophile. All samples were analyzed following Marfey's derivatization by Waters Acquity UPLC-MS using a BEH C18 column (Waters). To correct for small deviations in injection volume, an internal derivatization standard was included (0.098 mM

L-arginine). Derivatized amino acid product quantitation was performed by integrating chromatograph peaks at 340 nm and corrected by dividing by the internal standard peak area. To calculate product concentrations, a standard curve was generated by subjecting stock solutions of L-phenylalanine (50 mM – 0.4 mM) in water to the identical procedure used to process and derivatize enzymatic reaction solutions, in triplicate, with internal standard. These curves were used to calculate the concentrations of UstD products in solution, and subsequently total turnover numbers after dilution factor correction.

Enzymatic reactions

An Eppendorf tube was charged with electrophile (10 μmol, 1 equiv., 50 mM final concentration, 5% DMSO final concentration). Then, L-aspartate sodium salt monohydrate (50 μmol, 5 equiv., 250 mM final concentration), pyridoxal-5'-phosphate (trifluoromethyl ketones reactions used 10- molar equivalents all others used 50- molar equivalents to final enzyme concentration,) and buffer were aliquoted into the tube. Reactions were initiated by addition of UstD (0.1 mol% catalyst, 1,000 max turnover number) to bring the total reaction volume to 200 μL. The reactions were placed at 37 °C for 4 h. Reactions were quenched with 200 μL of acetonitrile (1 reaction volume) and diluted with 200 μL of 1:1 ACN:DI H_2O to homogenize reaction solutions. Denatured enzyme was removed by centrifugation at 16160 *xg* for 5 min. Marfey's derivatization of the clarified enzymatic reactions were performed to quantify amino acid yield.

Marfey's derivatization procedure

To a fresh 96-well plate, 10 μL of quenched reaction mixture (2.7 mM total amines in reaction, 1 equiv., ~0.8 μmol total amines in reaction) 140 μL of 10.41 mM $NaHCO_3$ with 0.21 mM of L-arginine as an internal standard (9.7 mM $NaHCO_3$ final concentration, 3.5 equiv. base,

2.9 µmol NaHCO$_3$), and 10 mM L-FDAA (5 mM final concentration, 1.8 equiv., 1.5 µmol) were added in each well. The derivatization reaction was allowed to proceed at 37 °C for 18 h. The reactions were quenched with 300 µL of 60 mM HCl in acetonitrile (1 reaction volume) and analyzed via UPLC-MS no later than 24 h after quenching. Note that the amino acid products are susceptible to formation of dehydration products upon addition of the acid required to quench the reactions. This results in two product peaks with masses differing by 18 mass units. The dehydration product could result from elimination of the alcohol to form an alkene or intramolecular lactonization.[41] Typically, the linear product will display mass peaks for both the molecular ion and a dehydrated ion (-18 mass units), while the dehydration products will only display a mass signal associated with dehydration (-18 mass units). Turnover numbers were calculated based on the total integration of linear and dehydrated amino acid product peaks at 340 nm.

Preparative Scale in vitro Biocatalytic Reactions

Procedure P1: *Preparative scale production of unprotected γ-hydroxy amino acids*

A 100-mL round bottom flask was charged with a given ketone (0.5 mmol, 1.0 equiv, 50 mM final concentration), which was then dissolved in an appropriate amount of MeOH (5% v/v final concentration). This solution was then diluted with 100 mM potassium phosphate buffer (pH 7.0) containing 100 mM sodium chloride. L-aspartate sodium salt monohydrate (2.5 mmol, 5.0 equiv, 250 mM final concentration) and 10-50 molar equivalents of pyridoxal-5'-phosphate (PLP) relative to final enzyme concentration were then added, followed by addition of 7G11 or 7B05 (0.1% mol cat). The total reaction volume was 10 mL. The reaction flask was placed in the dark at 37 °C for 4 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

(4,000 rpm, 15 min) to remove aggregated protein. The decanted supernatant was then

concentrated to ~2 mL by rotary evaporation and loaded onto a preparative reverse-phase C18

column pre-equilibrated at 1% methanol: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.

Procedure P2: *Preparative scale production of Fmoc-protected* ***4.7***

A 100-mL round bottom flask was charged with oxetanone (0.5 mmol, 1.0 equiv, 50 mM final

dimer concentration), which was then dissolved in an appropriate amount of MeOH (5% v/v

final concentration). This solution was then diluted with 100 mM potassium phosphate buffer

(pH 7.0) containing 100 mM sodium chloride. L-aspartate sodium salt monohydrate (2.5 mmol,

5.0 equiv, 250 mM final concentration) and 50 molar equivalents of pyridoxal-5'-phosphate

(PLP) relative to final enzyme concentration were then added, followed by addition of 7G11

(0.1% mol cat). The total reaction volume was 10 mL. The reaction flask was placed in the dark

at 37 °C for 4 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

(4,000 rpm, 15 min) to remove aggregated protein. The supernatant was collected in a 250-mL

round bottom and basified to pH ~10 using 6 M NaOH. Then Fmoc-Cl (3.7 mmol, 1.5 equiv of

total amines) was dissolved in 10 mL ACN and added to the round bottom. The reaction was

allowed to stir at room temperature for 4 h. The ACN was removed from the reaction by rotary

evaporation. The resulting aqueous layer was acidified to pH ~ 3 using HCl. An extraction of the

resulting solution was performed with EtOAc (3x25 mL). The combined organic layers were

washed with saturated NaHCO$_3$ (2x25 mL), then saturated NaCl (2x25 mL), and finally dried

over MgSO$_4$. The organic layer was gravity filtered then concentrated to dryness and redissolved

in a minimum amount of DCM. The resulting solution was loaded onto a preparative normal-

phase silica column. Purification was performed via gradient elution (hexane:EtOAc) 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 transferred to a pre-weighed vial, and dried on high vacuum.

Procedure P3: *Preparative scale production of Fmoc-protected* ***4.8***

A 100-mL round bottom flask was charged with 1,3-dihydroxyacetone dimer (0.5 mmol, 1.0

equiv, 25 mM final dimer concentration), which was then dissolved in an appropriate amount of

MeOH (5% v/v final concentration). This solution was then diluted with 100 mM potassium

phosphate buffer (pH 7.0) containing 100 mM sodium chloride. L-aspartate sodium salt

monohydrate (5 mmol, 10.0 equiv, 250 mM final concentration) and 50 molar equivalents of

pyridoxal-5'-phosphate (PLP) relative to final enzyme concentration were then added, followed

by addition of 7G11 (0.4% mol cat). The total reaction volume was 20 mL. The reaction flask

was placed in the dark at 37 °C for 4 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 (4,000 rpm, 15 min) to remove aggregated protein. The

supernatant was collected in a 250-mL round bottom and basified to pH ~10 using 300 μL of 6

M NaOH. Then Fmoc-Cl (3.7 mmol, 0.74 equiv of total amines, 93.7 mM final concentration)

was added and the reaction was allowed to stir at room temperature for 4 h, final pH = 6. The

ACN was removed from the reaction by rotary evaporation. The resulting aqueous layer was

acidified to pH ~ 4 using HCl which caused the formation of a gel-like solid. Addition of EtOAc

(100 mL) was added to the heterogeneous solution and the mixture was stirred vigorously until the solid dissolved completely. An extraction of the resulting solution was performed with EtOAc (3x50 mL). The combined organic layers were concentrated to dryness and redissolved in a minimum amount of DCM. The resulting solution was loaded onto a preparative normal-phase silica column. Purification was performed via gradient elution (hexane:EtOAc) 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 transferred to a pre-weighed vial, and dried on high vacuum.

Characterization of γ-hydroxy amino acid products

4.1 – Synthesis of (2S,4S)-2-amino-5,5,5-trifluoro-4-hydroxy-4-phenylpentanoic acid
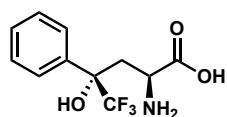
Prepared from 2,2,2-trifluoroacetophenone using procedure **P1**. **Enzyme:** 7G11 **Isolated yield:** 55% **^{1}H NMR** (500 MHz, Deuterium Oxide:MeOH-d^4) δ 7.69 (d, J = 7.6 Hz, 2H), 7.53 – 7.41 (m, 3H), 3.14 (d, J = 11.2 Hz, 1H), 2.70 (d, J = 14.5 Hz, 1H), 2.19 (dd, J = 14.4, 11.3 Hz, 1H). **^{13}C NMR** (126 MHz, Deuterium Oxide:MeOH-d^4) δ 181.6, 140.0z, 131.5, 131.4, 129.8, 128.1 (q, J = 284.4 Hz), 80.70 (q, J = 28.1 Hz) 55.4, 38.6. **^{19}F NMR** (377 MHz, Deuterium Oxide:MeOH-d^4) δ -80.84. **HRMS (ESI):** [M-H]$^-$ calcd. for $C_{12}H_{14}F_3NO_3$, 276.0853; found, 276.0855.

4.2 – Synthesis of (2S,4R)-2-amino-4-benzyl-5,5,5-trifluoro-4-hydroxypentanoic acid
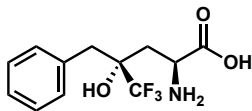
Prepared from 1,1,1-trifluoro-3-phenyl-2-propanone using procedure **P1**. **Enzyme:** 7B05
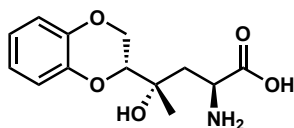
Isolated yield: 46% **^{1}H NMR** (500 MHz, Deuterium Oxide) δ 7.45 – 7.35 (m, 5H), 3.56 (dd, J = 7.7, 6.0 Hz, 1H), 3.19 (d, J = 14.1 Hz, 1H), 3.08 (d, J = 14.1 Hz, 1H), 2.17 (dd, J = 15.1, 6.3 Hz, 1H), 2.03 (dd, J = 15.2, 7.7 Hz, 1H). **^{13}C NMR** (126 MHz, Deuterium Oxide) δ 178.9, 135.0, 131.1, 128.4, 127.3, 126.8 (q, J = 286.3 Hz), 75.1 (q, J = 26.6 Hz), 51.7, 40.2, 34.4. **^{19}F NMR** (377 MHz, D$_2$O) δ -79.58. **HRMS (ESI):** [M-H]$^-$ calcd. for $C_{11}H_{12}F_3NO_3$, 262.0697; found, 262.0696.

4.3 – Synthesis of (2S,4S)-2-amino-5-(furan-2-yl)-4-hydroxy-4-methylpentanoic acid
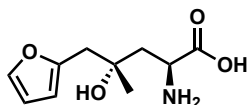
Prepared from 2-furylacetone using procedure **P1**. **Enzyme:** 7G11 **Isolated yield:** 39% **^{1}H NMR** (500 MHz, Deuterium Oxide) δ 7.50 (d, J = 1.2 Hz, 1H), 6.47 (dd, J = 3.2, 1.9 Hz, 1H), 6.30 (d, J = 3.1 Hz, 1H), 3.83 (dd, J = 8.3, 4.8 Hz, 1H), 2.96 (dd, J = 16.9, 2.1 Hz, 2H), 2.13 (dd, J = 15.0, 4.8 Hz, 1H), 1.86 (dd, J = 15.0, 8.4 Hz, 1H), 1.29 (s, 3H). **^{13}C NMR** (126 MHz, D$_2$O) δ 177.9, 152.0, 142.1, 110.6, 108.5, 72.7, 52.6, 41.9, 39.1, 26.7. **HRMS (ESI):** [M-H]$^-$ calcd. for $C_{10}H_{15}NO_4$, 212.0928; found, 212.0925.

4.4 – Synthesis of (2S,4S)-2-amino-4-((R)-2,3-dihydrobenzo[b][1,4]dioxin-2-yl)-4-hydroxypentanoic acid

Prepared from 1-(2,3-dihydro-1,4-benzodioxin-2-yl)ethanone using procedure **P1**. **Enzyme:** 7G11 **Isolated yield:** 31% **¹H NMR** (500 MHz, Deuterium Oxide:MeOH-d^4) δ 7.02 – 6.96 (m, 1H), 6.95 – 6.87 (m, 3H), 4.53 (dd, J = 10.8, 1.6 Hz, 1H), 4.14 – 4.01 (m, 2H), 3.76 (dd, J = 8.2, 5.2 Hz, 1H), 2.33 (dd, J = 14.9, 5.2 Hz, 1H), 1.81 (dd, J = 14.9, 8.2 Hz, 1H), 1.32 (s, 3H). **¹³C NMR** (126 MHz, Deuterium Oxide:MeOH-d^4) δ 179.7, 144.3, 143.8, 123.0, 122.7, 118.3, 117.8, 78.8, 73.2, 65.6, 53.1, 40.7, 23.4. **HRMS (ESI):** [M-H]⁻ calcd. for $C_{13}H_{17}NO_5$, 266.1034; found, 266.1034.

4.5 – Synthesis of (2S)-2-amino-3-(3-hydroxy-7-methyl-3,4-dihydro-2H-benzo[b][1,4]dioxepin-3-yl)propanoic acid
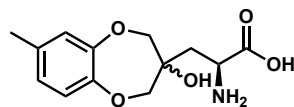
Prepared from calone using procedure **P1**. **Enzyme:** 7G11 **Isolated yield:** 96%, d.r. = 1:1 **¹H NMR** (500 MHz, Deuterium Oxide:MeOH-d^4) δ 6.90 (dd, J = 8.0, 3.9 Hz, 1H), 6.86 – 6.80 (m, 2H), 4.18 – 4.02 (m, 4H), 3.69 (dd, J = 8.7, 4.8 Hz, 1H), 2.24 (s, 3H), 2.21 (dd, J = 15.0, 4.8 Hz, 1H), 1.87 (dd, J = 15.0, 8.5 Hz, 1H). **¹³C NMR** (126 MHz, Deuterium Oxide:MeOH-d^4) δ 181.0, 152.38, 152.36, 150.44, 150.41, 137.0, 127.2, 123.94, 123.92, 123.36, 123.34, 80.7, 80.5, 79.7, 79.5, 76.7, 54.6, 39.4, 22.4. **HRMS (ESI):** [M-H]⁻ calcd. for $C_{13}H_{17}NO_5$, 266.1034; found, 266.1033.

4.6 – **Synthesis of (2*S*)-2-amino-3-(3-hydroxytetrahydrothiophen-3-yl)propanoic acid**
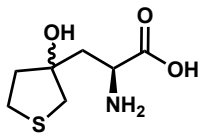

Prepared from 4,5-dihydro-3(*2H*)-thiophenone using procedure **P1**. **Enzyme:** 7G11 **Isolated**

yield: 59%, d.r. = 1:1 **¹H NMR** (500 MHz, Deuterium Oxide) δ 3.89 – 3.80 (m, 1H), 3.07 – 2.86

(m, 4H), 2.39 – 2.09 (m, 3H), 1.95 (dddd, *J* = 12.9, 10.1, 8.1, 2.0 Hz, 1H). **¹³C NMR** (126 MHz,

D₂O) δ 176.8, 176.7, 83.2, 82.9, 53.76, 53.71, 42.3, 42.1, 41.0, 40.8, 39.6, 28.25, 28.21. **HSQC:**

note a high-resolution HSQC is included to differentiate ¹³C signals for each diastereomer.

HRMS (ESI): [M-H]⁻ calcd. for C₇H₁₃NO₃S, 192.0689; found, 192.0687.

4.7 – **Synthesis of ((*S*)-2-(((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-3-(3-hydroxyoxetan-**

3-yl)propanoic acid
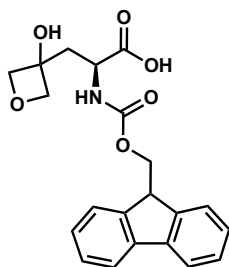
Prepared from 3-oxetanone using procedure **P2**. **Enzyme:** 7G11 **Isolated yield:** 62% **¹H NMR**

(500 MHz, DMSO-*d₆*) δ 7.93 – 7.86 (m, 3H), 7.68 (d, *J* = 7.5 Hz, 2H), 7.42 (t, *J* = 7.5 Hz, 2H),

7.34 (t, *J* = 7.4 Hz, 2H), 4.75 (d, *J* = 7.8 Hz, 1H), 4.68 – 4.62 (m, 2H), 4.60 (d, *J* = 7.8 Hz, 1H),

4.45 (dd, *J* = 18.1, 9.4 Hz, 1H), 4.39 – 4.31 (m, 2H), 4.23 (t, *J* = 6.7 Hz, 1H), 2.86 (dd, *J* = 13.0,

9.5 Hz, 1H), 2.32 (dd, *J* = 13.0, 10.0 Hz, 1H). **¹³C NMR** (126 MHz, DMSO) δ 173.9, 155.6,

143.72, 143.67, 140.7, 127.6, 127.1, 125.09, 125.05, 120.1, 82.0, 80.9, 80.7, 65.7, 49.7, 46.6,

35.9. **HRMS (ESI):** [M-H]⁻ calcd. for C₂₁H₂₁NO₆, 382.1296; found, 382.1299.

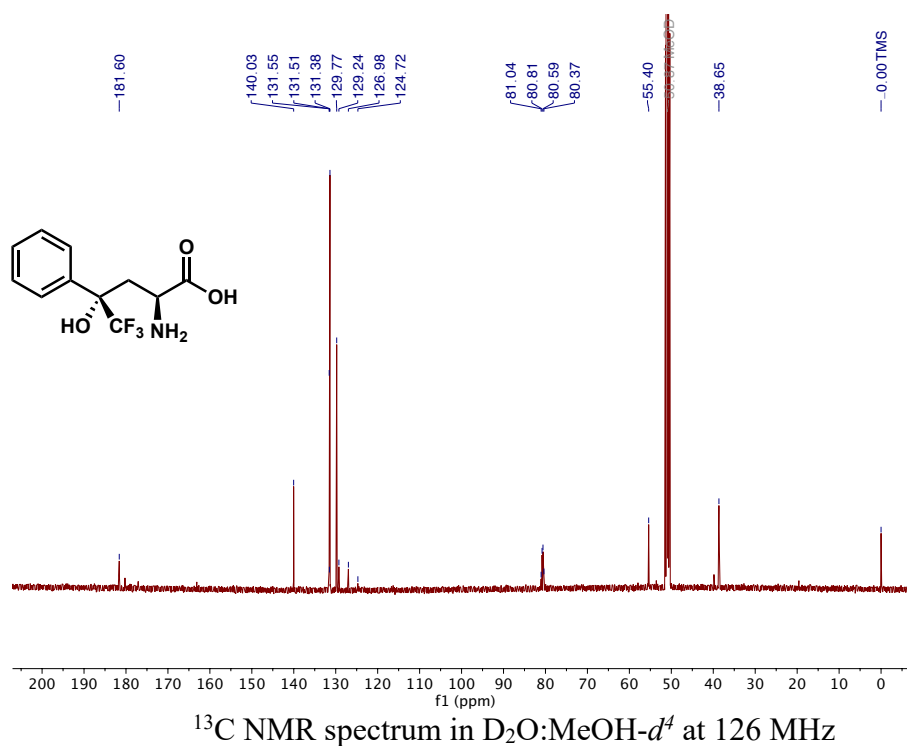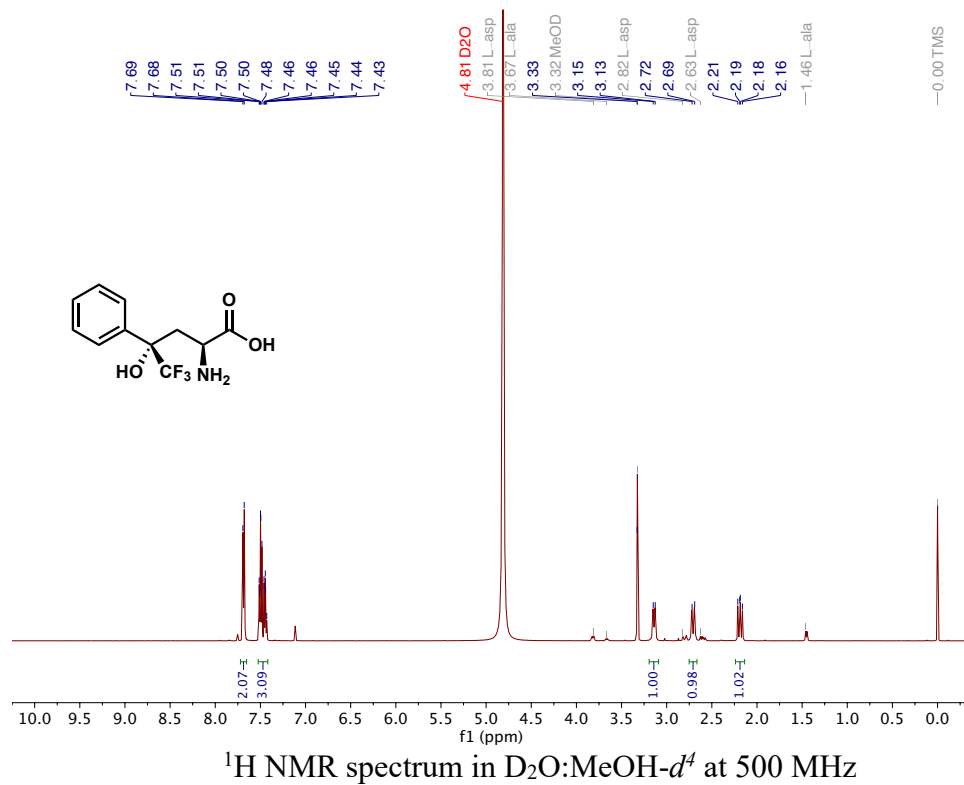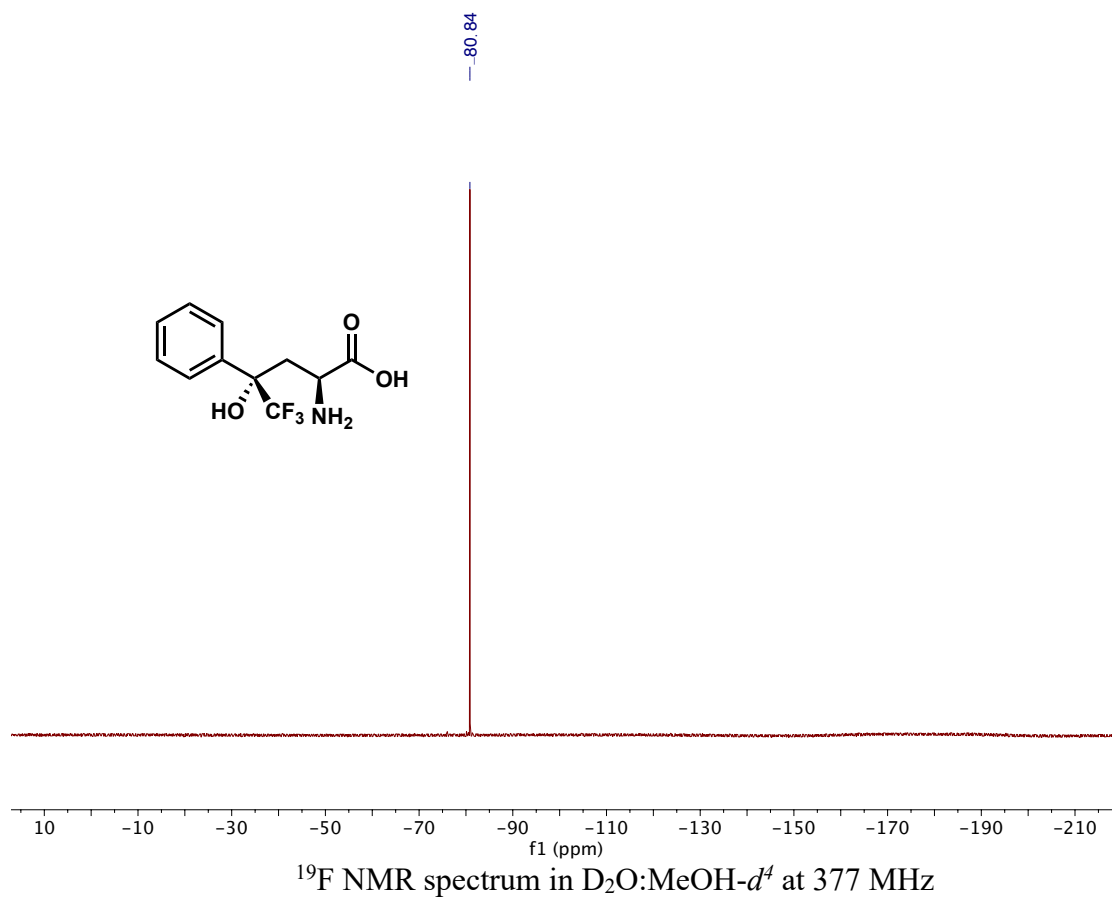### 4.8 – (*S*)-2-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-4,5-dihydroxy-4-(hydroxymethyl)pentanoic acid
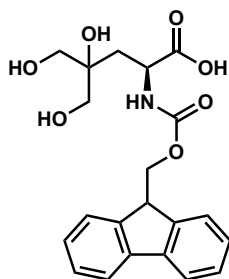
Prepared from 1,3-dihydroxyacetone using procedure **P3**. **Enzyme:** 7G11 **Isolated yield:** 71%

¹H NMR (500 MHz, Methanol-d_4) δ 7.79 (d, J = 7.5 Hz, 2H), 7.65 (d, J = 7.5 Hz, 2H), 7.39 (t, J = 7.5 Hz, 2H), 7.31 (t, J = 7.1 Hz, 2H), 4.66 (t, J = 10.2 Hz, 1H), 4.41 – 4.32 (m, 2H), 4.22 (t, J = 7.0 Hz, 1H), 3.75 (d, J = 12.0 Hz, 1H), 3.68 – 3.58 (m, 3H), 2.52 (dd, J = 12.7, 10.0 Hz, 1H), 2.13 (dd, J = 12.7, 10.5 Hz, 1H). **¹³C NMR** (126 MHz, MeOD) δ 177.4, 158.3, 145.22, 145.20, 142.6, 128.8, 128.2, 126.22, 126.21, 120.9, 87.9, 68.1, 65.7, 65.5, 52.6, 48.3, 33.4. **HRMS (ESI):** [M-H₂O+NH₄]⁺ calcd. for $C_{21}H_{23}NO_7$, 401.1707; found, 401.1700. [M-H₂O+CH₃CO₂]⁻ calcd. for $C_{21}H_{23}NO_7$, 442.1507; found, 442.1508. Note: ammonium acetate was added to samples during preparation for collecting HRMS data, the adducts detected reflect this addition.

NMR Spectra

4.1

^{1}H NMR spectrum in D$_2$O:MeOH-d^4 at 500 MHz

^{13}C NMR spectrum in D$_2$O:MeOH-d^4 at 126 MHz

−80.84

^{19}F NMR spectrum in D$_2$O:MeOH-d^4 at 377 MHz

4.2

^{1}H NMR spectrum in D$_2$O at 500 MHz

^{13}C NMR spectrum in D$_2$O at 126 MHz

−−79.58
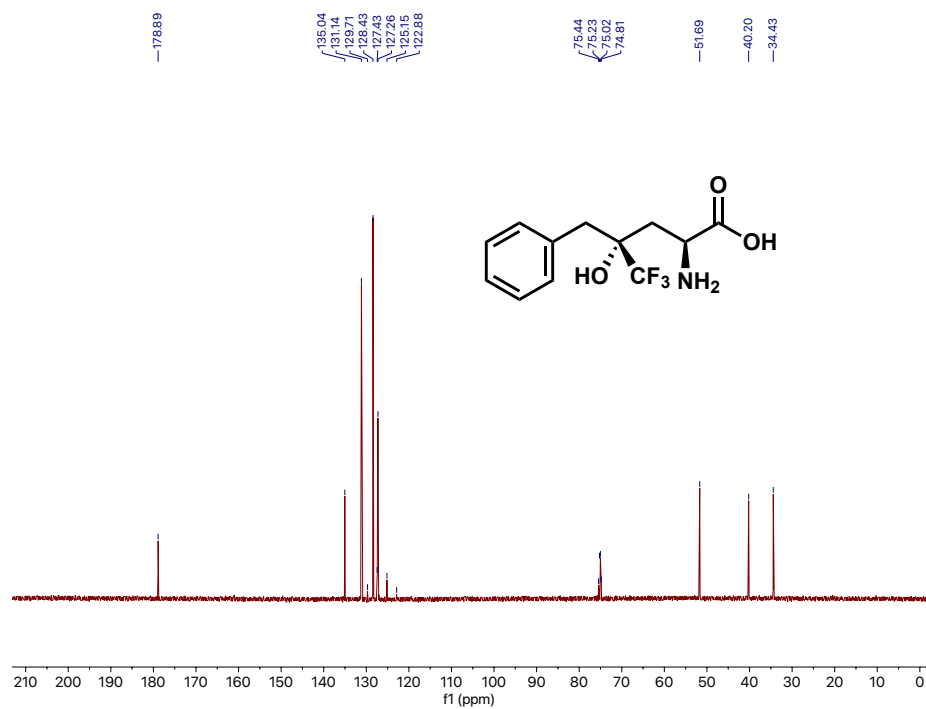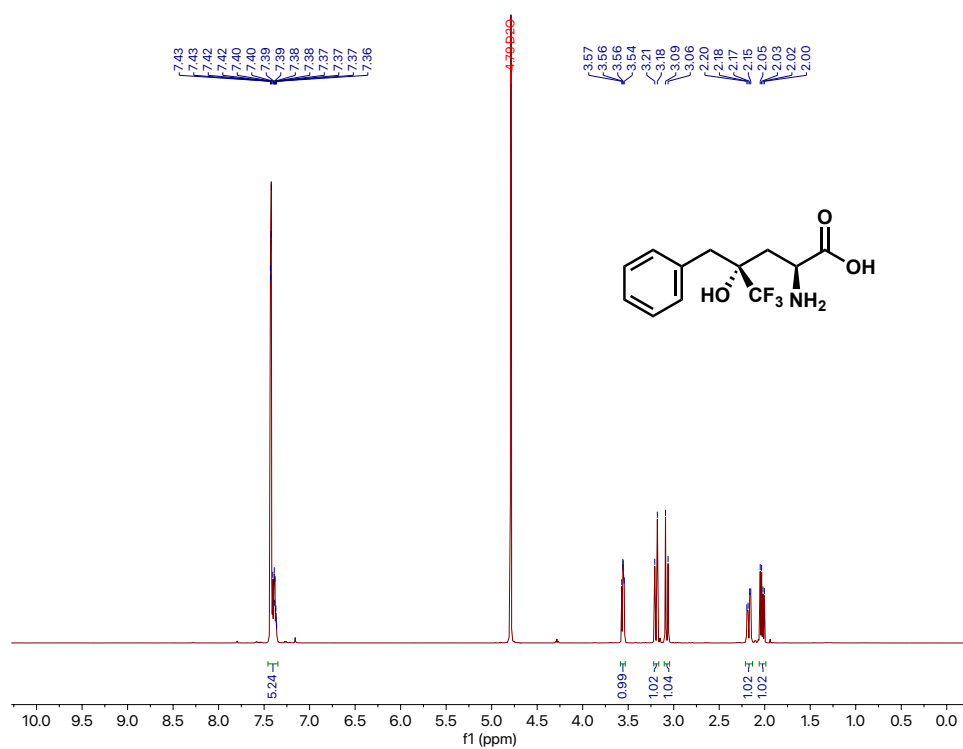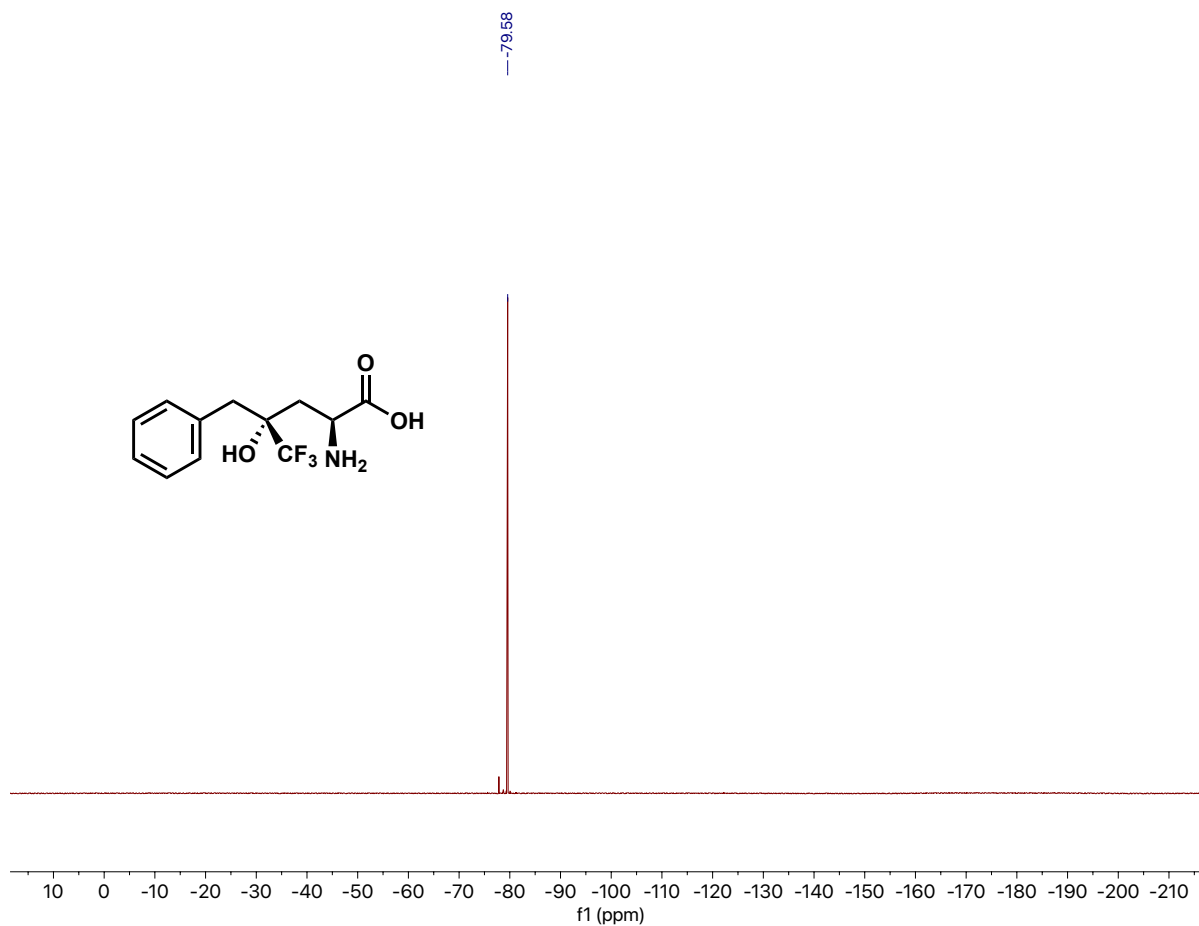
^{19}F NMR spectrum in D$_2$O at 377 MHz

4.3

7.50
7.50
6.47
6.47
6.46
6.46
6.30
6.29
4.79 D2O
3.84
3.83
3.83
3.82
2.99
2.96
2.96
2.93
2.15
2.14
2.12
2.11
1.89
1.87
1.86
1.84
1.29

0.94
1.01
0.97
1.00
2.08
1.01
1.00
3.11

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
f1 (ppm)

^{1}H NMR spectrum in D$_2$O at 500 MHz

177.89
152.02
142.14
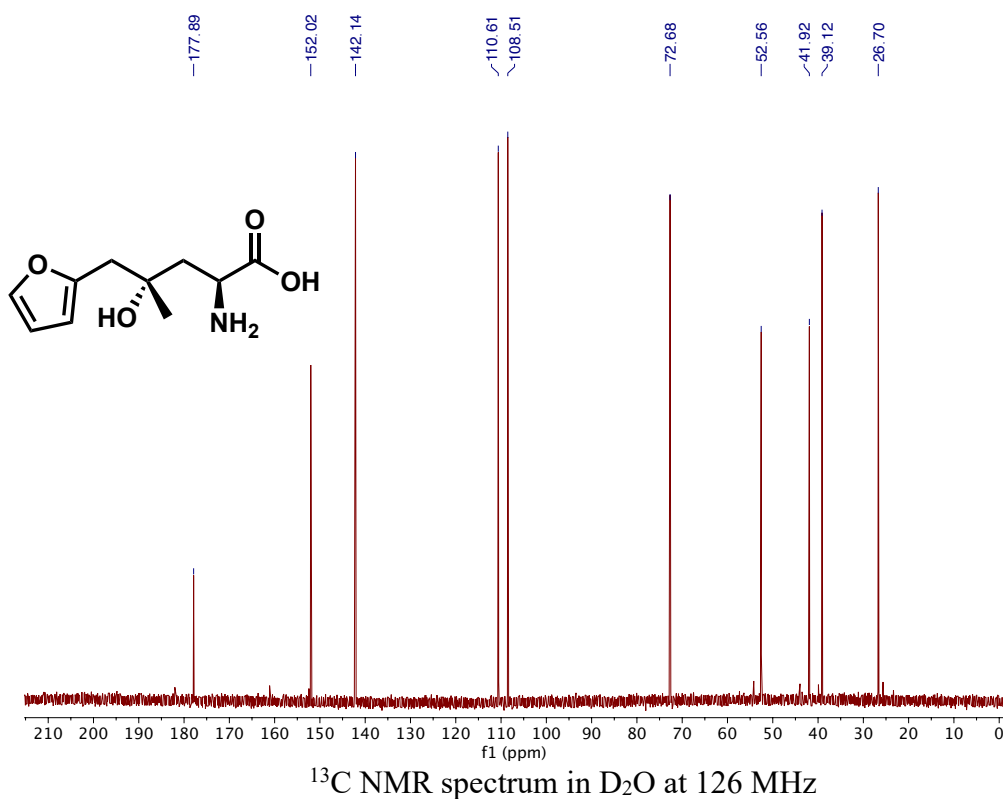110.61
108.51
72.68
52.56
41.92
39.12
26.70

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

^{13}C NMR spectrum in D$_2$O at 126 MHz

4.5

6.91
6.90
6.90
6.89
6.85
6.85
6.84
6.83
6.82
6.81
4.82 HDO
4.17
4.15
4.14
4.14
4.13
4.12
4.11
4.10
4.08
4.07
4.07
4.06
4.05
4.04
4.03
3.70
3.69
3.68
3.68
3.32 MeOH
2.24
2.23
2.22
2.20
2.19
1.91
1.89
1.88
1.86
1.85
0.00 TMS

1.04
2.09
4.26
1.00
3.13
1.00
1.05

^{1}H NMR spectrum in D$_2$O:MeOH-d^4 at 500 MHz

181.11
152.38
152.36
150.44
150.41
137.03
127.18
123.94
123.36
123.34
80.70
80.55
79.69
79.54
76.72
54.60
39.44
22.39
0.00 TMS

80.70
80.55
79.69
79.54

^{13}C NMR spectrum in D$_2$O:MeOH-d^4 at 126 MHz

4.4

^{1}H NMR spectrum in D$_2$O:MeOD-d^4 at 500 MHz
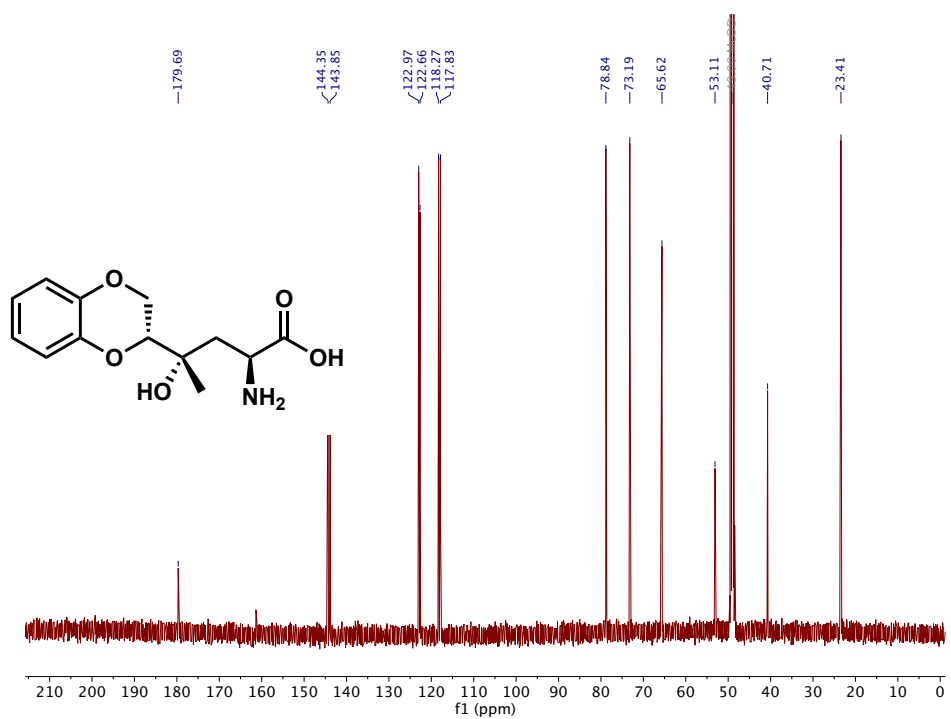
^{13}C NMR spectrum in D$_2$O:MeOH-d^4 at 126 MHz

4.6
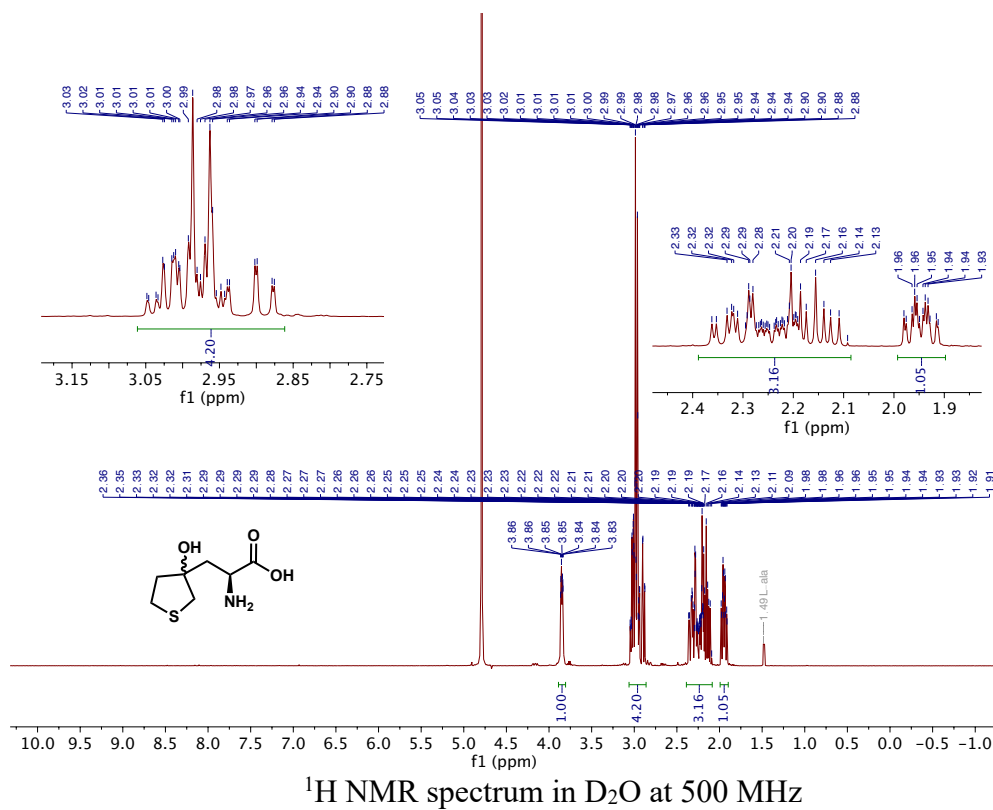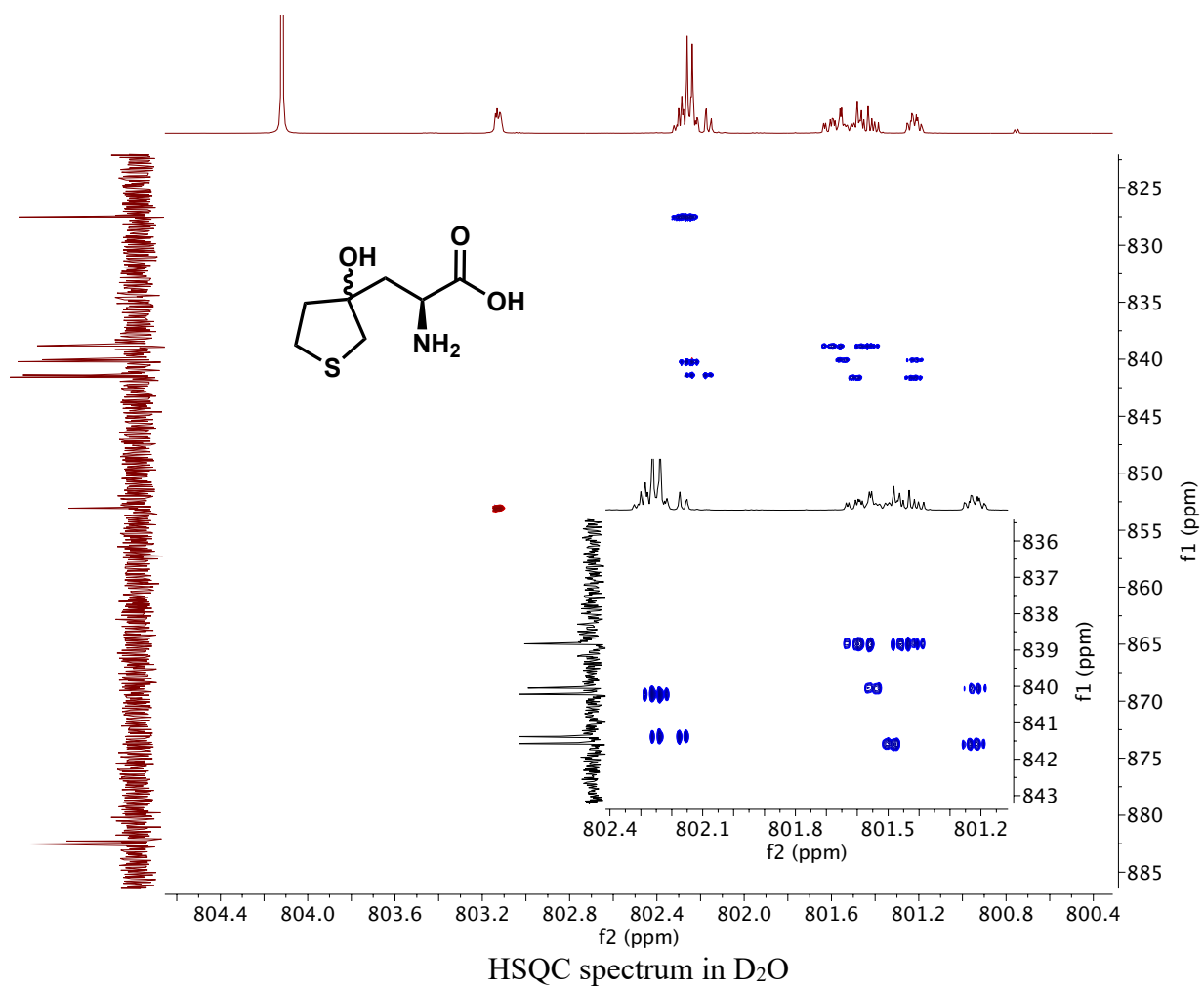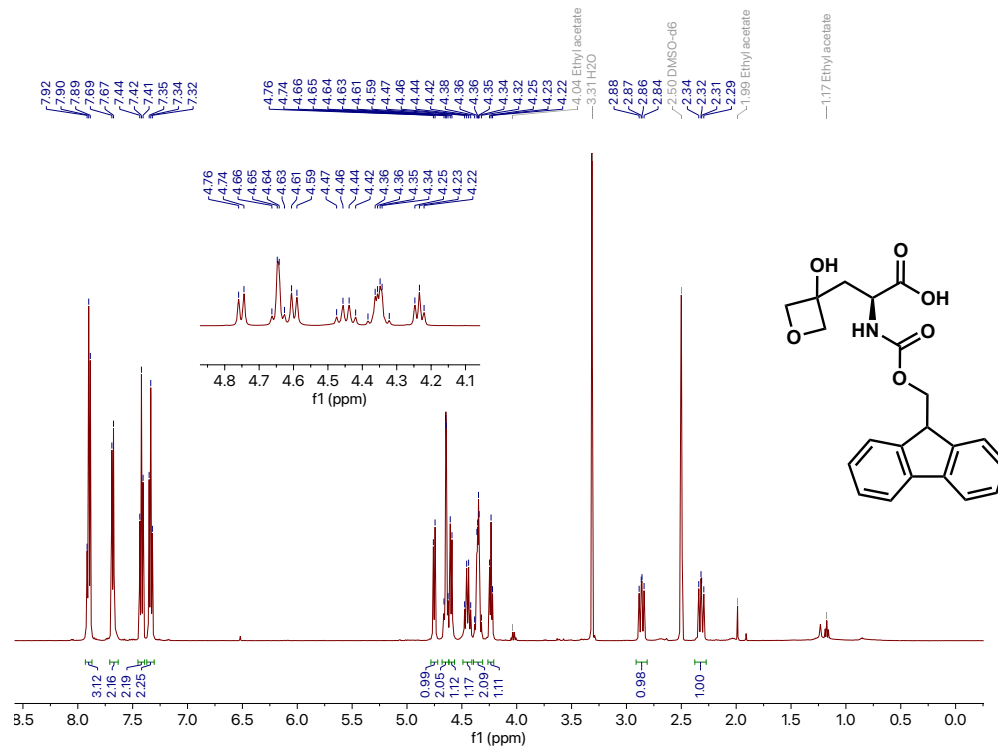
^{1}H NMR spectrum in D$_2$O at 500 MHz

^{13}C NMR spectrum in D$_2$O at 126 MHz

HSQC spectrum in D$_2$O

4.7

¹H NMR spectrum in DMSO-*d⁶* at 500 MHz
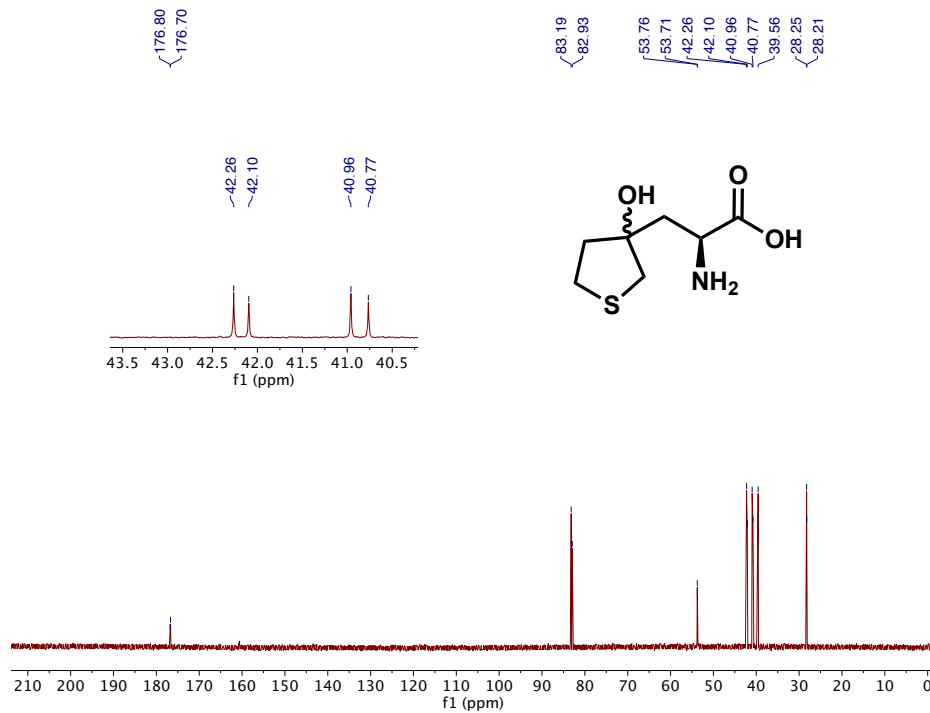

¹³C NMR spectrum in DMSO-*d⁶* at 126 MHz
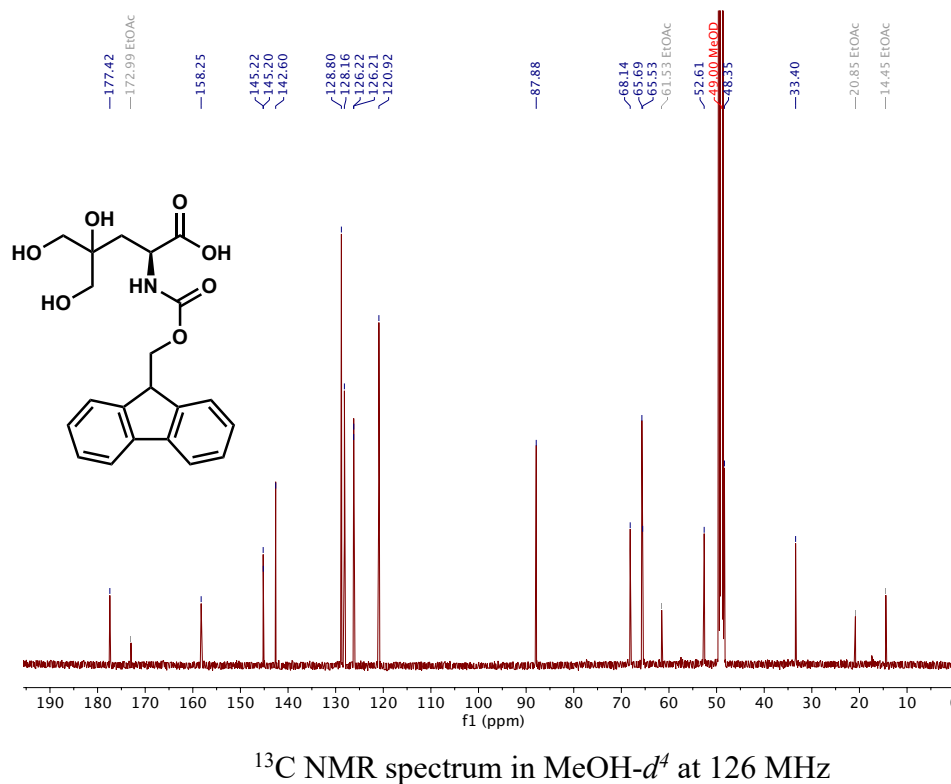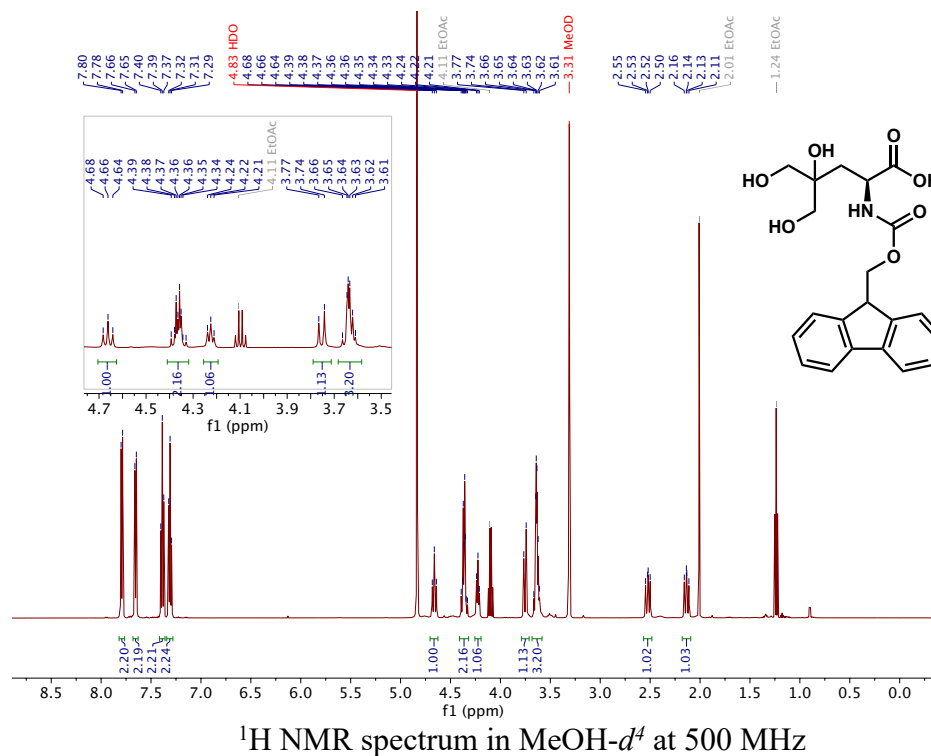
4.8

^{1}H NMR spectrum in MeOH-d^4 at 500 MHz
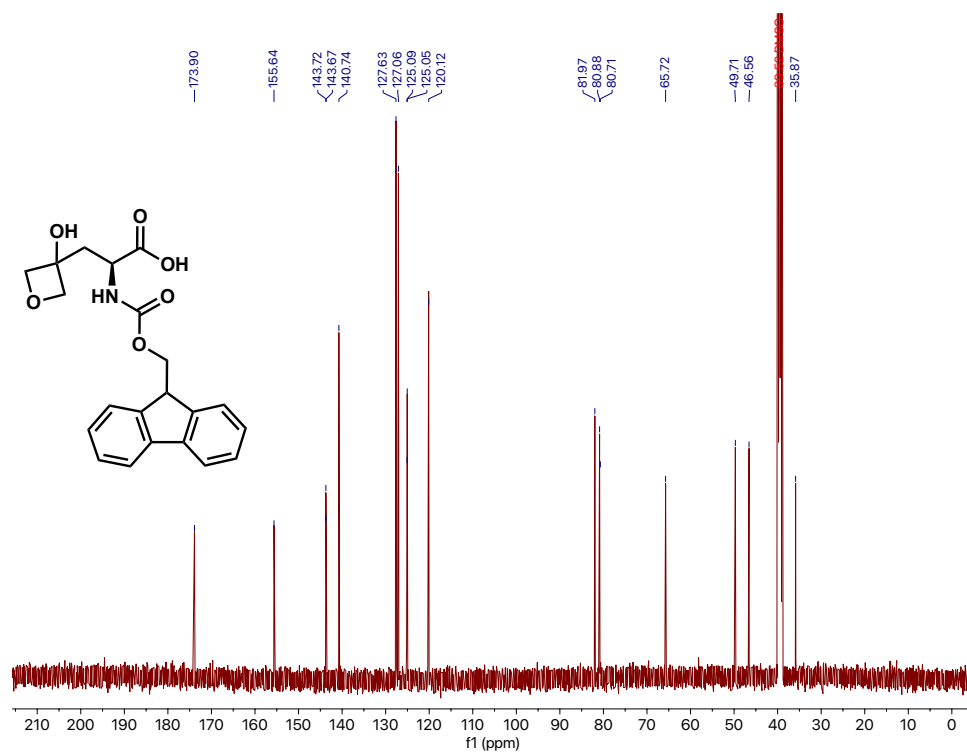
^{13}C NMR spectrum in MeOH-d^4 at 126 MHz

4.5 References

1. Chen, B. S. & Ribeiro de Souza, F. Z. Enzymatic synthesis of enantiopure alcohols: current state and perspectives. *RSC Adv.* **9**, 2102–2115 (2019).

2. Zhang, R. *et al.* Enzymatic Synthesis of Noncanonical α-Amino Acids Containing γ-Tertiary Alcohols. *Angew. Chemie Int. Ed.* **63**, e202318550 (2024).

3. Bruffy, S. K. *et al.* Biocatalytic asymmetric aldol addition into unactivated ketones. (2023). doi:10.26434/CHEMRXIV-2023-2V3CQ

4. Fujimori, D. G. *et al.* Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16498–16503 (2007).

5. Maharaj, V. J., Moodley, N. & Vahrmeijer, H. Characterization of natural monatin isomers, a high intensity sweetener from the plant Sclerochiton ilicifolius from South Africa. *South African J. Bot.* **115**, 37–43 (2018).

6. Yin, X. S., Qi, W. Y. & Shi, B. F. Synthesis of tryptophan-containing 2,5-diketopiperazines via sequential C–H activation: total syntheses of tryprostatin A, maremycins A and B. *Chem. Sci.* **12**, 13137–13143 (2021).

7. Ariza, J., Font, J. & Ortuño, R. M. An efficient and concise entry to (-)-4,5-dihydroxy-d-threo-l-norvaline. Formal synthesis of clavalanine. *Tetrahedron Lett.* **32**, 1979–1982 (1991).

8. Tamura, O. *et al.* Chelation controlled 1,3-dipolar cycloaddition of 5,6-dihydro-5-phenyl-1,4-oxazin-2-one N-oxide with allyl alcohols: A short-step synthesis of clavalanine intermediate. *Tetrahedron Lett.* **40**, 895–898 (1999).

9.   Kortet, S., Claraz, A. & Pihko, P. M. Catalytic Enantioselective Total Synthesis of (+)-Lycoperdic Acid. *Org. Lett.* **22**, 3010–3013 (2020).

10.   Moreno, C. J. *et al.* Synthesis of γ-Hydroxy-α-amino Acid Derivatives by Enzymatic Tandem Aldol Addition–Transamination Reactions. *ACS Catal.* **11**, 4660–4669 (2021).

11.   Pickl, M., Ebner, M., Gittings, S., Clapés, P. & Kroutil, W. Biocatalytic Transamination of Aldolase-Derived 3-Hydroxy Ketones. *Adv. Synth. Catal.* **365**, 1485–1495 (2023).

12.   Li, Z. *et al.* Kinetics and mechanism of oxirane formation by darzens condensation of ketones: Quantification of the electrophilicities of ketones. *J. Am. Chem. Soc.* **140**, 5500–5515 (2018).

13.   Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry*. (AIP Publishing, 2006).

14.   Liu, Y. L. & Lin, X. T. Recent Advances in Catalytic Asymmetric Synthesis of Tertiary Alcohols via Nucleophilic Addition to Ketones. *Adv. Synth. Catal.* **361**, 876–918 (2019).

15.   Palomo, C., Oiarbide, M. & García, J. M. Current progress in the asymmetric aldol addition reaction. *Chem. Soc. Rev.* **33**, 65–75 (2004).

16.   Yamashita, Y., Yasukawa, T., Yoo, W. J., Kitanosono, T. & Kobayashi, S. Catalytic enantioselective aldol reactions. *Chem. Soc. Rev.* **47**, 4388–4480 (2018).

17.   Adachi, S. & Harada, T. Asymmetric Mukaiyama aldol reaction of nonactivated ketones catalyzed by allo-threonine-derived oxazaborolidinone. *Org. Lett.* **10**, 4991–5001 (2008).

18.   Oisaki, K., Zhao, D., Suto, Y., Kanai, M. & Shibasaki, M. New chiral bis(diphenylphospholane) ligands: design, synthesis, and application to catalytic enantioselective aldol reaction to ketones. *Tetrahedron Lett.* **46**, 4325–4329 (2005).

19. Oisaki, K., Zhao, D., Kanai, M. & Shibasaki, M. Catalytic enantioselective aldol reaction to ketones. *J. Am. Chem. Soc.* **128**, 7164–7165 (2006).

20. Duan, Z. *et al.* Enantioselective synthesis of 3-hydroxy oxindoles by ytterbium-catalysed decarboxylative addition of β-ketoacids to isatins. *Org. Biomol. Chem.* **11**, 6456–6459 (2013).

21. Duan, Z. *et al.* A convenient enantioselective decarboxylative aldol reaction to access chiral α-hydroxy esters using β-keto acids. *Beilstein J. Org. Chem. 1095* **10**, 969–974 (2014).

22. Wei, A. J., Nie, J., Zheng, Y. & Ma, J. A. Ni-catalyzed highly chemo-, regio-, and enantioselective decarboxylative aldol reaction of β,γ-unsaturated α-ketoesters with β-ketoacids. *J. Org. Chem.* **80**, 3766–3776 (2015).

23. Gao, H. *et al.* Direct Catalytic Asymmetric Synthesis of β-Hydroxy Acids from Malonic Acid. *Org. Lett.* **17**, 5962–5965 (2015).

24. Luppi, G. *et al.* Dipeptide-catalyzed asymmetric aldol condensation of acetone with (N-alkylated) isatins. *J. Org. Chem.* **70**, 7418–7421 (2005).

25. Wang, F., Xiong, Y., Liu, X. & Feng, X. Asymmetric Direct Aldol Reaction of a-Keto Esters and Acetone Catalyzed by Bifunctional Organocatalysts. (2007). doi:10.1002/adsc.200700285

26. Xu, X. Y. *et al.* Asymmetric organocatalytic direct aldol reactions of ketones with α-keto acids and their application to the synthesis of 2-hydroxy-γ-butyrolactones. *J. Org. Chem.* **72**, 9905–9913 (2007).

27. France, S. P., Lewis, R. D. & Martinez, C. A. The Evolving Nature of Biocatalysis in

Pharmaceutical Research and Development. *JACS Au* **3**, 715–735 (2023).

28.   Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chemie Int. Ed.* **60**, 88–119 (2021).

29.   Müller, M. Enzymatic Synthesis of Tertiary Alcohols. (2014). doi:10.1002/cben.201300005

30.   Maruyama, K. *et al.* Cloning, sequencing, and expression of the gene encoding 4-hydroxy-4-methyl-2-oxoglutarate aldolase from Pseudomonas ochraceae NGJ1. *Biosci. Biotechnol. Biochem.* **65**, 2701–2709 (2001).

31.   Laurent, V. *et al.* Pyruvate Aldolases Catalyze Cross-Aldol Reactions between Ketones: Highly Selective Access to Multi-Functionalized Tertiary Alcohols. *ACS Catal.* **10**, 2538–2543 (2020).

32.   Laurent, V. *et al.* Synthesis of Branched-Chain Sugars with a DHAP-Dependent Aldolase: Ketones are Electrophile Substrates of Rhamnulose-1-phosphate Aldolases. *Angew. Chemie* **130**, 5565–5569 (2018).

33.   Liu, Z. Q., Xiang, Z. W., Shen, Z., Wu, Q. & Lin, X. F. Enzymatic enantioselective aldol reactions of isatin derivatives with cyclic ketones under solvent-free conditions. *Biochimie* **101**, 156–160 (2014).

34.   Lehwald, P., Richter, M., Röhr, C., Liu, H. W. & Müller, M. Enantioselective Intermolecular Aldehyde–Ketone Cross-Coupling through an Enzymatic Carboligation Reaction. *Angew. Chemie Int. Ed.* **49**, 2389–2392 (2010).

35.   Giovannini, P. P., Pedrini, P., Venturi, V., Fantin, G. & Medici, A. Bacillus

stearothermophilus acetylacetoin synthase: A new catalyst for C–C bond formation. *J. Mol. Catal. B Enzym.* **64**, 113–117 (2010).

36.   Stereocenter, M. Q. & Gaggero, N. Building Up Quaternary Stereocenters Through Biocatalyzed Direct Insertion of Carbon Nucleophiles on Ketones. *European J. Org. Chem.* **2019**, 7615–7628 (2019).

37.   Li, R., Li, Z. L., Zhou, H. Y., He, Y. H. & Guan, Z. Enzyme-catalyzed asymmetric construction of chiral tertiary alcohols via aldol reaction using proteinase. *J. Mol. Catal. B Enzym.* **126**, 90–98 (2016).

38.   Hibi, M. *et al.* Characterization of Bacillus thuringiensis L-isoleucine dioxygenase for production of useful amino acids. *Appl. Environ. Microbiol.* **77**, 6926–6930 (2011).

39.   Klein, C. & Hüttel, W. Tertiary alcohol preferred: Hydroxylation of trans-3-methyl-L-proline with proline hydroxylases. *Beilstein J. Org. Chem. 7193* **7**, 1643–1647 (2011).

40.   Plummer, C. M. *et al.* Synthesis of Saturated Benzodioxepinone Analogues: Insight into the Importance of the Aromatic Ring Binding Motif for Marine Odorants. *European J. Org. Chem.* **2015**, 486–495 (2015).

41.   Abu-Salem, Q. *et al.* Stepwise aldol addition and cyanhydrin formation with acetone and thallous cyanide. *Zeitschrift fur Naturforsch. - Sect. B J. Chem. Sci.* **63**, 1040–1044 (2008).