

# Biocatalytic synthesis of non-standard amino acids by a decarboxylative aldol reaction

Andrew Buller (✉ [arbuller@wisc.edu](mailto:arbuller@wisc.edu))

University of Wisconsin-Madison <https://orcid.org/0000-0002-9635-4844>

Jonathan Ellis

University of Wisconsin - Madison

Meghan Campbell

University of Wisconsin - Madison

Prasanth Kumar

University of Wisconsin - Madison

Eric Geunes

University of Wisconsin - Madison

Craig Bingman

University of Wisconsin-Madison <https://orcid.org/0000-0002-3073-5089>

---

Physical Sciences - Article

Keywords:

Posted Date: June 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-544226/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Catalysis on February 21st, 2022.  
See the published version at <https://doi.org/10.1038/s41929-022-00743-0>.

# Abstract

Enzymes are renowned for their catalytic efficiency and selectivity, but relatively few carbon-carbon bond forming enzymes have found their way into the biocatalysis toolbox. While engineering can overcome the challenges associated with C-C bond formation for some enzyme systems, the broader synthetic potential of biocatalysis is hindered by the lack of high-quality C-C bond forming transformations. Here we show that the enzyme UstD performs a highly selective decarboxylative aldol addition with diverse aldehyde substrates to make non-standard,  $\gamma$ -hydroxy amino acids. We increased the activity of UstD through three rounds of classic directed evolution and an additional round of computationally-guided engineering. The enzyme that emerged, UstD2.0, is efficient in a whole-cell biocatalysis format, which circumvents the need for enzyme purification, thereby facilitating its use in traditional organic settings. This new, highly stereoselective enzyme represents a unique expansion of the biosynthetic toolbox. The products are highly desirable, functionally rich bioactive  $\gamma$ -hydroxy amino acids that we demonstrate can be prepared stereoselectively on gram-scale. The X-ray crystal structure of UstD2.0 at 2.25 Å reveals the active site and the molecular basis for the remarkably promiscuity of this catalyst. Taking inspiration from the versatile reactivity of enamines in organic synthesis, we hypothesize that the enamine intermediate of UstD can be engineered to react with electrophiles other than aldehydes. The advent of structural information enabled by engineering of UstD2.0 provides a foundation for probing the unique mechanism of UstD and will guide efforts to expand the reactivity of this unique enzyme.

## Main Text

Major advances have been made in the practical use of enzymes for enantioselective functional group manipulations<sup>9</sup>. For example, enantioselective reduction of ketones or enantiospecific hydrolysis of racemic esters are now routine in process chemistry. There have also been impressive strides made in enzymatic C-H activation<sup>10</sup>. However, the development of enzymes to form C-C bonds on preparative scale lags far behind traditional synthetic organic methodology<sup>11</sup>.

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 microtubulin polymerization (Fig. 1A)<sup>12</sup>. This enzyme, UstD, decarboxylates the side chain of L-aspartate (**1**), forming a putative nucleophilic enamine intermediate (Fig. 1B). This enamine then attacks an aliphatic aldehyde appended to a cyclic tetrapeptide, resulting in the formation of a  $\gamma$ -hydroxy amino acid side chain. The loss of CO<sub>2</sub> renders this enantioselective C-C bond forming reaction irreversible. However, 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  $\gamma$ -hydroxy amino acids (Fig. 1B). Such non-standard amino acids (nsAAs) are found in bioactive natural products, such as caspofungin and clavalanine (Fig. 1A)<sup>13</sup>. While nature employs side chain hydroxylation to tune bioactivity, these nsAAs are virtually absent from medicinal chemistry<sup>14</sup> because they require multistep synthesis<sup>13</sup>. This challenge extends to biocatalysis, where an elegant multi-enzyme cascade was required to access  $\gamma$ -hydroxy nsAAs<sup>15</sup>. Beyond

their use in pharmaceuticals, nsAAs can be enabling for a host of synthetic and chemical biology applications<sup>16,17</sup>. 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.

We expressed *N*-His-UstD (wt-UstD) in *Escherichia coli* (Supplementary Figure 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 (**2a**) and were pleased to observe a successful decarboxylative aldol addition to afford the  $\gamma$ -hydroxy nsAA **3a** by UPLC-MS (Supplementary Figure 2A). 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 analysed the product from a reaction with 4-bromobenzaldehyde (**2b**). The crystal structure of the product (**3b**) revealed the aldol addition occurred with the same stereochemical outcome as the native reaction (Supplementary Figure 2B). These transformations indicated that wt-UstD has potential for organic synthesis, but the comparatively modest activity (<1000 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 **2a**), we hypothesized that directed evolution and reaction condition optimization could be used to increase the catalytic efficiency of UstD toward non-native substrates.

To inform our engineering process, we used a homology model of wt-UstD derived from a distantly related cysteine desulfurase (27% identity)<sup>18,19</sup>. Six residues in the predicted active site were chosen for saturation mutagenesis, and we used benzaldehyde (**2a**) as a model substrate for directed evolution (Fig. 2A). Mutation at positions predicted to form direct contacts with the cofactor resulted in inactivation of the catalyst, a common trend amongst PLP-dependent enzymes<sup>20</sup>. 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. 2B). Concurrently, we employed global random mutagenesis on wt-UstD to search throughout the protein sequence for activating mutations. Serendipitously, a second activating mutation was discovered, L393M, immediately adjacent to Cys392. We combined these mutations to yield the double variant UstD<sup>C392L, L393M</sup>, which had a further increase in activity to 4.9-fold above wild-type (Supplementary Figure 3). It is common for mutation of neighboring residues to display cooperativity<sup>21,22</sup>, and we chose to test additional mutations in this region of the sequence (Fig. 2B). 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<sup>TLM</sup> and UstD<sup>FVF</sup> (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<sup>TLM</sup>. Reaction mixtures were initially coloured yellow (Supplementary Figure 1) by the presence of PLP that co-purified with the enzyme but became colourless over time, suggesting the cofactor is degraded during the reaction. Gratifyingly, supplementation of PLP led to a large increase in product formation (Supplementary Figure 4A). We did not observe a significant change when the concentration of **1** was increased (Supplementary Figure 4B). However, we observed formation of l-alanine in reactions, indicating some **1** is lost to a non-productive protonation of the nucleophilic enamine intermediate<sup>12</sup>. We therefore used aldehyde as the limiting reagent and two equivalents of **1** for subsequent experiments, which identified an optimal initial pH of 7.0 (Supplementary Figure 4C). Lastly, we varied the catalyst loading and found that UstD<sup>TLM</sup> was capable of high conversion (~70%) with just 0.01 mol % catalyst loading (Supplementary Figure 4D). With these optimized conditions, we evaluated the performance of wt-UstD and both activated variants, UstD<sup>TLM</sup> and UstD<sup>FVF</sup>, 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.

Engineering enzymes for activity on a model substrate often leads to specialist catalysts with diminished activity on substrate analogs<sup>23,24</sup>. Initial comparisons among wt-UstD, UstD<sup>FVF</sup>, and UstD<sup>TLM</sup> with a small panel of aldehydes suggested that both variants had evolved towards improved overall activity (Supplementary Figure 5). 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<sup>25</sup>. Product formation was observed with virtually every substrate tested, from the large and hydrophobic biphenyl aldehyde (**2g**) to the small and hydrophilic glycolaldehyde (**2p**) (Fig. 2C). Generally, the variant UstD<sup>TLM</sup> performed the most turnovers and displayed excellent diastereoselectivity, typically forming a 95:5 ratio of diastereomers (dr). While UstD<sup>FVF</sup> typically performed fewer turnovers than UstD<sup>TLM</sup> with most substrates, UstD<sup>FVF</sup> generally had higher selectivity than wt-UstD or UstD<sup>TLM</sup> (Supplementary Table 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 (**2c**), but high activity was observed for the electron deficient *p*-NO<sub>2</sub>-benzaldehyde (**2d**) with both engineered enzymes. To better capture the maximum turnover number with **2d**, we repeated the reactions at lower catalyst loadings, which revealed that the engineered variants can perform ~34,000 turnovers (Supplementary Figure 6). Active site mutagenesis had little apparent impact on reactions with some highly hydrophobic substrates, such as the methoxynaphthyl (**2e**), 3,4-dichlorobenzyl (**2f**), and biphenyl (**2g**) aldehydes; reactivity in these cases may be limited by poor aqueous solubility (Fig. 2C). In contrast, reactivity on *o*-tolualdehyde (**2h**) and thiophene-3-carboxaldehyde (**2i**) increased dramatically during evolution. UstD<sup>TLM</sup> displayed a nine-fold increase in activity on **2i** and a remarkable 23-fold increase in turnovers with **2h** compared to wt-UstD. Activity with the imidazole substrate **2j** was demonstrated and was one of the few substrates for which

wt-UstD had the highest activity; the product is a previously unreported analog of histidine. Reactivity with the cinnamaldehyde (**2k**) improved with both variants relative to wt-UstD. Reactions proceeded smoothly with several aliphatic substrates, including isobutyraldehyde (**2l**), cyclopentylaldehyde (**2m**), and even 10-undecenal (**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 (**2o**), which resulted in formation of  $\gamma$ -hydroxyglutamate, an intermediate in hydroxyproline metabolism<sup>26</sup>. Lastly, we observed good reactivity with glycolaldehyde to yield the di-hydroxylated amino acid **4p**. Previously, a protected form of **4p** was identified as a key intermediate in the synthesis of clavulanine (Fig. 1B)<sup>13</sup>, an antibiotic that inhibits the biosynthesis of methionine<sup>27</sup>. Activity on **2p** increased two-fold, with improved diastereoselectivity and pristine enantioselectivity, for UstD<sup>TLM</sup> 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  $\gamma$ -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. 2D). 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. Bespoke manipulations, such as lactonization with the  $\gamma$ -hydroxy group, can also be employed to facilitate isolation<sup>15</sup>.

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<sup>-1</sup> culture) due to poor solubility (Supplementary Figure 1). While enzyme immobilization can be used to increase the utility of purified protein catalysts<sup>28</sup>, a complementary synthetic methodology would use whole-cell preparations of UstD; this latter approach is attractive to process chemists<sup>29</sup>. 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<sup>TLM</sup> 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 (Supplementary Figure 7). It is well known among protein crystallographers that removing surface Cys residues can increase soluble expression and increase the probability of crystallization<sup>30</sup>. 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. 3A, Supplementary Figure 7).

To efficiently navigate this sequence space, we employed linear regression modelling to predict sequence-activity relationships<sup>31</sup>. 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. 3A). 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 (Supplementary Figure 8). 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).

We evaluated the three most active variants predicted by the model, UstD<sup>TLM-ACASC</sup>, UstD<sup>TLM-ASCSC</sup>, and UstD<sup>TLM-ASASC</sup>. Comparisons of expression and whole cell activity were made between these variants, the parent enzyme, and most active variant identified from screening, UstD<sup>TLM-SCASC</sup>. We were delighted to find the expression titer was increased relative to UstD<sup>TLM</sup> for all variants, up to 48 mg protein L<sup>-1</sup> culture (Supplementary Figure 9). While purified enzyme activity is slightly decreased for the new variants, their overall activity in whole cells is significantly improved (Fig. 3A, Supplementary Figure 9). Tests at analytical scale showed, at 0.25% w/v cell loading, that UstD<sup>TLM</sup> formed **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<sup>TLM-ACASC</sup> (designated UstD<sup>2.0</sup>), produced **3a** in 31% yield, a 2.4-fold boost over UstD<sup>TLM</sup> and a cumulative 15-fold boost over wild type. Higher conversions were achieved by increasing the cell loading of UstD<sup>2.0</sup> to 1% w/v, which afforded **3a** in 78% yield on analytical scale (Fig. 3A). To demonstrate the utility of UstD<sup>2.0</sup>, large-scale reactions were carried out with **2a** and **2d**. Reaction with **2a** at 0.5% w/v catalyst loading afforded 0.80 g **3a** in 77% isolated yield with pristine stereoselectivity following purification by reverse-phase chromatography. Reaction with **2d** at just 0.1% w/v catalyst loading provided 1.4 g **3d** in 98% isolated yield with high stereoselectivity (see SI for details). Notably, these cell loadings are sufficient for process-scale biocatalytic reactions<sup>32</sup>, illustrating that UstD<sup>2.0</sup> can operate on the scale needed to meet the demands of practical organic synthesis.

While the engineering we report here produced a 'generalist' variant of UstD, structural information could guide more targeted engineering for the production of specific  $\gamma$ -hydroxy nsAAs. Despite extensive efforts, we were unable to produce crystals of wt-UstD. In contrast, UstD<sup>2.0</sup> readily crystallized, which we attribute to the decrease in surface Cys residues. The 2.25-Å crystal structure of UstD<sup>2.0</sup> was determined using

experimental phases from a Au(III) derivative (Fig. 3B, PDB ID: 7MKV). This structure revealed an active site at the dimer interface, which is common among fold-type I PLP-dependent enzymes<sup>33</sup>. 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 (Supplementary Figure 10).

In the future, we envision engineering UstD for increased activity with non-aldehyde substrates. As an initial demonstration, we showed that purified UstD<sup>2.0</sup> performs ~50 turnovers with the ketone substrate trifluoroacetone to produce a nsAA bearing a tertiary alcohol side chain (Supplementary Figure 11). 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 intermediates formed by TrpB and reactions with attenuated substrates were enabled by directed evolution that increased the lifetime of the reactive intermediate.<sup>34,35</sup> Hence, future engineering to decrease the rate of enamine protonation in UstD<sup>2.0</sup> may further expand the substrate scope.

Here, we engineered a C-C bond forming enzyme, UstD, that catalyzes a decarboxylative aldol addition using the loss of CO<sub>2</sub> from l-aspartate as a thermodynamic driving force to produce  $\gamma$ -hydroxy amino acids. This mechanism of action and innate aldehyde promiscuity 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<sup>FVF</sup> and UstD<sup>TLM</sup>, 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 new UstD<sup>2.0</sup> is poised to deliver a desirable new nsAA synthon for medicinal chemistry and the crystal structure of UstD<sup>2.0</sup> will facilitate future work to explore the mechanism and expand the reactivity of this unique enzyme.

## Declarations

**Acknowledgments:** We thank I. Guzei for small molecule x-ray structure determination and S.H. Gellman and members of the Buller group for critical reading of the manuscript. The crystal mounting and data collection were mediated by the Collaborative Crystallography Core, Department of Biochemistry, UW-Madison and data were collected at the Life Sciences Collaborative Access Team beamline 21ID-D at the Advanced Photon Source, Argonne National Laboratory and we thank Z. Wawrzak for technical assistance during data collection. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). This work

was supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison, Wisconsin Alumni Research Foundation, National Institute of Health (grant DP2-GM137417), Morgridge Institute for Research – Metabolism Theme Fellowship, and the NIH Biotechnology Training Grant (T32-GM008349). The Bruker AVANCE III-500 NMR spectrometers were supported by the Bender Fund. The Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. The Bruker D8 VENTURE Photon III X-ray diffractometer was partially funded by NSF Award (#CHE-1919350) to the UW–Madison Department of Chemistry

### Author contributions:

A.R.B., J.M.E. conceptualized the goals and aims of the project. J.M.E., M.E.C., P.K., E.P.G., C.A.B., A.R.B. carried out development of the chemistry and enzyme. J.M.E. developed code for data analysis and developed the linear regression model. J.M.E., M.E.C. verified results. J.M.E., M.E.C., P.K., A.R.B. prepared figures and data visualizations. A.R.B. secured funding for the project leading to this publication. A.R.B. coordinated team members for the development of the chemistry and enzyme evolution. C.A.B supervised data acquisition of protein crystals leading to a resolved crystal structure. A.R.B. supervised the research activity planning and execution. J.M.E, M.E.C., A.R.B. prepared the initial manuscript. J.M.E., M.E.C., P.K., A.R.B. reviewed and edited the initial manuscript providing critical commentary and revisions.

**Competing interests:** Authors declare that they have no competing interests.

### Additional Information:

**Supplementary Information** is available for this paper.

**Correspondence and requests for materials** should be addressed to A.R.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints)

## References

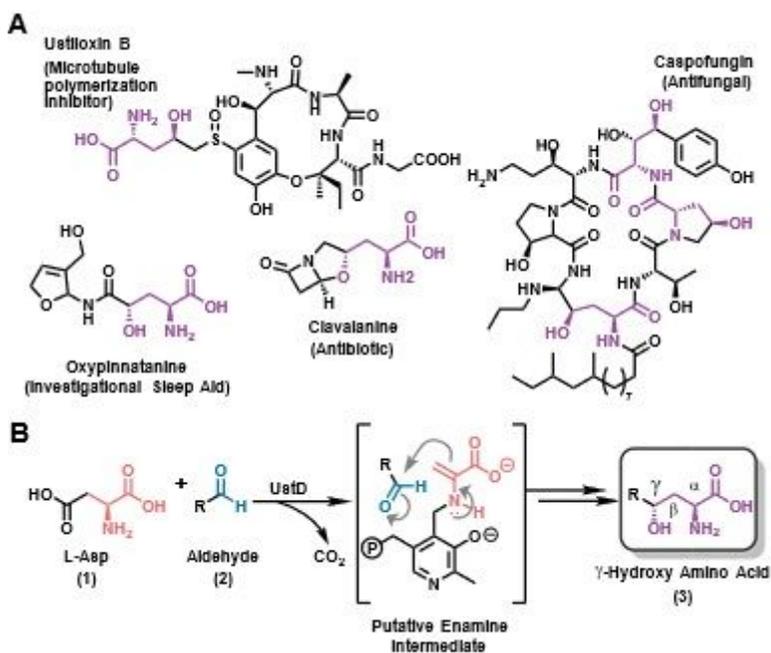
1. Schmidt, N. G., Eger, E. & Kroutil, W. Building Bridges: Biocatalytic C-C-Bond Formation toward Multifunctional Products. *ACS Catalysis* **6**, 4286–4311 (2016).
2. Fesko, K. & Gruber-Khadjawi, M. Biocatalytic Methods for C-C Bond Formation. *ChemCatChem* **5**, 1248–1272 (2013).
3. Fujii, I. Heterologous expression systems for polyketide synthases. *Natural Product Reports* **26**, 155–169 (2009).
4. 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).

5. 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).
6. 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).
7. Zetsche, L. E. & Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nature Reviews Chemistry* **4**, 334–346 (2020).
8. Mukherjee, S., Yang, J. W., Hoffmann, S. & List, B. Asymmetric enamine catalysis. *Chemical Reviews* **107**, 5471–5569 (2007).
9. 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).
10. Zhang, X. *et al.* Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science (80-. )*. **369**, 799–806 (2020).
11. 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).
12. 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).
13. 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).
14. Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *Journal of Medicinal Chemistry* **59**, 10807–10836 (2016).
15. Moreno, C. J. *et al.* Synthesis of  $\gamma$ -Hydroxy- $\alpha$ -amino Acid Derivatives by Enzymatic Tandem Aldol Addition–Transamination Reactions. *ACS Catal.* **11**, 4660–4669 (2021).
16. 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).
17. Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019).
18. Yang, J. *et al.* The I-TASSER suite: Protein structure and function prediction. *Nature Methods* **12**, 7–8 (2014).

19. Ho, T. H. *et al.* Catalytic Intermediate Crystal Structures of Cysteine Desulfurase from the Archaeon *Thermococcus onnurineus* NA1. *Archaea* **2017**, 1–11 (2017).
20. Kumar, P. *et al.* L -Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate. *ACS Chem. Biol.* **16**, 95 (2021).
21. 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).
22. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10**, 866–876 (2009).
23. 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).
24. 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).
25. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **49**, 591–596 (1984).
26. Wu, G. *et al.* Proline and hydroxyproline metabolism: Implications for animal and human nutrition. *Amino Acids* **40**, 1053–1063 (2011).
27. 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).
28. 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).
29. 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).
30. 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).
31. Fox, R. Directed molecular evolution by machine learning and the influence of nonlinear interactions. *J. Theor. Biol.* **234**, 187–199 (2005).

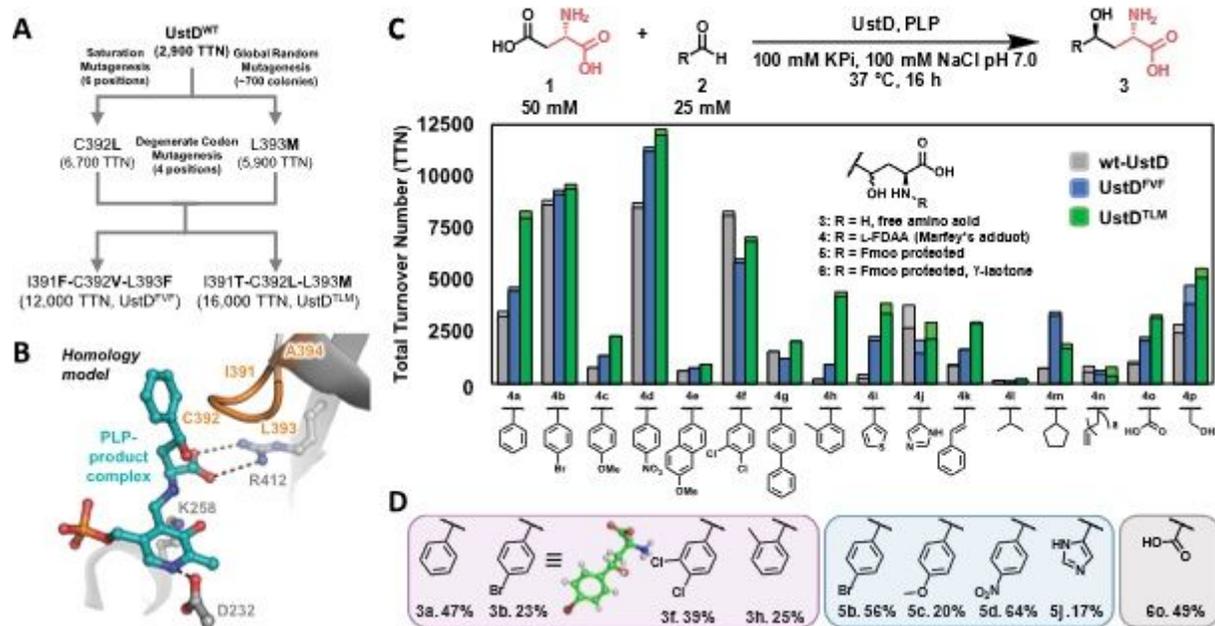
32. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* (80- ). **366**, 1255–1259 (2019).
33. Eliot, A. C. & Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **73**, 383–415 (2004).
34. 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).
35. Boville, C. E. *et al.* Engineered Biosynthesis of  $\beta$ -Alkyl Tryptophan Analogues. *Angew. Chemie - Int. Ed.* **57**, 14764–14768 (2018).

## Figures



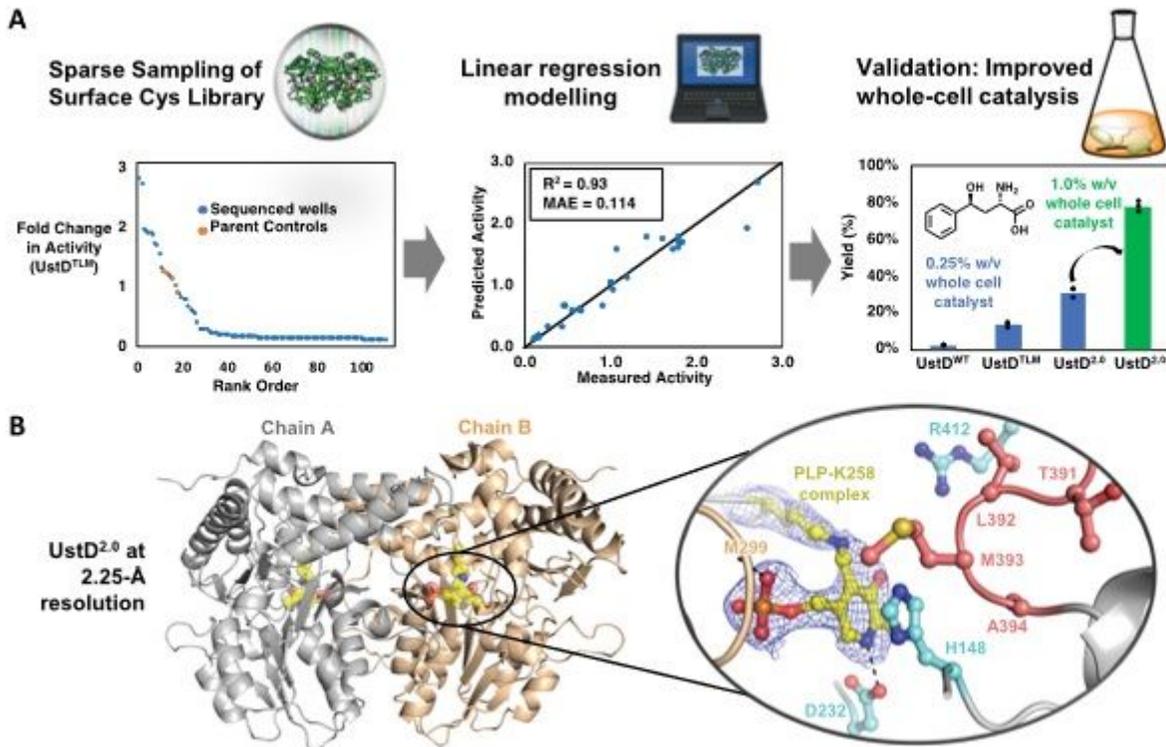
**Figure 1**

Relevance and mechanism of enzymatic C-C bond formation. A) Bioactive molecules with a  $\gamma$ -hydroxy amino acid motif shown in purple. The native product of UstD is Ustiloxin B. B) The generalized decarboxylative aldol reaction of UstD showing the putative enamine nucleophilic intermediate.



**Figure 2**

Directed evolution of UstD and evaluation of variants. A) Lineage of activated UstD variants. Standard screening conditions: 25 mM benzaldehyde, 50 mM L-aspartate, buffer (100 mM KPi, pH 7.0, 100 mM NaCl), 37 °C, 16 h. Catalyst activity measured by total turnover number (TTN). B) Computational model of UstD bound to 3a, derived through homology modeling. Active site residues are shown as sticks and loop residues targeted for mutagenesis are coloured in orange. Potential hydrogen bonds are shown as black dashes. C) Performance evaluation of UstD and activated variants measured by Marfey's analysis of the enzymatic products. Lighter coloured bar sections represent the amount of the other Cy epimer from which diastereomeric ratios are calculated. See Supporting Methods for details. D) Synthesis of select products at 0.2 mmol scale with isolated yields. The different purification strategies are denoted by the different colours, free amino acid (purple), Fmoc protected amino acid (blue), lactonization with Fmoc protection (grey). Note, reactions from which 3b was purified used wt-UstD.



**Figure 3**

Engineering UstD for increased crystalizability 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/activity data. The second stage builds a linear regression model to correlate sequences to activity. This regression model is then used to predict activated sequences which are validated in the last stage using whole cell catalyst. B) Cartoon representation of the overall structure of UstD2.0. Individual monomers are coloured grey (chain A) and brown (chain B). PLP-K258 complex is shown as semitransparent 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. TLMA loop residues are coloured in salmon. Hydrogen bonds are shown as black dashes.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SIFormattedNatureketoneadded.pdf](#)