

Substrate multiplexed protein engineering facilitates promiscuous biocatalytic synthesis

Authors: Allwin D. McDonald^{1,2}, Peyton M. Higgins^{1,2}, & Andrew R. Buller^{1,*}

¹*Department of Chemistry, University of Wisconsin–Madison, 1101 University Avenue, Madison, Wisconsin 53706, United States*

²*These authors contributed equally to this work*

* To whom correspondence should be addressed: arbuller@wisc.edu

Abstract

Enzymes with high activity are readily produced through protein engineering, but intentionally and efficiently engineering enzymes for an expanded scope is a contemporary challenge. Measuring reaction outcomes on mixtures of substrates, called here SUBstrate Multiplexed Screening (SUMS), has long been used to rigorously quantitate enzyme specificity. Despite the potential utility of SUMS to guide engineering of promiscuous enzymes, this approach has not found widespread adoption in biocatalysis. Here, we develop principles of how to design robust SUMS methods that, rather than assess absolute specificity, use heuristic readouts of substrate promiscuity to identify hits for further investigation. This rich information enables engineering of activity for multiple substrates simultaneously and identifies enzyme variants with altered promiscuity, even when overall activity is lower. We demonstrate the effectiveness of SUMS by engineering two enzymes to produce pharmacologically active tryptamines from simple indole precursors in a biocatalytic cascade. These advances leverage common laboratory equipment and represent a highly accessible and customizable method for enzyme engineering.

Biocatalysts are prized for their ability to perform well-defined transformations. However, the use of enzymes in practical chemical synthesis is often hampered by their small or poorly understood substrate scopes.¹ Using traditional protein engineering approaches, activity can readily be increased on a model compound.²⁻⁴ Advances in both smart library design⁵⁻⁸ and screening speed⁹⁻¹² have aided efforts to engineer enzymes. Many engineering campaigns using these approaches result in highly promiscuous catalysts.^{2,12-14} However, the scope of intermediates along evolutionary lineages are often unknown. Consequently, when protein engineering does yield a catalyst with a limited scope, evolution is often tediously repeated to generate activity with additional substrates.^{8,15-17} Screening for activity on a single substrate necessarily overlooks mutations that are activating for substrates not included in the screen and can *inadvertently lead to enzymes with high activity but narrow substrate scopes*.¹⁶⁻¹⁸ Methods that directly assess catalyst promiscuity would overcome this recurring barrier and enable the development and application of biocatalysts for organic synthesis, both as single enzymes and in multi-enzyme cascade settings.

An alternative to single-substrate screening is to obtain information on catalyst promiscuity by screening with multiple substrates, either iteratively or in competition. Previously, these approaches have gone by various names including fingerprinting, multi-substrate, or multiplexed assays.¹⁹⁻²² To avoid confusion as to whether substrates were screened in separate parallel reactions or in competition, we use the term substrate multiplexed screening (SUMS) to specifically refer to screens where substrates are in direct competition (Fig 1a). A classic application of SUMS has been for characterization of native enzyme specificity, the extent to which an enzyme distinguishes between substrates.^{21,23,24} In pioneering work, the Reymond group showed how careful assay design to maintain initial velocity conditions can enable the high-throughput characterization of lipase and esterase substrate specificities.¹⁹ SUMS is also the default modality for *in vivo* screens and selections, as the ability to discriminate between metabolites is often the key parameter.²⁵⁻²⁷ However, there are altogether different properties that make an enzyme attractive for organic synthesis. Ideally, one desires an enzyme that will react with a wide range of pure substrates and drive reactions to high yield. Recently, Knorrscheidt *et al.* demonstrated how a SUMS method using a cocktail of three substrates could successfully identify mutations that altered the activity, specificity, and regioselectivity of an unspecific peroxygenase.²⁸ Nevertheless, SUMS approaches for engineering biocatalysts are rare. The kinetic underpinnings of SUMS outside of initial velocity conditions are not well-described and the relationship between assay design and multiplexed reaction outcomes is uncertain. Hence, despite the rich history of

competition assays guiding enzymology, the broader applicability of multiplexed screening approaches for biocatalysis is still unknown.

We identified two enzymes for the systematic exploration of how SUMS could be used to monitor enzyme promiscuity directly during the engineering process. These enzymes, tryptophan synthase and tryptophan decarboxylase, can be combined in a two-enzyme cascade for the synthesis of substituted tryptamines from L-serine and indole as precursors (Fig 1b). The tryptamine products are desirable synthons of highly bioactive pharmacophores, and their formation is a committed step in biosynthesis of indole alkaloids.^{29,30} We chose the L-tryptophan (Trp) decarboxylase from *Ruminococcus gnavus* (*RgnTDC*), an enzyme that natively catalyzes the decarboxylation of Trp to form tryptamine.³¹ Previously, we showed that *RgnTDC* is an exceptional decarboxylase with many Trp analogs but struggles with the highly bioactive 4- and 5-substituted substrates.³⁰ Such poor activity with non-native substrates is a recurring limitation among biocatalysts - one that is especially limiting in a cascade setting. For the synthesis of these substituted Trp analogs, we selected a previously-engineered β -subunit of tryptophan synthase from the thermophilic archaeon *Pyrococcus furiosus* (2B9), which catalyzes the bimolecular condensation of L-serine (Ser) and indole.^{13,32} Although 2B9 has high activity

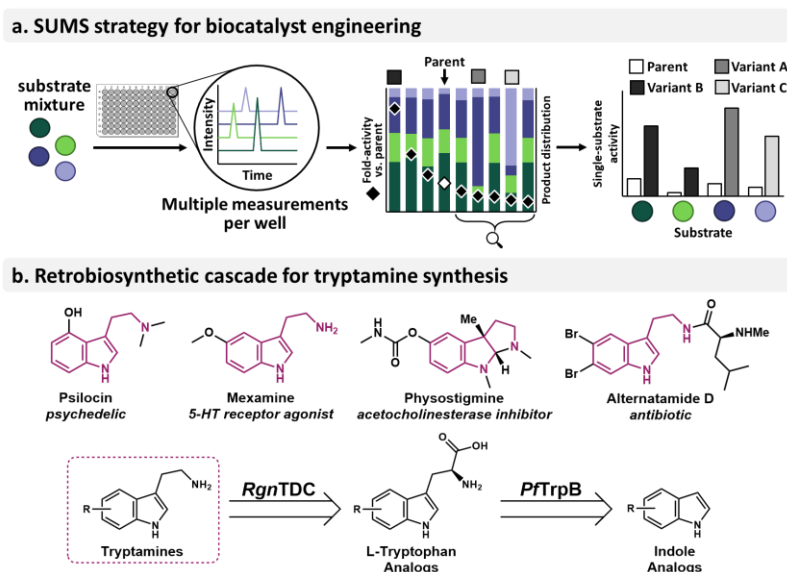


Figure 1. Substrate multiplexed screening (SUMS) and cascade synthesis to produce substituted tryptamines. **a.** SUMS measures the abundance of multiple products reacting in competition during protein engineering. **b. Top:** Tryptamine analogs with varied pharmacological properties. **Bottom:** Retrosynthesis of tryptamines from substituted indoles using the L-tryptophan decarboxylase from *Ruminococcus gnavus* (*RgnTDC*) and the *Pyrococcus furiosus* tryptophan synthase β -subunit (*PfTrpB*).

with a variety of substrate analogs at 75 °C, its activity decreases at lower temperatures, which cause a shift in the rate-limiting step (Fig S1).³³ Here, we develop SUMS approaches to rapidly assess the substrate scope of these distinct enzymes, form new hypotheses about their function, and efficiently construct bioactive molecules through a promiscuous, one-pot two-enzyme cascade.

Results

Analysis of underlying kinetics of substrate completion reactions

Before we began screening mutant libraries, we investigated many variables of SUMS, such as substrate choice, relative substrate concentrations, and assay duration, that can impact the observed product profile. To connect the SUMS output to the underlying kinetics, we used *RgnTDC* as a model system. In general, when multiple substrates are competing for an active site, *each substrate acts as a competitive inhibitor for all other substrates*.³⁴ For a unimolecular reaction under initial velocity conditions with equimolar substrates in competition with one another, the product abundances will be exactly proportional to the catalytic efficiencies (k_{cat}/K_M) of the individual reactions in isolation (Fig 2a).^{34,35} As has been described, this relationship holds true even when the individual substrate concentrations exceed their K_M 's.^{34,35} We measured traditional Michaelis-Menten parameters for *RgnTDC* with a variety of substituted Trp analogs (Table S1). Comparison of these data to results from multiplexed reactions showed that the ratio of the catalytic efficiencies is indeed deterministic of the product ratios (*see SI discussion*). As has long been appreciated in enzymology, such multiplexed activity measurements are a true measure of specificity and provide rich kinetic information about enzyme function.¹⁹ However, these relationships are restricted to initial velocity conditions and are an incomplete measure of synthetic utility.

To capture enzyme stability effects and achieve high conversions, effective screening conditions for biocatalysis applications often utilize longer reaction times beyond the initial velocity regime. When reactions are run to higher conversion, the product profile becomes uncoupled from the Michaelis-Menten kinetics and is, instead, a heuristic readout of reactivity that we show here can be tuned to match the goals of biocatalysis research (Fig. 2a, *see SI discussion*). We posited that by screening on a mixture with both highly reactive and inert substrates, we could identify catalysts that retain the ability to operate at high turnover numbers as well as identify desirable increases in activity with multiple sluggish substrates.

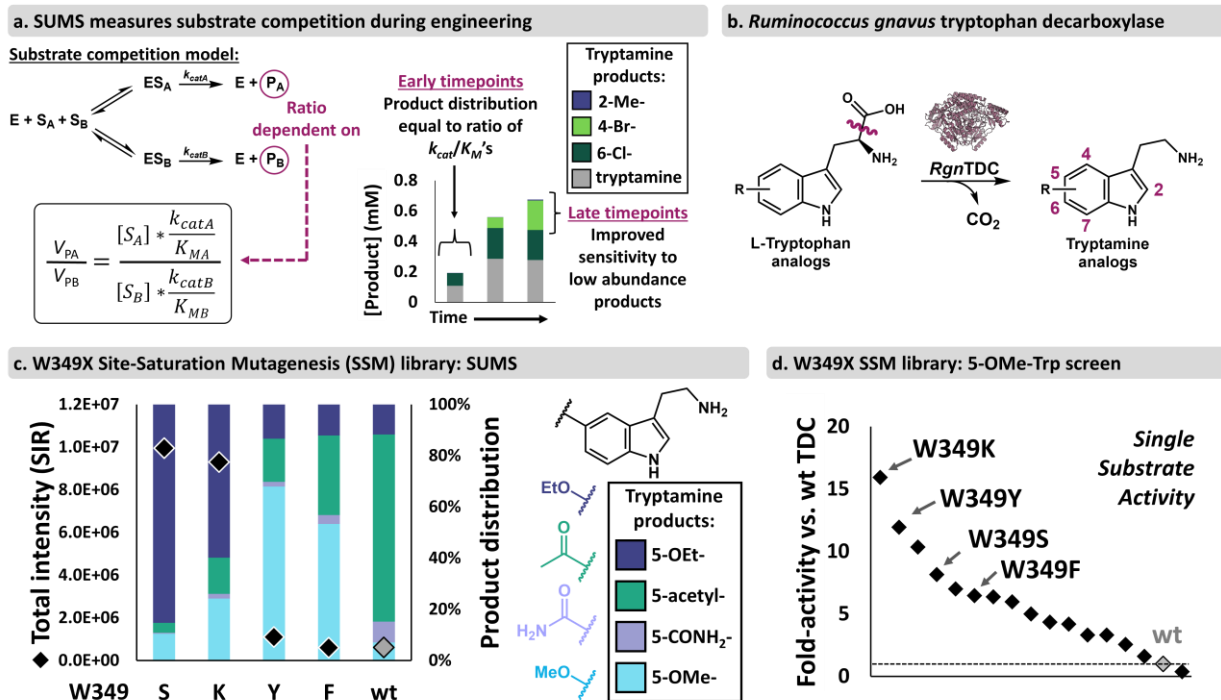


Figure 2. Substrate multiplexed screening (SUMS)-based engineering of *Ruminococcus gnavus* tryptophan decarboxylase (*RgnTDC*). **a. Left:** Substrate competition model with equation describing relative rates of product formation. **Right:** Timecourse of a substrate-multiplexed reaction of *RgnTDC* with 2-Me-, 4-Br-, 6-Cl-, and Trp. Full reaction conditions found in Fig S42. **b.** General reaction of *RgnTDC*, with the labile bond highlighted. **c.** SUMS results from a W349X library with 5-OEt-, 5-acetyl-, 5-CONH₂-, 5-OMe-, and 2-Me, 5-OMe-Trp. Colored bars indicate relative abundances of each product, and black diamonds indicate total intensity of single ion retention (SIR) for each product's unique m/z. No product was observed from 2-Me, 5-OMe-Trp. **d.** Fold-activity relative to wild-type from a single-substrate screen of the W349X library with 5-OMe-Trp corresponding to classical protein engineering techniques. Retention of function curves with full sequence analyses are shown in Fig S3-4.

SUMS assesses substrate promiscuity of enzyme variants

We began engineering for higher *RgnTDC* activity with 5-substituted Trp analogs, as structure-based modelling suggested the active site residue W349 forms preclusive steric interactions with these substrates (Fig S2). We screened a site-saturation mutagenesis (SSM) library, which exchanges the native residue for each other proteinogenic amino acid, at W349 with a mixture of five substrates. For most of the substrates, we found that many mutations increased activity, and that increases in activity varied among the different substrates (Fig S3). The structurally conservative mutations W349Y and W349F increased activity most with 5-OMe-Trp relative to other substrates, whereas the smaller W349S mutation had the highest activity increase with 5-OEt-Trp and produced the most total product. From this screen, W349K was identified as the most generally improved variant because it produced only

slightly less 5-OEt-tryptamine than W349S and formed the most product with all other substrates (Fig 2c).

To contrast the promiscuity information from SUMS with traditional approaches, we also performed a single-substrate screen with 5-OMe-Trp on the same W349 library (Fig 2d). As before, we found that almost any mutation increased activity with 5-OMe-Trp. However, there was a poor correlation between activity on 5-OMe-Trp and general activation on 5-substituted Trp analogs. Although W349K was the most activating mutation in both screens, mutations such as W349Y appeared to be highly reactive with 5-OMe-Trp but only poorly tolerated other Trp analogs. These results illustrate how SUMS can immediately identify shifts in both substrate promiscuity and activity with no greater screening effort than would be required for a more traditional, but less informative, approach.

While detailed structural analysis revealed W349 as a conspicuous site for improved activity on 5-substituted Trp analogs, such detailed hypotheses are not readily formed with all enzymes. We reasoned that SUMS could also be efficiently deployed in a setting where there is no specific hypothesis as to which residues govern activity with specific substrates. To simulate this common scenario, we screened a mixture of Trp analogs that were each substituted at a different position against a set of nine active site SSM libraries (Fig 3a, Fig S5). From these screens, we found that mutation at two positions, L126 and H120, had only modest impacts on activity and promiscuity. Mutation at L336 and T356 resulted in many catalytically feeble enzymes, and the variants that retained activity had promiscuity profiles that were similar to wild-type. For the other sites, mutation caused large changes to apparent promiscuity while retaining significant catalytic activity. For example, we observed > 50-fold activity increases with several *Rgn*TDC-Trp analog pairs, such as L355M with 4-Br-Trp and F98V with 2-Me-Trp (Fig 3b). Screening with this more diverse substrate mixture also revealed that W349K maintains high activity with non-5-substituted-Trp substrates like 6-Cl-Trp. Other mutations, such as V99A and L339V, were less strongly activating for 2-Me-Trp and 4-Br-Trp but retained broad activity for substituted Trp analogs.

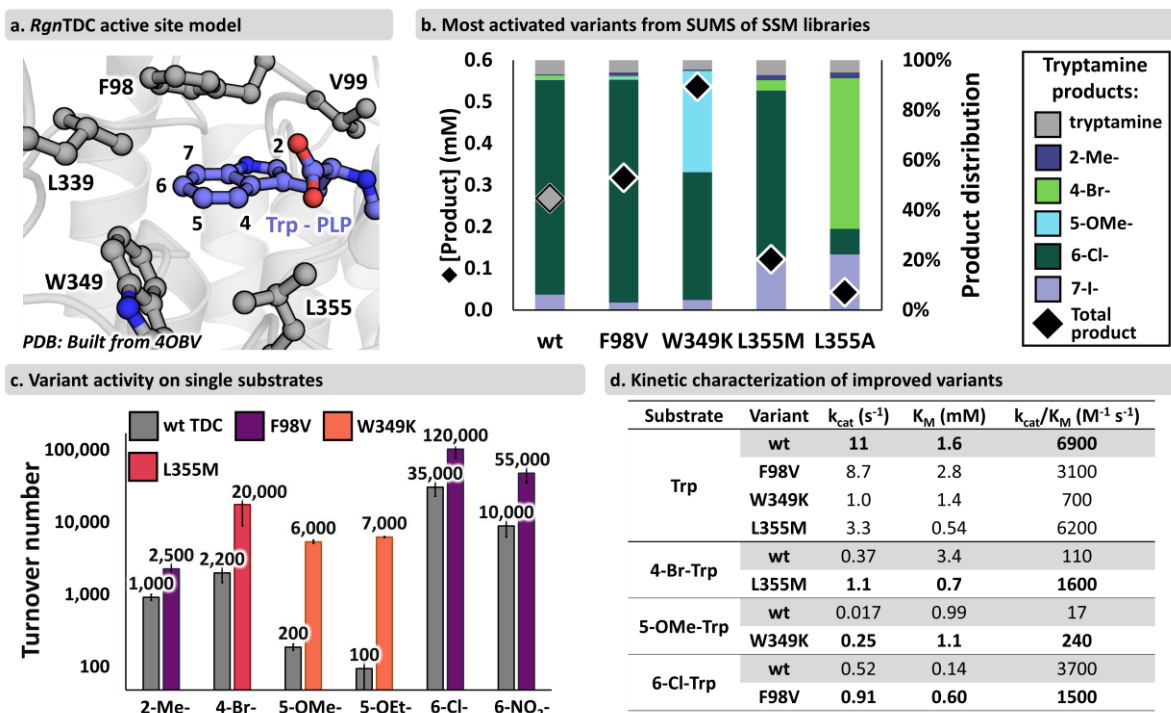


Figure 3. SUMS identifies *RgnTDC* active site mutations that improve activity for a range of substrates. **a.** Active site model of *RgnTDC* (built from PDB ID: 4OBV)³¹ with residues highlighted at which mutations were found that significantly altered promiscuity or improved activity. **b.** Select improved variants from active site libraries. Substrate screening conditions: 0.2 mM Trp and 7-I-Trp, and 2 mM 2-Me-Trp, 4-Br-Trp, 5-OMe-Trp, and 6-Cl-Trp, 4 h, 37 °C. Colored bars indicate relative abundances of each product, and black diamonds indicate the total product formed. Full screening results found in Fig S6-S14. **c.** Turnover numbers of wild-type *RgnTDC* and the top improved variant for each substrate. Different variants are depicted by different colored bars. Reactions were conducted in triplicate with standard deviation shown as a bar. **d.** Michaelis-Menten parameters for wild-type *RgnTDC* and activated variants for Trp and Trp analogs. Kinetic and turnover data were conducted in triplicate and complete data including error analysis are shown in Tables S1, S2, curves in Fig S17-S21.

Variants identified from SUMS have improved single-substrate activity

As with single-substrate library screening, validation of hits identified from SUMS is an essential step. While there are many confounding factors that make relative activity in competition distinct from activity on pure substrates, there is nevertheless no additional burden in the validation process, which we undertook with the *RgnTDC* variants. We were pleased to observe that turnover numbers from these pure substrate reactions trended well with multiplexed screening results, with the engineered variants showing large increases in single-substrate activity (Fig 3c, Table S2). Hence, in the same effort necessary to improve activity with one substrate, SUMS enabled the parallel engineering of *RgnTDC* variants for improved activity with multiple challenging substrates.

Characterization of *RgnTDC* variants identified from SUMS

To understand the kinetic determinants of substrate promiscuity shifts for *RgnTDC* variants, we measured Michaelis-Menten parameters (Fig 3d, Table S1). We found that k_{cat}/K_M values correlated well with observed activities in competition, even though we did not screen under initial velocity conditions. All activated variants showed higher k_{cat} values with their more reactive substrates when compared to wild-type. Notably, there was significant variation in changes to K_M values for activated *RgnTDC* variants, and such effects were difficult to rationalize for many mutations from structural analysis. The W349K mutation, for example, accelerates decarboxylation of 5-OMe-Trp exclusively by increasing k_{cat} , with minimal impact to K_M values (Fig 3d). The case of the L355M was notable. Molecular modeling indicates 4-substituted Trp analogs would form deleterious steric clashes with L355, and rational approaches to engineering would prescribe mutation to smaller sidechains. While the small L355A mutation improved activity on 4-Br-Trp, the conservative L355M mutation was even more activating and had a decreased K_M for both Trp and 4-Br-Trp compared to wild-type *RgnTDC* (Fig 3d). We highlight these unexpected findings as an advantage of interrogating active site libraries with SUMS, as such mutations could have been missed entirely by screening with the wrong pairings of substrate and mutational site.

SUMS can identify distally activating mutations for a bimolecular reaction

To further develop SUMS for biocatalysis, we next turned our attention to improving activity of the engineered tryptophan synthase β -subunit variant 2B9 on diverse indole analogs (Fig 4a). Whereas *RgnTDC* catalyzes a relatively simple unimolecular reaction, tryptophan synthase catalyzes a bimolecular reaction that is not well-described by simple kinetic models.³⁶ The ratio of the products from direct substrate competition can differ significantly from the ratio of the catalytic efficiencies measured in isolation.³⁴ Irrespective of the underlying kinetic phenomena, we reasoned stoichiometry can be leveraged to facilitate assay design of bimolecular reactions. By holding the invariant substrate as the limiting reagent (Ser) and providing an excess of the multiplexed reagents (indole analogs), information about specificity is maintained throughout the course of the reaction.

Because the parent enzyme, 2B9, already possesses modest activity on 4- and 5-substituted indole analogs, our engineering goal was to identify mutations that broadly increase activity at moderate temperatures with multiple indole substrates. *RgnTDC* engineering (above) utilized active site mutagenesis, where small structural perturbations are expected to have large effects on k_{cat} and K_M . However, residues that influence enzyme activity and substrate promiscuity can be distributed

throughout the enzyme scaffold,³⁷ and such distal mutations are known to modulate *PfTrpB* function.^{13,32,38} We therefore elected to screen a globally random mutagenesis library of 2B9 variants to determine whether a SUMS approach could lead to identification of residues beyond the active site that alter either activity or substrate promiscuity.

We screened a library of 2B9 variants against a panel of commercially available indole analogs bearing substituents with diverse steric and electronic properties at 25 °C (*see SI for in-depth discussion of assay optimization*). Decades of study have shown that most mutations to an enzyme have a neutral to deactivating impact on function.^{39,40} Correspondingly, we observed that nearly all variants displayed total activity that was either similar to or lower than 2B9 (Fig S23-25). A handful of variants appeared to increase overall product formation with little change in promiscuity (Fig 4b). We purified the most activated variant, I102T, which contains a single mutation outside the active site, and found it was as good or better than 2B9 with a variety of indole analogs under single-substrate conditions (Fig S26). SUMS can thus achieve a traditional goal of globally random mutagenesis – identifying distal, activating mutations – while simultaneously providing insights into the substrate scope of the improved enzyme.

A change in product distribution indicates a residue impacts the active site

A unique strength of SUMS is that the promiscuity of *all* variants, activated or deactivated, is assessed, providing an additional metric by which to evaluate variants. For example, we observed a variant with lower overall activity but with a significant shift in product distribution towards 2,3-dihydroiso-L-tryptophan (DIT), which is formed through C-N bond formation with indoline. Under multiplexed conditions this variant, H275R, reproducibly generated more DIT than 2B9 (Fig 4c). Curiously, under single-substrate conditions, H275R was *not* an activated variant and instead produced DIT more slowly than 2B9, leading us to investigate this apparent contradiction between SUMS results and activity on single substrates.

We turned to single-substrate kinetic analysis with indole, *PfTrpB*'s native substrate, and indoline to probe why, in some cases, improved activity in a multiplexed screen does not translate to improved catalysis with a pure substrate. 2B9 has a strong kinetic preference for indole compared to indoline, but the H275R mutation impacts the relative activity with each substrate asymmetrically. Relative to 2B9, H275R shows a dramatic >100-fold decrease in k_{cat}/K_M with indole but only a modest ~3-fold decrease in k_{cat}/K_M with indoline (Fig 4g). Consequently, when the H275R mutation is introduced, indole is a less effective competitive inhibitor and formation of DIT increases (Fig 4c, Fig S40, *see SI discussion*). This

analysis affirmed that the discrepancy between SUMS results and single-substrate activity can be resolved by considering the role of competitive inhibition in a multiplexed system. More importantly, the change in H275R promiscuity itself *immediately implies that the mutation impacts activity through cooperative interactions with the active site*, rather than a global enzymatic property like protein stability. We therefore hypothesized that a different mutation at H275 might increase activity, rather than reduce it. This hypothesis was further motivated by the location of H275, which is a ‘second sphere’ residue situated near the entrance to the enzyme’s active site (Fig 4d).

We screened a SSM library at H275 with the same substrate mixture as before and observed a range of enzyme activities and product distributions (Fig S27). Several variants possessed activity and promiscuity similar to 2B9. Other variants resembled H275R, exhibiting an overall decrease in product formation and a shift in distribution to favor DIT. We also observed mutations that resulted in a general increase in activity across all substrates screened, with H275E displaying the largest boost (Fig 4e). We subsequently validated that H275E has increased activity in single-substrate reactions, and these improvements extend to substrates that were not present in the original screen, such as the sterically bulky nucleophile 5-OEt-indole (Fig 4f). Notably, H275R was deactivated for all tested substrates, meaning no single-substrate screen could have identified the original H275R as a mutation of any interest. Critically, it was only by screening on a mixture of substrates and observing a shift in product distribution that the H275 site’s role in substrate discrimination was identified. Hence, information from SUMS enabled use of a low-activity variant, H275R, as an intermediate to access a broadly activated enzyme, H275E.

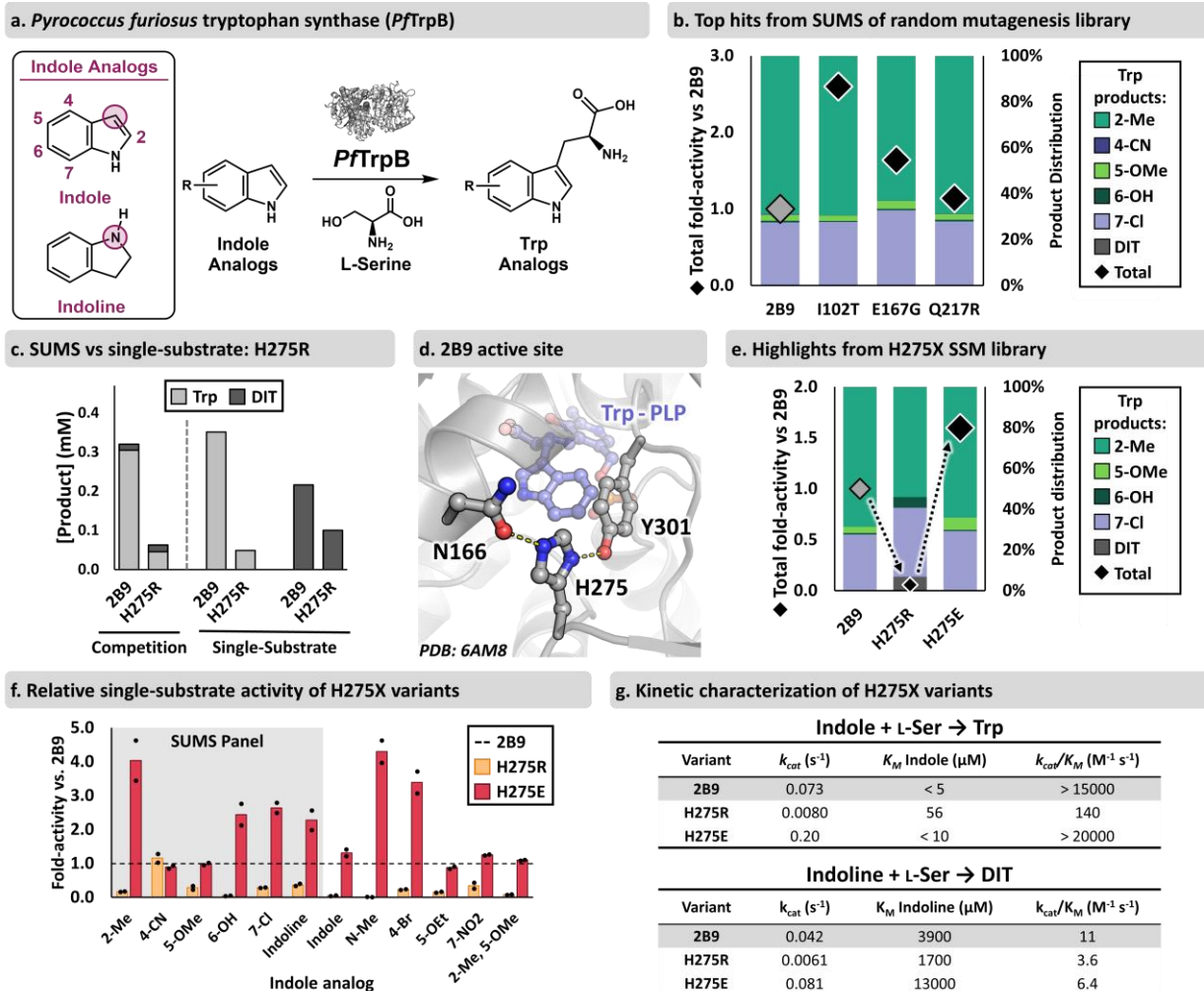


Figure 4. SUMS-based engineering of the β -subunit of *Pyrococcus furiosus* tryptophan synthase (*PfTrpB*) **a.** General reaction scheme for *PfTrpB*. Indole analogs were used to screen *PfTrpB* libraries. The nucleophilic atom is shown with a circle. **b.** SUMS results for generally activated variants detected during globally random mutagenesis library screening. Complete library results and experimental conditions are shown in Fig S23-25. **c.** Comparison of Trp and DIT production under competition and single-substrate reaction conditions, using purified 2B9 and H275R enzymes. Complete duplicate data are shown in Fig S28. **d.** H275 is a second-sphere residue that forms hydrogen bonds with neighboring residues, N166 and Y301 (PDB ID: 6AM8).³³ **e.** SUMS results for 2B9 and two variants from the H275X SSM library. 4-CN-indole was also included in reactions, but no product was observed. Complete library results are shown in Fig S27. **f.** Product formation in single-substrate reactions. Activity of H275R and H275E is shown relative to activity of 2B9 (black dashed line). **g.** Michaelis-Menten parameters for 2B9, H275R, and H275E with either indole or indoline as the nucleophilic substrate. Kinetic data were conducted in triplicate, and complete data including error analysis are shown in Fig S29.

SUMS leads to mechanistic insights

As we found with *RgnTDC*, SUMS-based engineering yielded thought-provoking results that raised new mechanistic questions about the causes of altered activity. Here, we were motivated to determine

whether the activation afforded by H275E mimicked the same effects as were previously found in the evolution of 2B9 for activity at 75 °C.^{13,33}

In the absence of substrates, the PLP cofactor of *Pf*TrpB is bound to K82 as an internal aldimine, E(Ain) (Fig S30). We solved the structure of H275E-E(Ain) at 2.1-Å resolution, which showed a significant conformational change of a subdomain of *Pf*TrpB, called the COMM domain, relative to the parent 2B9 (Fig 5a). Mutation at H275 disrupts a hydrogen bond network between two residues (Y181, Y301) that flank the active site and shifts the structure into the most extended-open conformational state of a TrpB observed to date. Catalysis is initiated by addition of Ser, which for the parent 2B9 results in accumulation of a mixture of the Ser external aldimine, E(Aex₁), and the electrophilic amino-acrylate intermediate, E(A-A) at 37 °C. The activating H275E mutation shifts the ratio of intermediates to favor E(A-A), which is poised to react with a nucleophilic substrate, such as indole (Fig 5b). Notably, the E(A-A) intermediate is also subject to a competing hydrolysis reaction. This shunt-reaction is 2.5-fold slower for H275E than 2B9 (Fig S31), indicating that the H275E mutation kinetically shields the reactive intermediate, affording more time for nucleophiles to react.

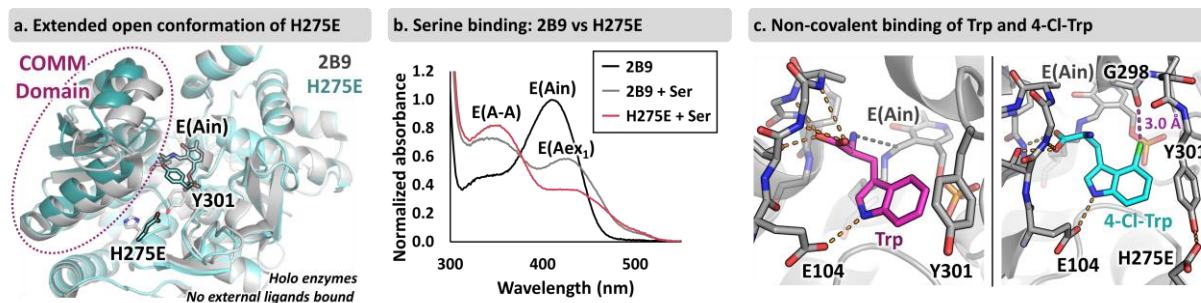


Figure 5. Crystallographic and spectroscopic characterization of H275E. **a.** Internal aldimine structure of H275E (light blue, PDB: 7RNQ) is superimposed with the corresponding structure of 2B9 (grey, PDB: 6AM7). **b.** Addition of 20 mM L-serine (Ser) to 2B9 (grey), results in two peaks corresponding to external aldimine E(Aex₁) and amino acrylate, E(A-A), intermediates. Addition of 20 mM Ser to H275E (pink) shows a dominant peak corresponding to E(A-A). A representative enzyme-only trace is shown in black. **c.** X-ray structures of H275E. *Left:* Trp binding is shown in magenta from PDB: 7ROF. *Right:* 4-Cl-Trp binding is shown in cyan from PDB: 7RNP. Hydrogen and halogen bonds are shown in orange and purple dashes, respectively.

Titration monitored by UV-vis spectroscopy shows H275E decreases the K_D for Trp while simultaneously promoting a non-covalent product binding mode (Fig S32). Previously, tighter binding has been associated with an increased population of covalently bound adducts.³³ To determine how products fit within the active site, we solved the structure of H275E with two ligands bound, Trp and 4-Cl-Trp, at 2.39 and 2.25-Å resolution, respectively (Fig 5c). The Trp-bound structure of H275E showed a new ligand binding pose, with the α -amine oriented for nucleophilic attack into the PLP. The structure of

H275E with 4-Cl-Trp bound showed no major conformational change is required to accommodate the 4-Cl group. Instead, there is a 3.0-Å halogen bond to G298. Together, these data show that the H275E mutation activates *Pf*TrpB through a distinct molecular mechanism.

Cascade catalysis is empowered by enzymes with complementary substrate scopes

Last, we sought to demonstrate the practical utility of the enzymes produced via SUMS. Many enzymes are more synthetically useful when employed in cascades, which can overcome thermodynamic limitations and obviate the need for purification of intermediates.⁴¹ While the use of multiple enzymes in concert can magnify the benefits afforded by biocatalysis, catalysts must have complementary substrate scopes to synthesize a diverse set of products.^{41–43} To this end, we demonstrate efficient cascade catalysis through the mmol syntheses of tryptamine analogs, including 5-OMe-tryptamine and 5-OEt-tryptamine, known serotonin receptor agonists,⁴⁴ and 2-Me-tryptamine and 4-Br-tryptamine, which were particularly challenging products for cascade reactions using the parent enzymes.³⁰ Each product was made in a telescoped biocatalytic cascade with H275E and an engineered *Rgn*TDC variant and isolated with improved yields compared to reactions with the parent enzymes (Fig 6). Although no *Rgn*TDC variant was identified with improved activity for *all* Trp analogs, the direct

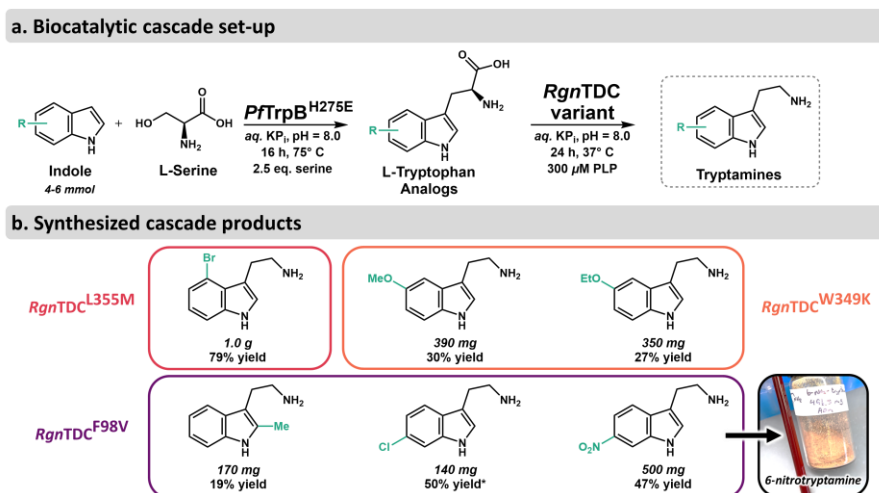


Figure 6. Engineered biocatalytic cascade for synthesis of tryptamine analogs. **a.** Utilized biocatalytic cascade for the telescoped biosynthesis of tryptamine analogs. **b.** Synthesized tryptamines, with the *Rgn*TDC variants used for different syntheses highlighted. *1.4 mmol substrate used for 6-chlorotryptamine synthesis.

assessment of substrate scope provided by SUMS allowed us to select an optimal catalyst for each tryptamine product. These reactions proceeded smoothly and afforded access to a variety of desirable tryptamines on preparative scale.

Discussion

A central limitation to the synthetic application of many enzymes is their hard-to-predict and too-often poor substrate scope when compared to organic methodology.⁴⁵ Traditional protein engineering excels at increasing activity on a single substrate but provides no selective pressure to improve activity across a broad substrate scope. By letting multiple substrates compete for the enzyme active site, we have shown how SUMS can facilitate identification of enzymes with increases in activity on multiple different substrates in parallel. We detail how factors such as relative substrate concentration, stoichiometry, and the time course of a reaction all influence the resultant product profile. By carefully constructing screening conditions, we showed that SUMS provides exceptional advantages when screening for increases in activity on multiple substrates and facilitates discovery of desirable biocatalysts.

Because activity in competition is not identical to activity on isolated substrates, the information gleaned from SUMS is more than the sum of its parts. We emphasize that the application of SUMS here does not rely on accurate measurement of true substrate specificity, as screening does not take place under initial velocity conditions. Instead, we use SUMS to identify changes in the ratios of the products for all variants screened, including those that appear neutral or even deactivating with respect to one or more substrates in the reaction. We identified a phenomenon through which mutations increase formation of a product in competition, even as the single-substrate reactions are universally slowed. The potential for this result arises when substrates with vastly different reactivities are included in competition. Because highly reactive substrates are also good competitive inhibitors, mutations that decrease activity with a preferred substrate can increase the amount of product formed from a poor substrate in competition. Hence, activity in competition can be increased through *repression of inhibition*. We note that these results are reproducible and understandable, and not a ‘false positive’. Indeed, because such mutations are altering the ratio of the products, we assert that they necessarily are influencing the active site and not some global feature like enzyme stability. We used this shift in promiscuity to identify H275 as a key site for altering activity of 2B9. Screening a site-saturation mutagenesis library at this position led to a generally activating mutation. This example highlights the engineering advantage of screening for both activity and promiscuity. While the repression of inhibition phenomenon is a consequence of universally applicable relative rate effects, the extent to which mutations cause this effect and how readily they may be leveraged to identify genuinely beneficial mutations is unknown and warrants future study.

Because SUMS is not limited to working with a particular enzyme or substrate class, many catalytic challenges can be tackled. Hence, we anticipate that SUMS will be adopted as a widely applicable method for protein engineering where catalysts with broad or well-defined substrate scopes are desired. The implementation of SUMS described here used LC-MS for detection of products, and in theory any chromatography or mass spectrometry that enables parallel resolution of products is compatible with SUMS. Indeed, such methods are already commonly used for screening libraries.^{2,4,9,15,28,46} Notably, because variants are identified based on *relative* changes in product distribution, SUMS data can be analyzed without quantitation of absolute product concentration. Last, the substrate mixtures used in this work featured compounds with which the parent enzyme already had detectable baseline activity. In principle, SUMS could also be applied with substrates that do not react with the parent enzyme to rapidly screen for gain of function mutations, facilitating discovery of altogether new reactions.

Conclusion

We show here the successful application of SUMS to engineer enzymes with improved activity on multiple compounds simultaneously. By directly assessing enzyme activity on substrates in competition, SUMS provides uniquely rich promiscuity information that has hitherto been underutilized during engineering campaigns. Importantly, just as knowledge of enzyme mechanism is not a pre-requisite for the effective application of directed evolution, *a priori* kinetic knowledge is not required for the design of multiplexed screens. While the resulting product profiles can be analyzed in terms of the underlying kinetics, such detailed follow-up studies are not necessary if the final goal is simply to develop a catalyst that is highly active on desirable substrates. These results and the generality of the assay design principles underscore the potential for SUMS to be applied to virtually any class of enzyme. Indeed, the ease of implementing SUMS in the context of existing protein engineering methods should minimize barriers to adoption of this budding approach.

Methods

Screening of *RgnTDC* site-saturation libraries

Cell pellets were thawed and then resuspended in lysis buffer: 50 mM potassium phosphate buffer (pH = 8.0), 1 mg/mL Hen Egg White Lysozyme (GoldBio), 0.2 mg/mL DNaseI (GoldBio), 1 mM MgCl₂, and 300 μM pyridoxal 5'-phosphate (PLP). A volume of 600 μL lysis buffer per well was used.

After 45 min of shaking at 37 °C, the resulting lysate was then spun down at 4000 *xg* to pellet cell debris. Then, 180 μ L of the resulting supernatant was added to 20 μ L of a substrate mixture in a separate reaction plate. Final substrate concentrations are as follows: **W349X 5-substituted-Trp screen**: 2 mM each of 5-methoxytryptophan, 5-ethoxytryptophan, 5-methoxy-2-methyltryptophan, 5-carboxamidotryptophan, and 5-acetyltryptophan; **W349X single-substrate screen**: 2 mM 5-methoxytryptophan; **active site site-saturation mutagenesis screens**: 2 mM each of 2-methyltryptophan, 4-bromotryptophan, 5-methoxytryptophan, and 6-chlorotryptophan; 0.2 mM 7-iodotryptophan and 0.2 mM tryptophan. Reactions were incubated at 37 °C for 4 h, quenched via addition of 150 μ L 1:1 acetonitrile:1 M HCl, and centrifuged at 4000 *xg* for 10 min. 200 μ L of the quenched reaction mixture supernatant was filtered into a 96-well plate for UPLC-MS analysis. Data were collected on an Acquity UHPLC with an Acquity QDA MS detector (Waters) using an Intrada Amino Acid column (Imtakt). Tryptamine product *m/z* ion counts were used to quantify product formation from the tryptophan reaction mixture from corresponding standard curves (Fig S15).

Screening of *PfTrpB* libraries

Lysis buffer was prepared as described above and used to resuspend cells expressing *PfTrpB* variants. Cells were lysed for 1 hour at 37 °C then heat treated at 75 °C for 15 min. After cooling on ice, lysate was spun down at 4000 *xg* at 4 °C for 20 min. A 96-well plate was loaded with 20 μ L substrate mixture (final concentration of 5 mM each 2-methylindole, 4-cyanoindole, 5-methoxyindole, 6-hydroxyindole, and indoline, plus 2.5 mM 7-chloroindole). All indole stocks were prepared in DMSO. For globally random mutagenesis library plate A, potassium phosphate buffer (50 mM, pH = 8.0, 160 μ L) containing L-serine (5 mM final concentration) was added, followed by heat-treated lysate (20 μ L). For subsequent plates, lysate volume was increased to 50 μ L and buffer volume reduced to 130 μ L. Reactions were set up such that the DMSO cosolvent comprised 10% of the final reaction volume (200 μ L). Reactions were run at room temperature (25 °C) for 2.5 h and were quenched with 200 μ L of acetonitrile containing 0.1 M HCl and 1 mM tryptamine (as internal standard). Plates were spun down at 4000 *xg* at 4 °C for 20 min. A 200 μ L aliquot of each quenched reaction was filtered into a 96-well plate for analysis by UPLC-MS. Product formation was quantified by integration of peaks on single ion retention (SIR) channels corresponding to each expected product, normalized against the tryptamine internal standard.

Cascade synthesis and isolation of tryptamines

4-6 mmol (1.4 mmol for 6-chloroindole) of the corresponding indole analog was added to a 1 L Erlenmeyer flask and dissolved in 20 mL MeOH. 12 mmol Ser was added, and the resulting solution was diluted up to just under 500 mL with 50 mM potassium phosphate buffer (pH = 8.0). PLP was added such that the final concentration was 300 μ M. Then, H275E was added at 0.05% mol catalyst relative to the indole analog. The solution was incubated at 75 °C for 16 h. (H275E was found to be activating at 75 °C, Fig S33). Following UPLC-MS analysis of conversion, the solution was cooled to 37 °C, upon which *RgnTDC* was added at 0.02 – 0.2% mol catalyst relative to the indole. The solutions were incubated at 37 °C for 24 h. Solutions were then evaporated down to 50 – 100 mL. To break emulsions, the solutions were acidified with 6 M HCl until pH < 1, 100 mL ethyl acetate (EtOAc) was added, and the resulting mixtures were centrifuged at 4000 *xg* for 10 min. These solutions were added to a separatory funnel,

the aqueous layer was drained, and the organic layer removed. This was repeated twice more, with 2 mL 6 M HCl added in between extractions. Then, the aqueous layer was alkalized with 6 M NaOH until pH > 12. Tryptamine products were then extracted 3x with 150 mL EtOAc, with 2 mL 6 M NaOH added in between extractions to the aqueous layer. Organic layers were pooled, dried with sodium sulfate, filtered, and evaporated down to 5-10 mL. Solutions were transferred to 20 mL scintillation vials, evaporated to near dryness (some tryptamines were observed as liquids at 50 °C), and dried under vacuum overnight. Dried samples were weighed and submitted for ¹H and ¹³C NMR analysis.

Acknowledgements

We thank Professor Samuel Gellman, Professor Patrick Willoughby, Lydia Perkins, Jon Ellis, and the Buller Buddies for the helpful discussion, insight, and editorial comments. We thank Dr. Craig Bingman for assistance with crystallographic data collection. This work was supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation and the NIH (DP2-GM137417) to A.R.B.; and an NSF Graduate Fellowship (DGE-1747503) to P.M.H. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The Bruker AVANCE III-500 NMR spectrometers were supported by the Bender Fund. Use of 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. In addition, we thank NIH Grant 1S10OD020022-1 for providing funding for the Q Extractive Plus Orbitrap used for high resolution mass spectrometry analysis of prepared compounds.

Author Contributions

A.D.M., P.M.H., and A.R.B. designed experiments, analyzed data, and prepared the manuscript; A.D.M. and P.M.H. conducted experiments.

References

1. Goodwin, N. C., Morrison, J. P., Fuerst, D. E. & Hadi, T. Biocatalysis in Medicinal Chemistry: Challenges to Access and Drivers for Adoption. *ACS Med. Chem. Lett.* **10**, 1363–1366 (2019).
2. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*. **329**, 305–309 (2010).

3. Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **8**, 476–480 (2017).
4. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science*. **366**, 1255–1259 (2019).
5. Sandström, A. G., Wikmark, Y., Engström, K., Nyhlén, J. & Bäckvall, J. E. Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *PNAS* **109**, 78–83 (2012).
6. Fox, R. J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338–344 (2007).
7. Amin, N. *et al.* Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.* **17**, 787–793 (2004).
8. 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).
9. Diefenbach, X. W. *et al.* Enabling Biocatalysis by High-Throughput Protein Engineering Using Droplet Microfluidics Coupled to Mass Spectrometry. *ACS Omega* **3**, 1498–1508 (2018).
10. McLaren, D. G. *et al.* High-Throughput Mass Spectrometry for Hit Identification: Current Landscape and Future Perspectives. *SLAS Discov.* **26**, 168–191 (2021).
11. Ye, L., Yang, C. & Yu, H. From molecular engineering to process engineering: development of high-throughput screening methods in enzyme directed evolution. *Appl. Microbiol. Biotechnol.* **2017 1022** **102**, 559–567 (2017).
12. Obexer, R. *et al.* Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
13. 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).
14. Tang, Q. *et al.* Directed Evolution of a Halide Methyltransferase Enables Biocatalytic Synthesis of Diverse SAM Analogs. *Angew. Chemie - Int. Ed.* **60**, 1524–1527 (2021).

15. Andorfer, M. C., Park, H. J., Vergara-Coll, J. & Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C-H bonds. *Chem. Sci.* **7**, 3720–3729 (2016).
16. 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).
17. 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).
18. Andorfer, M. C. *et al.* Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal.* **7**, 1897–1904 (2017).
19. Goddard, J., Reymond, J. & Uni, V. Enzyme Activity Fingerprinting with Substrate Cocktails. *JACS Commun.* 11116–11117 (2004).
20. Kim, H. *et al.* A multi-substrate screening approach for the identification of a broadly applicable Diels – Alder catalyst. *Nat. Commun.* 1–6 (2019). doi:10.1038/s41467-019-08374-z
21. Stanistic, A., Husken, A. & Kries, H. HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes. *Chem. Sci.* **10**, 10395–10399 (2019).
22. Zhang, K., El Damaty, S. & Fasan, R. P450 fingerprinting method for rapid discovery of terpene hydroxylating P450 catalysts with diversified regioselectivity. *J. Am. Chem. Soc.* **133**, 3242–3245 (2011).
23. Kuo, Y. M., Henry, R. A. & Andrews, A. J. Measuring specificity in multi-substrate/product systems as a tool to investigate selectivity in vivo. *Biochim. Biophys. Acta - Proteins Proteomics* **1864**, 70–76 (2016).
24. Joiner, C. M., Levine, Z. G., Aonbangkhen, C., Woo, C. M. & Walker, S. Aspartate Residues Far from the Active Site Drive O-GlcNAc Transferase Substrate Selection. *J. Am. Chem. Soc.* **141**, 12974–12978 (2019).
25. Desai, B. J. & Gonzalez, R. L. Multiplexed genomic encoding of non-canonical amino acids for labeling large complexes. *Nat. Chem. Biol.* **16**, 1129–1135 (2020).
26. Weeks, A. M. & Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol.* **14**, 50–57 (2018).

27. Wang, L., Xie, J. & Schultz, P. G. EXPANDING THE GENETIC CODE. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
28. Knorrscheidt, A. *et al.* Simultaneous screening of multiple substrates with an unspecific peroxygenase enabled modified alkane and alkene oxyfunctionalisations. *Catal. Sci. Technol.* 6058–6064 (2021). doi:10.1039/d0cy02457k
29. Fricke, J., Blei, F. & Hoffmeister, D. Enzymatic Synthesis of Psilocybin. *Angew. Chemie - Int. Ed.* **56**, 12352–12355 (2017).
30. McDonald, A. D., Perkins, L. J. & Buller, A. R. Facile in Vitro Biocatalytic Production of Diverse Tryptamines. *ChemBioChem* **20**, 1939–1944 (2019).
31. Williams, B. B. *et al.* Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. *Cell Host Microbe* **16**, 495–503 (2014).
32. Buller, A. R. *et al.* Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).
33. Buller, A. R. *et al.* Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. *J. Am. Chem. Soc.* **140**, 7256–7266 (2018).
34. Cornish-Bowden, A. Enzyme specificity: Its meaning in the general case. *J. Theor. Biol.* **108**, 451–457 (1984).
35. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* (W. H. Freeman and Company, 1999).
36. Heilmann, H. D. On the mechanism of action of Escherichia coli tryptophan synthase Steady-state investigations. *BBA - Enzymol.* **522**, 614–624 (1978).
37. Wrenbeck, E. E., Azouz, L. R. & Whitehead, T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8**, 1–10 (2017).
38. Watkins-Dulaney, E., Straathof, S. & Arnold, F. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBioChem* **22**, 5–16 (2021).
39. Matthews, B. W. Structural and genetic analysis of protein stability. *Annual Review of*

- Biochemistry* **62**, 139–160 (1993).
40. 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).
 41. Schrittwieser, J. H., Velikogne, S., Hall, M. & Kroutil, W. Artificial Biocatalytic Linear Cascades for Preparation of Organic Molecules. *Chemical Reviews* **118**, 270–348 (2018).
 42. Sattler, J. H. *et al.* Redox Self-Sufficient Biocatalyst Network for the Amination of Primary Alcohols. *Angew. Chemie* **51**, 9156–9159 (2012).
 43. Staudt, S. *et al.* Direct Oxidation of Cycloalkanes to Cycloalkanones with Oxygen in Water. *Angew. Chemie* **52**, 2359–2363 (2013).
 44. Glennon, R. A. Higher-end serotonin receptors: 5-HT₅, 5-HT₆, and 5-HT₇. *Journal of Medicinal Chemistry* **46**, 2795–2812 (2003).
 45. Reetz, M. T. What are the Limitations of Enzymes in Synthetic Organic Chemistry? *Chem. Rec.* **16**, 2449–2459 (2016).
 46. McIntosh, J. A. *et al.* Engineered Ribosyl-1-Kinase Enables Concise Synthesis of Molnupiravir, an Antiviral for COVID-19. *ACS Cent. Sci.* acscentsci.1c00608 (2021). doi:10.1021/acscentsci.1c00608