Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues

David K. Romney, [ORCID] Javier Murciano-Calles, Jöri E. Wehrmüller, and Frances H. Arnold*[ORCID]

Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, United States

S Supporting Information

ABSTRACT: Derivatives of the amino acid tryptophan (Trp) serve as precursors for the chemical and biological synthesis of complex molecules with a wide range of biological properties. Trp analogues are also valuable as building blocks for medicinal chemistry and as tools for chemical biology. While the enantioselective synthesis of Trp analogues is often lengthy and requires the use of protecting groups, enzymes have the potential to synthesize such products in fewer steps and with the pristine chemo- and stereoselectivity that is a hallmark of biocatalysis. The enzyme TrpB is especially attractive because it can form Trp analogues directly from serine (Ser) and the corresponding indole analogue. However, many potentially useful substrates, including bulky or electron-deficient indoles, are poorly accepted.

We have applied directed evolution to TrpB from *Pyrococcus furiosus* and *Thermotoga maritima* to generate a suite of catalysts for the synthesis of previously intractable Trp analogues. For the most challenging substrates, such as nitroindoles, the key to improving activity lay in the mutation of a universally conserved and mechanistically important residue, E104. The new catalysts express at high levels (>200 mg/L of *Escherichia coli* culture) and can be purified by heat treatment; they can operate up to 75 °C (where solubility is enhanced) and can synthesize enantiopure Trp analogues substituted at the 4-, 5-, 6-, and 7-positions, using Ser and readily available indole analogues as starting materials. Spectroscopic analysis shows that many of the activating mutations suppress the decomposition of the active electrophilic intermediate, an amino-acrylate, which aids in unlocking the synthetic potential of TrpB.
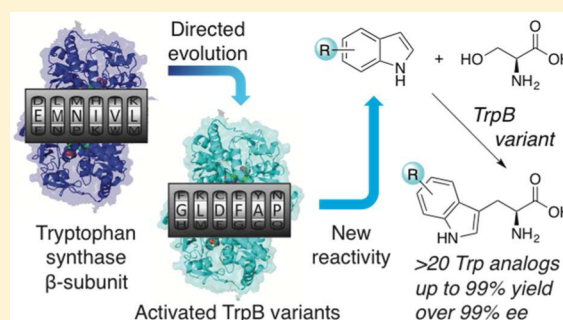
INTRODUCTION

To develop functional complex molecules, synthetic chemists must have access to a broad range of chemical building blocks. These include noncanonical amino acids, which are not found in proteins, but are often encountered as intermediates in biosynthesis. Derivatives of tryptophan (Trp) occupy a distinguished position in the echelons of noncanonical amino acids, because they are biosynthetic precursors to compounds that possess diverse activities,[1,2] including anticancer, antibiotic, immunosuppressant, antifungal, and phytotoxic properties. Tryptophan analogues have also been used as starting materials for chemical syntheses[3−9] as well as probes for chemical biology.[10−13] Therefore, efficient methods to synthesize Trp analogues are needed.

Enzymes have emerged as powerful new tools for chemical synthesis, catalyzing reactions with unrivaled rate acceleration and exceptional selectivity.[14−16] While nature has created a vast repertoire of enzymes that modify Trp through a plethora of transformations, including nitration, halogenation, and alkylation (Figure 1a),[2] only a few of these enzymes have been investigated for use in synthetic chemistry.[17−22] Furthermore, this approach is limited because the synthesis of each Trp analogue requires a different enzyme. We wished to develop an alternate biocatalytic approach, in which a single catalyst

platform would provide direct access to any Trp analogue from readily available starting materials.

Enzymes such as acylases[9,23] and transaminases[24] (Figure 1b) have been applied to synthesis of Trp analogues, but in these approaches, the majority of the product must be assembled in advance, with the enzymes mostly serving to set the stereochemistry at the end. Many methods, such as those that use esterases, rely on kinetic resolution, which limits the maximum theoretical yield of product to 50%. A notable exception is the use of tryptophan synthase (TrpS), which can assemble Trp analogues from L-serine (Ser) and the corresponding indole analogue with retention of enantiopurity (Figure 1c).[25−30]

While the chemical synthesis of Trp analogues is challenging, methods to prepare indole analogues abound.[31,32] Thus, the TrpS methodology has the potential to provide direct access to a wide array of products. Unfortunately, the activity of wild-type TrpS is highly sensitive to the electronic and steric properties of the substrates. Protein-engineering techniques such as directed evolution are an effective means to unlock new function, such as high activity with non-native substrates. However, directed
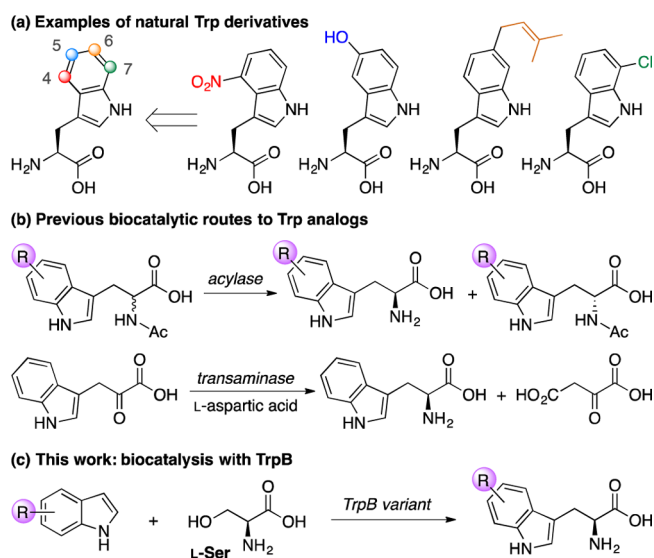
Figure 1. Synthesis of Trp analogues. (a) Examples of biosynthetic intermediates derived from Trp. (b) Previous synthetic methods using enzymes.[9,23,24] (c) Alternative biocatalytic route based on TrpB. Ac = acetyl.

evolution has scarcely been applied to TrpS, partly because this protein is a complex of two distinct proteins: an α-subunit (TrpA) and a β-subunit (TrpB). Although only TrpB is involved in the proposed synthesis of Trp analogues, TrpA is an allosteric actuator, whose absence severely attenuates TrpB activity. Recently, we developed variants of TrpB from *Pyrococcus furiosus*[33,34] and other organisms,[35] which do not require their TrpA partner proteins to achieve full activity. We therefore wished to expand this initial set of stand-alone variants into a general platform for the synthesis of Trp analogues, especially those on which TrpS is poorly active.

In an initial survey of substrates, our attention was immediately drawn to 4-substituted indoles, since this substitution pattern is present in natural products, but was poorly reactive in previous studies with TrpB. In particular, we focused on 4-nitroindole (Scheme 1) because the nitro

Scheme 1. Test Reaction for Optimization of TrpB
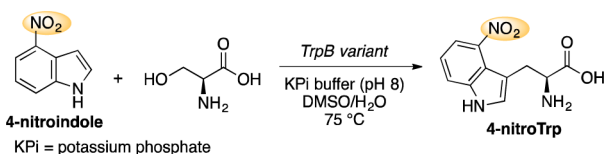
substituent not only creates a steric impediment to substrate binding, but also withdraws electron density from the indole moiety. Since the indole is a nucleophile in this reaction manifold, electron-withdrawing substituents are intrinsically deactivating. The compound 4-nitrotryptophan (4-nitroTrp) is a biosynthetic and chemical precursor to thaxtomin A, a potentially useful agrochemical (Chart 1).[4,36−38] Additionally, 4-nitroTrp is a chemical precursor to the tumor-promoter indolactam V.[5] However, the chemical synthesis of enantiopure 4-nitroTrp is complex,[4] and the natural enzymatic route is not amenable to application on a large scale.[39] Furthermore, we hypothesized that the evolutionary journey toward an optimized TrpB catalyst for 4-nitroTrp would yield catalysts with improvements for other challenging substrates.

Chart 1. Natural Products Synthesized from 4-NitroTrp
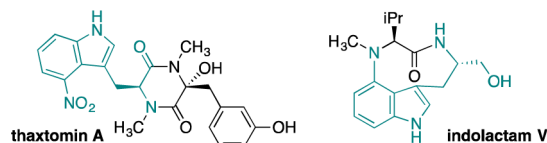
■ RESULTS

Initial Activity of Stand-Alone Variants toward 4-Nitroindole. TrpB enzymes from *P. furiosus* and *Thermotoga maritima* are optimal parents for directed evolution, due to their high thermostability.[40] We therefore tested the wild-type proteins (*Pf*TrpB and *Tm*TrpB), as well as the already-generated stand-alone variants for the production of 4-nitroTrp. As expected, the wild-type enzymes exhibited only trace activity (Figure S1). In addition, many of the variants formed a significant amount of a side product, which is putatively isotryptophan **1** (Figure 2). Fortunately, one variant, *Pf*2B9,
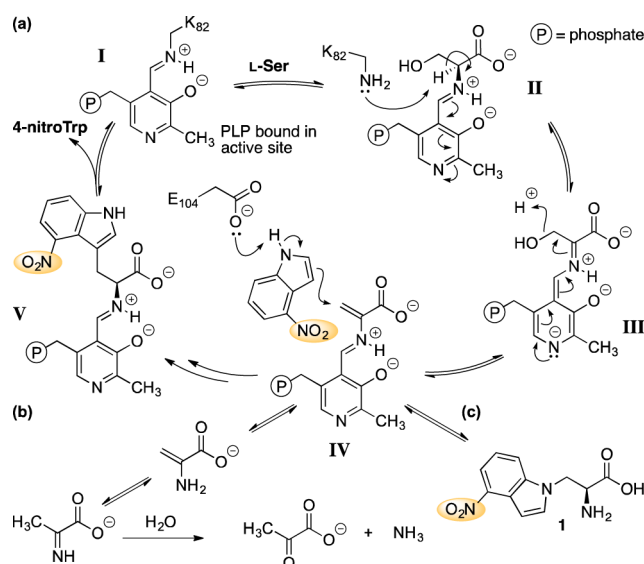

Figure 2. Putative pathways for reaction with 4-nitroindole. (a) Catalytic cycle for formation of 4-nitroTrp. (b) Enzymatic decomposition of Ser. (c) Competitive formation of isotryptophan **1**.

provided 18% conversion of 4-nitroindole to 4-nitroTrp. Notably, this variant, which has eight mutations from wild-type *Pf*TrpB, was initially evolved for activity with indole and threonine (Thr).[34] Thus, the fortuitous improvement for 4-nitroindole and Ser lent support to our hypothesis that optimizing a catalyst for production of 4-nitroTrp would provide simultaneous gains for other substrates.

We looked to the TrpB catalytic mechanism in order to identify what might be limiting conversion of 4-nitroindole. TrpB uses the cofactor pyridoxal phosphate (PLP), which is covalently bound to a lysine residue in the active site (Figure 2a, intermediate **I**). The lysine is displaced by Ser (intermediate **II**), which then undergoes α-deprotonation (intermediate **III**) and β-elimination to generate the active electrophile, amino-acrylate **IV**. Ideally, this would be attacked by the nucleophilic substrate, such as 4-nitroindole, to form the Trp product.

We observed that increasing catalyst loading had a negligible effect on production of 4-nitroTrp (Figure S2a), but that the Ser was almost completely consumed at the end of the reaction

period (Figure S3). By contrast, production of 4-nitroTrp was improved by addition of excess Ser (Figure S2b). These observations are consistent with the well-known side reaction in which the amino-acrylate is ejected from the PLP cofactor and undergoes hydrolytic decomposition to pyruvate (Figure 2b).[41] In addition, we sometimes observed formation of what we believe is isotryptophan 1 (Figure 2c and Figure S4a), in which 4-nitroindole adds to the amino-acrylate through the endocyclic nitrogen atom (N1) rather than the desired carbon atom (C3). While this reaction appears to be reversible, it undoubtedly slows the desired reaction. Our goal was to engineer a TrpB that would rapidly and quantitatively convert equimolar amounts of 4-nitroindole and Ser into 4-nitroTrp with perfect regio- and enantioselectivity.

Effects of Active-Site Mutagenesis. Lacking a high-throughput assay for production of 4-nitroTrp, we initially focused on small site-saturation mutagenesis libraries that could be screened by high-performance liquid chromatography (HPLC). Previously, an X-ray crystal structure was obtained of TrpB from *Salmonella enterica* (*Se*TrpB) with both the amino-acrylate and benzimidazole, an indole mimic, bound in the active site (PDB ID: 4HPX).[42] Although *Pf*TrpB and *Se*TrpB have modest sequence identity (59%), we previously observed a high degree of structural similarity between structures of *Pf*TrpB and TrpB from *Salmonella typhimurium* (*St*TrpB), which also have only 59% sequence identity.[33] We therefore constructed a homology model with the sequence of *Pf*TrpB and placed 4-nitroindole into the binding pose necessary to achieve C−C bond formation, in the hope of identifying steric clashes that could be alleviated by mutations (Figure 3). Unfortunately, the model suggested that the nitro
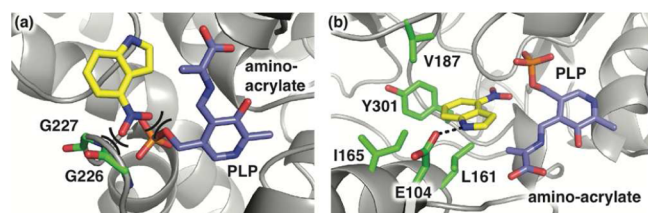

Figure 3. Model of 4-nitroindole (yellow) and the amino-acrylate in the active site of *Pf*TrpB (see Section 8.13 of Supporting Information). (a) Nitro group clashes with the protein backbone (green) and the PLP cofactor (purple). (b) Alternative view showing side-chains extending in to the active site and hydrogen bond with E104.

group was clashing with the protein main chain as well as with the PLP cofactor (Figure 3a). Nonetheless, we identified four residues whose side chains extended into the indole-binding pocket: L161, I165, V187, and Y301 (Figure 3b). The side chain of E104 also occupies the active site, but this residue is thought to bind indole through the NH moiety, thereby promoting attack from C3. Since this residue is universally conserved in TrpB homologues, and the enzymes already suffered from poor regioselectivity with 4-nitroindole, we elected to avoid mutagenesis at this position. We hypothesized that mutation at the other four positions could create space for 4-nitroindole to bind in an alternative pose that relieved the steric clashes but still allowed for attack of the amino-acrylate.

Mutations of the targeted residues were almost uniformly deleterious, with the exception of L161V, which boosted the yield of 4-nitroTrp to 25%. While this improvement is modest,

the mutation also suppressed formation of the side-product (Figure S4b). It is worth noting that although the side-chain of valine is indeed smaller than that of leucine, mutation of this position to alanine is deleterious. Thus, the effects of active-site mutagenesis cannot be rationalized by simple models based on sterics.

Screening of random-mutagenesis libraries. Having failed to achieve significant gains in activity from active-site mutagenesis, we hoped that random mutagenesis would provide beneficial mutations elsewhere in the protein scaffold. However, such libraries often have low frequencies of beneficial mutations, the result of which is that large numbers (hundreds to thousands) of variants must be screened to find improvements. In our initial generation of the stand-alone TrpB catalysts, we had exploited the spectroscopic shift between indole and Trp to screen initial rates in a high-throughput manner.[33] However, 4-nitroindole proved to be substantially more difficult. Unlike indole, 4-nitroindole has minimal solubility in water, even with 5% dimethyl sulfoxide (DMSO) as a co-solvent. Furthermore, the initial rate of even the most evolved catalyst was too low to screen reliably, a problem compounded by the fact that the spectral difference between 4-nitroindole and 4-nitroTrp is much smaller than that of indole and Trp. Instead, we chose to exploit the solubility differences between 4-nitroindole and 4-nitroTrp. Specifically, the reactions were run in aqueous conditions with 5% DMSO as a co-solvent, then the residual starting material was extracted in an organic solvent (ethyl acetate). The amount of 4-nitroTrp was determined by measuring the absorption of the aqueous phase in the visible spectrum.

Using the newly developed assay, we tested a random-mutagenesis library, generated by error-prone polymerase chain reaction, and identified a variant, with mutations M139L and L212P, that almost doubled the HPLC yield of 4-nitroTrp to 49% (Figure 4). We next tested a library which randomly
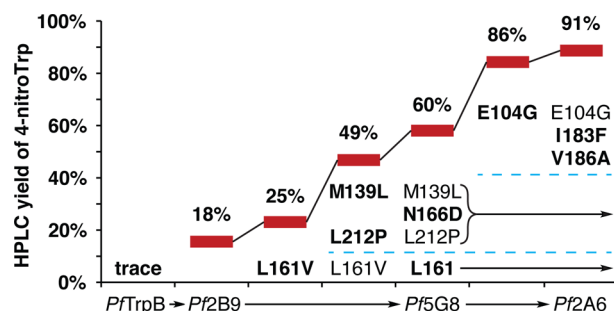

Figure 4. Evolutionary progression in production of 4-nitroTrp. Mutations in bold were added in the corresponding round of mutagenesis and screening. Dashed lines denote a new round of random mutagenesis. The horizontal axis indicates catalyst designations. See Section 8.7 of the Supporting Information for experimental details.

recombined those two mutations and N166D, a beneficial mutation identified in a separate study.[35] We also varied the active-site L161V mutation, since its effect had been comparatively minor. Indeed, in the best variant from this library (*Pf*SG8), the active-site mutation had reverted back to leucine and the other three mutations were retained. This variant formed 4-nitroTrp in 60% HPLC yield from equimolar amounts of 4-nitroindole and Ser.

Energized by that success, we generated another random-mutagenesis library with *Pf*5G8 as the parent protein. From this library, two mutations were found (I183F and V186A) that increased HPLC yield slightly, to ~70%. The most significant improvement, however, came from a variant bearing the mutation E104G, which increased HPLC yield to ~86% and eviscerated the earlier supposition that an H-bonding interaction with 4-nitroindole and the side-chain of E104 would promote the reaction. We then screened a recombination library and found that the E104G mutation recombined with I183F and V186A to produce 4-nitroTrp in 91% HPLC yield. Finally, we screened a site-saturation library at position 104 and found that glycine at this position was optimal, with Ala yielding similar, but slightly inferior results (Figure S5).

Optimization toward Isomeric Nitroindoles. With the new catalysts in hand, we wished to see if the evolution had created specificity for 4-nitroindole, or if it had improved activity for other substrates as well. We therefore tested catalysts *Pf*5G8 and *Pf*2A6 with 5-, 6-, and 7-nitroindole (Chart 2). Gratifyingly, *Pf*5G8 showed improvement for all substrates
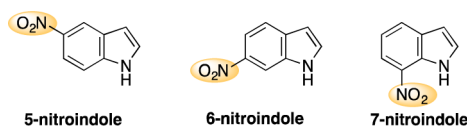
Chart 2. Nitro Substitution at Other Positions

5-nitroindole   6-nitroindole   7-nitroindole

Table 1. Optimizing Catalysts for Other Nitroindole Isomers

entry	catalyst[a]	HPLC yield of nitroTrp (%)[b]			
		4-nitro	5-nitro	6-nitro	7-nitro
1	*Pf*2B9	18	8	17	29
2	*Pf*5G8	60	64	64	64
3	*Pf*2A6	91	5	26	>99
4	*Tm*2F3		76		
5	*Tm*2F3 I184F		86		
6	*Pf*2B9 I165F Y301H			66	
7	*Pf*0A9			86	
8	*Pf*0A9 E104G			91	

[a]See Table 2 for catalyst designations. [b]Reactions used equimolar amounts of nitroindole and Ser. See Section 8.7 of the Supporting Information for experimental details.

compared to *Pf*2B9, forming all isomeric nitrotryptophans in about 60% HPLC yield (Table 1, entries 1 and 2). Enzyme *Pf*2A6, on the other hand, showed almost quantitative conversion of 7-nitroindole to the corresponding nitroTrp, but low activity with 5- and 6-nitroindole (Table 1, entry 3). Because we had previously identified catalysts that exhibited moderate activity with these substrates, we wondered if a subset of the mutations from *Pf*2A6 would further activate them for 5- and 6-nitroindole. The mutations associated with each catalyst designation are summarized in Table 2.

Previously, engineered variants of *Tm*TrpB had shown higher activity with 5-substituted indoles than their *Pf*TrpB homologues.[35] Additionally, it was observed that if certain beneficial mutations in *Pf*TrpB variants were transferred to the corresponding positions in *Tm*TrpB, then the activating effects were also transferred.[35] We therefore constructed a library in which the mutations of *Pf*2A6 were randomly recombined at

Table 2. Catalyst Designations and Associated Mutations

designation	species of origin	mutations
*Pf*2B9[a]	*P. furiosus*	I16V, E17G, I68V, F95L, T292S, F274S, T321A, V384A
*Pf*5G8	*P. furiosus*	*Pf*2B9 + M139L, N166D, L212P
*Pf*2A6	*P. furiosus*	*Pf*5G8 + E104G
*Pf*0A9	*P. furiosus*	*Pf*2B9 + M139L, I165F, N166D, Y301H
*Tm*2F3	*T. maritima*	P19G, I69V, K96L, P140L, N167D, L213P, T292S

[a]Variant from ref 34.

the corresponding positions in *Tm*TrpB. Screening for activity with 5-nitroindole revealed two variants that outperformed all previous catalysts in the production of 5-nitrotryptophan (5-nitroTrp). The first, *Tm*2F3, contained five mutations that were originally found in *Pf*2B9 plus all three mutations from *Pf*5G8; this variant formed 5-nitroTrp in 76% HPLC yield (Table 1, entry 4). The second variant was identical, but also contained one of the mutations found in *Pf*2A6 (I184F, according to numbering in *T. maritima*). This mutation gave a further boost in yield to 86% (Table 1, entry 5).

To improve activity with 6-nitroindole, we returned to the parent enzyme *Pf*2B9 and the active-site mutagenesis libraries from the beginning of this study. Although these libraries had not yielded significant gains for 4-nitroindole, we hypothesized that 6-nitroindole might respond differently, due to its distinct shape. Indeed, while no mutation at L161 was activating, the single mutations I165F and Y301H were both beneficial (Figure S6). Ultimately, the best variant contained both of these mutations and formed 6-nitroTrp in 66% HPLC yield (Table 1, entry 6). A random recombination library of the mutations from *Pf*2A6 revealed a new variant, *Pf*0A9, bearing mutations M139L and N166D, which increased the yield to 86% (Table 1, entry 7). Surprisingly, the mutation E104G also enhanced activity, albeit modestly, to 91% yield (Table 1, entry 8).

Substrate Scope of Catalyst Panel. To evaluate their generality and synthetic utility, we tested the newly optimized catalysts against a range of indole substrates (Table 3). We were pleased to find that the catalysts accepted essentially every indole analogue that we tested, often forming the corresponding Trp product in excellent yield. For preparative reactions, however, we chose to focus on halogenated and electron-deficient indoles, since historically these have been the most challenging.

Despite the tremendous improvement in 4-nitroTrp production, the catalysts still struggled with substituents at the 4-position. With 4-nitroindole, *Pf*2A6 can achieve ~5000 turnovers, but the reaction seems to slow considerably toward the end (Figure S7). As a result, a higher catalyst loading is required to achieve the high conversion observed in the catalyst evolution (Table 3, entry 1). Fortunately, the catalysts are expressed at high levels (>200 mg/L of *E. coli* culture) and can be used as heat-treated lysate, without additional protein purification. As a result, we were easily able to synthesize over a gram of 4-nitroTrp (73% isolated yield) using the protein from a 1-L bacterial culture (see Section 9.4 of Supporting Information). The enzymes could also form 4-fluorotryptophan in excellent yield (Table 3, entry 2), as well as 4-bromo- and 4-cyanotryptophan, albeit with modest activity (Table 3, entries 3 and 4, respectively). With indole-4-boronic acid, however, no product was detected (Table 3, entry 5). Strangely, the optimal

Table 3. Tryptophan Analogues Produced by Catalyst Panel[a]

Entry	Substrate	R	Catalyst	Isolated yield (%)
1	4-substituted indole	NO$_2$	*Pf*2A6	95[b]
2		F	*Tm*2F3	97
3		Br	*Tm*2F3	72
4		CN	*Tm*2F3 I184F	41[b]
5		B(OH)$_2$		ND
6	5-substituted indole	NO$_2$	*Tm*2F3 I184F	88[b]
7		CN	*Tm*2F3	79
8		CONH$_2$	*Tm*2F3	77
9		B(OH)$_2$	*Pf*0A9	37
10		I	*Pf*0A9	74[b]
11		CF$_3$	*Pf*2A6	19[b,c]

Entry	Substrate	R	Catalyst	Isolated yield (%)
12	6-substituted indole	NO$_2$	*Pf*0A9 E104G	91
13		Cl	*Pf*0A9	98
14		Br	*Pf*0A9	97
15		CN	*Pf*0A9	99
16		B(OH)$_2$	*Pf*0A9	49
17	7-substituted indole	NO$_2$	*Pf*2A6	98
18		CN	*Pf*2A6	98
19		Cl	*Pf*0A9	99
20		I	*Pf*0A9	91
21		Br	*Pf*5G8	53
22		B(OH)$_2$		ND
23	5,6-disubstituted indole (R^5, R^6)	5,6-Cl$_2$	*Pf*5G8	87
24	5,7-disubstituted indole (R^5, R^7)	5-Br-7-F	*Tm*2F3 I184F	56
25		5-Cl-7-I	*Tm*2F3	10

[a]Reactions used 0.02 mol % catalyst loading (maximum 5000 turnovers) and 1.1 equiv Ser relative to indole substrate. [b]Catalyst loading was 0.1 mol % (maximum 1000 turnovers). [c]Reaction gives alkylation at nitrogen. ND, not detected.

catalyst was different for almost every substrate, further underscoring the idiosyncrasy of 4-substitution.

As with 4-nitroindole, the reaction with 5-nitroindole slowed toward the end (Figure S8). Nonetheless, 5-nitroTrp could be obtained in 88% yield with a higher catalyst loading (Table 3, entry 6). Good results were also obtained with other electron-withdrawing substituents, such as nitrile, carboxamide, and boronate (Table 3, entries 7−9). Notably, we also observed promising activity with 5-iodo- and 5-trifluoromethylindole (Table 3, entries 10 and 11), on which previous catalysts had shown no detectable activity.

The 6-substituted indoles proved to be the best behaved and most predictable series of substrates. At this position, high yields were obtained with the nitro substituent (Table 3, entry 12), as well as halo substituents (Table 3, entries 13 and 14) and other electron-withdrawing substituents like nitrile (Table 3, entry 15) and boronate (Table 3, entry 16). Unlike with the 4- and 5-positions, the two enzymes evolved for activity with 6-nitroindole were always among the best catalysts.

Excellent activity was also observed with 7-substituted indoles, though the optimal catalyst was hard to predict. While 7-nitroindole and 7-cyanoindole favored *Pf*2A6 (Table 3, entries 17 and 18), *Pf*0A9 gave optimal activity for 7-chloro- and 7-iodoindole (Table 3, entries 19 and 20). Strangely, 7-bromoindole gave poorer results than the other haloindoles, though 7-bromotryptophan could be formed in modest yield with catalyst *Pf*5G8 (Table 3, entry 21). No product was detected with indole-7-boronic acid (Table 3, entry 22).

Finally, we explored the production of Trp analogues with multiple substituents. This capability is important both because poly-substituted Trp derivatives are precursors to many natural products and because poly-halogenated arenes are prevalent in bioactive compounds in general.[22,43] Fortunately, the 5,6-dichloro and 5-bromo-7-fluoro products could be obtained in

good yields using *Pf*5G8 and *Tm*2F3 I184F, respectively (Table 3, entries 23 and 24). The bulkier 5-chloro-7-iodo product was also accessible, though in low yield (Table 3, entry 25).

Kinetic Effects of the Mutations. Our hypothesis at the outset of this project had been that the conversion of slowly reacting substrates like 4-nitroindole was limited by competing hydrolysis of the amino-acrylate intermediate. Increasing the conversion at the reaction end point, therefore, would require accelerating the nucleophilic attack of the substrate, such as through binding in the active site, and increasing the persistence of the amino-acrylate intermediate. While the poor solubility of 4-nitroindole frustrated our ability to measure Michaelis−Menten kinetics, we could approximate the initial rate of 4-nitroTrp production under the reaction conditions by measuring conversion at short reaction times (Table 4). In

Table 4. Initial Rates Throughout Evolution[a]

entry	catalyst	initial turnover frequency (min^{-1})		
		to 4-nitroTrp	to pyruvate	to Trp
1	*Pf*TrpB	–	25.0 ± 0.3	19 ± 1.2
2	*Pf*2B9	1.25 ± 0.07	12.2 ± 0.5	60.9 ± 0.16
3	*Pf*5G8	1.8 ± 0.12	2.0 ± 0.2	9.9 ± 0.5
4	*Pf*5G8 E104G	3.5 ± 0.2	0.9 ± 0.12	7.03 ± 0.07
5	*Pf*2A6	7.0 ± 0.3	1.4 ± 0.10	17.6 ± 0.3

[a]See Sections 8.10, 8.11, and 8.12 of the Supporting Information.

addition, we measured the rate of Ser deamination by incubating the enzymes with Ser, in the absence of a nucleophilic substrate, and measuring the production of pyruvate. Finally, we measured the initial rate of Trp production under the same reaction conditions that we used for 4-nitroTrp.

10773

Of all the catalysts, wild-type TrpB has the highest deamination rate, even higher than the rate of Trp production, and produces negligible levels of 4-nitroTrp (Table 4, entry 1). The variant Pf2B9, which produces Trp at a rate similar to the wild-type complex,[33] shows both an increase in the rate of 4-nitroTrp production and a decrease in the rate of Ser deamination (Table 4, entry 2). The next variant, PfSG8, exhibits only a modest increase in the rate of 4-nitroTrp production, but a 6-fold decrease in the rate of Ser deamination and an almost 3-fold decrease in the rate of Trp production (Table 4, entry 3). The mutation E104G changes both 4-nitroTrp production and Ser deamination in approximately equal measure, but surprisingly exerts only a minor influence on Trp production (Table 4, entry 4). Finally, the addition of I183F and V186A actually increases the rate of Ser deamination, but increases the rate of the desired reaction even more (Table 4, entry 5). These last two mutations also return Trp production to the level of PfTrpB.

■ DISCUSSION

Organic Synthesis Using TrpB. In the synthesis of noncanonical amino acids, installation of the chiral amino-acid moiety is often the most challenging part, due to the stereochemistry at the α-carbon. Many synthetic approaches, therefore, involve the modification of existing amino acids, in which the stereochemistry is already established. Even then, the reactivity of the amine and carboxylate functional groups can interfere with many synthetic methodologies, and thus require protecting groups. In contrast, TrpB uses unprotected Ser as the amino acid donor and forms the products in almost perfect enantiopurity. The reaction is operationally straightforward (simply add substrates and enzyme to buffer and apply heat), and the products can often be purified by precipitation directly from the reaction mixture, or alternatively by a facile chromatographic separation. Due to their high expression level and thermostability, the enzymes can be obtained easily and handled without special precautions. Furthermore, the active electrophilic species, the amino-acrylate, is formed and persists only in the enzyme active site, thus minimizing the burden of product purification.

While the range of products formed by TrpB was previously restricted to a narrow set of Trp analogues, the results of this study demonstrate that the product scope can be greatly expanded through protein engineering. More generally, they show how TrpB can mediate efficient reactions with exceptionally weak nucleophiles like 4-nitroindole, which is electronically deactivated, sterically hindered, and poorly soluble. It is likely, therefore, that the product scope of TrpB can be expanded even further to include many types of noncanonical amino acids, in addition to Trp analogues.

Role of M139 and N166 in Open/Close Transition. TrpB is well known to adopt discrete conformational states, including a fully open state, which is inactive but permits substrate entry and product release, and a fully closed state, which promotes amino-acrylate formation and nucleophilic addition but blocks access to the active site. Previous studies have also shown that stabilization of the closed state is associated with a decreased rate of Ser deamination.[44] Since M139L and N166D lie outside the active site, we hypothesized that their beneficial effect, which is mainly to suppress Ser deamination, might arise from stabilization of the closed state. While the evolved variants from this study did not form diffracting crystals, we previously obtained structures of wild-

type PfTrpB in the open and partially closed states.[33] A comparison of the structures immediately reveals that the side-chain of N166, which is mutated to aspartate in PfSG8, forms a hydrogen bond in the closed state with the side-chain of H275, which undergoes a rotameric switch that closes the active site (Figure 5). It is therefore plausible that strengthening of this
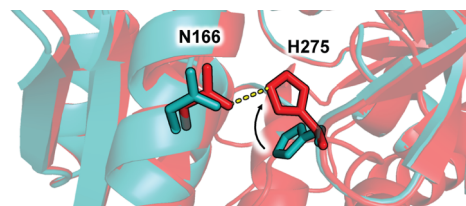

Figure 5. Overlaid crystal structures of PfTrpB (PDB ID: 5DVZ) in the open state (cyan) and partially closed state (PDB ID: 5DW0, red) showing the side-chain motion of N166, and H275.

interaction with the more basic aspartate would stabilize the closed state of the enzyme. Interestingly, the parent enzyme, Pf2B9, also has the mutation F274S, which is adjacent to H275, thus highlighting the importance of mutations in this region to activity. The effect of mutating M139 to L is more difficult to rationalize, but this residue is adjacent to the domain that mediates communication between TrpA and TrpB in the wild-type complex and which is known to undergo a large displacement in the transition from open to closed. Thus, it is plausible that a mutation at this position would influence the transition between the open and closed states.

Influence of Mutations on the Active Site. Unlike the previous stages of evolution, all the mutations from the final round (E104G, I183F, and V186A) occur at positions in the enzyme active site. In the case of I183 and V186, the side-chains do not interact directly with the substrates, nor do they undergo significant movement during the transition from the open to closed state. Thus, the beneficial effects of the mutations are likely due to subtle reshaping of the active site to accommodate the size of 4-nitroindole and bind the substrate in a more reactive conformation. This is consistent with the observation that addition of these two mutations greatly increases the rate of 4-nitroTrp production while exerting little effect on the rate of Ser deamination (Table 4, entries 4 and 5).

Various roles have been assigned to the side-chain of E104, including activation of the Ser β-hydroxyl group as a leaving group, as well as binding and activation of the indole nucleophile (see Figure 2, intermediates II and III).[45] Indeed, studies of TrpS from S. typhimurium (StTrpS) showed that mutation of the corresponding residue to alanine[46] eliminated activity with indole and Ser.[47] However, this activity was rescued by the introduction of certain monovalent cations, such as Cs+, indicating that neither of the aforementioned roles of E104 are essential. It therefore seems that its most significant role is in modulating the transition of the enzyme to the closed state. This was supported by the observation that increasing concentrations of CsCl shifted the catalytic steady state of the variant away from the external aldimine (Figure 2, intermediate II) toward the amino-acrylate as the major species, a trend that is a hallmark of closed-state stabilization. By contrast, PfSG8 and the E104G variants exhibit similar rates of Trp production, and the amino-acrylate predominates in the steady state of both variants (Figure S9). In fact, the E104G mutation appears to stabilize the closed state, as inferred from the 2-fold decrease in

deamination rate (Table 4, entries 3 and 4). Thus, it may be that the other mutations in *Pf*SG8 have changed the function of E104, or that the role of this residue is different in *Pf*TrpB and *St*TrpB. Studies are ongoing to elucidate the role and generality of this mysterious mutation.

Impact on Substrate Specificity. It is often thought that evolving an enzyme for activity with a particular substrate ultimately imparts specificity toward that substrate as well. This study, however, shows that directed evolution toward a particularly challenging substrate can be an effective strategy to improve activity for non-native substrates in general. Indeed, the intermediate catalyst, *Pf*SG8, exhibited concomitant increases in activity for all four nitroindoles (Table 1). While this enzyme was rarely the optimal catalyst, as indicated by its infrequent appearance in Table 3, it exhibited good activity for most substrates (see Table S1). The additional mutations in the most evolved enzyme, *Pf*2A6, appeared to lower the substrate generality, but nonetheless improved activity for multiple substrates, such as 7-nitroindole, in addition to the test substrate, 4-nitroindole. Furthermore, the mutations of *Pf*2A6 served as the basis to generate all other catalysts in this study, demonstrating that although a specific catalyst may not be optimal for all substrates, the mutations are activating in multiple contexts.

Our approach of creating a small panel of general catalysts likely succeeded because the conversion of most non-native substrates was limited by a single enzymatic process, Ser hydrolysis. Thus, directed evolution could select for mutations that curtailed Ser hydrolysis independently of a specific substrate. This is consistent with the results of Table 4, in which the rate of Ser deamination decreases even in the absence of a nucleophilic substrate. While these specific circumstances may not apply to all enzyme engineering problems, we would expect directed evolution to provide a general benefit where side activities that limit enzyme performance can be targeted independently.

CONCLUSION

By evolving for activity with nitroindoles, we have developed a panel of TrpB-derived biocatalysts that exhibit good to excellent activity with mono- and disubstituted indoles. The substrate scope includes indoles bearing electron-withdrawing groups, such as nitro and cyano, which are fundamentally deactivating in this reaction manifold. These results showcase the potency of the amino-acrylate as an electrophile, as well as the ability of the active site to protect the amino-acrylate from degradation, while promoting reactions with even the most stubborn nucleophiles. These qualities make TrpB catalysis a versatile and easy-to-use platform for the production of valuable synthetic building blocks.

The results also demonstrate that the universally conserved E104 residue is a key target for mutagenesis to improve activity with non-native substrates. This may provide a promising starting point to adapt TrpB for the synthesis of new product classes. More generally, this demonstrates how mutations at residues that seem crucial can in fact be the most valuable handles for tuning activity with new substrates.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b05007.

Additional figures (Figures S1−S9), tables, and experimental procedures (PDF)

NMR spectra (PDF)

AUTHOR INFORMATION

Corresponding Author

*frances@cheme.caltech.edu.

ORCID Ⓞ

David K. Romney: 0000-0003-0498-7597

Frances H. Arnold: 0000-0002-4027-364X

Notes

The authors declare the following competing financial interest(s): The contents of this paper are the subject of a patent application submitted by Caltech, and some authors are entitled to a royalty on revenues arising from that patent.

REFERENCES

(1) Jordan, P. A.; Moore, B. S. *Cell Chem. Bio.* **2016**, *23*, 1504−1514.

(2) Alkhalaf, L. M.; Ryan, K. S. *Chem. Biol.* **2015**, *22*, 317−328.

(3) Feng, Y.; Holte, D.; Zoller, J.; Umemiya, S.; Simke, L. R.; Baran, P. S. *J. Am. Chem. Soc.* **2015**, *137*, 10160−10163.

(4) Zhang, H.; Ning, X.; Hang, H.; Ru, X.; Li, H.; Li, Y.; Wang, L.; Zhang, X.; Yu, S.; Qiao, Y.; Wang, X.; Wang, P. G. *Org. Lett.* **2013**, *15*, 5670−5673.

(5) Xu, Z.; Zhang, F.; Zhang, L.; Jia, Y. *Org. Biomol. Chem.* **2011**, *9*, 2512−2517.

(6) Ma, B.; Banerjee, B.; Litvinov, D. N.; He, L.; Castle, S. L. *J. Am. Chem. Soc.* **2010**, *132*, 1159−1171.

(7) Artman, G. D.; Grubbs, A. W.; Williams, R. M. *J. Am. Chem. Soc.* **2007**, *129*, 6336−6342.

(8) Baran, P. S.; Guerrero, C. A.; Ambhaikar, N. B.; Hafensteiner, B. D. *Angew. Chem., Int. Ed.* **2005**, *44*, 606−609.

(9) Ley, S. V.; Priour, A.; Heusser, C. *Org. Lett.* **2002**, *4*, 711−714.

(10) Talukder, P.; Chen, S.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. *Biochemistry* **2015**, *54*, 7457−7469.

(11) Roy, A. D.; Goss, R. J. M.; Wagner, G. K.; Winn, M. *Chem. Commun.* **2008**, 4831−4833.

(12) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12088−12093.

(13) Hilaire, M. R.; Ahmed, I. A.; Lin, C.-W.; Jo, H.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 6005−6009.

(14) Choi, J.-M.; Han, S.-S.; Kim, H.-S. *Biotechnol. Adv.* **2015**, *33*, 1443−1454.

(15) Nestl, B. M.; Hammer, S. C.; Nebel, B. A.; Hauer, B. *Angew. Chem., Int. Ed.* **2014**, *53*, 3070−3095.

(16) Wohlgemuth, R. *Curr. Opin. Biotechnol.* **2010**, *21*, 713−724.

(17) Payne, J. T.; Andorfer, M. C.; Lewis, J. C. *Angew. Chem., Int. Ed.* **2013**, *52*, 5271−5274.

(18) Durak, L. J.; Payne, J. T.; Lewis, J. C. *ACS Catal.* **2016**, *6*, 1451−1454.

(19) Shepherd, S. A.; Menon, B. R. K.; Fisk, H.; Struck, A.-W.; Levy, C.; Leys, D.; Micklefield, J. *ChemBioChem* **2016**, *17*, 821−824.

(20) Latham, J.; Henry, J.-M.; Sharif, H. H.; Menon, B. R. K.; Shepherd, S. A.; Greaney, M. F.; Micklefield, J. *Nat. Commun.* **2016**, *7*, 11873.

(21) Dodani, S. C.; Kiss, G.; Cahn, J. K. B.; Su, Y.; Pande, V. S.; Arnold, F. H. *Nat. Chem.* **2016**, *8*, 419−425.

(22) Smith, D. R. M.; Uria, A. R.; Helfrich, E. J. N.; Milbredt, D.; van Pée, K.-H.; Piel, J.; Goss, R. J. M. *ACS Chem. Biol.* **2017**, *12*, 1281−1287.

(23) Blaser, G.; Sanderson, J. M.; Batsanov, A. S.; Howard, J. *Tetrahedron Lett.* **2008**, *49*, 2795−2798.

(24) Baldwin, J. E.; Dyer, R. L.; Ng, S. C.; Pratt, A. J.; Russell, M. A. *Tetrahedron Lett.* **1987**, *28*, 3745−3746.

(25) Corr, M. J.; Smith, D. R. M.; Goss, R. J. M. *Tetrahedron* **2016**, *72*, 7306−7310.

(26) Smith, D. R. M.; Willemse, T.; Gkotsi, D. S.; Schepens, W.; Maes, B. U. W.; Ballet, S.; Goss, R. J. M. *Org. Lett.* **2014**, *16*, 2622−2625.

(27) Perni, S.; Hackett, L.; Goss, R. J.; Simmons, M. J.; Overton, T. W. *AMB Express* **2013**, *3*, 66.

(28) Winn, M.; Roy, A. D.; Grüschow, S.; Parameswaran, R. S.; Goss, R. J. M. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4508−4510.

(29) Goss, R. J. M.; Newill, P. L. A. *Chem. Commun.* **2006**, 4924.

(30) Lee, M.; Phillips, R. S. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 1563−1564.

(31) Guo, T.; Huang, F.; Yu, L.; Yu, Z. *Tetrahedron Lett.* **2015**, *56*, 296−302.

(32) Inman, M.; Moody, C. J. *Chem. Sci.* **2013**, *4*, 29−41.

(33) Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 14599−14604.

(34) Herger, M.; van Roye, P.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. *J. Am. Chem. Soc.* **2016**, *138*, 8388−8391.

(35) Murciano-Calles, J.; Romney, D. K.; Brinkmann-Chen, S.; Buller, A. R.; Arnold, F. H. *Angew. Chem., Int. Ed.* **2016**, *55*, 11577−11581.

(36) Zhang, H.; Wang, Q.; Ning, X.; Hang, H.; Ma, J.; Yang, X.; Lu, X.; Zhang, J.; Li, Y.; Niu, C.; Song, H.; Wang, X.; Wang, P. G. *J. Agric. Food Chem.* **2015**, *63*, 3734−3741.

(37) Bourgault, J. P.; Maddirala, A. R.; Andreana, P. R. *Org. Biomol. Chem.* **2014**, *12*, 8125−8127.

(38) King, R. R.; Calhoun, L. A. *Phytochemistry* **2009**, *70*, 833−841.

(39) Barry, S. M.; Kers, J. A.; Johnson, E. G.; Song, L.; Aston, P. R.; Patel, B.; Krasnoff, S. B.; Crane, B. R.; Gibson, D. M.; Loria, R.; Challis, G. L. *Nat. Chem. Biol.* **2012**, *8*, 814−816.

(40) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5869−5874.

(41) Crawford, I. P.; Ito, J. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *51*, 390−397.

(42) Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F. *Biochemistry* **2013**, *52*, 6396−6411.

(43) Agarwal, V.; Miles, Z. D.; Winter, J. M.; Eustáquio, A. S.; El Gamal, A. A.; Moore, B. S. *Chem. Rev.* **2017**, *117*, 5619−5674.

(44) Fan, Y. X.; McPhie, P.; Miles, E. W. *Biochemistry* **2000**, *39*, 4692−4703.

(45) Dunn, M. F. *Arch. Biochem. Biophys.* **2012**, *519*, 154−166.

(46) E109A in the numbering of *S. typhimurium*.

(47) Ruvinov, S. B.; Ahmed, S. A.; McPhie, P.; Miles, E. W. *J. Biol. Chem.* **1995**, *270*, 17333−17338.