HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2024 August 01.

Published in final edited form as:

Nat Chem Biol. 2024 August ; 20(8): 1086–1093. doi:10.1038/s41589-024-01619-z.

The β-subunit of tryptophan synthase is a latent tyrosine synthase

Patrick J. Almhjell[1,†], **Kadina E. Johnston**[2,‡], **Nicholas J. Porter**[1,§], **Jennifer L. Kennemur**[1], **Vignesh C. Bhethanabotla**[1], **Julie Ducharme**[1,¶], **Frances H. Arnold**[1,2,*]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

Aromatic amino acids and their derivatives are diverse primary and secondary metabolites with critical roles in protein synthesis, cell structure and integrity, defense, and signaling. All *de novo* aromatic amino acid production relies on a set of ancient and highly conserved chemistries. Here we introduce a new enzymatic transformation for L-tyrosine synthesis by demonstrating that the β-subunit of tryptophan synthase (TrpB)—which natively couples indole and L-serine to form L-tryptophan—can act as a latent "tyrosine synthase". A single substitution of a near-universally conserved catalytic residue unlocks activity toward simple phenol analogs and yields exclusive *para* carbon–carbon bond formation to furnish L-tyrosines. Structural and mechanistic studies show how a new active-site water molecule orients phenols for a non-native mechanism of alkylation, with additional directed evolution resulting in a net >30,000-fold rate enhancement. This new biocatalyst can be used to efficiently prepare valuable L-tyrosine analogs at gram scales and provides the missing chemistry for a conceptually different pathway to L-tyrosine.

The aromatic amino acids (aroAAs) L-phenylalanine (Phe), L-tyrosine (Tyr), and L-tryptophan (Trp) serve as fundamental building blocks of proteins and diverse secondary metabolites across all life[1–4]. These molecules are produced by organisms on a massive global scale: lignin synthesis, for example, drives 20–30% of all photosynthetically fixed

*Corresponding author. frances@cheme.caltech.edu.
†Present address: Department of Biochemistry, Stanford University, Stanford, CA 94305, USA.
‡Present address: Merck & Co., Inc, South San Francisco, CA 94080, USA.
§Present address: Codexis, Inc., Redwood City, CA 94063, USA.
¶Present address: Quebec Government Office, Los Angeles, CA 90024, USA.

Author Contributions Statement

Conceptualization: P.J.A. and F.H.A.; Investigation: P.J.A., K.E.J., N.J.P., J.L.K., V.C.B.; Validation: P.J.A., K.E.J, N.J.P., V.C.B., J.L.K., J.D.; Writing – original draft: P.J.A.; Writing – review & editing: P.J.A., K.E.J., N.J.P., J.L.K., V.C.B., J.D., F.H.A.; Funding acquisition: F.H.A.; Supervision: F.H.A.

Competing Interests Statement

P.J.A. and F.H.A. are inventors on patent applications filed by the California Institute of Technology that cover enzymatic synthesis of tyrosine analogs from analogs of phenol and serine (US17/985,033 and PCT/US2022/049617, pending). The remaining authors declare no competing interests.

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to F.H.A.; plasmids encoding the enzymes reported in this study are available for research purposes under a material transfer agreement with the California Institute of Technology.

carbon through the pathways of aroAA biosynthesis[5]. Despite some differences in the arrangement of the biosynthetic pathways and their modes of regulation across organisms, all *de novo* biosynthesis of these aroAAs relies on a universally conserved set of chemical transformations to convert chorismate—the common aromatic precursor—into Phe, Tyr, and Trp[1,2,6] (Fig. 1a). These chemistries are ancient, and likely existed in some of the earliest ancestors in the tree of life[6,7].

Not only are these pathways the only ones known for the *de novo* production of the canonical aroAAs within primary metabolism, they are the only pathways possible with the currently known repertoire of enzyme activities, *in vivo* or *in vitro*. For example, although animals lack the shikimate pathway to produce chorismate, they are capable of synthesizing Tyr, but this occurs *via* hydroxylation of Phe obtained in their diets rather than from chorismate[2]. Other organisms can make their aroAAs from various aryl acids, but these molecules are themselves metabolites of previously produced aroAAs[4,8]. Certain enzymes that natively degrade aroAAs will operate in the reverse, synthetic direction but require large excesses of reagents that are infeasible *in vivo*[9,10]. The pyridoxal 5'-phosphate (PLP)-dependent enzymes tyrosine phenol-lyase (TPL)[11] and tryptophanase (Trpase)[12] have been shown to synthesize Tyr or Trp from phenol or indole and ammonium pyruvate, respectively, acting through a transient amino-acrylate intermediate[12] (Fig. 1b). These enzymes have found applications in biocatalysis, with TPL used to prepare valuable Tyr analogs for biological research[13,14] and for a significant portion of global L-DOPA production *via* the Ajinomoto process[3]. However, TPL and Trpase are not "synthases" and remain intrinsically limited by the thermodynamically favorable degradation of the amino-acrylate: with an estimated physiological K'_{eq} of 6.4×10^4, the β-elimination of Tyr to ammonium pyruvate greatly predominates under thermodynamic control[11,15].

We envisioned a new enzymatic chemistry that could overcome these limitations. The β-subunit of tryptophan synthase (TrpB) catalyzes the final step of all known *de novo* Trp biosynthesis *via* the Friedel-Crafts alkylation of indole[6,16] (Fig. 1a). Whereas Trpase and TPL catalyze the rapid breakdown of their amino-acrylate intermediates, TrpB avoids this reaction pathway and produces a kinetically stable amino-acrylate, which is generated from L-serine (Ser) (Fig. 1c and Extended Data Fig. 1a and 1b). This imparts a level of kinetic control over the reaction pathway[10] and renders TrpB able to favor Trp synthesis over its degradation in all known *de novo* Trp biosynthesis. The latter role is instead filled by Trpase in organisms like *Escherichia coli* (Extended Data Fig. 1b). This allows TrpB, rather than Trpase or TPL, to function efficiently within natural and synthetic metabolism[17], in addition to being used for biocatalytic synthesis of Trp analogs[18–21] and diverse noncanonical amino acids (ncAAs) *in vitro*[22–25]. However, whereas Trpase and TPL parallel one another for Trp and Tyr degradation, respectively, there is no equivalent to TrpB for Tyr synthesis: no enzyme has been shown to synthesize Tyr or its analogs directly from phenols under kinetic control, even though phenols are present in metabolism[26–29] (Fig. 1d) and environmental contamination[30]. We reasoned that it should be possible to engineer a TrpB-like "tyrosine synthase" (TyrS) that catalyzes the facile alkylation of phenols and avoids the degradation activity of TPL (Fig. 1e). Demonstration of this synthetic activity *in vitro* would support the feasibility of an alternative, TrpB-like enzymatic route to Tyr, a molecule central to all of life, and would be useful for preparation of Tyr analogs.

Results

Directed evolution of a tyrosine synthase

Given that TPL uses an amino-acrylate intermediate to accomplish regioselective Tyr synthesis (and degradation), we suspected that TrpB could be engineered to do the same, but differences between the two enzymes' native reactions highlighted potential challenges. Phenol is a small, symmetric molecule with a single heteroatom for coordination. Further, it has three unique nucleophilic positions: two carbon atoms *ortho* and one *para* to the hydroxyl, as well as the hydroxyl itself (Fig. 1e). Tyr synthesis requires a highly *para*-selective catalyst. In TPL, multiple residues coordinate the phenolic hydroxyl group to facilitate *para* C–C bond breakage[31], which also lowers the energy barrier for the reverse reaction (Extended Data Fig. 1c). In TrpB, a catalytic glutamate (E105) present in all characterized enzymes plays a similar role by coordinating indole for C–C bond formation during Trp synthesis (Extended Data Fig. 1d)[24]. These parallel strategies suggested that an efficient TyrS would require an analogous coordination mechanism to favor *para* C–C bond formation over *ortho* or O-alkylation.

When representative engineered TrpB variants were challenged with phenol and Ser, however, none of these three possible amino-acid products was observed. This prompted us to look for a substrate that might serve as an evolutionary stepping stone from indole to phenol and use a 'substrate walk' approach[32] to evolve further (Fig. 2a). We chose 1-naphthol as an electron-rich phenol analog similar enough to indole to bind in the active site and also be aligned for *para* C–C bond formation (Fig. 2b). To our delight, both TrpB variants tested—Tm9D8*[33], from *Thermotoga maritima* and engineered for 4-cyano-Trp formation at 37 °C, and *Pf*2B9[20], from *Pyrococcus furiosus* and engineered for β-methyl-Trp production at 75 °C—reacted with 1-naphthol and Ser to form an amino acid condensation product (Extended Data Fig. 2a). Tm9D8* was selected for further evolution because it retains high thermostability while also displaying good activity at 37 °C, which reduces oxidation of electron-rich substrates and would be important for future *in vivo* applications.

Under the presumption that 1-naphthol was binding in a similar orientation to the natural indole substrate of TrpB, it appeared likely that the catalytic glutamate which coordinates indole was not optimal for this non-native substrate (Fig. 2b and Extended Data Fig. 2b). Substitutions at this position are generally detrimental, as is expected for mutations made to a conserved catalytic residue: previous studies showed, for example, that mutation to alanine can impair formation of the amino-acrylate in *Salmonella typhimurium* TrpB[34]. Additionally, mutating this residue to glycine in native TrpS and engineered TrpB enzymes (including *Tm*TrpB) reduced the rate of Trp synthesis by 2–20 fold and effectively abrogated activity with azulene, a non-indole nucleophile[24]. However, given that we had also seen that mutation of this residue could increase activity toward recalcitrant substrates[21], we performed site-saturation mutagenesis and screening at this position (E105) and identified that substitutions to glycine, alanine, and serine were highly activating. Substitution to glycine (E105G) provided the largest rate enhancement (18-fold over Tm9D8*). Preparative-scale synthesis using this enzyme allowed us to confirm that the *para*-alkylation product

β-(1-naphthol-4-yl)-L-alanine (NaphAla) was the sole product of this enzymatic reaction (Extended Data Fig. 2c).

We continued the substrate walk approach to evolve Tm9D8* E105G for activity on phenol by first increasing activity on 1-naphthol and then moving to substrates that are progressively more similar to phenol as new activities were detected (Fig. 2, a and c). We expected that this strategy would also simultaneously yield TyrS variants suitable for the synthesis of valuable Tyr analogs with diverse structures *en route* to achieving our primary goal of new enzymatic Tyr synthesis. This proved effective, and over an extensive evolution campaign utilizing global random mutagenesis, site-saturation mutagenesis, and recombination (see Supplementary Notes, Evolutionary Strategies), we created six new TmTyrS enzymes (*T. maritima*-derived tyrosine synthase enzymes 1–6; see Supplementary Table 1 for substitutions and Supplementary Information, Enzyme DNA Sequences for sequences). The final variant, TmTyrS6, accrued 22 substitutions from Tm9D8* and synthesizes Tyr at an apparent turnover frequency (TOF) of 0.23 min^{-1} (50 mM each substrate, 37 °C).

These evolved enzymes are all at least 99.5% enantio- and regioselective for Tyr synthesis: we detected no D-Tyr, consistent with the TrpB mechanism (Extended Data Fig. 3a), and no *ortho*-alkylation was detected even when the concentration of enzymatically prepared Tyr was >1,000-fold higher than the limit of quantification (Extended Data Fig. 3b). This regioselectivity is remarkable considering the fact that phenol presents multiple nucleophilic positions and there is no *a priori* expectation that the enzyme should so strongly prefer the *para* position. The approximate turnover frequencies are presented on a logarithmic scale in Fig. 2d, along with a conservative lower bound for the amount of enzymatically prepared Tyr under the given conditions (see Extended Data Fig. 3e and Supplementary Notes, Determination of limit for detectable turnover frequency). Because Tm9D8* does not make Tyr at this threshold, the evolution presented here represents at least a 30,000-fold increase in activity from Tm9D8* to TmTyrS6. Importantly, these reactions remain kinetically controlled: incubation of 10 μM TmTyrS6 with 1 mM Tyr at 37 °C for 20 hours results in complete retention of the Tyr and no detectible release of phenol, and the same results are obtained for other TyrS variants with Tyr analogs (Extended Data Fig. 4).

The effect and prevalence of E105G

The single E105G substitution imparted a significant rate enhancement in Tm9D8* for conversion of 1-naphthol (18 fold) and phenol (>100 fold), but it was unclear how general this effect is across other TrpBs and other substrates. We tested the generality by installing the equivalent E104G substitution in the stand-alone TrpB from *P. furiosus, Pf*2B9, which had earlier demonstrated activity with 1-naphthol when tested alongside Tm9D8*. The E104G substitution increased tyrosine synthase activity 7.8 fold at *Pf*2B9's optimal temperature of 75 °C (Fig. 3a). Variants lacking the catalytic glutamate were less active and less regioselective for native indole alkylation, as previously observed[21,24]. (N-alkylation increased from undetectable to approximately 1:4 and 1:1 N-:C-alkylation for Tm9D8* and *Pf*2B9, respectively; Extended Data Fig. 5a.) These levels were below the activity and

regioselectivity for 1-naphthol, demonstrating that this single substitution is sufficient to invert selectivity from indole to a sterically similar phenol analog in both TrpB homologs.

Removal of the glutamate sidechain also enhanced activity on simple phenol analogs under all tested conditions (Fig. 3a). E104G effected a >40-fold increase in the activity of *Pf*2B9 toward 2-chlorophenol and a 77-fold increase toward 2-iodophenol. Tm9D8* E105G saw an even more impressive >420- and an 800-fold increase over Tm9D8* when provided 2-chlorophenol and 2-iodophenol, respectively. In both the Tm9D8* and *Pf*2B9 variants, this sole glutamate-to-glycine substitution was sufficient to enable detectable Tyr formation at higher enzyme and substrate concentrations (Fig. 3a and Extended Data Fig. 5b).

The native catalytic glutamate is strictly conserved in all characterized TrpB enzymes within the human-annotated SwissProt database (451 sequences). To examine the conservation of this residue in more detail, we analyzed 18,051 TrpB-like sequences with 14–93% aligned sequence identity to Tm9D8*. Of these, 98.28% (17,741) contained the catalytic glutamate (Extended Data Fig. 6a). Three other amino acids occurred with a non-negligible frequency: alanine (0.59%; 107), aspartate (0.53%; 95), and glycine (0.43%; 77). Aspartate provides a similar functionality to the native glutamate[35], but the presence of alanine and glycine in otherwise TrpB-like sequences is surprising. Notably, these sequences are further grouped into two distinct classes. Sequences with the E105A-equivalent are found across plants, primarily in the malvids, and are homologous to TrpB enzymes found in native Trp biosynthesis. Those with the E105G-equivalent are found predominantly in soil-dwelling *Actinobacteria* like *Actinomyces* and *Streptomyces* (which have diverse secondary metabolic pathways and can produce phenols[27]) and share homology with TrpB2 enzymes, which are stand-alone Trp synthases (i.e., no TrpA binding partner) suggested to scavenge free indole and return it to metabolism in the form of Trp due to their low-nanomolar K_M for indole[36].

We examined the other positions that were altered during evolution and found only one with equivalent conservation. The substitution G229A, which was highly activating for phenol activity in TmTyrS5, occurs at a position that is conserved in 98.20% (17,726) of the sequences in this dataset. Remarkably, differences at this position are highly correlated with the identity of the amino acid at position 105 (aligned) and organism: the G229S/T-equivalent is found in plants (E105A) while the G229A-equivalent is found among bacteria (E105G) (Extended Data Figs. 6b and 6c). Phylogenetic analysis supports this correlation and demonstrates that these are unique, uncharacterized clades of TrpB-like sequences (Fig. 3b and Extended Data Fig. 6d). It remains to be seen whether these sequences encode efficient Trp synthases, unidentified Tyr/Tyr analog synthases (given their association with phenol and lignin-producing/assimilating species), or whether they have other biological roles. One uncharacterized plant TrpB-like sequence is found in the model organism *Arabidopsis thaliana*. This gene (AT5G28237) has been mentioned in a number of reports—often encoding a “putative Trp synthase”—but has not been experimentally characterized[37–39]. Our results suggest that, at minimum, its product is unlikely to be an efficient Trp synthase.

Structure and mechanism of TyrS

To understand this new enzymatic activity, especially the role of the E105G substitution, we pursued structural characterization of TyrS enzymes in various catalytically relevant states, including in complex with substrate analogs. We found that TmTyrS1 readily formed crystals that provided high-resolution experimental X-ray crystal structures for this study (Extended Data Fig. 7 and Supplementary Table 2). This variant represents an intermediate between TrpB and TyrS activity, and the substrate preference for indole or 1-naphthol can be determined by the presence of E105 or G105, respectively (see Extended Data Fig. 8). Previous reports showed that TrpB crystals formed in the resting state (known as the internal aldimine, E(Ain), state) can be used to access the reactive amino-acrylate complex, E(A-A)[40]. This is also the case for TyrS crystals, as soaking Ser into the E(Ain) crystals readily led to formation of a stable E(A-A) complex in both subunits of the asymmetric unit (Fig. 4b inset and Extended Data Fig. 7c). The ability to observe this reactive intermediate highlights its stability within the TrpB scaffold. In contrast, the E(A-A) species in Trpase and TPL is transient and can only be observed spectroscopically by kinetic trapping[12].

In both the E(Ain) and E(A-A) structures of TmTyrS1, removal of the E105 sidechain made space for the coordination of a new water molecule in the active site, which was well resolved in all structures. This water makes four hydrogen bonds, one to each of the following: G105 (backbone N, donor), V183 (backbone O, acceptor), Y182 (hydroxyl, undetermined donor or acceptor), and the fourth to another water molecule (Extended Data Fig. 7d and Fig. 4a). Active-site water molecules can enable catalysis through electrostatic interactions with the functional groups of substrates[41,42]. If the fourth hydrogen-bonding interaction to the additional water is replaced by one with the hydroxyl group of phenol, only the reactive carbon *para* to this group would extend to the reactive amino-acrylate C_β, rationalizing how these TyrS enzymes perform this non-native reaction with exquisite regioselectivity. To obtain more conclusive evidence of this, the structure of TmTyrS1 in the E(A-A) state was determined in complex with two non-reactive 1-naphthol analogs, 4-hydroxyquinoline (4-HQ) and quinoline *N*-oxide (QNO). Both analogs bound within the active site and formed the anticipated hydrogen bond with the active-site water (Extended Data Figs. 7g and 7h, see Supplementary Notes, Crystallographic observations). While QNO was oriented non-productively, 4-HQ bound in a reactive position with the atom *para* to its hydroxyl group oriented towards the amino-acrylate, supporting the role of the active-site water in directing regioselective bond formation in TmTyrS1 (Fig. 4a).

We next aimed to determine rate-limiting step(s) (RLSs) for this new reaction, which would provide insight into the mechanism of alkylation and identify the step(s) that were implicitly optimized during directed evolution. We determined kinetic isotope effects (KIEs) for the two relevant C–H bond-breaking steps in the reaction (see Extended Data Fig. 1a), and for the entire evolutionary lineage as previous work demonstrated that the RLS of TrpB can change over the course of directed evolution[40] (Supplementary Table 3). We observed no significant KIE for deprotonation of Ser-C_α *en route* to E(A-A), demonstrating that this step is not rate-limiting, even in TmTyrS1 when using the fast substrate 1-naphthol. Instead, we found deprotonation of phenol-C_4 (the carbon *para* to the hydroxyl) during the alkylation reaction to be significantly rate-limiting, with primary KIEs from 1.6–3.2 for 2-chlorophenol

and 1.9–4.0 for 2-methylphenol when protiated and deuterated substrates were measured in competition. These two substrates showed comparable activity to one another across the lineage of TyrS variants despite their electronic differences. Observation of these KIEs for all variants indicates that cleavage of the C4–H bond—the step that restores aromaticity lost during the nucleophilic attack from phenol-C_4—was the specific step optimized during evolution from Tm9D8* E105G.

This primary KIE has only been observed in Trpase and TPL, which use a concerted mechanism of alkylation[12,31]. In the direction of alkylation, this means that C–C bond formation at the reactive carbon occurs simultaneously with deprotonation of that carbon (see Extended Data Fig. 1a). In contrast, TrpB exhibits no primary KIE for this step and follows a step-wise mechanism of alkylation, first catalyzing C_3–C_β bond formation between indole and the amino-acrylate to form a discrete “indolenine” intermediate, which is followed by deprotonation of the sp^3-hybridized C_3 atom (Extended Data Fig. 1a). It has been suggested that this specific mechanistic difference imparts to TrpB its characteristic kinetic control[12]. If TyrS were to use a concerted pathway, as suggested by the primary KIEs, then some other aspect of TrpB’s functional architecture must be critical for TyrS to achieve its kinetically controlled Friedel-Crafts alkylation with phenols (e.g., the stable amino-acrylate, as described above).

We turned to theoretical calculations to investigate these alternative mechanisms and to obtain more conclusive evidence of the role of the active-site water in catalysis. We first calculated the structures of the intermediates involved in the step-wise mechanism of indole and phenol alkylation, using the initial conformations from the X-ray crystal structure and simulating a truncated system consisting of the PLP-bound amino-acrylate, substrate, and catalytic lysine at the quantum mechanical (QM) level of theory (a “minimal model”, see Methods). Consistent with the known stability of indolenine species[12], the indole-based intermediate occurring after C_3–C_β bond formation is indeed stable, even in the minimal model without explicit enzyme interactions to stabilize it (Extended Data Fig. 9a). In contrast, we found that the analogous intermediate with phenol does not converge to a stable species *in silico* (Extended Data Fig. 9b). To confirm this was not an artifact of the minimal model, we implemented a full quantum mechanical/molecular mechanical (QM/MM) model to capture any potentially stabilizing active-site interactions (see Methods), but obtained the same results. We therefore could not examine a step-wise pathway of phenol alkylation. We could, however, calculate the structure of the concerted transition state mediated by the interaction with the new active-site water and the N_ε of the catalytic lysine K83 (Fig. 4b).

Remarkably, the catalytic and reactive groups in this structure are nearly superimposable with the analogous ones in the co-crystal structure obtained with TmTyrS1 E(A-A) and 4-HQ (RMSD of 0.884 Å over 10 atoms, Fig. 4c), indicating that this non-reactive analog is acting similarly to a transition-state analog for the concerted reaction. Furthermore, the crystal structures with and without 4-HQ show minimal differences (RMSD of 0.185–0.415 Å over 140 active-site atoms, Fig. 4c; see also Extended Data Fig. 7i and Supplementary Table 4), suggesting that TyrS is leveraging active-site pre-organization of the E(A-A) state to catalyze this reaction[43]. These results present a reasonable mechanism by which a single mutation unlocks a new activity *via* bypass of an unstable intermediate along the native

reaction coordinate, pre-organizing the active site to accommodate the concerted step and thereby opening a latent pathway for Tyr synthesis in TrpB.

Efficient and scalable ncAA synthesis

Finally, we used TyrS to perform gram-scale biocatalytic ncAA synthesis in a manner similar to that described for its progenitor enzyme, TrpB[21,24,33]. Given that different substrates were targeted over the course of TyrS evolution, variants in this lineage should serve as biocatalysts for efficient synthesis of structurally diverse noncanonical Tyr analogs. Although high concentrations of phenolic substrates destabilize the enzyme (e.g., above 50 mM 2-methylphenol, 25 mM 2-chlorophenol, or 10 mM 1-naphthol), this can be overcome by slow addition of the phenolic substrate. We used this approach to synthesize NaphAla, a commercially unavailable blue-fluorescent ncAA whose applications have been limited by its challenging synthesis[44], at multi-gram scale. We prepared the enzyme catalyst by simply heat-treating lysate from a 1-L *Escherichia coli* expression culture and lyophilizing the clarified lysate to yield a benchtop-stable powder, requiring no column purification. Over the course of 24 hours, 1-naphthol ($0.14 / g, Millipore-Sigma) was slowly added to a solution of resuspended TmTyrS1 powder and Ser ($0.77 / g) to generate the NaphAla product, which precipitated from solution toward the end of the substrate addition (Fig. 5a). The resultant solid was collected over a filter, washed with ice-cold water and ethyl acetate to remove buffer salts and unreacted substrates, and subsequently dried *in vacuo*, affording 5.5 g of pure NaphAla without significant reaction optimization (see Supplementary Methods, Gram-scale syntheses of Tyr analogs for full details).

We used a similar approach to synthesize 3-methyl-Tyr (Fig. 5b). While chlorinated Tyr analogs are easily accessible from Tyr (along with other halogenated Tyr analogs), synthesis of Tyr analogs bearing other substitutions like methyl are not. Biological production of 3-Me-Tyr is known to occur naturally during the biosynthesis of saframycin A[45], in which Tyr is directly methylated by a radical *S*-adenosyl methionine (SAM) enzyme. It is otherwise available commercially at high cost (~$1,600 / g) or prepared synthetically *via* cross coupling of tetramethyltin and 3-iodo-Tyr occurring over six days at 70 °C[46] or using TPL followed by chromatographic purification[11]. Given the similar activity of TyrS variants towards 2-methylphenol and 2-chlorophenol (as determined from the rates obtained during KIE determination), we reasoned that TyrS would provide a convenient means of synthesizing this compound. Two sequential additions of 50 mM 2-methylphenol (<$0.1 / g) resulted in product precipitation at >90% conversion, allowing us to isolate 1.13 g of 3-methyl-Tyr without chromatographic purification (see Supplementary Methods, Gram-scale syntheses of Tyr analogs for full details). Given the ease of purification and low cost of the reagents and catalyst, as well as irreversibility, TyrS-based synthesis of tyrosine analogs is highly practical and efficient, requiring no specialized equipment.

Discussion

The fact that TrpB is responsible for all known Trp biosynthesis but had never been reported to catalyze the formation of Tyr presented us with a compelling challenge: could we discover a new solution to Tyr synthesis that is fundamentally different from that used

in nature? Throughout life, the only known route for the *de novo* biosynthesis of Tyr subjects chorismate—the precursor to all aroAAs—to the same three chemical steps: a Claisen rearrangement, dehydrogenation, and transamination[6] (see Fig. 1a). This chemistry parallels that used for Phe biosynthesis, which uses a dehydration in place of dehydrogenation. Both of these pathways diverge logically from that used for Trp biosynthesis, in which indole is first created from chorismate and then coupled to an amino-acrylate to furnish Trp. We reasoned that the Tyr synthesis paradigm could be inverted and a Trp-like approach could generate Tyr by either identification or engineering of an enzyme that could catalyze the irreversible and regioselective Friedel-Crafts alkylation of phenol to Tyr. Notably, while preparation of indole requires five enzymatic steps before it is biologically available as a nucleophile, phenol is accessible in just two enzymatic steps from chorismate[47] (Fig. 1d). Given that native Tyr biosynthesis occurs in three steps from chorismate, a Trp-like pathway for Tyr synthesis would use an identical number of steps to the native one.

Our results show that this final step is possible—it is a latent and highly evolvable activity of TrpB. Although TrpB is a clear choice for catalyzing the Friedel-Crafts alkylation of phenol, given its ability to do the same with indole, features of this new, biologically relevant reactivity make this result particularly remarkable. Despite being unknown in nature, we found that a single substitution is sufficient to increase this non-native activity by up to two orders of magnitude. It has been observed numerous times that solutions discovered through evolution can be surprising and unintuitive, and that mutation of conserved residues can greatly enhance non-native activities[48,49]—this study characterizes a new and a compelling example. In addition, not only do these enzymes alkylate phenol, but they do so with extraordinary regio- and chemoselectivity to furnish exclusively Tyr analogs, as opposed to *ortho*-Tyr or *O*-phenyl-Ser analogs. Through the E105G mutation, it appears that the binding of a new water molecule in place of the catalytic glutamate functional group is sufficient to facilitate a concerted pathway of C_4–C_β bond formation and deprotonation of the phenolic substrate, bypassing an unstable step-wise intermediate in this reaction that is stable during Trp formation. This single substitution improves activity toward 1-naphthol while dramatically decreasing it toward indole, suggesting that the presence or absence of this near-universally conserved sidechain can determine the optimal catalytic machinery for each reaction: its presence favors C_3-alkylation of the pyrrole group of indole, while its absence favors C_4-alkylation of the phenolic group in the sterically similar 1-naphthol.

It may seem curious that tyrosine synthases have not been observed in nature, given the ease with which this activity is unlocked. While our phylogenetic analysis suggests that this activity—or some related one—could already exist, we suspect that any such natural enzymes are likely involved in pathways of secondary metabolism, forming Tyr analogs instead. For a TyrS to function in the *de novo* biosynthesis of Tyr, a selective advantage would need to make it competitive with the existing prephenate-based pathway that has been optimized for billions of years. Among many factors, this would be difficult to achieve using phenol as an endogenous metabolite given the significant biological innovations cells employ to retain indole for Trp synthesis. For example, Trp biosynthesis from indole is supported by the allosterically activating α-subunit of tryptophan synthase (TrpA), which produces free indole and shuttles it to the TrpB active site, as well as the indole-scavenging TrpB2s[16,19,36]. However, unconstrained by the adaptive pressures inherent in

natural evolution, we suspect that future laboratory evolution, engineering, and design could be effective in spanning the fitness valley between prephenate-based Tyr synthesis and a hypothetical TyrS-based mode of Tyr synthesis, allowing us to explore new, uncharted areas of the fitness landscape of primary metabolism.

Methods

General experimental methods.

Chemicals were purchased from commercial sources and used without additional purification. Analytical LCMS was performed on an Agilent 1260 Infinity II LC/MSD iQ equipped with a reversed-phase Poroshell 120 EC-C18, 4.6x50 mm, 2.7 μm column using a gradient of H_2O/MeCN with 0.1% acetic acid by volume. Unless otherwise stated, the gradient applied was 1–95% MeCN over 3 minutes, then held for 0.5 minutes, then immediately dropped to 1% MeCN for 0.5 minutes. NMR spectra were collected on a Bruker 400 MHz (100 MHz) spectrometer equipped with a cryogenic probe. Proton chemical shifts are reported in parts per million (ppm) relative to tetramethylsilane and calibrated using the residual solvent resonance (H_2O/HDO in D_2O, 4.79 ppm). Regioselectivities were determined by LCMS comparison to known standards when they could be chromatographically separated and by performing NMR experiments when sufficient product could be isolated.

Cloning, expression, and preparation of enzyme catalysts.

Genes coding for all enzyme variants were cloned into pET22b(+) between the *Nde*I and *Xho*I restriction sites, in frame with the Lac-inducible T7 promotor and C-terminal 6xHis tag for expression in transformed BL21(DE3) *Escherichia coli* cells. Mutagenized constructs were prepared by site-saturation mutagenesis (SSM) *via* a QuikChange™-like PCR using Phusion® polymerase[50] and a modified implementation of the "22-codon trick"[51], error-prone PCR mutagenesis as described previously[24], or recombination *via* a modified version of Staggered Extension Process (StEP) PCR[52] (see Supplementary Methods, Detailed mutagenesis techniques for complete details). PCR constructs were treated with *Dpn*I at 37 °C for 1 hour, isolated *via* gel extraction, and assembled into a circular vector *via* Gibson Assembly®[53], which was then used to transform chemically competent *E. coli*.

Single colonies of *E. coli* harboring gene variants were isolated on Lysogeny Broth (LB) agar medium supplemented with 100 μg/mL carbenicillin. For large-scale expression, a single colony was transferred to 5 mL of LB with 100 μg/mL carbenicillin (LB_{carb}) and grown to stationary phase at 37 °C and 230 rpm. The culture was then diluted 1:250 into 250 mL Terrific Broth (TB) supplemented with 100 μg/mL carbenicillin (TB_{carb}) and grown for 6 hours at 37 °C at 250 rpm, yielding a dense, uninduced culture. Protein expression was then induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and proceeded at 30 °C for 24 hours. Cells were harvested *via* centrifugation at >5,000*g* for 10 minutes, the supernatant discarded, and then the cells were stored at −20 °C until needed.

To purify, thawed pellets were resuspended to 10 mL with a lysis buffer containing 25 mM potassium phosphate, 100 mM NaCl, and 20 mM imidazole, pH 8.0 (Buffer A),

then supplemented with 100 μM pyridoxal 5'-phosphate (PLP), 0.02 mg/mL DNase I, and BugBuster® at 1/10th the manufacturer's recommendation. Cell lysis proceeded at 37 °C for 1 hour with shaking at 220 rpm, at which point the lysate was heat treated at 75 °C for 30–60 minutes. The lysate was clarified by centrifugation for 20 minutes at 14,000*g*, and the supernatant was collected. The lysate was run over a gravity column prepared with 1–2.5 mL Ni-NTA Agarose (Qiagen) pre-equilibrated with Buffer A. (*Note*: in many cases the enzymes used in this study could not be purified on an FPLC system, as they appeared to form a strongly bound complex at the top of the column that over-pressurized FPLC systems.) The bound protein was then washed with 10 column volumes (CVs) of Buffer A, and protein was eluted with a mixture of 50% Buffer A and 50% 25 mM potassium phosphate, 100 mM NaCl, and 500 mM imidazole, pH 8.0 (Buffer B) and collected. An additional 1 mM PLP was added to the collected protein solution to ensure full cofactor incorporation, and the solution was then buffer exchanged into 50 mM potassium phosphate, pH 8.0 (KPi; *Note:* "KPi" used in this text always refers to 50 mM, pH 8.0) by dialysis. The purified protein was then flash frozen in ~20-μL drops in liquid nitrogen and stored at −80 °C. Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (ThermoFisher) according to the manufacturer's recommendations.

Alternatively, protein catalyst could be prepared as a heat-treated lysate and used directly for preparative-scale reactions, obviating the need for any chromatography. In these instances, thawed cell pellets were resuspended in a volume of KPi containing 100 μM PLP that was appropriate for the given reaction, usually 50–100 mL. The dilute resuspension was then heat treated at 75 °C for >1 hour to efficiently lyse the cells and denature the *E. coli* proteins, and then clarified by centrifugation at 14,000*g* for 20 minutes.

For plate-based expression, single colonies of a desired variant or library were transferred into the wells of a 96-well plate containing 300 μL LB_{carb}. The cultures were covered with a sterile, breathable film and grown to stationary phase at 37 °C and 220 rpm. From these plates, 20 μL of each culture were transferred to new 96-well plates containing 630 μL of TB_{carb} and grown for 6 hours at 37 °C and 220 rpm. Protein expression was induced with the addition of 50 μL of 14 mM IPTG in TB_{carb} for a final concentration of 1 mM IPTG in a total volume of 700 μL of TB_{carb}. Expression proceeded for 24 hours at 30 °C, 250 rpm, at which point cells were harvested by centrifugation at 4,500*g* for 10 minutes, discarding the supernatant, and, unless preparing lysate immediately, covered with a non-breathable film and frozen at −20 °C until needed. Heat-treated lysate was prepared by resuspending thawed cells in 300 μL of KPi supplemented with 100 μM PLP, heat treating for 15–90 minutes (depending on the desired thermostability challenge) at 75 °C, and clarifying by centrifugation at 4,500*g* for 10 minutes. Lysate was then added directly to preprepared reaction plates as described below. Because the concentration of enzyme is not known or measured in heat-treated lysate, improvements in expression, stability, or other factors are allowed to manifest as improvements in the catalyst.

Absorbance-based screening.

Enzyme variants were tested for activity by combining heat-treated lysate (prepared as described above) with Ser and an appropriate nucleophile (e.g., indole or a phenol analog),

along with 5% EtOH by volume as cosolvent, directly into a UV-transparent assay plate and measuring the change in absorbance over time at a given wavelength that has been validated to increase over the course of the reaction. For example, alkylation of the indole chromophore red-shifts its absorbance spectrum as it is converted to Trp, and the increase in absorbance can be correlated to Trp formation at 290 nm[54]. Likewise, the 1-naphthol chromophore red-shifts its absorbance spectrum as it is converted to the ncAA product, resulting in an increase between 284 nm and ~350 nm (Extended Data Fig. 7a). Changes in activity for evolution were quantified by either by endpoint measurement or initial rate determination.

LCMS screening.

The enzyme activities and absorbances of 2-chlorophenol and phenol were not suitable for absorbance-based screening in lysate, and therefore were screened *via* LCMS. In this case, heat-treated lysate was combined with Ser and the phenol analog, along with 5% EtOH by volume as cosolvent, and allowed to react overnight (typically ~18 hours) at 37 °C. The reactions were worked up with 300 μL of 1:1 1 M aq. HCl/MeCN and then filtered through a 0.2-μm 96-well filter plate (Pall AcroPrep #8019) *via* centrifugation at 5,000*g* until the soluble reaction components were collected in the wells of a 96-well assay plate. This plate was then sealed and run on a suitable LCMS method and column that can separate reaction components sufficiently for quantification by UV and/or MS. For medium-throughput screening, a C18 guard column provided sufficient separation of substrate and product for reliable screening of enzyme activity over a 0.65-minute method with 1.2 minutes between injections (2 mL/min flow rate; 0.00 min: 1% MeCN; 0.01 min: 95% MeCN; 0.26 min: 1% MeCN; hold to 0.65 min; post-time: 0.25 min).

Analytical-scale vial reactions.

Analytical reactions were performed in 2-mL glass vials in a total reaction volume of 200 μL. Vials were first charged with 10 μL of 20X stock of the nucleophile in EtOH (final concentration of 5% EtOH by volume), to which 190 μL of a mixture of Ser and purified enzyme in KPi were added. The reactions were generally protected from light (primarily 1-naphthol, which is a photoacid) and allowed to react at 37 °C. At the end of the reaction time, the 200-μL reactions were worked up with 800 μL of 1:1 1 M aq. HCl/MeCN, transferred to a microcentrifuge tube, and clarified by centrifugation at 14,000*g*. A 200-μL aliquot of this mixture was then collected and analyzed *via* LCMS as described in the general experimental methods.

Determination of enantiopurity.

Enantiopurity was assessed using Nα-(5-fluoro-2,4-dinitrophenyl)-L-alaninamide (Marfey's reagent), a common chiral derivatization agent[55]. To assess the enantiopurity of an enzymatic reaction without purification, an analytical-scale reaction was first carried out as described above. After the reaction time, 100 μL of 1 M aq. $NaHCO_3$ were added. Alternatively, for purified product, a ~10 mM solution in KPi was prepared which was then mixed 2:1 with 1 M aq. $NaHCO_3$. From these solutions, 125 μL were transferred into a new 2-mL vial along with 33 μL of a 33-mM solution of Marfey's reagent in acetone. This

mixture was incubated at 37 °C and 220 rpm for 2 hours, and then the mixture was diluted with 600 μL 1:1 1 M aq HCl/MeCN.

The products of the reaction were analyzed by LCMS using the following gradient of MeCN: 25–45% over 7 minutes. Products were monitored by MS in single-ion mode selected for the expected molecular ion of the S_NAr product (e.g., 434 m/z for the Tyr product). This provided baseline separation of DL-tyrosine peaks (Extended Data Fig. 4a). Putative O-alkylation product of the Tyr and Marfey's reagent, with the same mass as the desired N-alkylation product, was seen as an early single peak. A similar O-alkylation product of unreacted phenol and Marfey's reagent was also seen in enzymatic reactions with leftover phenolic substrate. In all cases, enzymatic products derivatized with Marfey's reagent were identified to have only a single peak corresponding to the L-enantiomer.

To confirm enantiopurity in the absence of accessible D-enantiomers of these compounds, the D-enantiomer of Marfey's reagent was used to prepare a racemic mixture of this reagent. Enzymatic Tyr products were derivatized with racemic and enantiopure Marfey's reagent mixtures which resulted in two or one peaks, respectively, and confirmed the "L" configuration as the only observable enzymatic product (see Supplementary Methods, Gram-scale syntheses of Tyr analogs for LCMS traces).

Phylogenetic analysis of TrpB-like sequences.

All human-annotated TrpB sequences from the SwissProt database were obtained and aligned to obtain a multiple sequence alignment (MSA) referenced to the sequence of Tm9D8*. The catalytic glutamate was conserved at position 105 in all 451 sequences. To probe this deeper, we obtained a multiple sequence alignment (MSA) of 18,719 TrpB-like sequences from the EVcouplings software[56]. We discarded 693 variants that did not contain an appropriately positioned catalytic lysine (K83) or had an insertion at position 105 in the MSA (i.e., from an improperly aligned and/or non-TrpB-like sequence), leaving 18,051 aligned sequences. The identities of the amino acid at position 105 and 229 were obtained and compared (Extended Data Figs. 6a–6c). To construct a phylogenetic tree, this dataset was sub-sampled to afford computational tractability. First, all E105G/A sequences were retained, as well as those from *A. thaliana, S. albulus, E. coli*, and *T. maritima* (select paralogs), and 100 randomly sampled sequences (i.e., native TrpB and TrpB2s). Tree construction was performed using IQ-TREE[57] (version 1.6.12) with ModelFinder[58]. The best-performing model (LG+R7) was then used to analyze a larger dataset comprised of 1,000 randomly sampled sequences (1,158 sequences total). The tree was visualized and annotated using the interactive Tree of Life (iTOL) from EMBL[59]. Two additional trees were also constructed: one tree using all paralogous E105-containing sequences (~500 sequences), as well as those from sub-strains of *E. coli, Saccharomyces cerevisiae*, and *T. maritima*, and 600 additional randomly sampled E105-containing sequences; and one tree using all TrpB-like 18,719 sequences. Both trees supported the original conclusions on the unique nature of the E105A- and E105G-based clades.

Crystallization, X-ray data collection, and structure determination.

For the crystallization of TmTyrS1, protein was purified as described above. For initial screening, protein was thawed from −80 °C to room temperature and diluted to 10 and 20 mg/mL in KPi. Using a Crystal Gryphon robot (Art Robbins Instruments), sparse matrix screening was performed using the Wizard HT 1 & 2 (Rigaku), JCSG+ (Molecular Dimensions), Index and PEGRx (Hampton Research) crystallization screens in Intelli-Plate 96-2 drop crystallization plates (Art Robbins Instruments) using 0.2 μL drops of precipitant followed by 0.2 μL of protein solution. Plates were sealed with transparent adhesive covers and incubated at room temperature. After 2 days, crystals were observed in well C3 of the Wizard Screen (1.2 M NaH_2PO_4/0.8 M K_2HPO_4, 0.1 M *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS), 0.2 M Li_2SO_4), which served as the precipitant for all crystallization presented here.

These crystals were then optimized by drop ratio variation in 24-well CrysChem M Plates (Hampton Research) using 1–6 μL protein drops and 2–5 μL precipitant drops. Yellow crystals with an atypical morphology (Extended Data Fig. 7a) appeared in all wells after 1–3 days, with larger crystals generally observed at higher protein and lower precipitant concentrations. Conditions for cryoprotection and small-molecule soaking experiments are provided in Supplementary Methods, Crystal cryoprotection and small-molecule soaking experiments.

Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 12-2. Data reduction and integration were carried out using XDS[60] and scaled using Aimless in the CCP4 suite of programs[61]. For the structure of holo TmTyrS1, molecular replacement (MR) was performed using the structure of a holo TrpB from *Pyrococcus furiosus* (*Pf*TrpB; PDB 5DVZ) as a search model in Phaser[62]. For all other structures, the protein chain of holo TmTyrS1 was used for MR. Model building and modification in the electron density was performed using Coot[63] and structure refinement was performed using Phenix[64]. Other ligands, specifically 4-HQ and quinoline *N*-oxide, as well as water molecules and ethylene glycol, were added during later stages of refinement. Occasionally, spurious electron density peaks were present in the active site, dimer interface, and COMM domain that could not be unambiguously modeled by alternative protein conformations, solvent, or other additives applied during the procedure, so these were left uninterpreted. The quality of the final models was evaluated with MolProbity[65] and PROCHECK[66]. Data collection and refinement statistics are presented in Supplementary Table 2.

Determination of kinetic isotope effects.

Kinetic isotope effects (KIEs) were measured as analytical-scale vial reactions with a few minor changes. The cosolvent used was DMSO rather than EtOH, because stocks of the deuterated phenols were previously prepared in commercial DMSO-d_6 for long-term storage. KIEs were measured in direct competition with equimolar concentrations of the standard and deuterated substrate at a total concentration of 5 mM phenolic substrate and 50 mM Ser. Reactions were analyzed by LCMS, extracting the ions corresponding to the appropriate product masses and comparing their ratios. Ser KIEs were measured with Ser-d_3

using 2-chlorophenol or 1-naphthol as the phenolic substrate. (*Note*: the +2 isotope of chlorine yielded a 30% relative abundance of a 218 m/z product for 3-chloro-Tyr, which also corresponded with the di-deuterated product of Ser-d_3 and 2-chlorophenol. Reported ratios of 216/218 ion counts are therefore not exactly the 3-chloro-Tyr 216 m/z ion over the deuterated 3-chloro-Tyr 218 m/z ion, which would give a value of slightly lower than 1 for a true KIE of 1.) KIEs are provided in Supplementary Table 3.

Computational modeling.

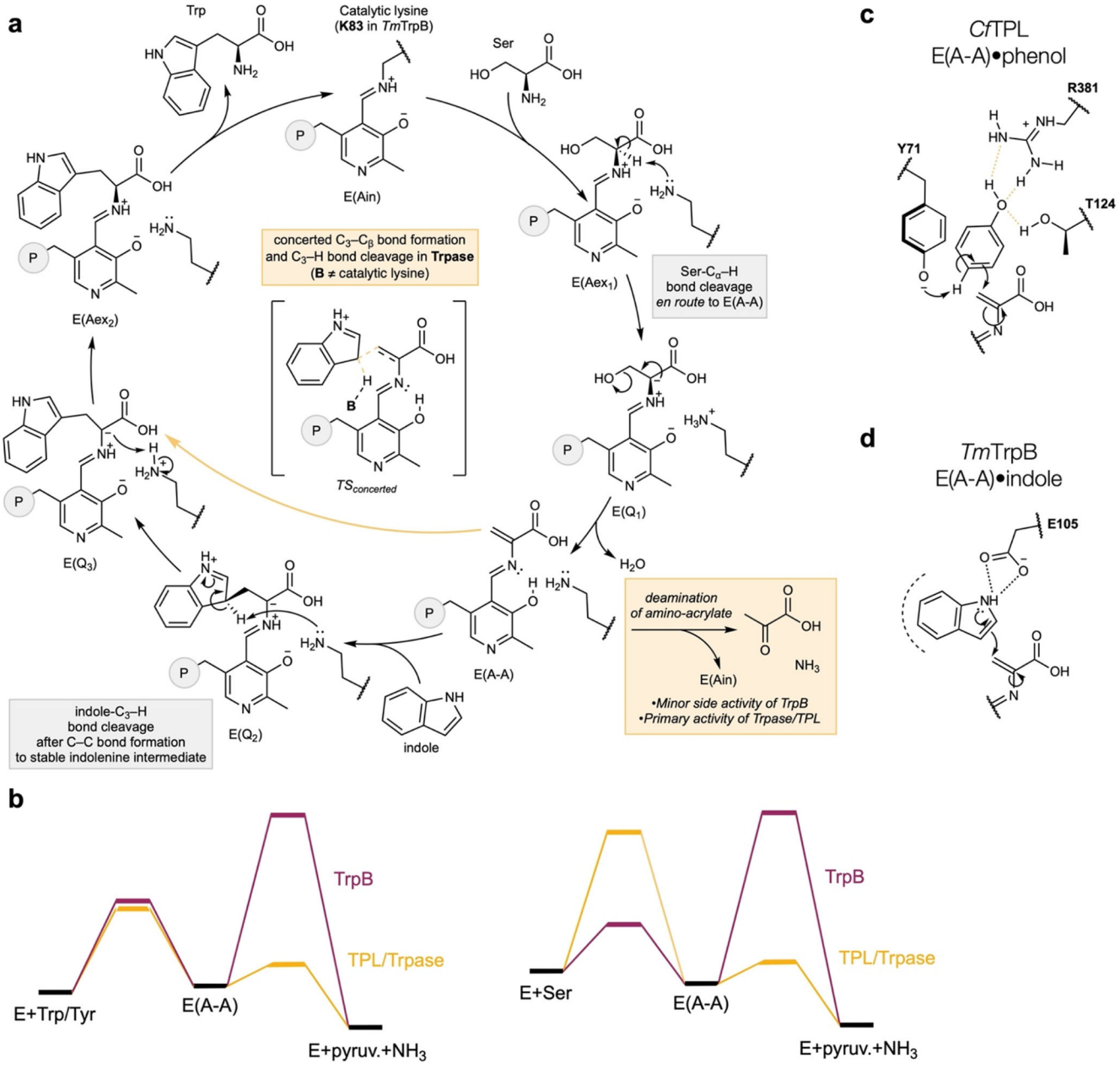
Computational models were calculated using the ORCA quantum chemistry package[67] (version 5.0.4). Initial input structures were generated using the X-ray crystal structure of the enzyme with bound 4-hydroxyquinoline (4-HQ; PDB 8EH1). The 4-HQ was replaced with the appropriate substrates, phenol and indole for Tyr and Trp synthesis, respectively, to study the proposed reaction mechanisms. All QM calculations were conducted at the DFT level using the B3LYP functional with D3 dispersion corrections and the augmented ma-def2-SVP basis set[68–72]. Calculations were also reproduced using the ωB97X functional with similar conclusions[71]. To examine the stability of proposed intermediates, models were constructed at two levels of complexity. The minimal model was defined to include only the substrate, PLP-bound amino-acrylate, catalytic lysine, and active-site water hypothesized to be important for the observed catalysis. The highly charged phosphate group of the PLP cofactor was modeled as a neutral methyl group to reduce coulombic interactions that are otherwise mitigated by the active site. The solvent-like effect of the enzyme was approximated using a conductor-like implicit solvent model with a dielectric constant of 7.25[73]. The step-wise intermediate structure with phenol was further examined by implementing a QM/MM scheme, where a larger sphere of coordinating residues near the active site were modeled at the molecular mechanics level using the OpenMM package interfaced with ORCA through the open-source package ASH (https://github.com/RagnarB83/ash,version0.9) using the CHARMM36 and CGenFF force fields[74–76]. The positions of the point charges of the MM level were included in the QM calculation using the point charge module in the ORCA package. Fixed hydrogen link-atoms were used where single bonds were broken in the QM embedding of the system. All atoms within 10 Å of the ligands were allowed to relax in the QM/MM model, and solvent molecules were modeled explicitly. Due to agreement between the minimal QM model and QM/MM model on the unfavorability of a step-wise pathway, we employed the more efficient minimal QM model to calculate the structure of the concerted transition state for phenol alkylation. All B3LYP- and ωB97X-derived structures are provided as Supplementary Data, with the B3LYP-derived structures used for the figures.

Alignment of experimental and transition state structures.

The calculated transition state structure was superimposed to the co-crystal structure of TmTyrS1 in the E(A-A) form with 4-HQ bound in the active site at the following atoms: the active-site water oxygen, the substrate phenolic oxygen, the atom *para* to the substrate hydroxyl (nitrogen for 4-HQ, carbon for phenol), the terminal nitrogen of the catalytic lysine, and the six atoms of the amino-acrylate (nitrogen, C_α, C_β, and the carboxylic acid). The crystal structures of TmTyrS1 E(A-A) with and without 4-HQ were aligned using the “match” command in ChimeraX[77] (version 1.16), comparing each subunit within

the asymmetric unit independently. Active-site residue RMSDs were determined using the "align" command specifying the atoms of all residues within 5 Å of the bound 4-HQ in the aligned structures (140 atom pairs matched by atom name and number; alignment shown in Extended Data Fig. 7i). RMSDs are provided in Supplementary Table 4.

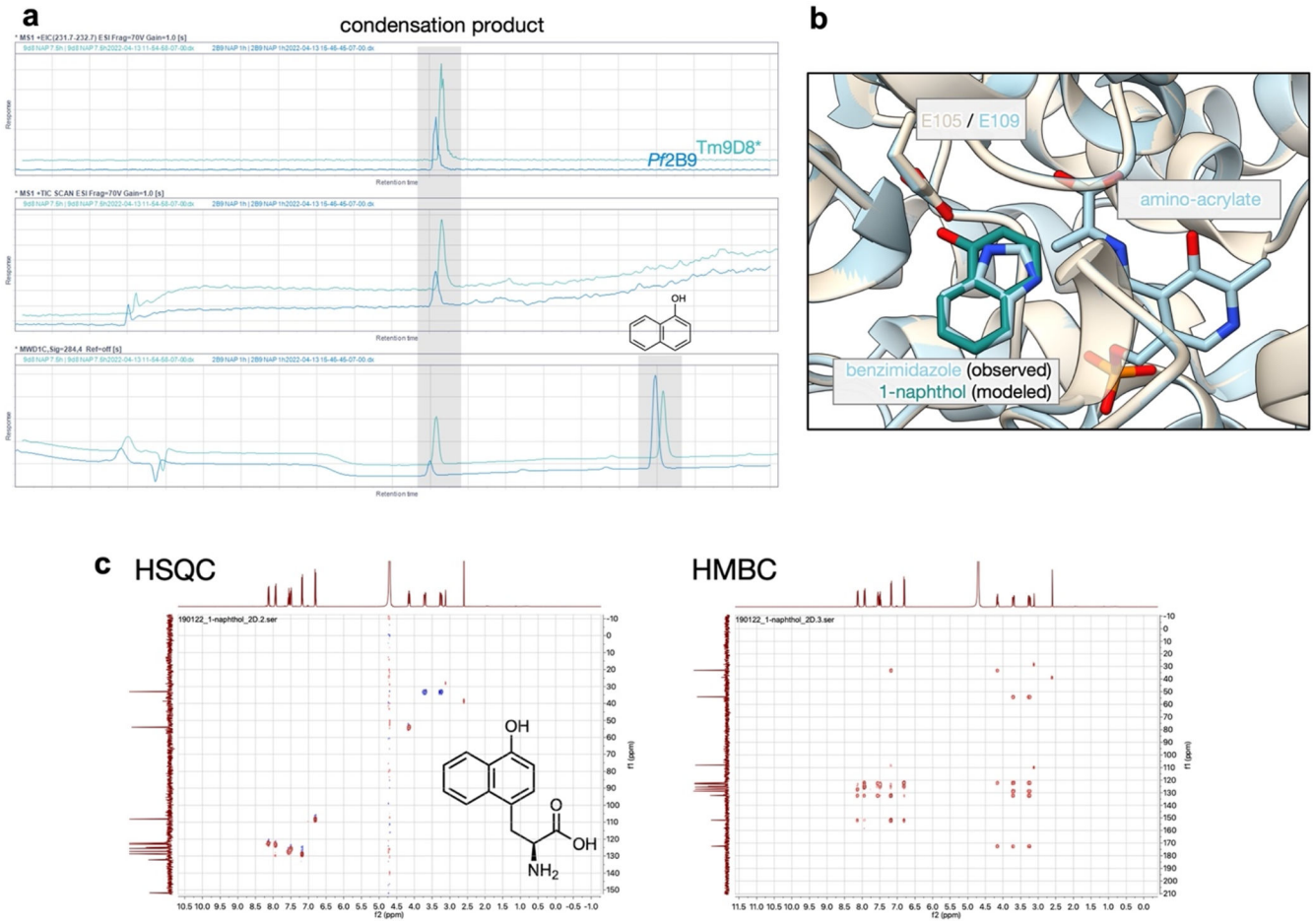
Extended Data



Extended Data Figure 1. Mechanistic details of TrpB and Trpase/TPL.

a. Catalytic cycle of *Thermotoga maritima* TrpB (*Tm*TrpB, black) with details of Trpase shown in yellow. Reaction arrows are shown as unidirectional in the synthetic direction for conceptual simplicity. **b.** Hypothetical reaction coordinates for β-elimination reactions

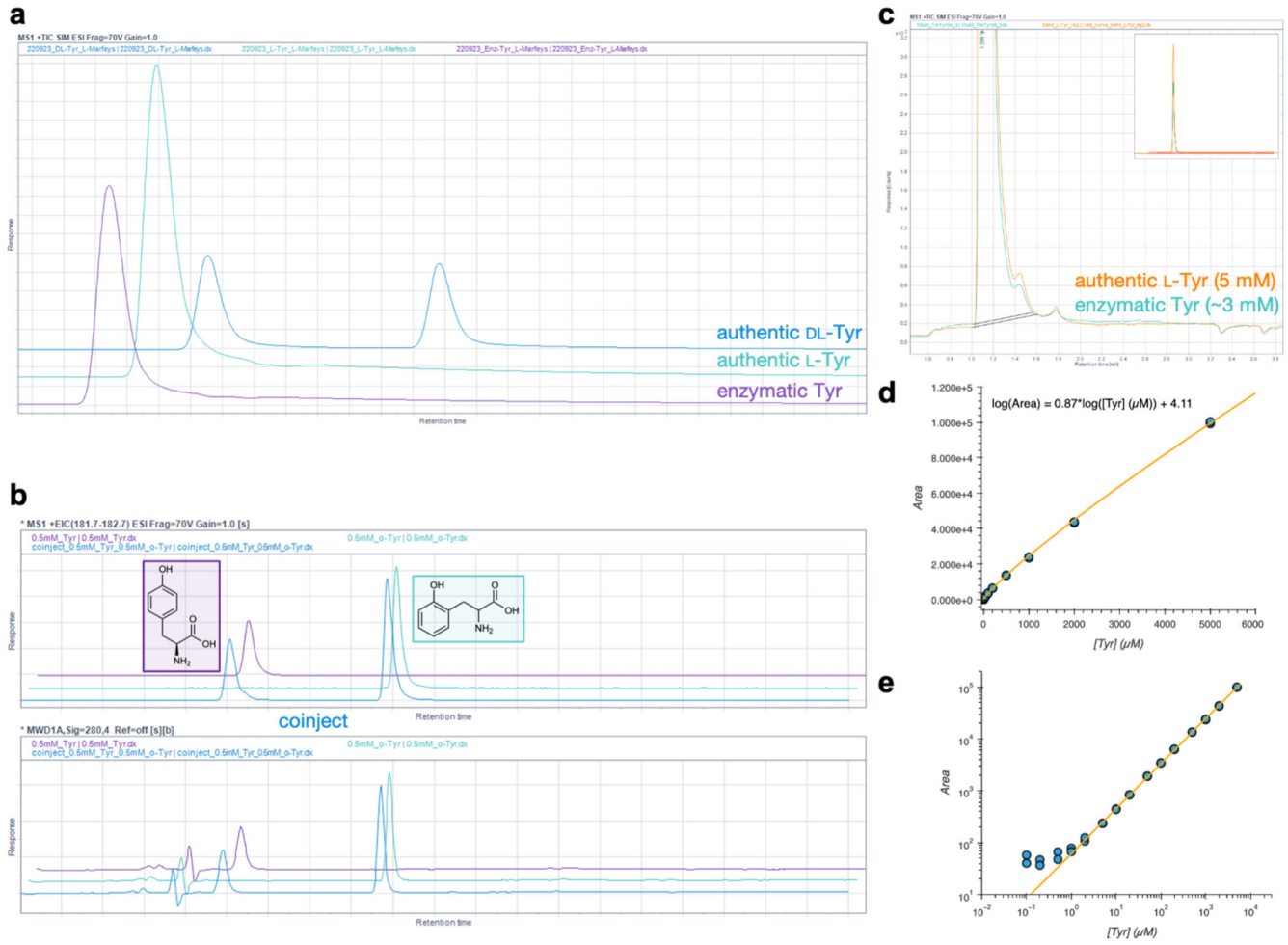
that access and degrade the enzyme-bound amino-acrylate intermediate (E(A-A), shown as the same energy on the enzymes) by TrpB (purple) and TPL/Trpase (yellow). *Left:* A comparable rate of L-tryptophan (Trp)/L-tyrosine (Tyr) β-elimination by an enzyme (E) will show accumulation of E(A-A) in TrpB and transience in TPL/Trpase if the barriers for conversion to pyruvate and ammonia are sufficiently different. Experimentally, a different intermediate accumulates in TrpB that occurs prior to E(A-A) formation, generally E(Q_3) or E(Q_2). *Right:* TrpB accesses E(A-A) from L-serine (Ser) and the E(A-A) species accumulates in the absence of a nucleophile. In the presence of indole, the left half of each reaction coordinate is connected, providing a direct pathway from Ser to Trp with minimal degradation of the amino-acrylate, despite the thermodynamic driving force of pyruvate and ammonia formation (see main text). **c.** Active-site coordination and activation of phenol by *Citrobacter freundii* tyrosine phenol lyase (*Cf*TPL) for reversible alkylation (and Tyr degradation). **d.** Active-site coordination and activation of indole for irreversible alkylation by *Tm*TrpB.



Extended Data Figure 2. Identification and characterization of the ncAA condensation product of TrpB and 1-naphthol.

a. Stand-alone variants of *Tm*TrpB (Tm9D8*, teal) and *Pyrococcus furiosus* TrpB (*Pf*2B9, blue) synthesize a condensation product when provided 1-naphthol and Ser. **b.** Alignment

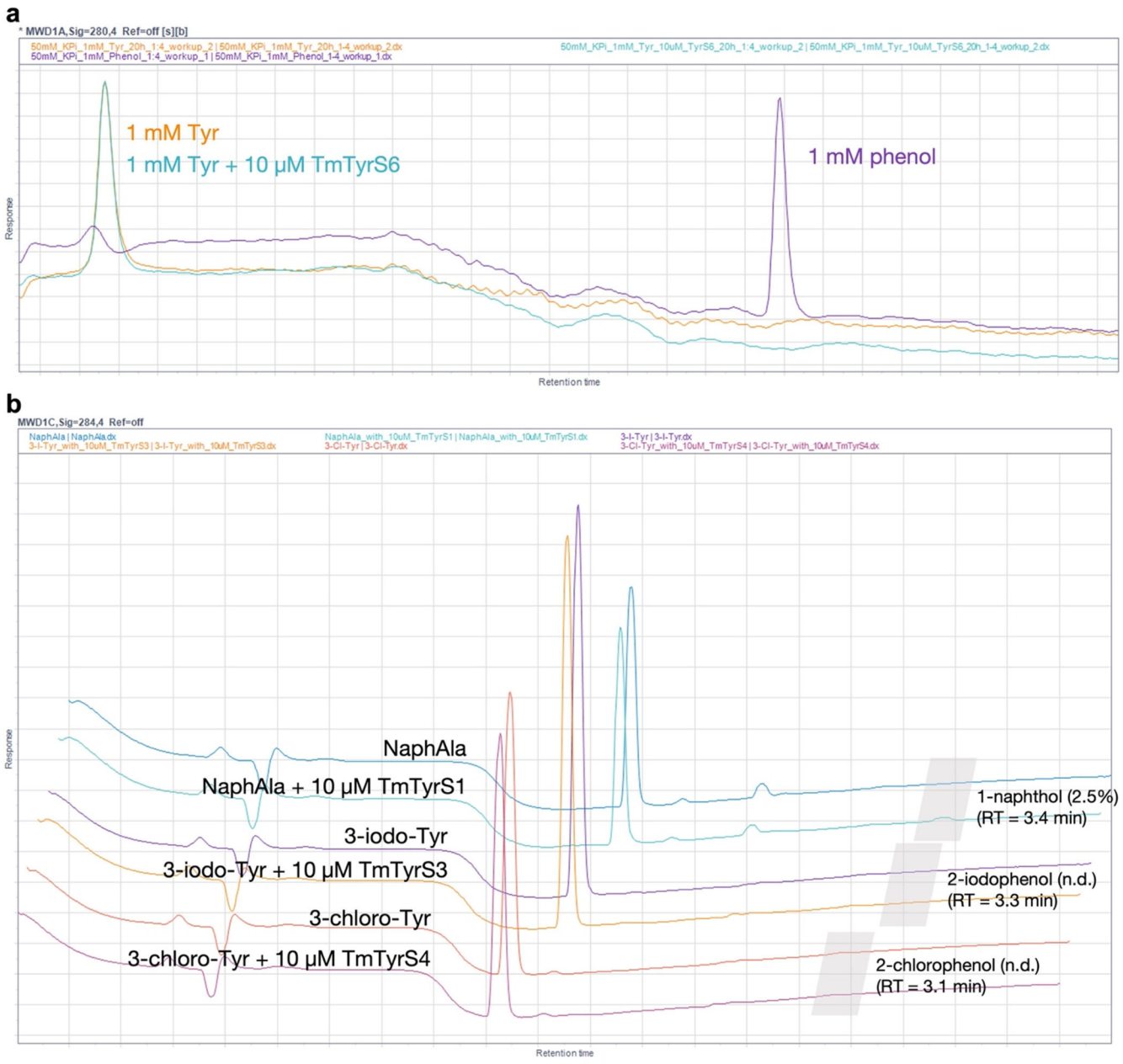
of a co-crystal structure of *Salmonella typhimurium* TrpS (*St*TrpS, PDB 4HPX) bound with the non-reactive inhibitor benzimidazole and a homology model of Tm9D8*. Pairwise alignment of 1-naphthol to benzimidazole aligned for *para* C–C bond formation demonstrates that the conserved catalytic residue E105 likely is not optimal for the extended phenolic group of 1-naphthol and other phenol analogs. **c.** Preparative-scale synthesis and 2D NMR characterization of the 1-naphthol product of Tm9D8* E105G reveals that the *para* alkylation product NaphAla is the only reaction product. HSQC = heteronuclear single-quantum correlation; HMBC = heteronuclear multiple bond correlation.



Extended Data Figure 3. Characterization of tyrosine synthase activity by enzymes engineered from TrpB.

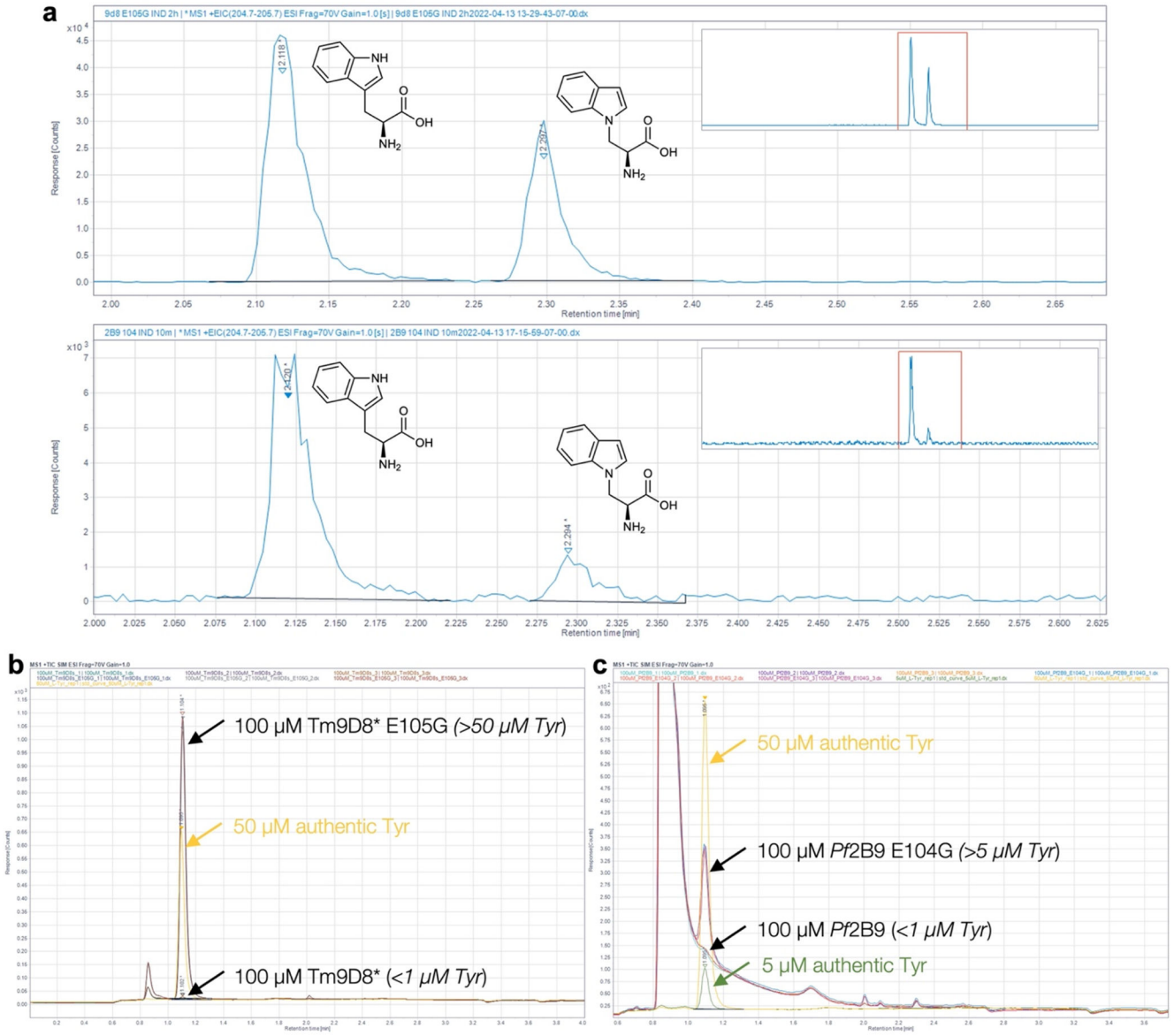
a. Catalytic cycle of *Thermotoga maritima* TrpB (*Tm*TrpB, black) with details of Trpase shown in yellow. Reaction arrows are shown as unidirectional in the synthetic direction for conceptual simplicity. **b.** Hypothetical reaction coordinates for β-elimination reactions that access and degrade the enzyme-bound amino-acrylate intermediate (E(A-A), shown as the same energy on the enzymes) by TrpB (purple) and TPL/Trpase (yellow). *Left:* A comparable rate of L-tryptophan (Trp)/L-tyrosine (Tyr) β-elimination by an enzyme (E) will show accumulation of E(A-A) in TrpB and transience in TPL/Trpase if the barriers for

conversion to pyruvate and ammonia are sufficiently different. Experimentally, a different intermediate accumulates in TrpB that occurs prior to E(A-A) formation, generally E(Q_3) or E(Q_2). *Right:* TrpB accesses E(A-A) from L-serine (Ser) and the E(A-A) species accumulates in the absence of a nucleophile. In the presence of indole, the left half of each reaction coordinate is connected, providing a direct pathway from Ser to Trp with minimal degradation of the amino-acrylate, despite the thermodynamic driving force of pyruvate and ammonia formation (see main text). **c.** Active-site coordination and activation of phenol by *Citrobacter freundii* tyrosine phenol lyase (CfTPL) for reversible alkylation (and Tyr degradation). **d.** Active-site coordination and activation of indole for irreversible alkylation by *Tm*TrpB.



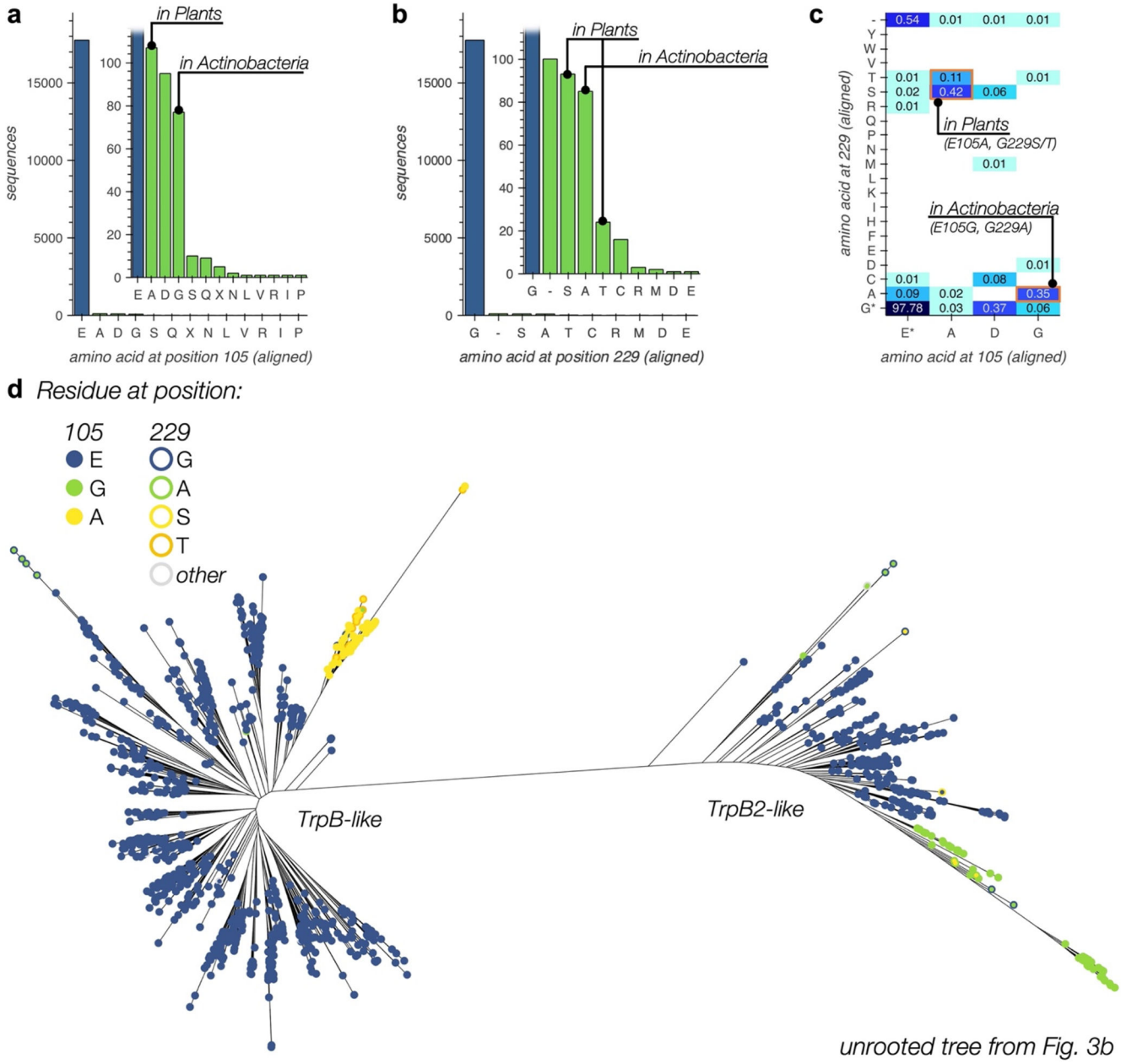
Extended Data Figure 4. TyrS variants remain kinetically controlled.

a. A 1 mM solution of authentic Tyr (approximately saturating) was incubated with and without 10 µM TmTyrS6 for 20 hours, showing no significant degradation. (1 mM phenol shown for reference.) **b.** Like in **a.**, saturated solutions of authentic NaphAla, 3-iodo-Tyr, and 3-chloro-Tyr were incubated with 10 µM of a TyrS variant that displayed high activity toward the respective phenol analog. Under thermodynamic control, as with Trpase and TPL, this enzyme would efficiently degrade the respective Tyr analog. Approximate areas in which the phenol analogs elute are designated in gray boxes. Only NaphAla shows very minor degradation at a rate far less than its rate of NaphAla synthesis, indicating that the reaction of TyrS remains under kinetic control for Tyr analogs as well.



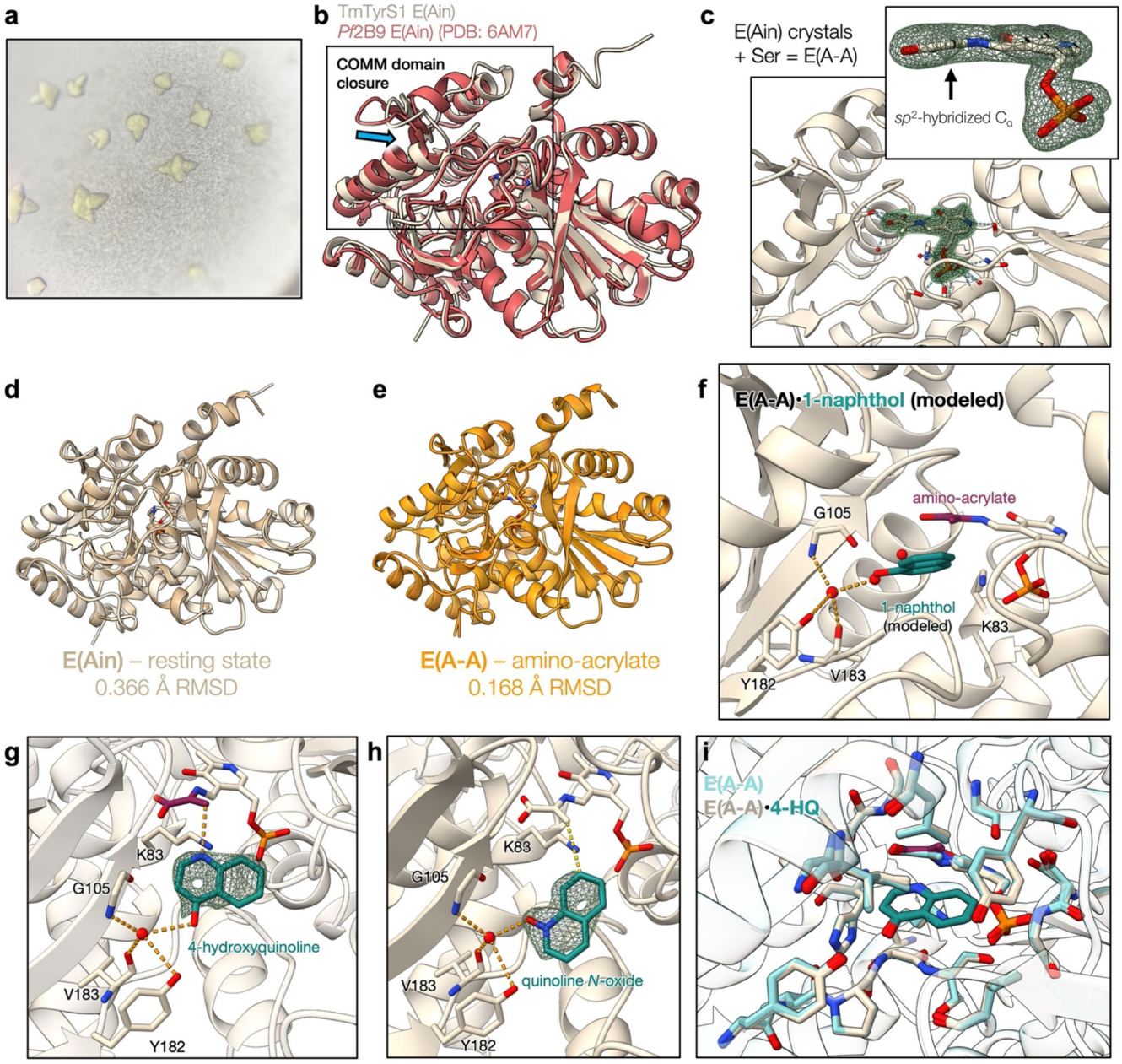
Extended Data Figure 5. Effect of E105G on Trp and Tyr synthesis.

a. In addition to reducing the rate of Trp synthesis, the E105G mutation also reduces regioselectivity of indole alkylation from >99.5%. *Upper*: Tm9D8* E105G with indole forms nearly 1:1 Trp to isoTrp, the N-alkylation product. *Lower*: *Pf*2B9 E104G retains some regioselectivity, but still markedly decreased from its previously unquantifiable levels. **b.** 100 μM Tm9D8* does not make Tyr above the limit of quantification (1 μM) in 24 hours, but the equivalent amount of Tm9D8* E105G makes >50 μM Tyr. **c.** The same analysis for *Pf*2B9 E104G, with the addition of a 5 μM standard of Tyr for reference, as *Pf*2B9 E104G at 75 °C is slower than Tm9D8* E105G at 37 °C.



Extended Data Figure 6. Phylogenetic analysis of TrpB sequences.

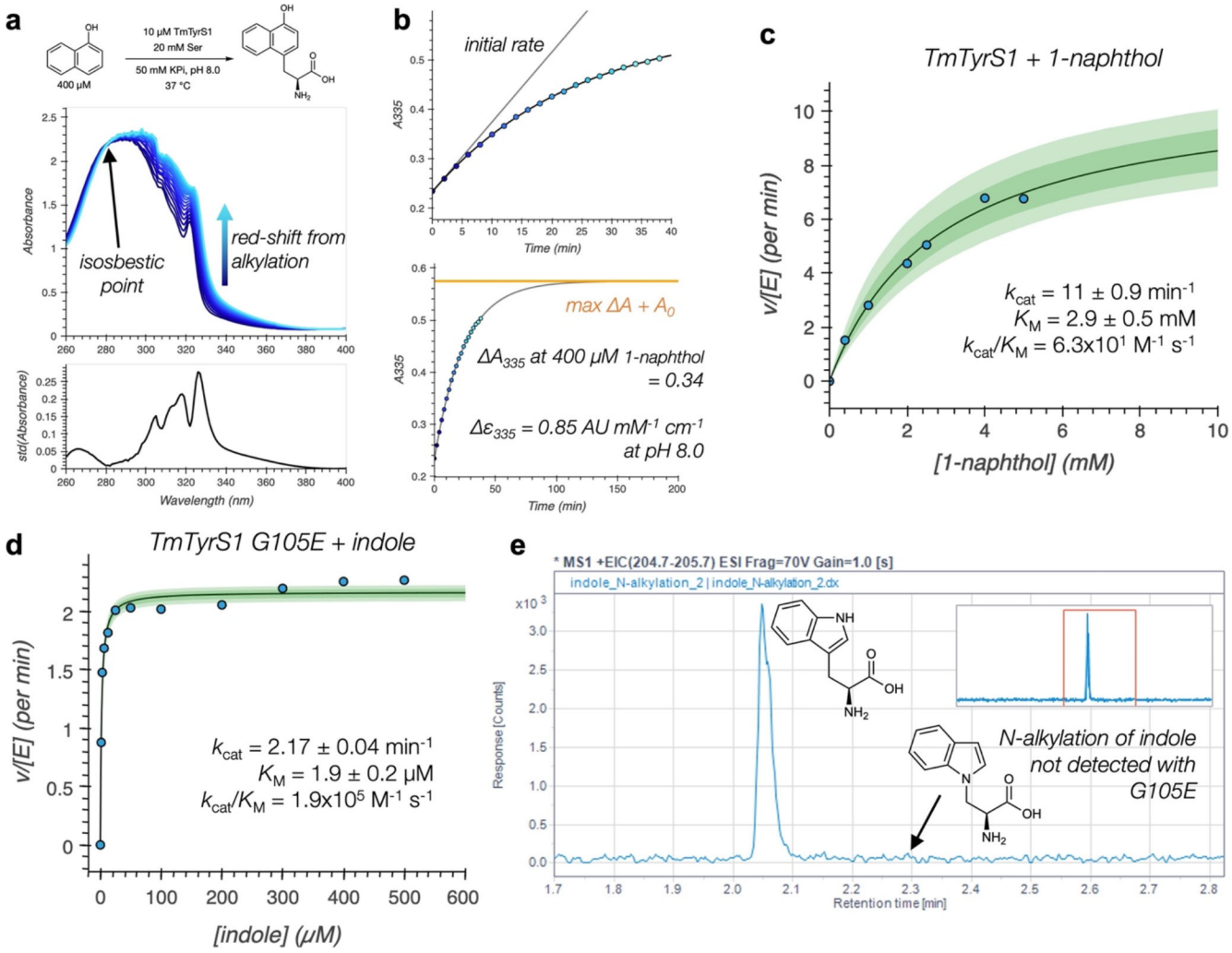
a. Conservation of E105 in 18,051 aligned TrpB-like sequences. *Inset*: axis-adjusted view. **b.** Conservation of G229 in 18,051 aligned TrpB-like sequences. *Inset*: axis-adjusted view. **c.** Pairwise correlations between mutations at E105 and G229. Those with E105G and G229A are TrpB2-like enzymes found primarily in *Actinobacteria*. Those with E105A and G229S/T are TrpB-like enzymes found in plants, primarily in malvids. **d.** Unrooted phylogenetic tree from Fig. 3b, built from 1,158 TrpB- and TrpB2-like sequences (all those with E105G/A, selected paralogs, and 1,000 randomly sampled sequences). Nodes are colored by residue identity at position 105 (fill) and 229 (outline).



Extended Data Figure 7. X-ray crystallographic analysis of TyrS variants.

a. TmTyrS1 crystals. The morphology was fairly amorphous but yielded excellent diffraction and were amenable to small-molecule soaking experiments. **b.** Comparison of the internal aldimine E(Ain) forms of TmTyrS1 and another stand-alone TrpB (*Pf*2B9). The TmTyrS1 structure adopts a more "closed" conformation with a disrupted COMM domain helix, but it is unclear if these are significant for TyrS, *Tm*TrpB/Tm9D8*, or at all. **c.** Polder omit map contoured at 5σ for the amino-acrylate, formed by soaking E(Ain) crystals with Ser. The COMM domain helix is re-formed. This reduces the conformational heterogeneity between the two TrpB subunits in the asymmetric unit between the E(Ain) state **(d)** and the Ser-soaked E(A-A) state **(e)**. **f.** The removal of E105 allows for the binding of new active-

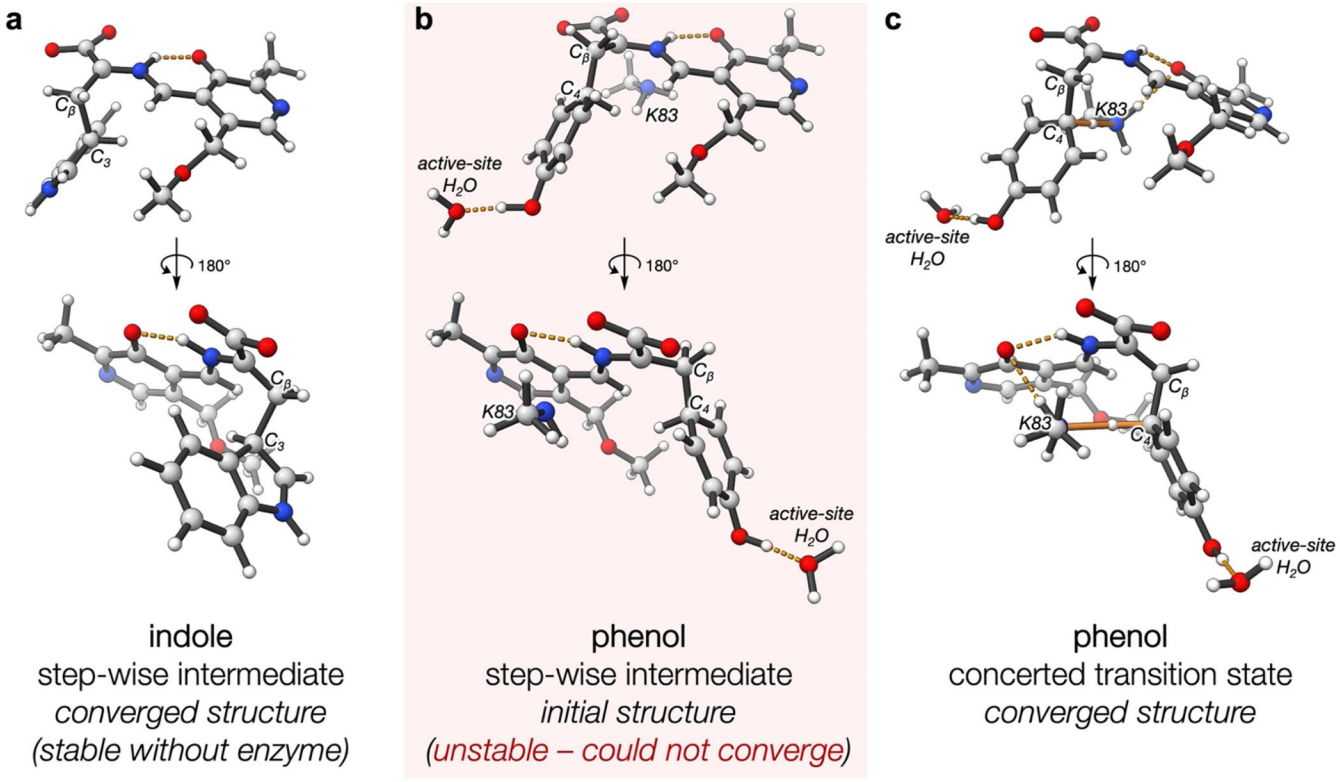
site water molecule, coordinated by G105, V183, and the sidechain of Y182. It additionally makes an interaction with a separate water molecule. Placement of 1-naphthol into this structure (as performed in Extended Data Fig. 2b) overlays the hydroxyl of 1-naphthol with the second water molecule, which would align 1-naphthol for C–C bond formation. **g.** Polder omit map countoured at 7σ for TmTyrS1 E(A-A) bound with 4-hydroxyquinoline (4-HQ). **h.** Polder omit map countoured at 7.6σ for TmTyrS1 E(A-A) bound with quinoline *N*-oxide. **i.** Alignment of each subunit in the E(A-A) state and E(A-A)•4-HQ state (4 separate structures), showing minimal reorganization when 4-HQ binds. See Supplementary Table 4 for pairwise RMSDs.



Extended Data Figure 8. TmTyrS1 kinetics and regioselectivity.

a. The 1-naphthol chromaphore is red-shifted as it is converted to NaphAla, with an isosbestic point near 284 nm, and a maximal difference near 330 nm. **b.** The absorbance change at 335 nm can be used to monitor reaction progress (*upper*) and to determine the change in molar absorptivity as 1-naphthol is converted to NaphAla at pH 8.0 (*lower*, 0.85 AU mM^{-1} cm^{-1}). **c.** Michaelis-Menten analysis of TmTyrS1 and 1-naphthol. The

k_{cat} value agrees with those determined by LCMS at saturating conditions (TOFs in Fig. 3a). Line represents the best-fit parameters, while colored areas represent fits determined from the 75 and 95% credible regions of k_{cat} and K_M. **d.** Michaelis-Menten analysis of TmTyrS1 G105E. While the k_{cat} is ~10–30-fold lower than other stand-alone TrpBs and 5-fold lower than TmTyrS1 with 1-naphthol, its K_M is quite low (nearly 2,000-fold lower than TmTyrS1 with 1-naphthol). Line represents the best-fit parameters, while colored areas represent fits determined from the 75 and 95% credible regions of k_{cat} and K_M. **e.** While all TyrS variants (and Tm9D8* E105G/*Pf*2B9 E104G) lose regioselectivity for indole alkylation (see Extended Data Fig. 5a and Fig. 3a Source Data), replacing the catalytic glutamate in TmTyrS1 (TmTyrS1 G105E) restores regioselectivity for indole alkylation.



Extended Data Figure 9. Computational modeling of intermediates and transition states.

a. Quantum mechanics (QM)-calculated structure of the step-wise intermediate for indole alkylation, after C_3–C_β bond formation but before deprotonation (C_3–H bond breakage). This structure readily converges, even in the absence of enzyme-mediated interactions. **b.** Initial structure of the step-wise intermediate for phenol alkylation, after C_4–C_β bond formation but before deprotonation (C_4–H bond breakage) for QM optimization. This did not converge to a stable structure, regardless of interactions modeled. **c.** QM-calculated transition-state structure for concerted C_4–C_β bond formation and C_4–H bond breakage (solid orange bonds) of phenol. This is mediated by the catalytic lysine (K83) and the active-site water installed with the E105G mutation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank many past and present members of the Arnold laboratory for helpful discussions on the manuscript and experimental assistance: Prof. Andrew Buller, Dr. Tina Boville, Dr. David Romney, Dr. Ella Watkins-Dulaney, Dr. Sabine Brinkmann-Chen, Dr. Nat Goldberg, among many others. We thank Siyuan Du for helpful discussions and Dr. Jens Kaiser from the Caltech Molecular Observatory for crystallography support. This work was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award no. R01GM125887. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Individual support comes from: Dr. Nagendranath Reddy Graduate Fellowship (P.J.A); Caltech Biotechnology Leadership Program (BLP; NIH Training Grant no. 5T32GM112592-5) (K.E.J.); Caltech AI4Science/Amazon AWS Fellowship (K.E.J.); Merck-Helen Hay Whitney Postdoctoral Fellowship (N.J.P.); Ruth L. Kirchstein NIH Postdoctoral Fellowship no. 1F32GM143797-01A1 (J.L.K.); National Science Foundation Graduate Research Fellowship (GRFP) Grant No. DGE-1745301 (V.C.B.); Fonds de recherche du Quebec Nature et Technologie (FRQNT) (j.D.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Data Availability

All data, including the 18,719 annotated TrpB-like sequences, are available in the main text or supplementary materials. Full protein crystallographic data have been deposited with the Protein Data Bank (PDB, https://www.rcsb.org) under accession codes 8EGY, 8EGZ, 8EH0, and 8EH1.

References

1. Lynch JH & Dudareva N Aromatic amino acids: A complex network ripe for future exploration. Trends Plant Sci. 25, 670–681 (2020). [PubMed: 32526172]
2. Rodriguez A et al. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. Microb. Cell Fact 13, 126 (2014). [PubMed: 25200799]
3. Lütke-Eversloh T, Santos CNS & Stephanopoulos G Perspectives of biotechnological production of L-tyrosine and its applications. Appl. Microbiol. Biotechnol 77, 751–762 (2007). [PubMed: 17968539]
4. Dodd D et al. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature 551, 648–652 (2017). [PubMed: 29168502]
5. Yoo H et al. An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. Nat. Commun 4, 2833 (2013). [PubMed: 24270997]
6. Merino E, Jensen RA & Yanofsky C Evolution of bacterial trp operons and their regulation. Curr. Opin. Microbiol 11, 78–86 (2008). [PubMed: 18374625]
7. Richards TA et al. Evolutionary origins of the eukaryotic shikimate pathway: Gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryot. Cell 5, 1517–1531 (2006). [PubMed: 16963634]
8. Porat I, Waters BW, Teng Q & Whitman WB Two biosynthetic pathways for aromatic amino acids in the archaeon *Methanococcus maripaludis*. J. Bacteriol 186, 4940–4950 (2004). [PubMed: 15262931]
9. Parmeggiani F, Weise NJ, Ahmed ST & Turner NJ Synthetic and therapeutic applications of ammonia-lyases and aminomutases. Chem. Rev 118, 73–118 (2018). [PubMed: 28497955]
10. Almhjell PJ, Boville CE & Arnold FH Engineering enzymes for noncanonical amino acid synthesis. Chem. Soc. Rev 47, 8980–8997 (2018). [PubMed: 30280154]
11. Nagasawa T et al. Syntheses of L-tyrosine-related amino acids by tyrosine phenol-lyase of Citrobacter intermedius. Eur. J. Biochem 117, 33–40 (1981). [PubMed: 7262088]

12. Phillips RS, Sundararaju B & Faleev NG Proton transfer and carbon–carbon bond cleavage in the elimination of indole catalyzed by *Escherichia coli* tryptophan indole-lyase. J. Am. Chem. Soc 122, 1008–1014 (2000).
13. Seyedsayamdost MR, Reece SY, Nocera DG & Stubbe J Mono-, di-, tri-, and tetra-substituted fluorotyrosines: New probes for enzymes that use tyrosyl radicals in catalysis. J. Am. Chem. Soc 128, 1569–1579 (2006). [PubMed: 16448128]
14. Romei MG, Lin C-Y, Mathews II & Boxer SG Electrostatic control of photoisomerization pathways in proteins. Science 367, 76–79 (2020). [PubMed: 31896714]
15. Beber ME et al. eQuilibrator 3.0: A database solution for thermodynamic constant estimation. Nucleic Acids Res. 50, D603–D609 (2022). [PubMed: 34850162]
16. Watkins-Dulaney E, Straathof S & Arnold F Tryptophan synthase: Biocatalyst extraordinaire. ChemBioChem 22, 5–16 (2021). [PubMed: 32677310]
17. Rix G et al. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. Nat. Commun 11, 5644 (2020). [PubMed: 33159067]
18. Smith DRM et al. The first one-pot synthesis of L-7-iodotryptophan from 7-iodoindole and serine, and an improved synthesis of other L-7-halotryptophans. Org. Lett 16, 2622–2625 (2014). [PubMed: 24805161]
19. Buller AR et al. Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. Proc. Natl. Acad. Sci. USA 112, 14599–14604 (2015). [PubMed: 26553994]
20. 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). [PubMed: 27355405]
21. Romney DK, Murciano-Calles J, Wehrmüller JE & Arnold FH Unlocking reactivity of TrpB: A general biocatalytic platform for synthesis of tryptophan analogues. J. Am. Chem. Soc 139, 10769–10776 (2017). [PubMed: 28708383]
22. Romney DK, Sarai NS & Arnold FH Nitroalkanes as versatile nucleophiles for enzymatic synthesis of noncanonical amino acids. ACS Catal. 9, 8726–8730 (2019). [PubMed: 33274115]
23. Dick M, Sarai NS, Martynowycz MW, Gonen T & Arnold FH Tailoring tryptophan synthase TrpB for selective quaternary carbon bond formation. J. Am. Chem. Soc 141, 19817–19822 (2019). [PubMed: 31747522]
24. Watkins EJ, Almhjell PJ & Arnold FH Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine. ChemBioChem 21, 80–83 (2020). [PubMed: 31513332]
25. Watkins-Dulaney EJ et al. Asymmetric alkylation of ketones catalyzed by engineered TrpB. Angew. Chem. Int. Ed 60, 21412–21417 (2021).
26. Saito Y, Sato T, Nomoto K & Tsuji H Identification of phenol- and *p*-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. FEMS Microbiol. Ecol 94, fiy125 (2018). [PubMed: 29982420]
27. Álvarez-Rodríguez ML et al. Degradation of vanillic acid and production of guaiacol by microorganisms isolated from cork samples. FEMS Microbiol. Lett 220, 49–55 (2003). [PubMed: 12644227]
28. Iwasaki Y et al. Novel metabolic pathway for salicylate biodegradation via phenol in yeast *Trichosporon moniliiforme*. Biodegradation 21, 557–564 (2010). [PubMed: 20020317]
29. Kuzuyama T, Noel JP & Richard SB Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. Nature 435, 983–987 (2005). [PubMed: 15959519]
30. Krastanov A, Alexieva Z & Yemendzhiev H Microbial degradation of phenol and phenolic derivatives. Eng. Life Sci 13, 76–87 (2013).
31. Milić D et al. Crystallographic snapshots of tyrosine phenol-lyase show that substrate strain plays a role in C–C bond cleavage. J. Am. Chem. Soc 133, 16468–16476 (2011). [PubMed: 21899319]
32. Chen Z & Zhao H Rapid creation of a novel protein function by *in vitro* coevolution. J. Mol. Biol 348, 1273–1282 (2005). [PubMed: 15854660]
33. Boville CE, Romney DK, Almhjell PJ, Sieben M & Arnold FH Improved synthesis of 4-cyanotryptophan and other tryptophan analogues in aqueous solvent using variants of TrpB from *Thermotoga maritima*. J. Org. Chem 83, 7447–7452 (2018). [PubMed: 29651849]

34. Ruvinov SB, Ahmed SA, McPhie P & Miles EW Monovalent cations partially repair a conformational defect in a mutant tryptophan synthase α2β2 complex (β-E109A). J. Biol. Chem 270, 17333–17338 (1995). [PubMed: 7615535]
35. Wittmann BJ, Johnston KE, Almhjell PJ & Arnold FH evSeq: Cost-effective amplicon sequencing of every variant in a protein library. ACS Synth. Biol 11, 1313–1324 (2022). [PubMed: 35172576]
36. Busch F et al. TrpB2 enzymes are *O*-phospho-L-serine dependent tryptophan synthases. Biochemistry 53, 6078–6083 (2014). [PubMed: 25184516]
37. Tzin V & Galili G The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. arbo.j 2010, (2010).
38. Nonhebel HM Tryptophan-independent indole-3-acetic acid synthesis: Critical evaluation of the evidence. Plant Physiol. 169, 1001–1005 (2015). [PubMed: 26251310]
39. Yin R, Frey M, Gierl A & Glawischnig E Plants contain two distinct classes of functional tryptophan synthase beta proteins. Phytochemistry 71, 1667–1672 (2010). [PubMed: 20701934]
40. Buller AR et al. Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. J. Am. Chem. Soc 140, 7256–7266 (2018). [PubMed: 29712420]
41. Kraut DA, Sigala PA, Fenn TD & Herschlag D Dissecting the paradoxical effects of hydrogen bond mutations in the ketosteroid isomerase oxyanion hole. Proc. Natl. Acad. Sci 107, 1960–1965 (2010). [PubMed: 20080683]
42. Pinney MM et al. Parallel molecular mechanisms for enzyme temperature adaptation. Science 371, eaay2784 (2021). [PubMed: 33674467]
43. Smith AJT et al. Structural reorganization and preorganization in enzyme active sites: Comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. J. Am. Chem. Soc 130, 15361–15373 (2008). [PubMed: 18939839]
44. Knör S, Laufer B & Kessler H Efficient enantioselective synthesis of condensed and aromatic-ring-substituted tyrosine derivatives. J. Org. Chem 71, 5625–5630 (2006). [PubMed: 16839142]
45. Tang M-C, Fu C-Y & Tang G-L Characterization of SfmD as a heme peroxidase that catalyzes the regioselective hydroxylation of 3-methyltyrosine to 3-hydroxy-5-methyltyrosine in saframycin A biosynthesis. J. Biol. Chem 287, 5112–5121 (2012). [PubMed: 22187429]
46. Schmidt EW, Nelson JT & Fillmore JP Synthesis of tyrosine derivatives for saframycin MX1 biosynthetic studies. Tetrahedron Lett. 45, 3921–3924 (2004).
47. Thompson B, Machas M & Nielsen DR Engineering and comparison of non-natural pathways for microbial phenol production. Biotechnol. Bioeng 113, 1745–1754 (2016). [PubMed: 26804162]
48. Contente ML, Roura Padrosa D, Molinari F & Paradisi F A strategic Ser/Cys exchange in the catalytic triad unlocks an acyltransferase-mediated synthesis of thioesters and tertiary amides. Nat. Catal 3, 1020–1026 (2020).
49. Coelho PS et al. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins *in vivo*. Nat. Chem. Biol 9, 485–487 (2013). [PubMed: 23792734]

Methods-only References

50. Xia Y. Chu W. Qi Q & Xun L. New insights into the QuikChange™ process guide the use of Phusion DNA polymerase for site-directed mutagenesis. Nucleic Acids Res. 43. e12 (2015) . [PubMed: 25399421]
51. Kille S. et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Synth. Biol 2. 83–92 (2013). [PubMed: 23656371]
52. Zhao H. Giver L. Shao Z. Affholter JA & Arnold FH Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nat. Biotechnol 16, 258–261 (1998). [PubMed: 9528005]
53. Gibson DG et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6. 343–345 (2009). [PubMed: 19363495]
54. Lane AN & Kirschner K The catalytic mechanism of tryptophan synthase from *Escherichia coli*. Eur. J. Biochem 129, 571–582 (1983). [PubMed: 6402362]

55. Bhushan R & Brückner H Use of Marfey's reagent and analogs for chiral amino acid analysis: Assessment and applications to natural products and biological systems. J. Chromatogr. B 879. 3148–3161 (2011).
56. Hopf TA. et al. The EVcouplings Python framework for coevolutionary sequence analysis. Bioinformatics 35. 1582–1584 (2019). [PubMed: 30304492]
57. Nguyen L-T, Schmidt HA, von Haeseler A & Minh BQ IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol 32, 268–274 (2015). [PubMed: 25371430]
58. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A & Jermiin LS ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589 (2017). [PubMed: 28481363]
59. Letunic I & Bork P Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296 (2021). [PubMed: 33885785]
60. Kabsch W XDS. Acta Cryst. D 66, 125–132 (2010). [PubMed: 20124692]
61. Winn MD et al. Overview of the CCP4 suite and current developments. Acta Cryst. D 67, 235–242 (2011). [PubMed: 21460441]
62. McCoy AJ et al. Phaser crystallographic software. J. Appl. Cryst 40, 658–674 (2007). [PubMed: 19461840]
63. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Cryst. D 66, 486–501 (2010). [PubMed: 20383002]
64. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst. D 66, 213–221 (2010). [PubMed: 20124702]
65. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Cryst. D 66, 12–21 (2010). [PubMed: 20057044]
66. Laskowski RA, MacArthur MW, Moss DS & Thornton JM PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst 26, 283–291 (1993).
67. Neese F, Wennmohs F, Becker U & Riplinger C The ORCA quantum chemistry program package. J. Chem. Phys 152, (2020).
68. Becke AD Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys 98, 5648–5652 (1993).
69. Lee C, Yang W & Parr RG Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B 37, 785–789 (1988).
70. Zheng J, Xu X & Truhlar DG Minimally augmented Karlsruhe basis sets. Theor. Chem. Acc 128, 295–305 (2011).
71. Grimme S, Antony J, Ehrlich S & Krieg H A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys 132, 154104 (2010). [PubMed: 20423165]
72. Chai J-D & Head-Gordon M Systematic optimization of long-range corrected hybrid density functionals. J. Chem. Phys 128, 084106 (2008). [PubMed: 18315032]
73. Barone V & Cossi M Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. J. Phys. Chem. A 102, 1995–2001 (1998).
74. Best RB et al. Optimization of the additive CHARMM alL-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. J. Chem. Theory Comput 8, 3257–3273 (2012). [PubMed: 23341755]
75. Vanommeslaeghe K et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem 31, 671–690 (2010). [PubMed: 19575467]
76. Eastman P et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLOS Comput. Biol 13, e1005659 (2017). [PubMed: 28746339]
77. Pettersen EF et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82 (2021). [PubMed: 32881101]

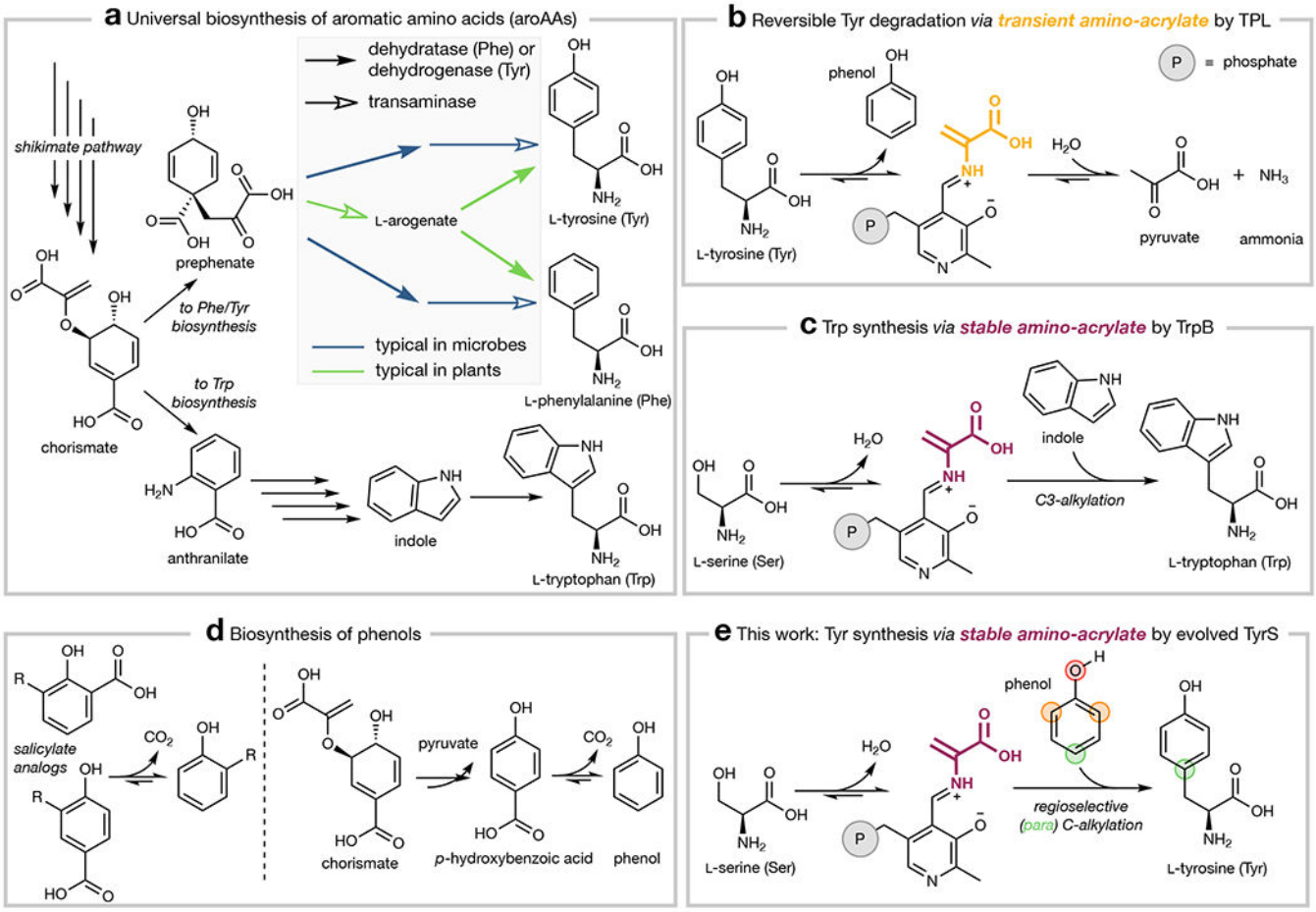


Fig. 1. Enzymatic synthesis of aromatic amino acids.

a. All known aromatic amino acid (aroAA) biosynthesis occurs by conversion of chorismate, the final product of the shikimate pathway, to the aroAAs using a limited and conserved set of chemistries. The aroAAs can then be used in protein synthesis, natural product biosynthesis, or further derivatized to noncanonical amino acids (ncAAs) by other enzymes. **b.** Reversible degradation of Tyr by tyrosine phenol lyase (TPL), a pyridoxal 5'-phosphate (PLP)-dependent enzyme, can be used for biocatalytic Tyr analog synthesis. **c.** The final step of L-tryptophan (Trp) biosynthesis by TrpB, which irreversibly alkylates indole to form Trp, can be used for the biocatalytic synthesis of Trp analogs. **d.** Many phenol analogs are prepared from decarboxylation of hydroxybenzoic acid (salicylic acid) analogs. Phenol itself can be biosynthesized in two enzymatic steps from chorismate. **e.** An engineered "tyrosine synthase" (TyrS) uses the synthetic prowess of TrpB and the substrate preference of TPL to irreversibly form Tyr from phenol and Ser regioselectively in the presence of multiple reactive atoms (colored circles).

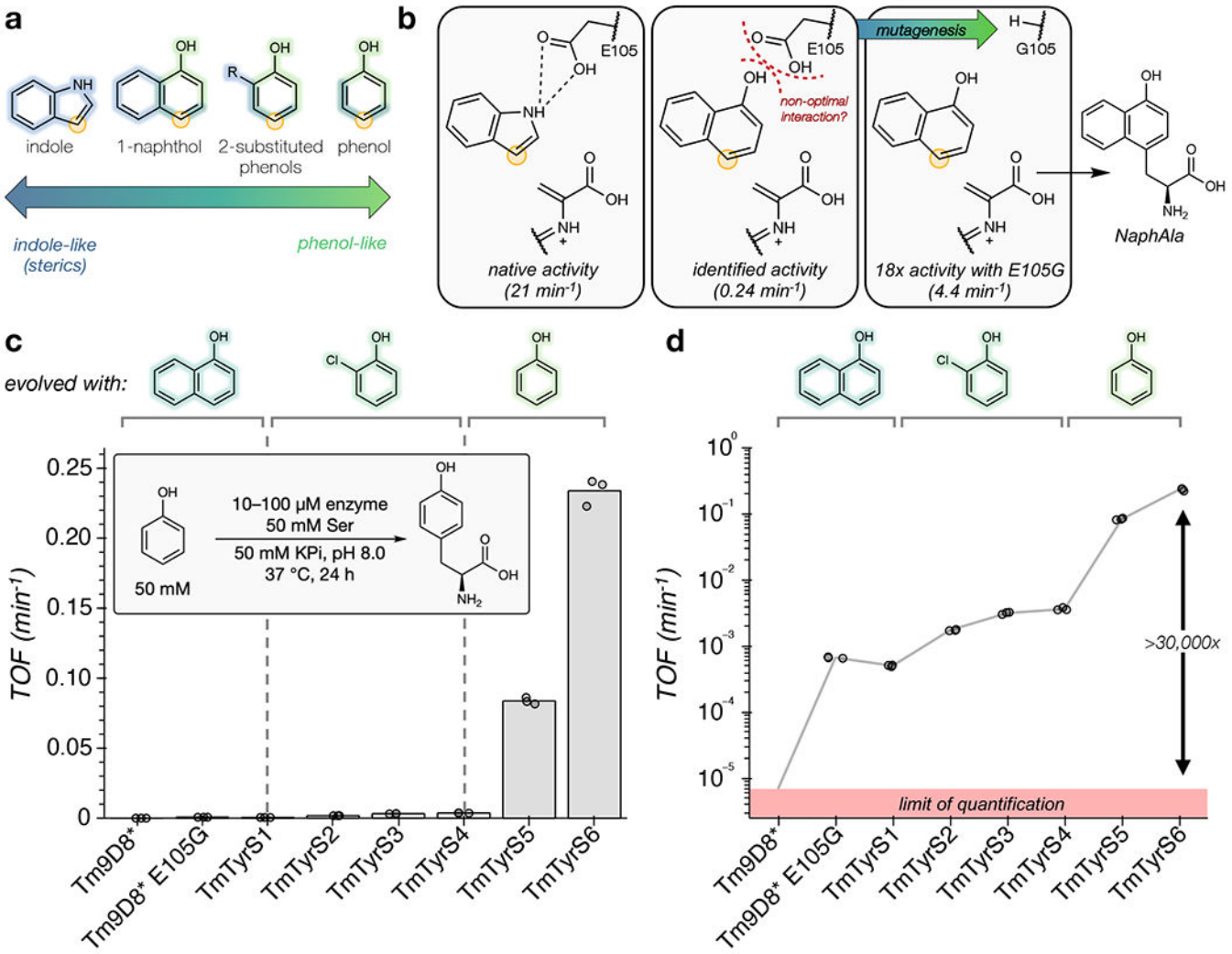


Fig. 2. Directed evolution of a tyrosine synthase.

a. Proposed 'substrate walk' from the native substrate of TrpB, indole, to phenol, with the colored circles representing the desired sites of alkylation. **b.** The universally conserved catalytic glutamate (E105) sidechain participates in interactions important for C3-alkylation of indole but which may not be optimal for *para*-alkylation of 1-naphthol (see also Extended Data Fig. 2b). Mutating this glutamate to glycine (G105) enhanced activity 18-fold. The reaction exclusively forms the *para*-alkylation product β-(1-naphthol-4-yl)-L-alanine (NaphAla). **c.** Approximate turnover frequencies (TOFs, min^{-1}) for conversion of 50 mM phenol to Tyr by TyrS lineage. The screening substrate used during evolution is shown above the chart, with substrate change demarcated by vertical dashed lines. Performed in n=3 technical replicates. **d.** The TOFs presented on a log scale with a quantified limit for the turnover frequency required to detect Tyr under the presented experimental conditions.

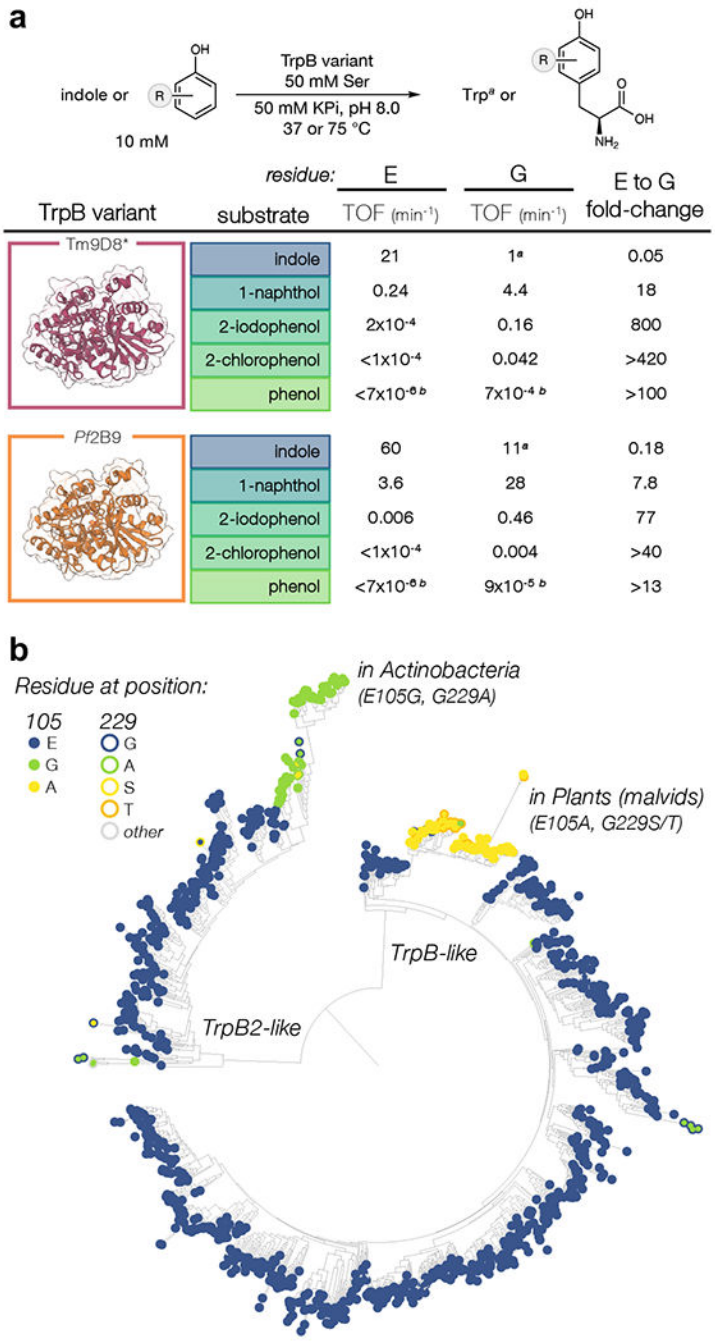


Fig. 3. Substitutions of the conserved catalytic glutamate in TrpBs.

a. Rates of conversion of different substrates in two TrpB variants with the catalytic glutamate sidechain (residue = E) or without it (residue = G). Tm9D8* tested at 37 °C; Pf2B9 tested at 75 °C. [a]Rate of indole conversion calculated using both Trp and isoTrp formation. [b]Performed with 50 mM substrate (Extended Data Fig. 5b). Reactions performed in technical duplicate. **b.** Phylogenetic analysis of TrpB-like sequences identifies two unique, uncharacterized clades with mutations at positions 105 and 229, which are otherwise >98% conserved across 18,051 sequences. Those with E105G and G229A

(green) are TrpB2-like enzymes found primarily in *Actinobacteria*. Those with E105A and G229S/T (yellow/orange) are TrpB-like enzymes found in plants, primarily in malvids. The phylogenetic tree was built from 1,158 sequences: those containing these mutations, selected paralogs, and 1,000 other randomly sampled TrpB sequences. The tree is midpoint rooted for visual clarity; see Extended Data Fig. 6d for the unrooted tree.

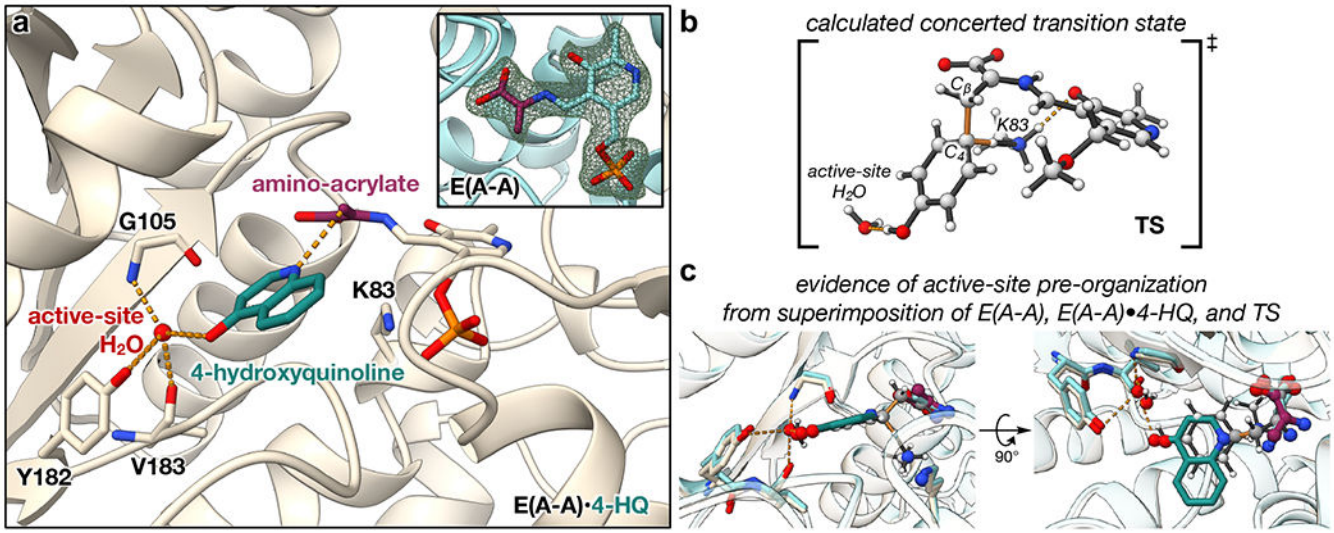


Fig. 4. Structural and mechanistic investigation of TyrS.

a. Crystal structure of amino-acrylate-bound TmTyrS1 species. *Inset*: polder omit map contoured at 5σ for the amino-acrylate in crystals soaked only with Ser. This E(A-A) species (blue) is highly stable. *Main*: enzyme in complex with the non-reactive 1-naphthol analog 4-hydroxylquinoline (4-HQ, teal) coordinated to a new active-site water in place of E105, which is present in all TyrS crystal structures obtained in this study. This structure suggests the water is sufficient to orient 1-naphthol and other phenolic nucleophiles for *para* C–C bond formation, explaining the regioselectivity of these reactions. **b.** Computational modeling of the alkylation reaction reveals that this active-site water can coordinate phenol through a concerted pathway of C_4–C_β bond formation and C_4–H bond cleavage by K83 (transition-state bonds colored in orange). This mechanism would bypass an unstable step-wise intermediate for phenol alkylation, which is otherwise stable for indole alkylation. The PLP phosphate group modeled as a methyl group. Hydrogen-bonding interactions are shown as dashed bonds. **c.** Alignments of the E(A-A) structure, the E(A-A)•4-HQ structure, and the calculated transition state (TS) structure. The TS and E(A-A)•4-HQ were superimposed at their catalytic and reactive groups (RMSD = 0.884 Å over 10 atoms, shown as large spheres). The high degree of alignment between these structures suggests that 4-HQ acts as a TS analog and that the active-site water installed by the E105G mutation pre-organizes the active site for this non-native reaction.

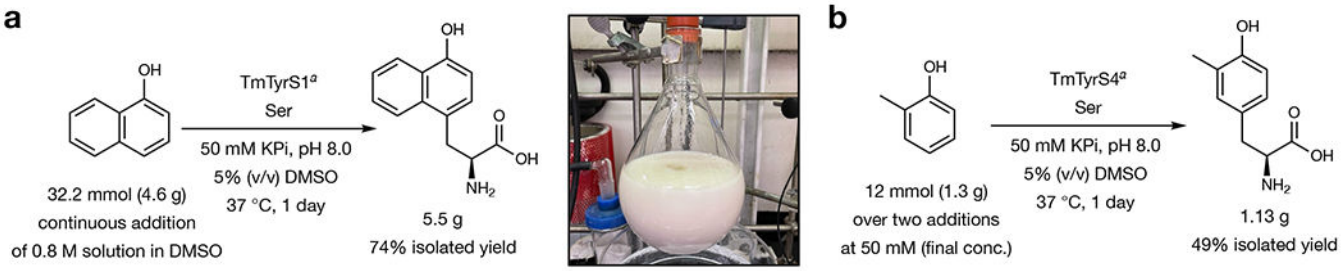


Fig. 5. Facile gram-scale syntheses of valuable Tyr analogs by TyrS.

a. Preparation of NaphAla from 1-naphthol and Ser at multi-gram scale. A solution of 1-naphthol in DMSO was continuously added to the reaction until the solution was saturated, precipitating pure NaphAla within the flask (as shown), which could be collected over a filter and washed. **b.** Preparation of 3-Me-Tyr from 2-methylphenol and Ser at gram scale. The increased solubility of 2-methylphenol (vs. 1-naphthol) allowed for a simpler batch-wise addition of 50 mM substrate. [a]Catalyst prepared as a lyophilized powder from heat-treated *E. coli* lysate.