Discovery, Engineering and Application of Transaminases in Biocatalysis

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Universität Greifswald

vorgelegt von

Chao Xiang

Greifswald, 06.07.2022

Dekan: Prof. Dr. Gerald Kerth

1 .Gutachter: Prof. Uwe. T. Bornscheuer

2. Gutachter: Prof. Florian Rudroff

Tag der Promotion: 09. 09. 2022

Table of Contents

Abbreviations

AADH	Amino acid dehydrogenase
API	Active pharmaceutical ingredient
ATA	Amine transaminase
BCAT	Branched-chain amino acid aminotransaminase
Boc-3APP	1-Boc-3-aminopiperidine
CAS	Computational alanine scanning
Cv-ATA	ATA from *C. violaceum*
DAAO	D-Amino acid oxidase
DATA	D-Amino acid aminotransaminase
*e*BCAT	BCAT from *E. coli*
ee	Enantiomeric excess
E. coli	*Escherichia coli*
FDH	Formate dehydrogenase
GDH	Glucose dehydrogenase
iPA	Isopropylamine/2-propylamine
IRED	Imine reductase
LBP	Large binding pocket
LDH	Lactate dehydrogenase
MAO	Monoamine oxidase
MD	Molecular dynamics
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate) reduced
$NAD(P)^+$	Nicotinamide adenine dinucleotide (phosphate) oxidized
PAL	Ammonia lyase
PEA	Phenylethylamine

PSOA	1-(3-Ethoxy-4-methoxyphenyl)-2-(methylsulfonyl)ethanamine
RedAm	Reductive aminase
SBP	Small binding pocket
Vf-ATA	ATA from *V. fluvialis*
KR	Kinetic resolution
AS	Asymmetric synthesis
PLP	Pyridoxal-5′-phosphate
PMP	5'-Pyridoxamine phosphate
HPLC	High performance liquid chromatography
PCR	Polymerase chain reaction
THIQ	N-[(1*R*)-1-[(4-Chlorophenyl)methyl]-2-[4-cyclohexyl-4-(1H-1,2,4-trazol-1-ylmethyl)-1-piperidinyl]-2-oxoethyl]-1,2,3,4-tetrahydro-3-isoquinolinecarboxamide

In addition, the genetic code and the one and three-letter codes for amino acids were used as well as SI or SI-derived units

Outline

With the aim to discover and create suitable biocatalysts for the synthesis of chiral amines in a faster and more efficient way, this thesis includes protein engineering studies (**Article I**), explores transaminase substrate specificities (**Articles II and IV**), and an ultrahigh-throughput growth system-based for the directed evolution of amine-forming enzymes (**Article III**).

The protein engineering studies described in **Article I** deal with the creation of a (*R*)-amine transaminase activity in the α-amino acid transaminase scaffold to expand our knowledge of the evolutionary relationship between amine transaminase and α-amino acid transaminase. **Article II** describes the broadening of the limited substrate scope of transaminases to enable the conversion of bulky substrates. In **Article III**, a growth selection system is described for an ultra-high throughput screening strategy to accelerate the identification of desired mutants, which can be widely applied to the directed evolution of amine-forming enzymes.

Article I. Creation of (*R*)-amine transaminase activity within an α-amino acid transaminase scaffold provides insights into enzyme evolution

M. Voss*, C. Xiang*, J. Esque, A. Nobili, M. J. Menke, I. André, M. Höhne, U. T. Bornscheuer*, ACS Chem. Biol*. 2020, **15**, 416-424.

* Equal contribution

In the past decade, most of the research for the discovery of new (*R*)-amine transaminases focussed on gene mining, but it was rarely explored to evolved this activity from other types of transaminases. We proposed that related proteins with similar overall scaffolds have the potential to catalyze a variety of different chemical transformations. In this article, we describe the successful introduction of (*R*)-amine transaminases activity in an α-amino acid aminotransferase (DATA) by only subtle changes in the enzyme's active site. Interestingly, none of these mutations caused significant (*R*)-amine acceptance on its own, but their combinatorial impact was predicted by a bioinformatic analysis employing the Rosetta programme. These results substantially expand our knowledge about the relationship between amine transaminases and α-amino acid transaminases and adds new candidates to the

toolbox of synthetically useful transaminases.

Article II. Directed evolution of an amine transaminase for the synthesis of an Apremilast intermediate via kinetic resolution

C. Xiang, S. Wu, U.T. Bornscheuer, *Bioorg. Med. Chem.*, 2021, **43**, 116271.

Kinetic resolution of racemates is an alternative strategy to asymmetric synthesis for chiral amine synthesis. Apremilast is an important pharmaceutical. In this article, we established a new method to obtain this key chiral amine intermediates for Apremilast synthesis using directed evolution of the *Vibrio fluvialis* transaminase (Vf-TA), which resulted in an effective double-mutant. The best three variants were successfully applied to produce the key chiral amine intermediate in enantiopure form (>99% *ee*) and at 49% conversion.

Article III. A growth selection for the directed evolution of amine forming/converting enzymes

S. Wu*, C. Xiang*, Y. Zhou, M. S. H. Khan, W. Liu, C. G. Feiler, R. Wei, G. Weber, M. Höhne, U. T. Bornscheuer, *Nature Commun.*, 2022, submitted.

* Equal contribution

The adaption of transaminase to target substrates is often realized by directed evolution, therefore the screening of large mutant libraries is always necessary. In this article, we described a growth selection system to enable ultrahigh-throughput screening of mutant libraries of *E. coli*. The principle of the methodology is based on the targeted chiral amine as the sole nitrogen source for cell growth in a chemically-defined medium. By substituting four small constitutive promoters with different strengths on the vectors to modulate the expression levels of the desired enzyme, the selection pressure for *E. coli* could be fine-tuned. This concept is simple, low equipment dependent, and represents a general screening strategy that is applicable to the directed evolution of amine-forming enzymes. For a transaminase, in only two rounds of evolution, we obtained a triple-mutant which showed >110-fold increase in its activity towards (*R*)-1-*Boc*-3-aminopiperidine. This enabled the asymmetric synthesis of this key chiral amine drug intermediate. We also demonstrated that growth selection can significantly improve the catalytic activity of a monoamine oxidase (270-fold) and of an ammonia lyase (26-fold) in only one or two rounds of evolution.

1 Introduction

1.1 Chiral Amines

1.1.1 Introduction and applications of chiral amines

Chirality was defined by Lord Kelvin in 1893 as follows: “I call any geometrical figure, or groups of points, chiral, and say it has chirality, if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself”. A chiral molecule exists as two stereoisomers that are mirror images of each other, called enantiomers. Usually, only one of the enantiomers has desirable (biological) activity, whereas the other enantiomer may causes undesired side-effects.[1] Therefore, the enantiomeric purity of compounds is important for the synthesis of pharmaceuticals, agrochemicals, flavors and fragrances.[2][3] Over the past few decades, great efforts have been made to develop different methods for preparing optically active compounds, *e.g.*, via chromatography, crystallization, but especially through chemical or biocatalytic synthesis methods.[4] There are two general strategies used to obtain optically active compounds: one is the asymmetric synthesis, and the other one is kinetic resolution of a racemic mixture into individual enantiomers.[5]

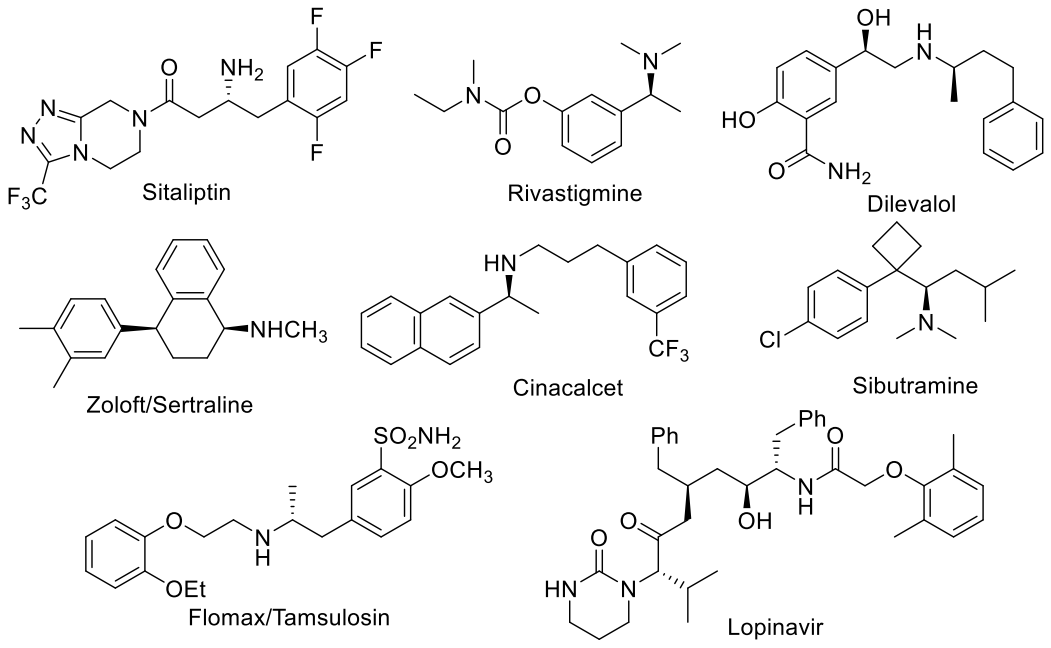


Figure 1. Examples of chiral amine functionality containing pharmaceutical drugs.

Chiral amines are indispensable functional groups for a large number of bioactive pharmaceuticals and agrochemicals probably because of their propensity for

hydrogen bond formation.[6] According to reports, approximately 40% of drug structures contain chiral amine functions, e.g., the anti-hyperglycemic drug sitagliptin, the Alzheimer's drug rivastigmine, the adrenergic antagonist dilevalol, and the antiretroviral drug lopinavir (Figure 1).[7][8][9] The most abundant chemical synthesis routes for chiral primary amines are asymmetric hydrogenation and asymmetric addition. The chirality of the amine product is determined by the chirality present in the ligand of the metal catalyst or the covalent incorporation of chiral auxiliaries in the prochiral substrate. As chiral amine compounds play an irreplaceable role in active pharmaceutical Ingredients and industries, there is an increasing desire to develop new and efficient methods to gain optically pure amines, especially through biocatalytic asymmetric synthesis.[10]

1.1.2 Biocatalytic routes to chiral amines

Synthetic methods of chiral amine can be divided into two categories: chemical or biocatalytic synthesis.[9] The predominant chemical strategy for the introduction of chirality is through chemical resolutions, transition metal, organometallic or asymmetric hydrogenation catalysis.[11][12] Since conventional chemical methods for the asymmetric synthesis of amines suffer from different limitations, *e.g.*, low efficiency, low chemo-, regio- and especially stereo-selectivity, and high environmental impact, also with the rising demand for high pure enantiopure amines, many studies have been carried out to develop alternative biocatalytic routes during the last three decades.[13][14] The advantages of biocatalysis over chemical synthesis are that enzyme-catalyzed reactions are and usually highly stereo- and regioselective and usually take place at ambient temperatures and atmospheric pressure. Furthermore, microbial cells and enzymes can be immobilized and reused for many cycles, which makes biocatalytic processes more economically and efficient.[15] As a result, biocatalysis has been widely studied and found numerous applications in various fields as an alternative to chemical catalysis, especially to make chiral amine compounds for pharmaceuticals and the chemical and flavor industries.[16]

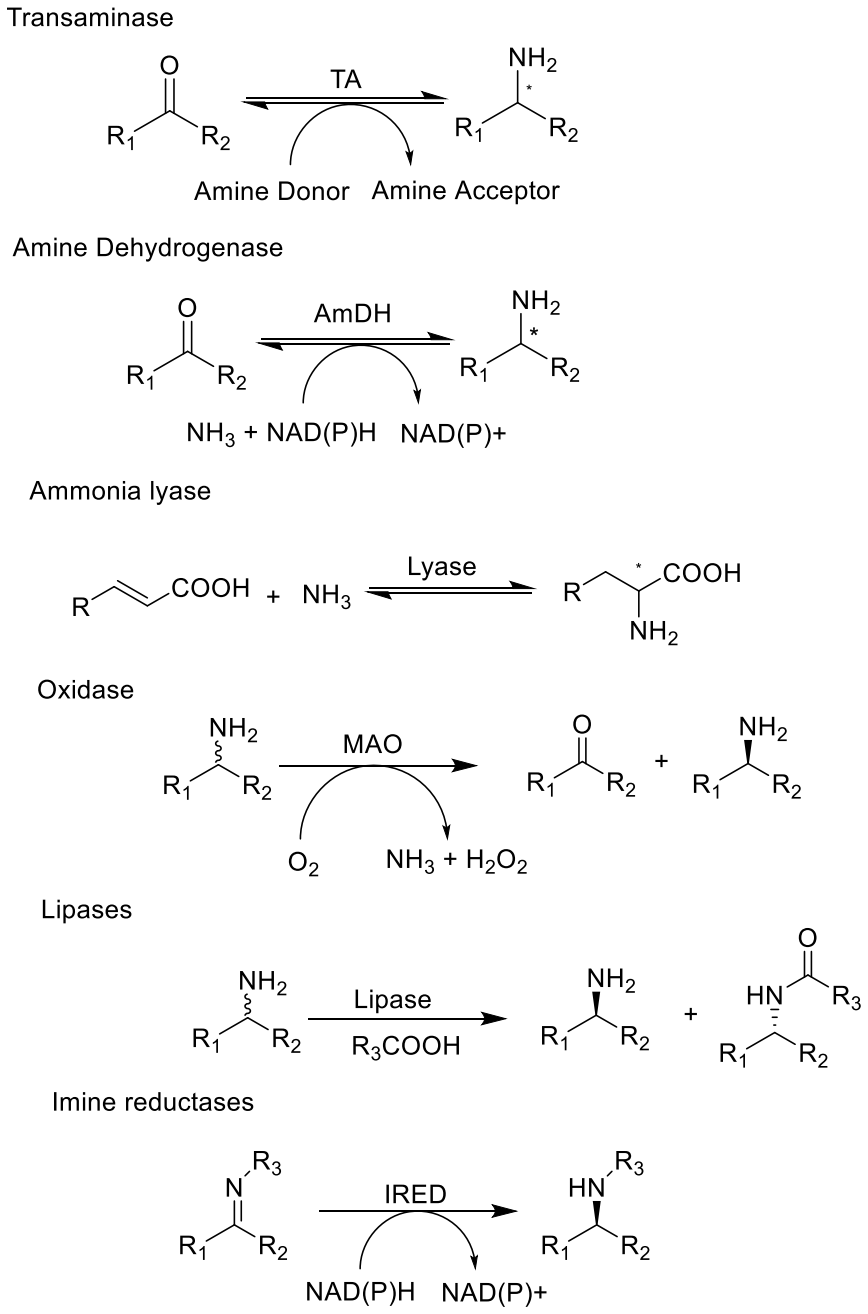


Figure 2. Different classes of enzymes that catalyze the enantioselective synthesis of amines.

For the synthesis of chiral primary amines, different types of enzymatic synthesis routes have been described using lipases, amine dehydrogenases, monoamine oxidases, amine dehydrogenase, ammonia lyases, and transaminases (Figure 2). Based on the reaction process, there are two main strategies used for chiral amine synthesis with transaminases: kinetic resolution of racemates or asymmetric synthesis from prochiral carbonyl precursors (Figure 3).[17]

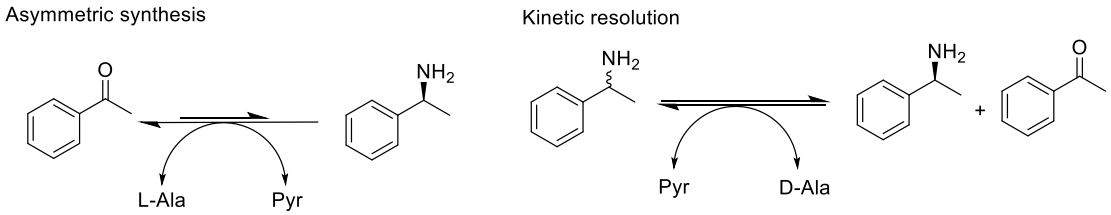


Figure 3. Transaminases can be applied either for the asymmetric synthesis of amines starting from the corresponding ketones or for the kinetic resolution of racemic amines.

1.1.2.1 Kinetic resolutions of racemic primary amines

Hydrolases were the first enzymes reported to be used in the kinetic resolution of amines. Due to the excellent properties of hydrolases in the kinetic resolution of alcohols and carboxylic acids, the application of hydrolases was further extended to the kinetic resolution of racemic amines, such as lipases.[18] The lipase from *Candida antarctica* lipase (CAL-A and -B) is one of the most efficient and widely used enzyme in aminolysis, ammonolysis reactions, esterification, transesterification, and alcohol and amine resolutions since Reetz and Schimossek reported the first dynamic kinetic resolutions of N-(1-phenylethyl) acetamide employing CAL-B as a racemization catalyst in 1996.[19] The company BASF introduced already in the 90ies the commercial synthesis of a broad range of chiral amines on industrial scale using a lipase originating from *Burkholderia plantarii*. This lipase catalyzes the highly stereoselectivity kinetic resolution of racemic amines using methoxyacetate ester as efficient acyl donor.[20]

Monoamine oxidases (MAO) are widely used in different amines resolution such as the MAO from *Aspergillus niger* (MAO-N).[21] In 2002, Marina *et al.* described a method for the deracemization of a range of α-amino acids through repeated cycles of MAO-N D1-catalyzed oxidation followed by nonselective chemical reduction.[22] Then, a large number of experiments using MAO-N and mutants for the deracemization of a wide range of amines were reported, such as using immobilized MAO-N D3 to obtain (*R*)-2-phenylpyrrolidine,[23] MAO-N D5 was used to obtain *N*-methoxy-1-cyclohexylethylamine,[24] [25][26] MAO-N AC1 can make enantiopure mexiletine,[27] MAO-N D9 and D11 have been applied to make enantiopure THIQ and β-carboline derivatives.[28][29]

Transaminases were described as enzyme for the kinetic resolution of amines by converting one enantiomer of the racemic amine to a low-value ketone coproduct, allowing for theoretical yields of 50% of the non-converted then optically pure remaining amine enantiomer.[30] Since the 1990s, the Celgene Corporation successfully used kinetic resolution for the synthesis of enantiopure amines,[31] this concept has been extensively applied in the last few years.[13] For example, the ω-TA from *Ochrobactrum anthropi* allowed for efficient kinetic resolution of α-PEA (500 mM),[32] and the (*R*)-ω-TA from *Mycobacterium vanbaalenii* was applied to the kinetic resolution of a series of aliphatic and aromatic amines.[33]

1.1.2.2 Asymmetric synthesis of chiral primary amines

There are several enzyme classes reported to be useful for the asymmetric synthesis of chiral amine compounds, including imine reductases, amine dehydrogenases, ammonia lyases, and transaminases.[6] [34]

Imine reductases have been described as enzymes that can catalyze the asymmetric reduction of various imines to the corresponding amine products utilizing NAD(P)H as cofactor.[35] Interestingly, in 2017 Turner's group reported a new imine type of a reductase discovered in *Aspergillus oryzae* (AspRedAm) that not only catalyzes the asymmetric reduction of various imines to the corresponding amine products but also converts various ketone and amine substrates to primary, secondary and tertiary amines. Moreover, this AspRedAm does not require a large excess of the amine donor, but can catalyze the formation of chiral amines with stoichiometric amounts of ketone and amine donor.[36] Similar to imine reductases, amine dehydrogenases (EC 1.4.99.3) have been described as enzymes that catalyze the reduction of ketones to chiral amines using NADH.[37] In 2012, Abrahamson *et al.* used an α-amino acid dehydrogenase as the scaffold to successfully create an amine dehydrogenase which can perform a reductive amination without the usually requirement of a carboxylic acid function next to the carbonyl group.[38] Ammonia-lyases (EC 4.3.1.X) catalyze the reversible cleavage of C−N bonds, typically of α-amino acids, generating ammonia and an unsaturated or cyclic derivative. Ammonia lyases represent a very large protein family as 31 different EC subclasses have been identified with remarkable differences in structure, function and mechanism.[39] The potential synthetic use of this class of enzymes is appealing and in 1976 a patent was filed by

Pfizer using immobilized RtPAL from *Rhodosporidium* to convert the unsubstituted cinnamic acid with a concentration of 500 mM in the presence of concentrated buffered ammonia to result in 90% yield of L-phenylalanine;[40] in 2014, Sarah *et al.* used the AvPAL discovered from *Anabaena variabilis* in the synthesis of 2-chloro- and 4-trifluoromethyl-phenylalanine on a preparative scale, with excellent conversion and enantiopurity and a moderate isolated yield of 42%.[41] Transaminase can catalyze asymmetric synthesis, enabling the direct asymmetric conversion of prochiral ketones to the corresponding chiral amines, with the possibility of obtaining the desired enantiopure amine form in 100% yield.[42]

1.2 Transaminases

Transaminases (EC 2.6.1.X) are pyridoxal-5′-phosphate- (PLP) dependent enzymes that catalyze the transfer of an amino group from an amino donor to an amino acceptor (Figure 4).[30]

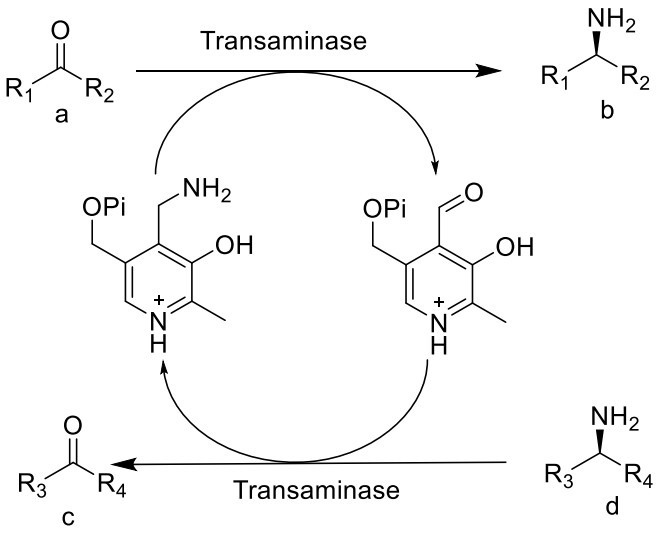


Figure 4. General scheme for transamination reactions: a, Amine acceptor; b, Amine acceptor product; c, Amine donor; d: Amine donor product.

Transaminases are widespread in nature and play an irreplaceable role in various cellular processes.[43] They are one of the most important enzymes used for the biosynthesis of chiral primary amines. Since their first industrial applications at the end of the last century, their excellent enantio- and regioselectivity, outstanding catalytic efficiency and stability, concise reaction and environmental friendliness have brought TAs to be one of the most promising biocatalysts in organic synthesis for compounds containing chiral amine units.[44] One of most important subgroups in

transaminases are ω-TAs, which can accept a large variety of carbonyl compounds as substrates.[45] The desirable ability of ω-TAs to directly transfer amino functions between amines and ketones, offers an immense potential application for the production of high-value chiral amines in APIs.[46] However, the industrial application of transaminases is limited by a disfavored reaction equilibrium in asymmetric synthesis, limited substrate scope, and substrate/product inhibition.[47][18][48] Therefore, in spite of the many advantages of transaminase catalyzed reactions, making transaminase processes more available for the production of a wider range of amines are still current research hotspots.[49]

1.2.1 Classification of transaminases

Depending on the position of the transferred amino function to the carboxylate group, transaminases have initially be classified into α-amino acid transaminases (α-TA), ω-amino acid transaminases (ω-TA). [50] α-TAs exclusively convert α-amino acids. They only accept a substrate with a carboxylic acid group in the α-position of the amino function and hence only allow the formation of α-amino acids.[51] Furthermore, transaminase can be assigned to six subfamilies, based on their different sequence and structure similarity.[30][50] Compared with α-TAs, ω-TAs are more useful, as they can convert substrates with a more distant carboxylate group, for instance they can catalyze the conversion of the ε-amino group in lysine. Transaminases which even lack the requirement of a carboxylic group in the substrate molecule are called amine transaminases (ATA).[52]

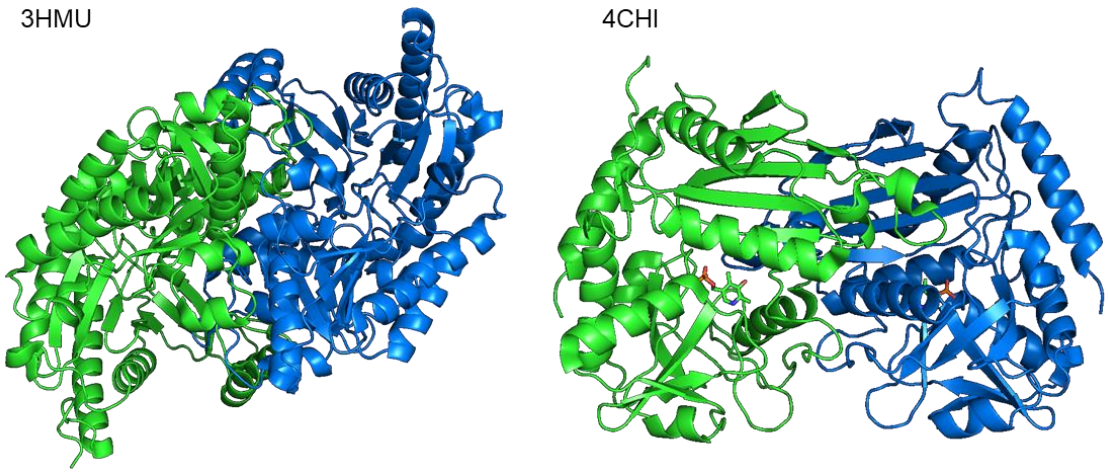


Figure 5. Structural comparison of transaminases of Fold I and IV of PLP-dependent enzymes. The different monomers are colored in green and blue. Left: fold type I transaminase, exemplified for the

amine transaminase from *Silicibacter pomeroyi* (PDB: 3HMU). Right: fold type IV transaminase, exemplified for the amine transaminase from *Aspergillus fumigatus* (PDB: 4CHI).

Transaminases (TA) belong to the class of pyridoxal-5′-phosphate- (PLP) dependent enzymes.[53] The family of PLP-dependent enzymes are widespread in nature since this enzyme family is able to catalyze many distinct reactions in cells including transamination, racemization, decarboxylation, elimination, aldol synthesis and many more.[54] In general, PLP-dependent enzymes are classified into seven major groups according to their fold type, with transaminases belonging to fold types I and IV.[55] According to the solved transaminase crystal structures, both fold types of the transaminases are homodimers, where the active site is located at the interface between the two subunits of the dimer, wherein each monomer participates in the active site.[56]

1.2.1.1 Fold-type I transaminases

The PLP-fold type I is a structural motif of a broad family of proteins that even with low sequence homology still form similar tertiary structures.[57] Therefore, even though fold type I transaminases (also known as the aspartate transaminase superfamily) have a large number of members with very different amino acid sequences and substrate specificities, these enzymes are all (*S*)-specific having similar subunits structural and active sites.[58] With more crystal structures reported, the structure-function relationship of (*S*)-selective transaminases has been understood deeper, and this has been well explained. By structural comparison, Käck *et al*. found that despite similar folding in this structural family, however, only some small anti-parallel *β* fragments near the N-terminal end of the small structural domain are conserved.[59] Steffen-Munsberg *et al*. analyzed the sequence-function relationships of PLP-dependent enzymes by a 3DM database sequence alignment and found that the location of the PLP in the active site of these enzymes is virtually identical and their active sites comprise one turn of a left-handed helix at the *si*-face of PLP.[55]

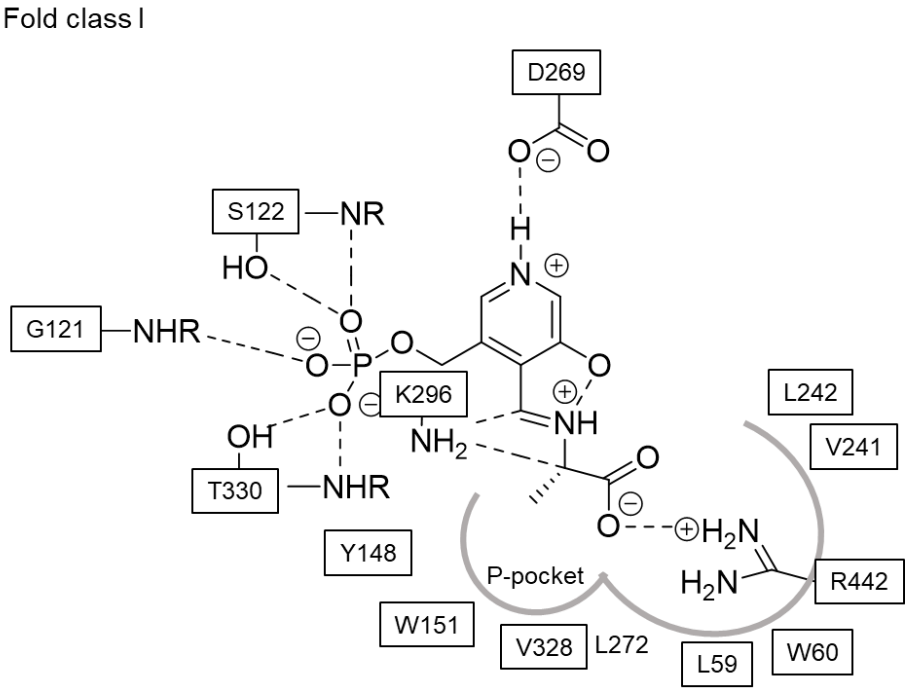


Figure 6. Schematic diagram showing the hydrogen bonds and electrostatic interactions stabilizing the bound PLP-Ala (external aldimine state) in the active site of BM-ATA84 (PDB: 5G2Q) belonging to fold type I.

Among these, (*S*)-ATAs are one of the most relevant enzymes in this subfamily for industrial applications, since the synthesis of chiral amines is feasible with high enantiomeric excess.[60] The exploration of (*S*)-ATA for biocatalytic applications started from the first sequence described in 2003, and then a large number of (*S*)-selective amine transaminases were discovered through enrichment culture and by sequence homology searches.[61] Up to now, a large number of (*S*)-selective aminotransferases have been well studied, the most famous are the TAs from *Vibrio fluvialis* (Vf-TA), *Bacillus megaterium* (BM-ωTA), *Chromobacterium violaceum* (CV-TA), *Ruegeria* sp. TM1040 (Rs-TA), and *Silicibacter pomeroyi* (Sp-TA) (Figure 6).[62][63][64]

1.2.1.2 Fold-type IV transaminases

The PLP-fold type IV consists of three subclasses of aminotransferases: (*S*)-selective L-branched chain amino acid aminotransferase (BCAT), (*R*)-selective D-amino acid aminotransferase (DAAT) and (*R*)-selective amine transaminase (*R*-ATA).[65][66][67] In contrast to their overall homolog fold, the specificity of these three transaminases is focused on their substrate, with -COOH in the α-position to the carbonyl group in a

molecule (α-amino and α-keto acids) or no -COOH in the molecule (*R*)-primary amines).[68][69] At the same time, there are no reported activities of class IV TAs towards amino acids which have one or more carbon atoms between the carbonyl group and COOH in the molecule.

L-Branched chain amino acid aminotransferases are the basic enzymes for the synthesis and degradation of branched chain amino acids and are the only fold type IV TAs among PLP-dependent (*S*)-selective TAs.[53][55][65] The (*S*)-selectivity of BCATs is caused by the α-COOH groups of substrates bound in the P-pocket, while the side chains are bound in the O-pocket. The substrate scope of BCATs corresponds to branched-chain L-amino acids such as L-threonine, L-leucine, L-glutamate and their keto analogs.[70][71] They only show activity towards α-amino acids probably because the negatively charged α-COOH is located near the phosphate group of the cofactor and the electrostatic repulsion between the α-COOH group and the phosphate probably is important for substrate reactivity.[72]

D-amino acid aminotransferases (DAATs) catalyze the transfer of amino groups of D-amino acids to ketone acceptors. In bacteria, DAATs are involved in the biosynthesis of D-amino acids, such as D-glutamate, which is an important component of cell wall peptidoglycans.[73] In the catalytic dimer of DAATs and BCATs, the structure as well as most loops are well superposed, while the different hydrophobicity of binding pockets gives the two transaminases opposite selectivity.[74] Therefore, Peisach *et al.* studied the structure of DAAT (PDB IDs: 1DAA, 3DAA) from strain YM-1 in detail. They introduced three polar side-chains of residues, Tyr31, Arg98, and His100, forming a “carboxylate trap” for the α-COOH group of the substrate in the O-pocket, thus determining the substrate scope limited strictly to D-amino acids (Figure 7).[75]

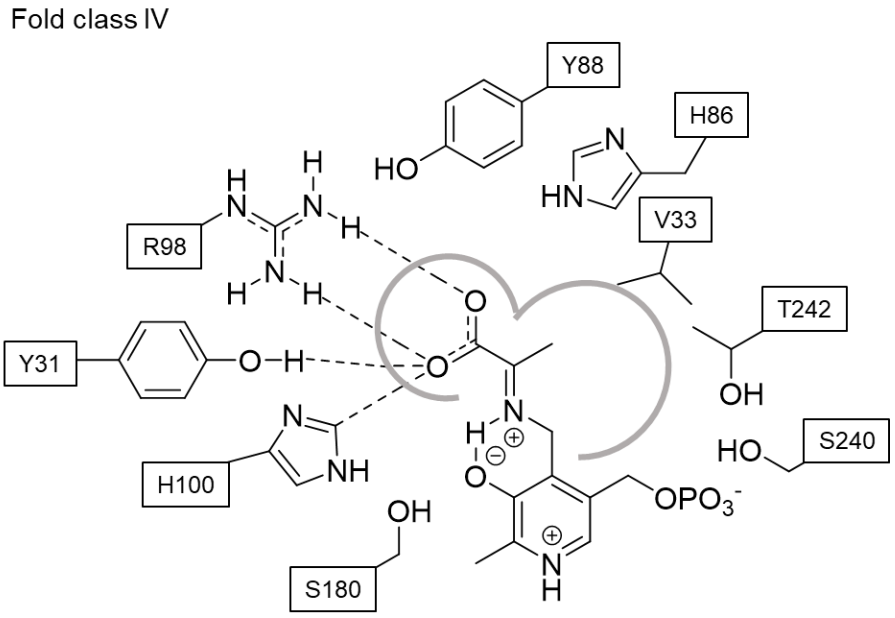


Figure 7. Activity site composition of the DATA (PDB: 3DAA) with D-alanine bound to the active site.

(*R*)-selective amine transaminases catalyze the transfer of an amino group between (*R*)-aromatic/aliphatic primary amines and carbonyl compounds, the main distinction between the active sites of *R*-TAs and BCATs or DATAs is an organization of the P-pocket formed only by hydrophobic residues that have no binding sites for the COOH group. Compared with D-amino acid aminotransferases and (*S*)-selective amine transaminases, (*R*)-selective amine transaminases can convert compounds lacking any carboxyl groups as substrates to directly synthesize (*R*)-amines of high interest to industry.[76] Most transaminases screened among bacterial enrichment cultivation have (*S*)-selectivity, most notably the first (*R*)-selective transaminase reported by Iwasaki *et al*. in 2006, from *Arthrobacter sp.* KNK16833 (Ar-TA) was identified in this manner, which was used for the microbial asymmetric synthesis of (*R*)-3,4-dimethoxyamphetamine.[77] In 2010, Savile *et al.* published the amino acid sequence of the ATA-117 (Figure 7), a homologous (*R*)-selective transaminase from *Arthrobacter sp.*, which was used for the asymmetric synthesis of an (*R*)-amine from small ketones as reported by Codexis.[78] Also in 2010, Höhne *et al.* identified 17 (*R*)-ω-TA by developing an in-silico strategy motif-based searching for the sequence in protein databases based on rational assignments.[79] By comparing the protein sequences and active site structures of DATAs, BCATs and ADCLs, a careful analysis of important residues around the active site of each type of enzyme followed, assuming that the members of fold class IV group of enzymes present certain

flexibility in the architecture of the active site. Based on this assumption they developed an algorithm that easily discarded all BCAT, D-ATA and ADCL from a protein database alignment. Finally, they applied the developed algorithm to public protein databases, identified 21 protein sequences, 17 of them could be characterized and confirmed as (*R*)-ATAs. The binding of PLP is shown in Figure 8. The identified (*R*)-ATAs were widely studied and applied in the asymmetric synthesis of a range of (*R*)-amines.[80]

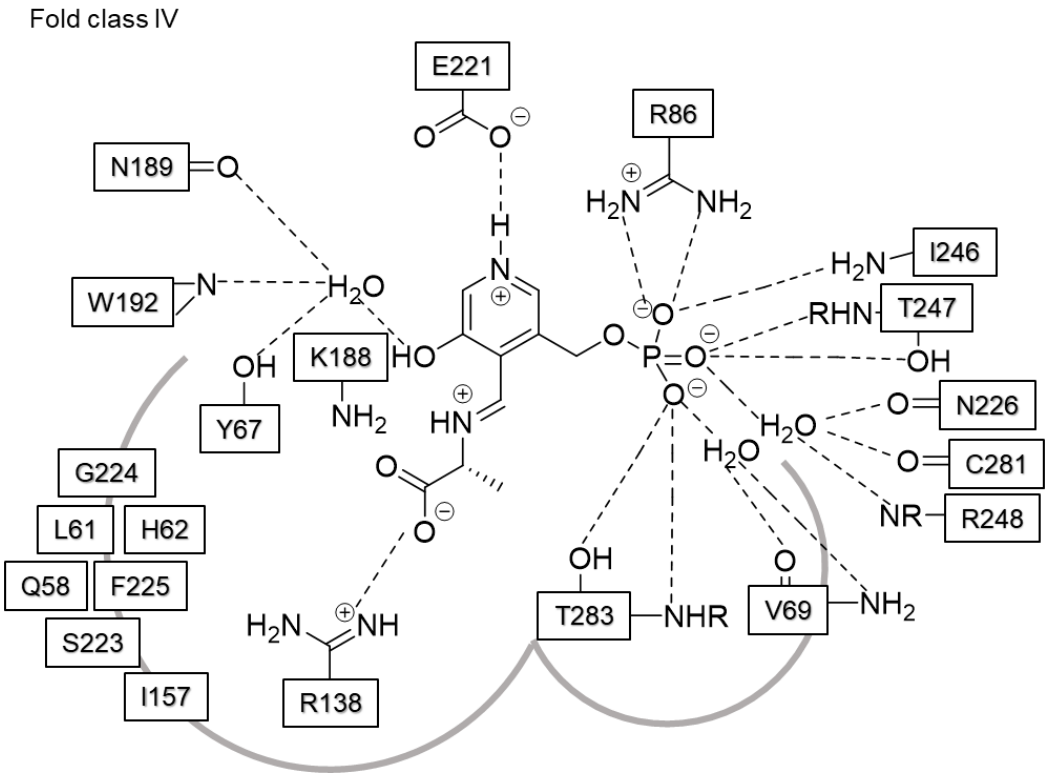


Figure 8. Schematic diagram showing the hydrogen bonds and electrostatic interactions stabilizing the bound PLP-Ala (external aldimine state) in the active site of ATA-117 (fold type IV).

1.2.1.3 Reaction mechanism of transaminases

Previous structural studies have shown that the active site pocket of TAs is composed of two pockets, the pocket near the phosphate group of PLP is the P-pocket, and the pocket in the vicinity of the O3′ atom of PLP is the O-pocket.[81] A PLP molecule positioned at the bottom of the active site constitutes the spatial standard for the active site pocket, which has a profound impact on the structure and function of the TAs.[68] Although this characteristic architecture of the active site limits the substrate scope, it also contributes to the high stereopreference of TAs. The reaction mechanism of transaminases was described as a Ping-Pong Bi-Bi

mechanism, where the reaction occurs through two sequential half-reactions via PLP.[82] In the resting state, PLP forms a Schiff base by covalently binding to an adjacent Lys residue. The binding site of PLP is located somewhat far from the entrance of the active site, resulting in a relatively deep funnel towards the active site.[83]

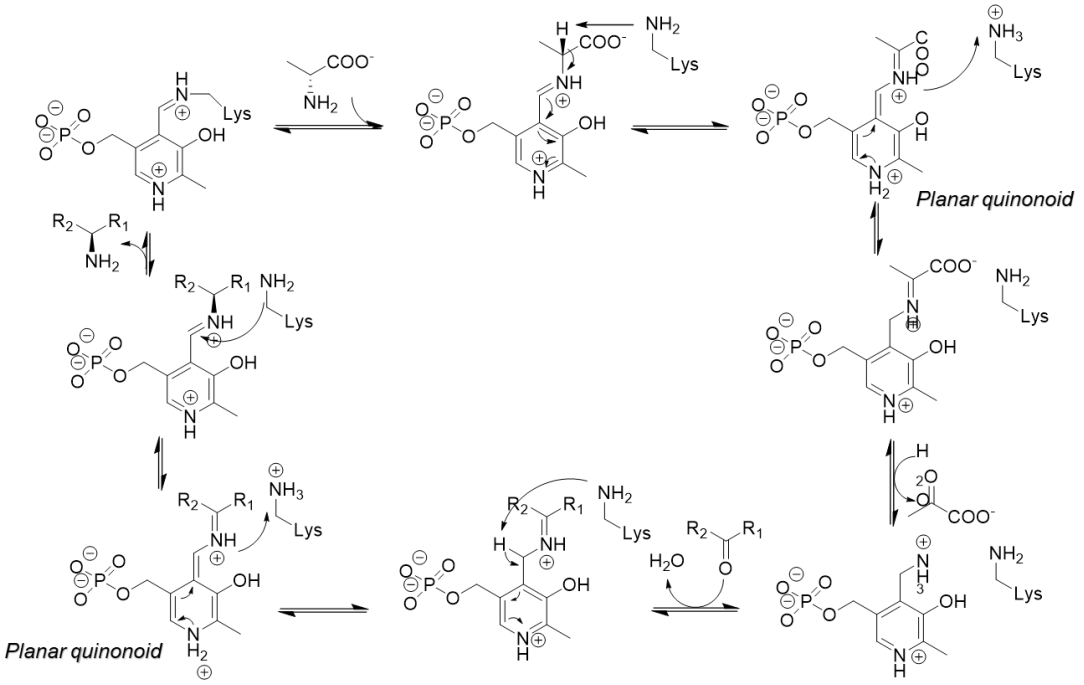


Figure 9. Mechanism of the deamination (clockwise) and amination (anti-clockwise) process with alanine as amine donor and pyruvate as amine acceptor, respectively. Lys = lysine

The first half-reaction of the transamination reactions starts with the PLP covalently bound to the enzyme in the active site of the transaminase (internal aldimine). First, the C=N bond of the Schiff base is attacked by the substrate amine, leading to imine formation between substrate amine and PLP (external aldimine). Then, after a suitably orientated lysine residue abstracts the proton at the amine carbon, the carbanion is formed via the planar and resonance-stabilized quinonoid intermediate. The newly formed imine is hydrolyzed, releasing the ketone product and an enzyme-PMP complex, completing the first reaction. In the second half-reaction, because all steps in the mechanism are reversible, which is transferring the amino group to another ketone substrate by the reverse pathway, completing the transamination reaction (Figure 9).[84]

1.2.2 High-throughput screening assays for transaminases

Even though transaminases are a class of enzymes for the production of high value

chiral amines, the limited substrates range of wild-type enzymes has often been regarded as the bottleneck for industrial applications.[85] In such situations, protein engineering techniques can be employed to generate mutated transaminases with desired characteristics. How to identify the desired variants in libraries from protein engineering become a crucial step.[86] Therefore, the robustness, reliability, and sensitivity of high-throughput screening methods are of great interest to accelerate the identification of optimized proteins.[87] Transaminases have the enzymatic property to exploit two different kinds of substrates (amino donor and amino acceptor), only one of them is the target compound, whereas the other one is the co-substrate.[88] Detection of these common co-substrates or their corresponding co-products by various approaches is generally used to measure the activity or design high-throughput screening methods for transaminases.[89]

α-Amino acids such as alanine with its corresponding α-keto acid (pyruvate) are most commonly employed as co-substrate, and several approaches to enzyme screening were based on the detection of the resulting amino acids either directly or coupled with other enzymes. Except for the direct detection of the resulting α-amino acids by capillary electrophoresis[90], a spectrophotometric detection or coupling with dehydrogenase to monitor the nicotinamide adenine dinucleotide phosphate (NAD(P)H) cofactor consumption by spectrophotometry at 340 nm have been used.[30] In 2004, Kim *et al.* reported a colorimetric high-throughput screening method that utilizes the blue color produced by α-amino acids with $CuSO_4$ and MeOH, which can be detected at 595 nm.[91] Truppo *et al.* developed a method based on a pH change where the pyruvate produced is reduced by lactate dehydrogenase (LDH) accompanied by the oxidation of glucose to gluconic acid by glucose dehydrogenase (GDH), and monitored the pH change with the aid of a pH indicator dye.[92] Another approach is a colorimetric detection of alanine production during the transamination reaction by a combination of amino acid oxidase (AAO) and horseradish peroxidase (HRP) in a cascade reaction system.[57][93] The DAAO coupled assay was described by Barber *et al.*, 2014. This assay is based on the detection of D-alanine by the D-amino acid oxidase from *T. variabilis* (Tv_DAAO) and the subsequent detection of the released H_2O_2 by the HRP. The Tv_DAAO exhibits only low activity towards D-glutamate, in contrast to the high activity towards D-alanine. This enables the measurement of the native DATA reaction. Barber *et al.* employed *o*-dianisidine as

HRP substrate, but one can also use the Ampliflu™ Red system.

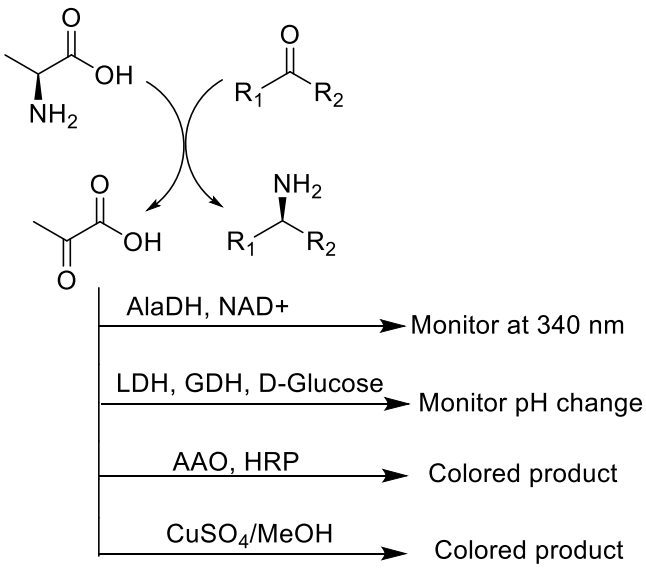


Figure 10. High-throughput screening methods for transaminase activity employing UV or pH analysis.

A number of specialized compounds have been developed and applied for high throughput screening of TA-mediated reactions.[94][95] The most common among those is 1-PEA, which was developed as a high throughput screening strategy for TA reactions due to the high UV absorbance of the corresponding product acetophenone (245 nm).[96] Later compounds such as 2-(4-nitrophenyl)ethan-1-amine, *o*-xylylene diamine, 1-(6-methoxynaphth-2-yl)ethylamine, (*S*)-2-aminotetraline derivatives, *o*-amino benzaldehyde and 1-pyrroline or 1-piperidine were applied for high-throughput screening of TA-mediated reactions since their products result in a color change or fluorescence absorbance.[97][98][99] The advantage of these method is their sensitivity and that they can be applied in solid phase screening. A disadvantage is that these substrates may not always be accepted by the target enzymes due to either steric or catalytic incompatibility.

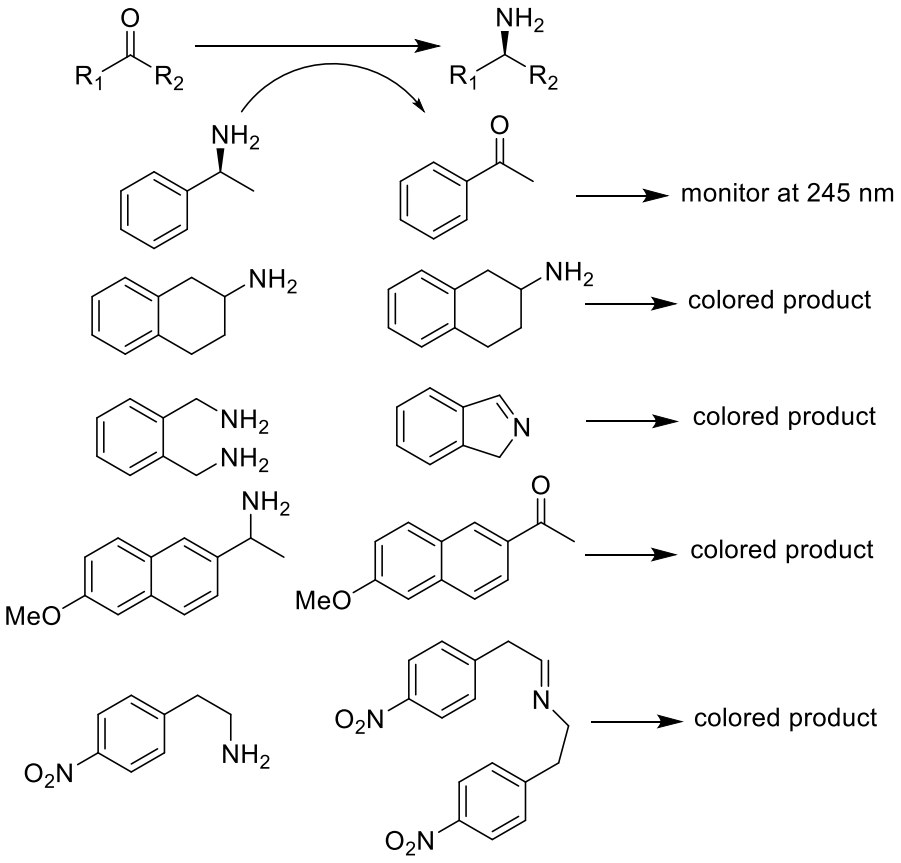


Figure 11. High throughput screening strategies to identify transaminase activity.

1.3 Protein Engineering

Enzymes are capable of catalyzing a large number of reactions, but most wild-type enzymes are not suitable for direct applications in industrial processes. Their major limitations are insufficient operational stability at high substrate concentrations, harsh chemical reaction conditions or low turnover numbers for non-natural substrates.[100] To make the enzyme process highly efficient or to adapt it to non-natural substrates, it is necessary to modify their catalytic properties.[101] Protein engineering is a practical and efficient strategy for the optimization of biocatalysts process, which can modify the enzyme to meet the desired catalytic properties. These two approaches are mainly used in the engineering of enzymes are: directed evolution or rational design. These two different techniques are often flexibly chosen for protein engineering based on prior knowledge of the enzyme structure-function relationship, biochemical activity, reaction mechanism and positioning of the substrate in the active site (Figure 12).

Table 1. Selected examples of recent transaminase protein engineering studies.

TAm sources	TA improvements	References
Variovorax paradoxus	Thermostability improvement	[102]
Chromobacterium violaceum	Enantiospecificity switch and improvement	[103]
Chromobacterium violaceum	Specificity improvement for (*S*)-1-phenylethylamine and 4’-substituted acetophenones	[104]
Chromobacterium violaceum	Activity improvement towards serine	[105]
Chromobacterium violaceum	Thermostability improvement	[106]
Chromobacterium violaceum	Specificity toward bulky substrates	[67][107]
Vibrio fluvialis	Enantiospecificity switch and improvement	[108]
Vibrio fluvialis	Activity improvement	[62]
Vibrio fluvialis	Specificity toward 2-acetylbiphenyl substrates	[109]
Vibrio fluvialis	Specificity toward bulky substrates	[110][111]
Aspergillus terreus	Thermostability improvement	[112][113][114]
Arthrobacter citreus	Enantiospecificity switch and improvement	[115]
Ruegeria sp. *TM1040*	Exo-3-amino-8-aza-bicyclo[3.2.1]oct-8-yl-phenylmethanone and other bulky substrates	[116][117]
Ruegeria sp. *TM1040*	Specificity toward bulky substrates	[118]
Halomonas elongata	Specificity toward acetophenone derivatives	[119]
Bacillus megaterium	Activity improvement	[120]
Ochrobactrum anthropi	Specificity toward bulky substrates and activity improvement	[121][122][123]
Ochrobactrum anthropi	Activity improvement	[124]
Exophiala xenobiotica	Specificity toward bulky substrates and activity improvement	[125]
Asperguillus fumigatus	Conversion rate improvement	[126]
Pseudomonas jessenii	Thermostability improvement	[127]

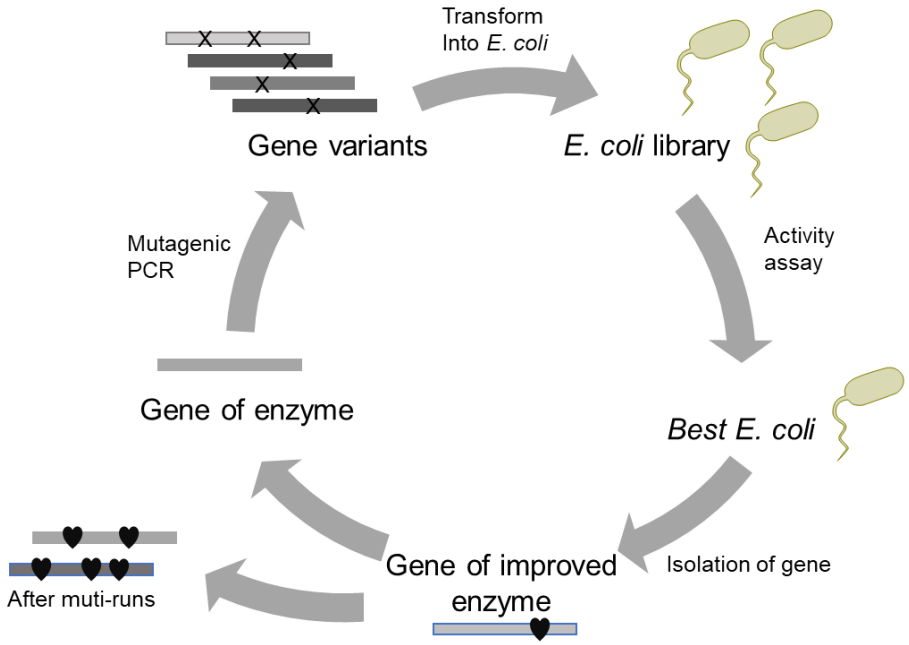


Figure 12. General strategies for directed evolution.

1.3.1 Directed evolution and rational design

Directed evolution is a powerful technique for the engineering of enzymes. The method mimics nature by performing iterative cycles of random mutagenesis and screening or selection for improved protein properties according to obtain a selective enzyme variant dedicated to a desired purpose.[128] In 2018, for the importance of this technique and its broad applicability in biocatalysts, Frances Arnold received the Nobel Prize in Chemistry for pioneering the field of directed evolution.[129] The major advantage of directed evolution is that information about the enzyme structure, the reaction mechanism, and the substrate orientation are unnecessary. Variants are selected based on the screening procedure applied to the mutant library. A widely applied technique for the randomized introduction of mutations is error-prone PCR, which has been also employed in the engineering of TAs. For instance, the specific activity of *Athrobacter citreus* TA towards substituted (*S*)-aminotetralin was improved from 5.9 to 1582.8 IU*g^{-1} by six rounds of random mutagenesis.[130]

The major bottleneck of random mutagenesis based directed evolution is the required of a high-throughput screening method for the analysis of the vast number of generated variants, most of them without a beneficial impact on the desired outcome.[67] To circumvent the large amount of screening work necessary for random mutagenesis derived libraries, and to increase the chance of identifying beneficial

variants, rational strategies based on the protein structure have been applied to protein engineering.[131] Rational design was initially used to improve the thermal stability and solvent tolerance parameters of a given enzyme with the aim of improving the stability of the entire process.[132] With the development of biocatalysis, this strategies was used for improved protein properties more and more, such as to create specificity activity for bulky substrates or to change the enantioselectivity etc.[133] The advantage of rational design is that by adopting the protein design process, like sequence comparison, molecular docking and molecular dynamics simulation, distinct protein variants can be created rationally without excessive screening of a huge mutant library.[134][109] Computer programmes such as CAVER, HotSpot Wizard or RosettaDesign allow the *in silico* variant design and screening. These computational approaches rely on high computational power and fine-tuned algorithms to model the physical motion of substrates and enzymes, as well as the energetic contribution of mutants to reaction catalysis.[135][136] With advances in computer algorithms, rational strategies based on protein structure have been implemented to increase the stability and increase/alter substrate scope of transaminases.

1.3.2 Evolution of transaminase stability and activity

High activity and stability are desired for transaminases applied in industrial processes, but transaminases with both properties hardly occur as natural wild-type enzymes. Therefore, it is necessary to modify existing enzymes in order to combine these two properties for industrial processing requirements. For example, by error-prone PCR-based directed evolution, Martin *et al.* engineered a mesophilic *Athrobacter citreus* TA to a thermostable TA mutant after 3 rounds of screening, the best mutant exhibited good activity at an optimal temperature of 50 °C.[130] Based on predictions calculated using FoldX software, the ω-transaminase from *Variovorax paradoxus* was mutated to improve process stability, and the best mutant (G98M) had an increased melting point (Tm) to 59.3°C, while the mutant fully retained its specific activity.[137] In 2015, Nobili and coworkers reported two combination mutants of the transaminase from *Vibrio fluvialis* (F85L/V153A and Y150F/V153A), which showed 30-fold increased activity toward (*S*)-phenylbutylamine and (*R*)-phenylglycinol from their corresponding ketones.[62] In 2019, Xie *et al.* reported that a stable variant of the (*R*)-selective aminotransferase (AT-ATA) from *Aspergillus terreus*

was constructed by consensus mutagenesis, with the best variant having 16.6-fold improved thermal stability at $t_{1/2}$ and 11.8°C higher T_{50}^{10} compared to the wild-type.[112] Huang *et al.* obtained four stabilized mutants (T130M, T130F, E133F and D134L) with improved stability from 19 candidates by site-directed mutagenesis.[138]

1.3.3 Expanding the substrate scope of transaminase

All known wild-type TAs have active sites with a small and a large binding pocket. In almost all cases, the small binding pocket is restricted to accommodate only a methyl group.[113] Therefore, engineering TAs to expand the substrate scope, especially for bulky substrates, is of high synthetic relevance. In general, based on the crystal structure of the protein, these enzymes can be easily improved by further engineering using structure-based protein design. Substitution of the residues with a bulky side chain in the active site to relieve steric and torsional constraints is a very effective method to expand the substrate spectrum for bulky substrates. Pavlidis *et al.* identified a (*S*)-transaminase (3FCR) and with four mutations (Y59W/Y87F/Y152F/T213A), they were able to synthesize a set of chiral bulky amines on a preparative scale with excellent conversion, isolated yield and enantiomeric purity.[139] In another work, Genz *et al.* used bioinformatic tools to design a mutant of the *Vibrio fluvialis* ATA (L56V/W57C/F85V/V153A) capable of catalyzing the asymmetric synthesis of 2,2-dimethyl-1-phenylpropan-1-amine, with 100% conversion and an enantiomeric excess value >99% *ee*.[111] One of the most notable examples is the manufacturing of (*R*)-sitagliptin, by structure-based engineering of both substrate-binding pockets, the substrate binding pocket of ATA-117 was successfully opened up for the bulky prositagliptin ketone which was hardly accepted by the wild-type enzyme (only ~0.4% conversion).[140] As shown above, structure-based protein engineering has been utilized successfully to expand the TA substrate spectrum to substrates with bulky side chains.[118][117][141]

1.3.4 Asymmetric synthetic applications of transaminases

Transaminases were originally used in the preparation of amino acids, but with the discovery of ω-TA and ATA, they have been gradually used for the production of a wide range of chiral amines, including mainly chiral amino acids, chiral amino alcohols and chiral amines.[18][142][4][143] In order for TA to be applied in a more efficient way in asymmetric synthesis, many strategies have been developed to improve the yields. For example, as reversible reactions, TA-mediated transamination suffer from

equilibrium issues, especially in asymmetric amine syntheses.[144] To optimize TA performance in organic synthesis, a number of physical and chemical strategies have been developed and described to displace the reaction equilibrium to the product side, such as excess of co-substrate, removal of product or co-product.[145][146] Combining transaminases with one or even more biocatalysts for the production of a target compound, enzymatic cascades have been demonstrated to be advantageous in shortening reaction routes, avoiding unstable or toxic intermediates, increasing the atom efficiency and reducing the amount of waste[147][148][149] Taking advantage of this method, TAs could be applied in organic synthesis in a more efficient fashion.[150]

2 Results

2.1 Creation of a (*R*)-amine transaminase activity within an α-amino acid transaminase scaffold provides insights into enzyme evolution

The use of (*R*)-transaminases to access pharmaceuticals or building blocks with chiral amine functionalities is an attractive alternative to asymmetric chemical synthesis. Although many studies focussed on enzyme discovery, protein engineering and the application of (*R*)-selective amine transaminases ((*R*)-ATA) in biocatalysis, little is known about the actual *in vivo* role and especially how these enzymes have evolved from the ubiquitous α-amino acid transaminases (α-AAT). Furthermore, we believe that related proteins with similar overall scaffolds have the potential to catalyze a variety of different chemical transformations. To shed light on the evolutionary process, we used a combination of sequence analysis and computational protein design studies to attempt introduction of (*R*)-amine transaminase activity into the scaffold of a BCAT from *Escherichia coli* (eBCAT) and a DATA from *B. subtilis* as α-AAT scaffolds. These enzymes are well-studied and their crystal structures had been solved, but they do not show activity towards the typical ATA benchmark substrate 1-phenylethylamine (PEA).

First, we tried eBCAT as a potential scaffold for the introduction of (*R*)-ATA activity, because the overall active site is composed like those in (*R*)-ATAs. Four residues (G38, R40, Y95 and R97) were chosen for mutagenesis in eBCAT. Unfortunately, none of the single mutants showed detectable activity in the classical acetophenone assay with (*R*)-PEA as amino donor and pyruvate as acceptor. Therefore, we focused our efforts on the engineering of the D-amino acid transferases (DATA) as an alternative and complementary scaffold to the eBCAT. Even though eBCAT and DATA both belong to fold-type IV transaminases, the "carboxylate trap" of DATA is located in the O-site in contrast to an eBCAT, where they are in the P-site. Consequently, similar as for R-ATA, DATA can accept D-alanine as the native substrate. In comparison to the BCAT, the DATA is restricted to only accommodate D-amino acids, since the O-site only accepts the α-carboxylic function of the substrate. In comparison to our engineering efforts with BCAT, the challenge with DATA is the

modification of the O-site to accept phenyl functions but not be completely destroyed the coordination of the α-carboxylate.

In eBCAT, the R97E variant was the only beneficial mutant for the introduction of (*R*)-PEA acceptance. The corresponding position in the DATA scaffold is Y88, and the glutamate substitution in Y88 resulted 12 mU mg^{-1} activity in the acetophenone assay with (*R*)-PEA as amine donor and pyruvate as amino acceptor. Mutating residues F26 and Y31 also led to initial activities in the acetophenone assay. In contrast, other single mutations or combinations did not further increase the activity of (*R*)-PEA. Based on the gained knowledge that the DATA scaffold in principle can be evolved towards (*R*)-PEA, we decided to employ the RosettaDesign server for enzyme design. Then, we undertook the computational redesign of the enzyme's active site using RosettaDesign to adapt it to the recognition of PLP-(*R*)-phenylethylamine. From the results, we derived five top-scoring mutants (M1-M5) for experimental testing. These sequences were found to contain seven or eight mutations, essentially composed of aromatic and aliphatic amino acid substitutions. Unfortunately, none of the mutants showed the desired (*R*)-PEA activity in the acetophenone assay. Interestingly, variant M2 showed the acceptance of (*R*)-PEA when we assayed PMP formation after incubating the purified variant with (*R*)-PEA. The lack of activity in the acetophenone assay can be explained by the fact that the mutations might have had a detrimental effect on D-alanine coordination. From there, we employed the 3DM database to identify mutations causing the detrimental effect regarding the pyruvate acceptance from the alignment statistics. Then, we removed three amino acid substitutions from the M2 variant, the resulting variant M2-4 (Y31F/H86F/S180A/T242I) showed an activity of 18 mU mg^{-1}. Even though this variant showed low activity, the variant did not show fast PMP formation upon incubation with (*R*)-PEA as observed with M2. So, the Y88F substitution was included again [resulting in variant M2-5 (Y31F/H86F/Y88F/S180A/T242I)], rapid (*R*)-PEA-based PMP formation was restored and the activity increased to 185 mU mg^{-1} for M2-5. A further increase in activity was achieved by combining the M2-5 variant with the H100L mutation, which is also present in 15% of the aligned sequences in the superfamily of the 3DM data set. The activity of this final variant was 326 mU mg^{-1} [M2-6 (Y31F/H86F/Y88F/H100L/S180A/T242I)]. These results substantially expanded our knowledge about the relationship between amine transaminase and ω-amino acid transaminases and adds new candidates to the toolbox of suitable transaminases.

2.2 Directed evolution of an amine transaminase for the synthesis of an Apremilast intermediate via kinetic resolution

Apremilast is an important active pharmaceutical developed by Celgene for the treatment of autoimmune and inflammatory diseases. Since the (*S*)-enantiomer of Apremilast is much more potent than the (*R*)-enantiomer, the (S)-enantiomer was further developed as a drug. The chemical synthesis route of Apremilast is the resolution of racemic 1-(3-ethoxy-4-methoxyphenyl)-2-(methylsulfonyl)ethanamine (*rac*-**1**) to (*S*)-**1** with stoichiometric use of *N*-Ac-L-leucine. Alternatively, the asymmetric hydrogenation approach also can produce the (*S*)-enantiomers from the corresponding enamines or ketone with precious and toxic rhodium or ruthenium complexes (Figure 13). The only one biocatalytic approach reported for the synthesis of the (*S*)-enantiomer Apremilast is resolution of a racemic mixture via acylation by a lipase (Novozym 435), yet the *ee* of acylated product was not perfect. To provide a new catalytic and enzymatic process for Apremilast, we performed the directed evolution of the amine transaminase from *Vibrio fluvialis*.

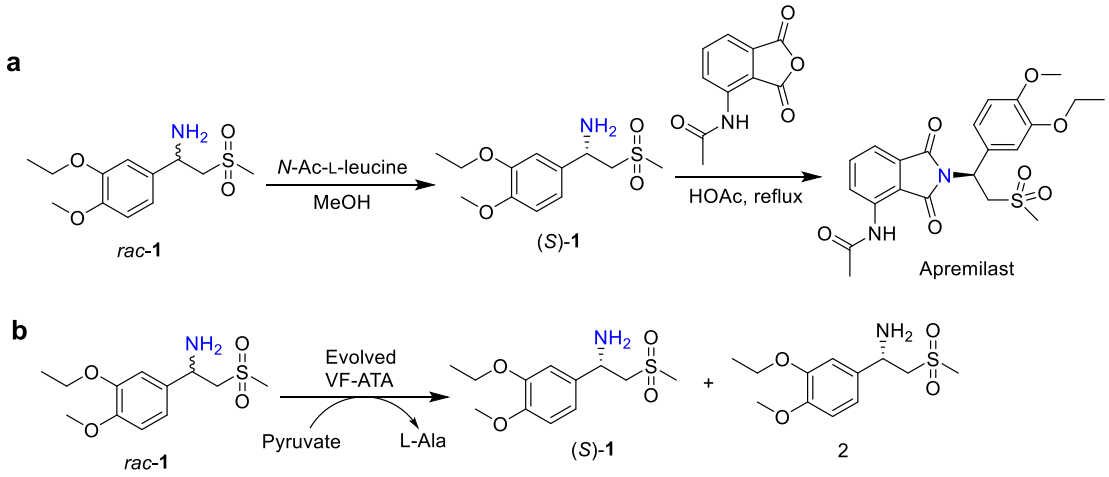


Figure 13. The original synthetic route to Apremilast involves resolution of *rac*-**1** to (*S*)-**1**. **b** Kinetic resolution of *rac*-**1** to (*S*)-**1** by the evolved VF-ATA.

Considering that kinetic resolution methods usually have low atom efficiency and the theoretical yield is only 50%, we chose asymmetric synthesis at the beginning. We screened a range of wild-type ATAs and some of their mutants available in our laboratory, and analyzed the reaction products by HPLC. The only ATAs with detectable activity was a double mutant of the ATA from *Vibrio fluvialis* (F85L/V153A, abbreviated as VF-2M), ATA117-Rd11 and 3DAA from *Bacillus subtilis* (Y31F). The

activity was determined using purified proteins of ATA117-Rd11and was 0.4 mU/mg, without by-products; 0.2 mU/mg (3DAA-Y31F) with by-products. So, we focused our effect on the engineering of the 3DAA enzyme as the starting scaffold. The NNK screening in the LBP of the DATA identified the three bulky amino acids Y31, H86, and Y88 as important hotspots to improve the *(R)*-ATA activity against (*S*)-PSOA. However, most of the variants showed an un-desired side reaction, which means that they cannot be used for asymmetric synthesis. Almost all of the variants showed the side reaction except for variant Y31H/H86V/Y88A/loop, but here acceptance of the amine acceptor molecule was a problem.

Therefore, we focused our efforts on the engineering VF-2M and its specific activity towards *rac*-PSOA for a kinetic resolution. This was determined by the acetophenone assay to be 0.62 mU/mg with pyruvate as amine acceptor. The wild-type of VF-ATA had no detectable activity (< 0.1 mU/mg), demonstrating the importance of the F85L/V153A mutations for the initial activity. To improve the activity towards *rac*-PSOA, we analyzed the crystal structure of VF-ATA (PDB: 4E3Q) and focused on the residues forming the substrate binding pocket (Figure 14). Using VF-2M as the starting template, single site-saturated mutagenesis of these 26 residues was performed with the Q5 mutagenesis method with NNK codons. For each site, 94 clones were picked and cultured in 96-deep well plates. After six rounds of evolution, we got three best variants, VF-6M-E (15.40 mU/mg), VF-7M-E (32.34 mU/mg), and VF-8M-E (39.43 mU/mg); these were selected for further characterization.

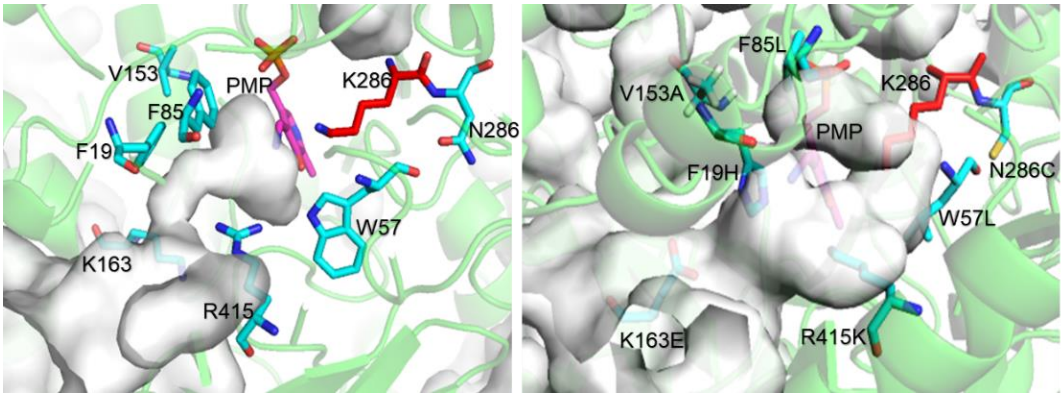


Figure 14. Structural study of VF-wt and VF-8M-E. Left, The substrate binding pocket of VF-wt; Right, the substrate binding pocket in the homology model of VF-8M-E.

Then, the best three variants were applied for the kinetic resolution of *rac*-PSOA under optimal conditions. As shown in Figure 3, 20 mM *rac*-PSOA was resolved by VF-8M-E in 24 h to reach 51% conversion, while VF-7M-E and VF-6M-E took 48 h to

reach >50% conversion. The optical purity of the remaining (*S*)-PSOA at 48 h was determined to be >99% *ee*. These results demonstrate that highly enantioselective VF-ATA variants could be designed and applied to produce the enantiopure amine intermediate for Apremilast.

2.3 A growth selection method for the directed evolution of amine forming/converting enzymes

The traditional approach for the discovery of transaminase is by enrichment culture, based on screening bacterial isolates expressing genes that can actively metabolize amines of interest as the sole nitrogen source with sugar-based media. But this concept may be limited by activity, expression levels, cell membrane mass transfer, or pH. Thus, only a few aminotransferases could be discovered by this approach (Figure 15). On the other hand, it might be possible to identify enzymes with increased activity via the growth rate of the host using the sole nitrogen source released by transaminase activity in sugar-based media. This could create or improve the activity of a transaminase for some substrates. Therefore, we hoped to use this concept for the directed evolution of enzymes for the synthesis of non-natural chiral amines.

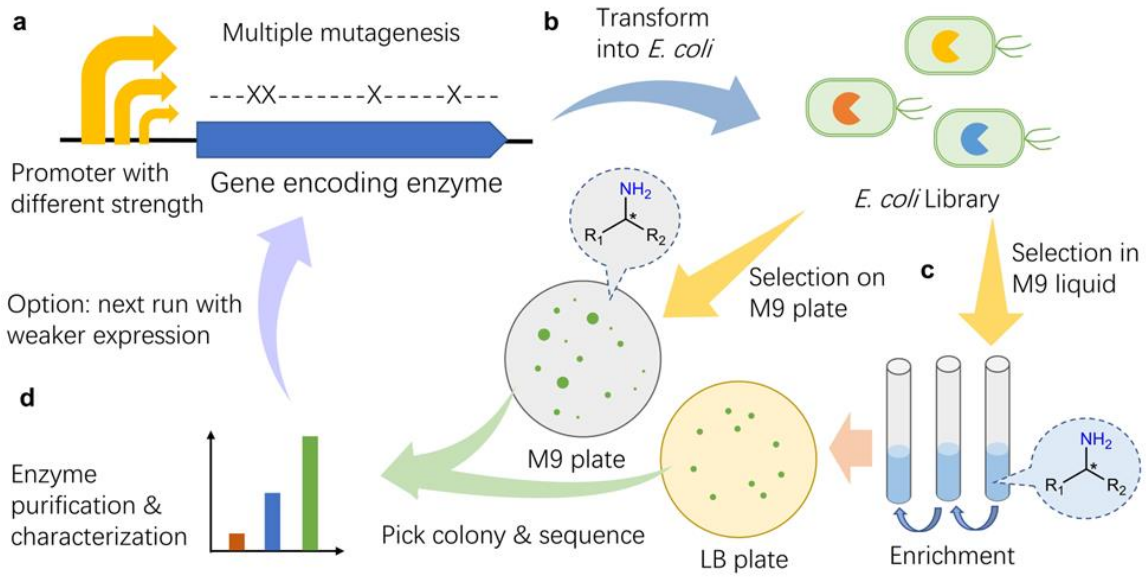


Figure 15. General procedure of the growth-based directed evolution of amine-forming/converting enzymes.

To demonstrate growth selection for the evolution of TAs, we selected the (*R*)-selective AtTA from *Aspergillus terreus* for the synthesis of (*R*)-Boc-3APP (Figure 16),

the key chiral synthon for Linagliptin, Trelagliptin, and Alogliptin. To verify the feasibility of growth selection, we cloned the AtTA(wt) gene under the control of four promoters with different strengths, and assayed the growth of the resulting transformed *E. coli* cells on M9 agar plates with (*R*)-Boc-3APP as the only nitrogen source. The results showed that all *E. coli* were unable to grow on a M9 plate supplement with (*R*)-Boc-3APP. It was expected here that the *E. coli* cell containing an active ATA variant should stand out by forming a colony on the plate if the activity of AtTA towards (*R*)-Boc-3APP was improved.

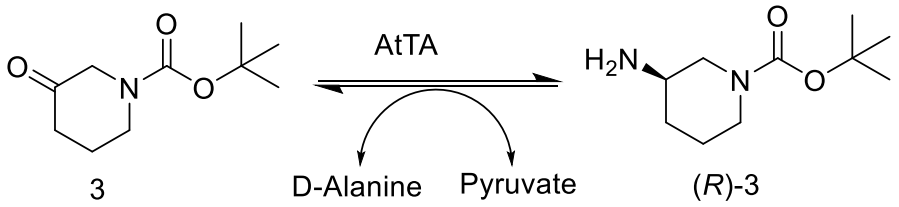


Figure 16. Targeted reaction converting **3** to *(R)*-**3** in an asymmetric synthesis.

To solve this problem, we first – according to the crystal structure of AtTA – designed three NNK libraries by addressing three hotspots (AtTA-1: H55NNK, Y60NNK, V62NNK; AtTA-4: L182NNK, W184NNK, L187NNK; AtTA-5: T274NNK, T275NNK, A276NNK) in parallel with up to 8000 substitutions (20x20x20 possible combinations); two NNK libraries by addressing two hotspots (AtTA-2: F115NNK, E117NNK, AtTA-3: V147NNK, V149NNK) in parallel with up to 400 substitutions (20x20 possible combinations). Then the libraries were screened using several different substrates as the sole nitrogen sources to identify mutants with high activity by the growth assay. By this concept, mutants with high activity for (*R*)-Boc-3APP were successfully identified.

In the first round of growth screening, we used a vector with a strong expression level promoter to produce the proteins (S-A1, S-A2, S-A3, S-A4, S-A5) resulting in thousands of different sized colonies in the screening agar plate. These obviously cannot be sequenced and confirmed one by one for activity. Thus, we transferred the library to a vector with a medium expression promoter (M-A1, M-A2, M-A3, M-A4, M-A5), and then screened again. We got colonies of different sizes on the M-A2 plate; no colonies grew on other plates. We picked 12 colonies for sequencing and identified 9 different mutants. We purified the proteins and tested the activity of all variants by the DAAO assay, the results are shown in Fig 4B. The variants F115M,

E117C(MC); F115M, E117V(MV); D5Y, F115H, E117C(YHC) showed the best activity. The best three variants were retransformed into *E. coli* to check if they can grow again on the plate with (*R*)-Boc-3APP as the sole nitrogen source. MC, MV, YHC were then used as the starting templates for the second round of mutagenesis. But, similar to the first round of growth screening, even though we used the vector with a weak expression level promoter to express the protein, thousands of different sized colonies were growing on the screening agar plate. Then, we tested the growth states of three starting templates using the vectors with different expression level promoters (medium-M, weak-w, very weak-w1, very very weak-w2). Based on these results, we choose w2 as the optimal vector to express the protein. So, we transferred the library to this vector with a very weak expression promoter (W2-YHC-1, W2-YHC-3, W2-YHC-4, W2-YHC-5), and then performed screening again.

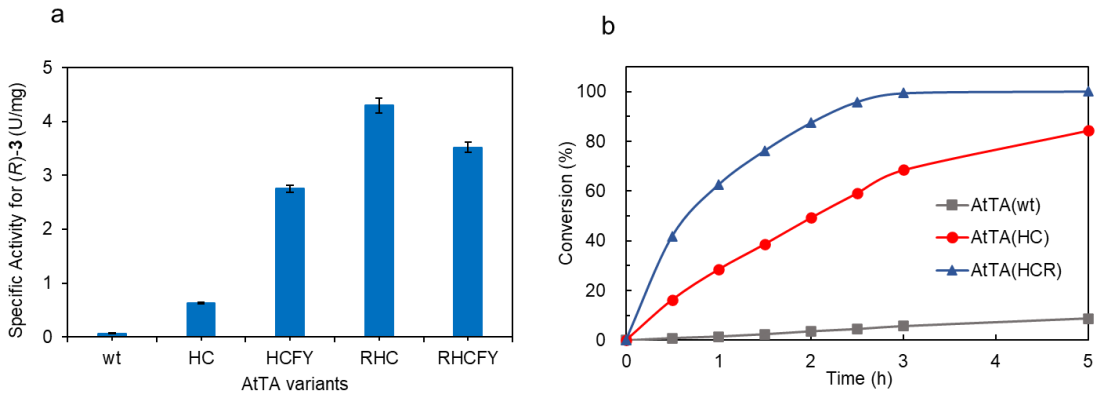


Figure 17. a, specific activity of purified AtTA variants in the conversion of (*R*)-**3** with pyruvate. b, time course of the asymmetric synthesis of *(R)*-**3** from **3** by purified AtTA variants.

Colonies with higher activity were send for sequencing and we identified 15 different mutants. Because with the w2 expression level almost no recombinant protein was produced, we cloned the gene into the pRSF-plasmid, purified the protein, tested the activity of all variants by DAAO assay, and then combined the best mutants together as well as tested the mutants (D5Y) which were deleted by random substitution. D5Y, H55R, F115H/E117C, V147F/V149Y were identified as the most important mutations. Unfortunately, the combination of the beneficial H55R and V147F/V149Y mutations did not result in an additive activity effect. The resulting variants were evaluated and AtTA(RHC) with mutations H55R/F115H/E117C gave the highest activity of 4.2 U/mg (110-fold over the wild-type) while maintaining very high enantioselectivity. The best AtTA variants were applied for the asymmetric synthesis of (*R*)-Boc-3APP: AtTA(RHC)

produced 50 g/L (*R*)-Boc-3APP (98% *ee*) in quantitative conversion in 5 h, while AtTA(wt) gave only (*R*)-Boc-3APP in 8.7% conversion under the same conditions. The preparative scale synthesis was performed with purified AtTA(RHC) to obtain (*R*)-Boc-3APP (98% *ee*) in 98% isolated yield (Figure 17).

To further demonstrate the applicability of growth selection, we have studied two other enzyme classes. We chose the cyclohexylamine oxidase (CHAO) from *Brevibacterium oxydans* and the phenylalanine ammonia lyase from *Petroselinum crispum* (PcPAL) as starting scaffolds for evolution. With proper structure-guided selection of mutation residues, a single round of evolution with growth selection was sufficient to significantly (>200-fold) improve the activity of CHAO towards the intermediate chiral amine of Cinacalcet. Combining the growth selection with protein engineering, the specific activity of a PcPAL variant towards non-natural amino acid increased 26-fold to 0.55 U/mg. Considering its outstanding features of high throughput, simple and low-equipment dependent, the growth selection can be widely adopted in academia and industry to develop suitable biocatalysts for various applications.

3 Conclusions and Outlook

Transaminases have been successfully applied to the synthesis of complex pharmaceutical intermediates due to their high enantioselectivity during the transformation. The most famous example being the synthesis of the drug Sitagliptin (to treat Diabetes type II) by Merck & Co. and Codexis from the prochiral precursor pro-Sitagliptin published in 2009, this fully demonstrates the great potential of transaminases for industrial synthesis.[140] Soon after the transaminase toolbox available for the synthesis of chiral amines was expanded in 2010 with the identification of 17 novel (*R*)-ATAs, which represented a breakthrough for the biocatalytic asymmetric synthesis of chiral amines, as only one *(R)*-transaminase had been reported before[79]. In 2015, the (*S*)-transaminase variant 3FCR-4M was reported, this variant showed highly stereoselectivity in the asymmetric synthesis of a set of bulky chiral amines making it suitable for the synthesis of a wide array of potential intermediates for pharmaceuticals.[139] Recent attention has focused on the need to overcome problems in the discovery of new transaminases as the use of TAs for the asymmetric synthesis of chiral amines from its ketone is hindered by the reaction equilibrium, nonenzymatic competitive substrate and/or product inhibition, and a poor substrate tolerance profile.

To prove that α-amino acid aminotransferases are the evolutionary origin of amine transaminases, we studied two members of the α-AAT-family of the fold-type IV of PLP-dependent enzymes, eBCAT and DATA and their respective mutants towards the acceptance of the benchmark amine (*R*)-PEA (**Article I**). By employing an in-silico approach combining computational protein design and sequence co-evolution analysis, we finally identified a DATA variant harboring six mutations, yielding in a total activity of 0.3 U mg^{-1} towards (*R*)-PEA, while maintaining overall stability. These results substantially expanded our knowledge about the relationship between amine transaminases and α-amino acid transaminases demonstrating that (*R*)-ATAs could indeed have evolved in nature from α-AAT.

Although a broad range of TA enzymes is now available, there are still some substrates that are difficult to accept. In **Article II**, to provide a new enzymatic access to the bulky chiral amine intermediate (*S*)-PSOA for the synthesis of Apremilast, we

identified a double mutant of the ATA from *Vibrio fluvialis*, which had initial activity for the conversion of *rac*-PSOA. Six rounds of directed evolution resulted in the VF-8M-E variant with a >60-fold increased specific activity. The purified VF-8M-E was successfully applied in the kinetic resolution of *rac*-PSOA to produce enantiopure (*S*)-PSOA at 49% conversion. The application of this ATA variant represents an alternative method for the synthesis of the key chiral amine intermediate for the manufacturing of Apremilast.

For the implementation of enzymatic processes, faster enzyme screening, process development and process intensification are the biggest challenges. In **Article III**, we developed a facile growth selection system generally applicable for different classes of amine-forming/converting enzymes. The principle of the methodology was based on the targeted chiral amine as the sole nitrogen source for cell growth in a chemically-defined medium. By fine-tuning a set of suitable promoters, the growth selection was demonstrated to result in a significant improvement of catalytic activity of a TA. We are convinced that the concept and method described in this work could be generally applicable to develop other enzymes (transaminases, amine oxidases, ammonia lyases, amine dehydrogenases, imine reductases, etc.) for the synthesis of non-natural chiral amines, amino acids, amides, and other nitrogen-containing molecules.

Looking ahead, further advances in transaminases and reaction engineering, machine learning, dynamic enzyme reaction modelling, and predictive retrosynthetic tools, will embed enzymatic synthesis of chiral amines as a green, sustainable, cost- and atom-efficient method for the manufacture of the ever-increasing molecular complexity across the chemical, pharma, and food industries.

4 References

[1] D. Psimadas, P. Georgoulias, V. Valotassiou, G. Loudos, *J. Pharm. Sci.* **2012**, *101*, 2271–2280.

[2] E. J. Ariëns, E. W. Wuis, E. J. Veringa, *Biochem. Pharmacol.* **1988**, *37*, 9–18.

[3] E. J. Ari, *Eur. J. Clin. Pharmacol.* **1984**, *26*, 663–668.

[4] R. N. Patel, *ACS Catal.* **2011**, *1*, 1056–1074.

[5] Y. Okamoto, T. Ikai, *Chem. Soc. Rev.* **2008**, *37*, 2593–2608.

[6] T. C. Nugent, M. El-Shazly, *Adv. Synth. Catal.* **2010**, *352*, 753–819.

[7] D. Ghislieri, N. J. Turner, *Top. Catal.* **2014**, *57*, 284–300.

[8] E. E. Ferrandi, D. Monti, *World J. Microbiol. Biotechnol.* **2018**, *34*, 1–10.

[9] M. Höhne, U. T. Bornscheuer, *ChemCatChem* **2009**, *1*, 42–51.

[10] Z. Wang, *Molecules* **2019**, *24*, 3412–3450.

[11] T. Yasukawa, R. Masuda, S. Kobayashi, *Nat. Catal.* **2019**, *2*, 1088–1092.

[12] D. J. C. Constable, P. J. Dunn, J. D. Hayler, G. R. Humphrey, J. L. Leazer, R. J. Linderman, K. Lorenz, J. Manley, B. A. Pearlman, A. Wells, A. Zaks, T. Y. Zhang, *Green Chem.* **2007**, *9*, 411–42.

[13] G. W. Zheng, J. H. Xu, *Curr. Opin. Biotechnol.* **2011**, *22*, 784–792.

[14] J. W. Song, J. H. Lee, U. T. Bornscheuer, J. B. Park, *Adv. Synth. Catal.* **2014**, *356*, 1782–1788.

[15] R. N. Patel, *Biomol. Eng.* **2001**, *17*, 167–182.

[16] S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chemie Int. Ed.* **2021**, *60*, 88–119.

[17] A. M. Bezborodov, N. A. Zagustina, *Appl. Biochem. Microbiol.* **2016**, *52*, 237–249.

[18] M. Höhne, K. Robins, U. T. Bornscheuer, *Adv. Synth. Catal.* **2008**, *350*, 807–812.

[19] R. N. Lima, C. S. dos Anjos, E. V. M. Orozco, A. L. M. Porto, *Mol. Catal.* **2019**, *466*, 75–105.

[20] C. Lamberth, *J. für Prakt. CHemie* **1994**, *336*, 632–633.

[21] V. F. Batista, J. L. Galman, D. C. Pinto, A. M. S. Silva, N. J. Turner, *ACS Catal.* **2018**, *8*, 11889–11907.

[22] M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, *Angew.*

Chemie - Int. Ed. **2002**, *41*, 3177–3180.

[23] R. Carr, M. Alexeeva, A. Enright, T. S. C. Eve, M. J. Dawson, N. J. Turner, *Angew. Chemie - Int. Ed.* **2003**, *42*, 4807–4810.

[24] T. S. C. Eve, A. Wells, N. J. Turner, *Chem. Commun.* **2007**, *2*, 1530–1531.

[25] C. J. Dunsmore, R. Carr, T. Fleming, N. J. Turner, *J. Am. Chem. Soc.* **2006**, *128*, 2224–2225.

[26] D. Ghislieri, A. P. Green, M. Pontini, S. C. Willies, I. Rowles, A. Frank, G. Grogan, N. J. Turner, *J. Am. Chem. Soc.* **2013**, *135*, 10863–10869.

[27] Z. Chen, Y. Ma, M. He, H. Ren, S. Zhou, D. Lai, Z. Wang, L. Jiang, *Appl. Biochem. Biotechnol.* **2015**, *176*, 2267–2278.

[28] I. Rowles, K. J. Malone, L. L. Etchells, S. C. Willies, N. J. Turner, *ChemCatChem* **2012**, *4*, 1259–1261.

[29] D. Ghislieri, D. Houghton, A. P. Green, S. C. Willies, N. J. Turner, *ACS Catal.* **2013**, *3*, 2869–2872.

[30] I. Slabu, J. L. Galman, R. C. Lloyd, N. J. Turner, *ACS Catal.* **2017**, *7*, 8263–8284.

[31] I. D. I. Stirling, L. Zeitlin, G. Brook, W. George, N. J, *Enantiomeric Enrichment and Stereoselective Synthesis of Chiral Amines*, **1990**, 4950606.

[32] E. S. Park, J. S. Shin, *Appl. Environ. Microbiol.* **2013**, *79*, 4141–4144.

[33] G. Shin, S. Mathew, H. Yun, *J. Ind. Eng. Chem.* **2015**, *23*, 128–133.

[34] V. Steck, D. M. Carminati, N. R. Johnson, R. Fasan, *ACS Catal.* **2020**, *10*, 10967–10977.

[35] M. Lenz, N. Borlinghaus, L. Weinmann, B. M. Nestl, *World J. Microbiol. Biotechnol.* **2017**, *33*, 1–10.

[36] G. A. Aleku, S. P. France, H. Man, J. Mangas-Sanchez, S. L. Montgomery, M. Sharma, F. Leipold, S. Hussain, G. Grogan, N. J. Turner, *Nat. Chem.* **2017**, *9*, 961–969.

[37] M. Iwaki, T. Yagi, K. Horiike, Y. Saeki, T. Ushijima, M. Nozaki, *Arch. Biochem. Biophys.* **1983**, *220*, 253–262.

[38] M. J. Abrahamson, E. Vázquez-Figueroa, N. B. Woodall, J. C. Moore, A. S. Bommarius, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3969–3972.

[39] F. Parmeggiani, N. J. Weise, S. T. Ahmed, N. J. Turner, *Chem. Rev.* **2018**, *118*, 73–118.

[40] E. Sanganyado, Z. Lu, Q. Fu, D. Schlenk, J. Gan, *Water Res.* **2017**, *124*, 527–542.

[41] S. L. Lovelock, N. J. Turner, *Bioorganic Med. Chem.* **2014**, *22*, 5555–5557.

[42] M. Höhne, S. Kühl, K. Robins, U. T. Bornscheuer, *ChemBioChem* **2008**, *9*, 363–365.

[43] A. Gomm, E. O'Reilly, *Curr. Opin. Chem. Biol.* **2018**, *43*, 106–112.

[44] P. Tufvesson, J. Lima-Ramos, J. S. Jensen, N. Al-Haque, W. Neto, J. M. Woodley, *Biotechnol. Bioeng.* **2011**, *108*, 1479–1493.

[45] E. S. Park, M. Kim, J. S. Shin, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 2425–2435.

[46] S. Schätzle, M. Höhne, E. Redestad, K. Robins, U. T. Bornscheuer, *Anal. Chem.* **2009**, *81*, 8244–8248.

[47] M. D. Truppo, J. D. Rozzell, J. C. Moore, N. J. Turner, *Org. Biomol. Chem.* **2009**, *7*, 395–398.

[48] M. D. Truppo, N. J. Turner, J. D. Rozzell, *Chem. Commun.* **2009**, 2127–2129.

[49] M. D. Truppo, J. D. Rozzell, J. C. Moore, N. J. Turner, *Org. Biomol. Chem.* **2009**, *7*, 395–398.

[50] S. Mathew, H. Yun, *ACS Catal.* **2012**, *2*, 993–1001.

[51] L. J. Guan, J. Ohtsuka, M. Okai, T. Miyakawa, T. Mase, Y. Zhi, F. Hou, N. Ito, A. Iwasaki, Y. Yasohara, M. Tanokura, *Sci. Rep.* **2015**, *5*, 1–8.

[52] P. K. Mehta, T. I. Hale, P. Christen, *Eur. J. Biochem.* **1993**, *214*, 549–561.

[53] R. Percudani, A. Peracchi, *BMC Bioinformatics* **2009**, *10*, 273.

[54] S. Schätzle, M. Höhne, K. Robins, U. T. Bornscheuer, *Anal. Chem.* **2010**, *82*, 2082–2086.

[55] F. Steffen-Munsberg, C. Vickers, H. Kohls, H. Land, H. Mallin, A. Nobili, L. Skalden, T. van den Bergh, H. J. Joosten, P. Berglund, M. Höhne, U. T. Bornscheuer, *Biotechnol. Adv.* **2015**, *33*, 566–604.

[56] S. W. Han, E. S. Park, J. Y. Dong, J. S. Shin, *Appl. Environ. Microbiol.* **2015**, *81*, 6994–7002.

[57] M. S. Weiß, I. V. Pavlidis, C. Vickers, M. Hohne, U. T. Bornscheuer, *Anal. Chem.* **2014**, *86*, 11847–11853.

[58] G. Schneider, H. Käck, Y. Lindqvist, *Structure* **2000**, *8*, 1–6.

[59] H. Käck, J. Sandmark, K. Gibson, G. Schneider, Y. Lindqvist, *J. Mol. Biol.* **1999**, *291*, 857–876.

[60] F. Guo, P. Berglund, *Green Chem.* **2017**, *19*, 333–360.

[61] D. Koszelewski, K. Tauber, K. Faber, W. Kroutil, *Trends Biotechnol.* **2010**, *28*, 324–332.

[62] A. Nobili, F. Steffen-Munsberg, H. Kohls, I. Trentin, C. Schulzke, M. Höhne, U. T. Bornscheuer, *ChemCatChem* **2015**, *7*, 757–760.

[63] U. Kaulmann, K. Smithies, M. E. B. Smith, H. C. Hailes, J. M. Ward, *Enzyme Microb. Technol.* **2007**, *41*, 628–637.

[64] I. V. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding, U. T. Bornscheuer, *Nat. Chem.* **2016**, *8*, 1076–1082.

[65] K. Inoue, S. Kuramitsu, K. Aki, Y. Watanabe, T. Takagi, M. Nishigai, A. Ikai, H. Kagamiyama, *J. Biochem.* **1988**, *104*, 777–784.

[66] A. Iwasaki, K. Matsumoto, J. Hasegawa, Y. Yasohara, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 1563–1573.

[67] M. Voss, D. Das, M. Genz, A. Kumar, N. Kulkarni, J. Kustosz, P. Kumar, U. T. Bornscheuer, M. Höhne, *ACS Catal.* **2018**, *8*, 11524–11533.

[68] K. Hirotsu, M. Goto, A. Okamoto, I. Miyahara, *Chem. Rec.* **2005**, *5*, 160–172.

[69] E. Y. Bezsudnova, K. M. Boyko, V. O. Popov, *Biochem.* **2017**, *82*, 1572–1591.

[70] E. S. Venos, M. H. Knodel, C. L. Radford, B. J. Berger, *BMC Microbiol.* **2004**, *4*, 1–14.

[71] E. Y. Bezsudnova, T. N. Stekhanova, D. A. Suplatov, A. V. Mardanov, N. V. Ravin, V. O. Popov, *Arch. Biochem. Biophys.* **2016**, *607*, 27–36.

[72] M. Goto, I. Miyahara, H. Hayashi, H. Kagamiyama, K. Hirotsu, *Biochemistry* **2003**, *42*, 3725–3733.

[73] M. J. Pucci, J. A. Thanassi, H. T. Ho, P. J. Falk, T. J. Dougherty, *J. Bacteriol.* **1995**, *177*, 336–342.

[74] S. Sugio, G. A. Petsko, J. M. Manning, K. Soda, D. Ringe, *Biochemistry* **1995**, *34*, 9661–9669.

[75] D. Peisach, D. M. Chipman, P. W. Van Ophem, J. M. Manning, D. Ringe, *Biochemistry* **1998**, *37*, 4958–4967.

[76] M. D. Patil, G. Grogan, A. Bommarius, H. Yun, *Catalysts* **2018**, *8*, DOI 10.3390/catal8070254.

[77] A. Iwasaki, Y. Yamada, N. Kizaki, Y. Ikenaka, J. Hasegawa, *Appl. Microbiol. Biotechnol.* **2006**, *69*, 499–505.

[78] D. Koszelewski, D. Clay, D. Rozzell, W. Kroutil, *European J. Org. Chem.* **2009**, 2289–2292.

[79] M. Höhne, S. Schätzle, H. Jochens, K. Robins, U. T. Bornscheuer, *Nat. Chem. Biol.* **2010**, *6*, 807–813.

[80] H. Kohls, F. Steffen-Munsberg, M. Höhne, *Curr. Opin. Chem. Biol.* **2014**, *19*, 180–192.

[81] S. Kwon, H. H. Park, *Comput. Struct. Biotechnol. J.* **2019**, *17*, 1031–1039.

[82] K. E. Cassimjee, M. S. Humble, V. Miceli, C. G. Colomina, P. Berglund, *ACS Catal.* **2011**, *1*, 1051–1055.

[83] M. S. Humble, K. E. Cassimjee, M. Hãkansson, Y. R. Kimbung, B. Walse, V. Abedi, H. J. Federsel, P. Berglund, D. T. Logan, *FEBS J.* **2012**, *279*, 779–792.

[84] M. Fuchs, J. E. Farnberger, W. Kroutil, *European J. Org. Chem.* **2015**, *2015*, 6965–6982.

[85] C. Sayer, R. J. Martinez-Torres, N. Richter, M. N. Isupov, H. C. Hailes, J. A. Littlechild, J. M. Ward, *FEBS J.* **2014**, *281*, 2240–2253.

[86] J. Hopwood, M. D. Truppo, N. J. Turner, R. C. Lloyd, *Chem. Commun.* **2011**, *47*, 773–775.

[87] S. Mathew, G. Shin, M. Shon, H. Yun, *Biotechnol. Bioprocess Eng.* **2013**, *18*, 1–7.

[88] T. Scheidt, H. Land, M. Anderson, Y. Chen, P. Berglund, D. Yi, W. D. Fessner, *Adv. Synth. Catal.* **2015**, *357*, 1721–1731.

[89] L. Skalden, M. Thomsen, M. Höhne, U. T. Bornscheuer, W. Hinrichs, *FEBS J.* **2015**, *282*, 407–415.

[90] N. N. Samsonova, S. V Smirnov, I. B. Altman, L. R. Ptitsyn, **2003**, *10*, 1–10.

[91] B. Y. Hwang, B. G. Kim, *Enzyme Microb. Technol.* **2004**, *34*, 429–436.

[92] M. D. Truppo, N. J. Turner, *Org. Biomol. Chem.* **2010**, *8*, 1280–1283.

[93] J. Hopwood, M. D. Truppo, N. J. Turner, R. C. Lloyd, *Chem. Commun.* **2011**, *47*, 773–775.

[94] A. Gomm, W. Lewis, A. P. Green, E. O'Reilly, *Chem. - A Eur. J.* **2016**, *22*, 12692–12695.

[95] M. T. Gundersen, P. Tufvesson, E. J. Rackham, R. C. Lloyd, J. M. Woodley, *Org. Process Res. Dev.* **2016**, *20*, 602–608.

[96] S. Schätzle, M. Höhne, E. Redestad, K. Robins, U. T. Bornscheuer, *Anal. Chem.* **2009**, *81*, 8244–8248.

[97] A. P. Green, N. J. Turner, E. O'Reilly, *Angew. Chemie - Int. Ed.* **2014**, *53*, 10714–10717.

[98] D. Baud, N. Ladkau, T. S. Moody, J. M. Ward, H. C. Hailes, *Chem. Commun.* **2015**, *51*, 17225–17228.

[99] T. Scheidt, H. Land, M. Anderson, Y. Chen, P. Berglund, D. Yi, W. D. Fessner, *Adv. Synth. Catal.* **2015**, *357*, 1721–1731.

[100] T. Börner, S. Rämisch, E. R. Reddem, S. Bartsch, A. Vogel, A. M. W. H.

Thunnissen, P. Adlercreutz, C. Grey, *ACS Catal.* **2017**, *7*, 1259–1269.

[101] S. W. Han, E. S. Park, J. Y. Dong, J. S. Shin, *Adv. Synth. Catal.* **2015**, *357*, 1732–1740.

[102] O. Buß, D. Muller, S. Jager, J. Rudat, K. S. Rabe, *ChemBioChem* **2018**, *19*, 2241.

[103] M. S. Humble, K. E. Cassimjee, V. Abedi, H. J. Federsel, P. Berglund, *ChemCatChem* **2012**, *4*, 1167–1172.

[104] K. E. Cassimjee, M. S. Humble, H. Land, V. Abedi, P. Berglund, *Org. Biomol. Chem.* **2012**, *10*, 5466–5470.

[105] D. Deszcz, P. Affaticati, N. Ladkau, A. Gegel, J. M. Ward, H. C. Hailes, P. A. Dalby, *FEBS J.* **2015**, *282*, 2512–2526.

[106] H. Land, J. C. Campillo-Brocal, M. Svedendahl Humble, P. Berglund, *ChemBioChem* **2019**, *20*, 1297–1304.

[107] H. Land, F. Ruggieri, A. Szekrenyi, W. D. Fessner, P. Berglund, *Adv. Synth. Catal.* **2020**, *362*, 812–821.

[108] L. Skalden, C. Peters, J. Dickerhoff, A. Nobili, H. J. Joosten, K. Weisz, M. Höhne, U. T. Bornscheuer, *ChemBioChem* **2015**, *16*, 1041–1045.

[109] D. F. A. R. Dourado, S. Pohle, A. T. P. Carvalho, D. S. Dheeman, J. M. Caswell, T. Skvortsov, I. Miskelly, R. T. Brown, D. J. Quinn, C. C. R. Allen, L. Kulakov, M. Huang, T. S. Moody, *ACS Catal.* **2016**, *6*, 7749–7759.

[110] M. Genz, C. Vickers, T. van den Bergh, H. J. Joosten, M. Dörr, M. Höhne, U. T. Bornscheuer, *Int. J. Mol. Sci.* **2015**, *16*, 26953–26963.

[111] M. Genz, O. Melse, S. Schmidt, C. Vickers, M. Dörr, T. van den Bergh, H. J. Joosten, U. T. Bornscheuer, *ChemCatChem* **2016**, *8*, 3199–3202.

[112] D. F. Xie, J. X. Yang, C. J. Lv, J. Q. Mei, H. P. Wang, S. Hu, W. R. Zhao, J. R. Cao, J. L. Tu, J. Huang, L. H. Mei, *J. Biotechnol.* **2019**, *293*, 8–16.

[113] J. Huang, D. F. Xie, Y. Feng, *Biochem. Biophys. Res. Commun.* **2017**, *483*, 397–402.

[114] X. Gao, X. Zhang, N. Zhu, Y. Mou, H. Zhang, X. Liu, P. Wei, *Appl. Microbiol. Biotechnol.* **2020**, *104*, 3959–3969.

[115] M. Svedendahl, C. Branneby, L. Lindberg, P. Berglund, *ChemCatChem* **2010**, *2*, 976–980.

[116] M. S. Weiß, I. V. Pavlidis, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding, U. T. Bornscheuer, *Org. Biomol. Chem.* **2016**, *14*, 10249–10254.

[117] M. S. Weiß, I. V Pavlidis, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding, U. T. Bornscheuer, **2017**, 1022–1026.

[118] O. BuB, M. Voss, A. Delavault, P. Gorenflo, C. Syldatk, U. Bornscheuer, J. Rudat, *Molecules* **2018**, *23*, 1–11.

[119] M. L. Contente, M. Planchestainer, F. Molinari, F. Paradisi, *Org. Biomol. Chem.* **2016**, *14*, 9306–9311.

[120] N. Van Oosterwijk, S. Willies, J. Hekelaar, A. C. Terwisscha Van Scheltinga, N. J. Turner, B. W. Dijkstra, *Biochemistry* **2016**, *55*, 4422–4431.

[121] S. W. Han, J. Kim, H. S. Cho, J. S. Shin, *ACS Catal.* **2017**, *7*, 3752–3762.

[122] S. W. Han, E. S. Park, J. Y. Dong, J. S. Shin, *Appl. Environ. Microbiol.* **2015**, *81*, 6994–7002.

[123] S. W. Han, J. S. Shin, *Biotechnol. Bioprocess Eng.* **2019**, *24*, 176–182.

[124] H. G. Kim, S. W. Han, J. S. Shin, *Adv. Synth. Catal.* **2019**, *361*, 2594–2606.

[125] A. Telzerow, J. Paris, M. Håkansson, J. González-Sabín, N. Ríos-Lombardía, M. Schürmann, H. Gröger, F. Morís, R. Kourist, H. Schwab, K. Steiner, *ACS Catal.* **2019**, *9*, 1140–1148.

[126] A. W. H. Dawood, J. Bassut, R. O. M. A. de Souza, U. T. Bornscheuer, *Chem. - A Eur. J.* **2018**, *24*, 16009–16013.

[127] Q. Meng, N. Capra, C. M. Palacio, E. Lanfranchi, M. Otzen, L. Z. Van Schie, H. J. Rozeboom, A. M. W. H. Thunnissen, H. J. Wijma, D. B. Janssen, *ACS Catal.* **2020**, *10*, 2915–2928.

[128] G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, *Angew. Chemie - Int. Ed.* **2020**, *59*, 13204–13231.

[129] F. H. Arnold, *Angew. Chemie - Int. Ed.* **2019**, *58*, 14420–14426.

[130] A. R. Martin, R. DiSanto, I. Plotnikov, S. Kamat, D. Shonnard, S. Pannuri, *Biochem. Eng. J.* **2007**, *37*, 246–255.

[131] F. Steffen-Munsberg, C. Vickers, A. Thontowi, S. Schätzle, T. Meinhardt, M. SvedendahlHumble, H. Land, P. Berglund, U. T. Bornscheuer, M. Höhne, *ChemCatChem* **2013**, *5*, 154–157.

[132] V. G. H. Eijsink, A. Bjørk, S. Gåseidnes, R. Sirevåg, B. Synstad, B. Van Den Burg, G. Vriend, *J. Biotechnol.* **2004**, *113*, 105–120.

[133] E. S. Park, J. Y. Dong, J. S. Shin, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 651–660.

[134] K. E. Cassimjee, B. Manta, F. Himo, *Org. Biomol. Chem.* **2015**, *13*, 8453–8464.

[135] A. Jurcik, D. Bednar, J. Byska, S. M. Marques, K. Furmanova, L. Daniel, P. Kokkonen, J. Brezovsky, O. Strnad, J. Stourac, A. Pavelka, M. Manak, J. Damborsky, B. Kozlikova, *Bioinformatics* **2018**, *34*, 3586–3588.

[136] J. Brezovsky, E. Chovancova, A. Gora, A. Pavelka, L. Biedermannova, J.

Damborsky, *Biotechnol. Adv.* **2013**, *31*, 38–49.

[137] O. Buß, D. Muller, S. Jager, J. Rudat, K. S. Rabe, *ChemBioChem* **2018**, *19*, 379–387.

[138] J. S. Shin, B. G. Kim, *J. Org. Chem.* **2002**, *67*, 2848–2853.

[139] I. V. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding, U. T. Bornscheuer, *Nat. Chem.* **2016**, *8*, 1076–1082.

[140] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science (80-.).* **2010**, *329*, 305–309.

[141] A. W. H. Dawood, M. S. Weiß, C. Schulz, I. V Pavlidis, H. Iding, **2018**, 3943–3949.

[142] M. Fuchs, D. Koszelewski, K. Tauber, W. Kroutil, K. Faber, *Chem. Commun.* **2010**, *46*, 5500–5502.

[143] S. Schätzle, F. Steffen-Munsberg, A. Thontowi, M. Höhne, K. Robins, U. T. Bornscheuer, *Adv. Synth. Catal.* **2011**, *353*, 2439–2445.

[144] N. Richter, R. C. Simon, H. Lechner, W. Kroutil, J. M. Ward, H. C. Hailes, *Org. Biomol. Chem.* **2015**, *13*, 8843–8851.

[145] X. Wu, M. Fei, Y. Chen, Z. Wang, Y. Chen, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7399–7408.

[146] A. P. Green, N. J. Turner, E. O'Reilly, *Angew. Chemie - Int. Ed.* **2014**, *53*, 10714–10717.

[147] H. Land, P. Hendil-Forssell, M. Martinelle, P. Berglund, *Catal. Sci. Technol.* **2016**, *6*, 2897–2900.

[148] G. Rehn, B. Ayres, P. Adlercreutz, C. Grey, *J. Mol. Catal. B Enzym.* **2016**, *123*, 1–7.

[149] M. S. Malik, E. S. Park, J. S. Shin, *Green Chem.* **2012**, *14*, 2137–2140.

[150] H. Mallin, M. Höhne, U. T. Bornscheuer, *J. Biotechnol.* **2014**, *191*, 32–37.

Curriculum vitae

Basic information

Name:	Chao XIANG
Date of birth:	Sep 18th,1992
Gender:	Male
Nationality:	Chinese
Place of birth:	Shandong, China
Email:	xiangchaozjut@yahoo.com

Education

10/2018-09/2022	University of Greifswald PhD-student (Biochemistry) In the group of Prof. Uwe T. Bornscheuer
09/2015-06/2018	Zhejiang University of Technology Master of Science in Engineering (Biochemical Engineering) In the group of Prof. Yuguo Zheng
09/2011-06/2015	Zhejiang University of Technology Bachelor of Science in Engineering (Biological Engineering)
09/2008-06/2011	Jiaonan No. 2 High School

Research Experience

10/2018-9/2022	PhD thesis in the group of Prof. Uwe T. Bornscheuer, Institute of Biochemistry, University of Greifswald *Discovery, Engineering and Application of Transaminases in Biocatalysis*
09/2015-6/2018	Master thesis in the group of Prof. Yuguo Zheng, Institute of Biotechnology and Engineering Screening and engineering of ω-transaminase and its application in biosynthesis of Sitagliptin
09/2011-6/2015	Bachelor thesis in the group of Dr. Yuguang Wang, Institute of Biotechnology and Engineering

Publications & Patents

[1] Shuke Wu, **Chao Xiang**, Yi Zhou, Mohammad Saiful Hasan Khan, Weidong Liu, Christian G. Feiler, Gert Weber, Matthias Höhne, and Uwe T. Bornscheuer, **A growth selection for the directed evolution of amine forming/converting enzymes,** *Nature Communications*, 2022, submitted.

[2] **Chao Xiang**, Shuke Wu, UweT. Bornscheuer, **Directed evolution of an amine transaminase for the synthesis of an Apremilast intermediate via kinetic resolution**, *Bioorg. Med. Chem.*, 2021, **43**, 116271.

[3] Moritz Voss*, Chao Xiang*, Jeremy Esque, Alberto Nobili, Marian J. Menke, Isabelle Andre, Matthias Höhne, and Uwe T. Bornscheuer, **Creation of (*R*)-amine transaminase activity within an α-amino acid transaminase scaffold provides insights into enzyme evolution**, *ACS Chem. Biol*. 2020, **15**, 416-424.

[4] Cheng Feng; **Xiang Chao**; Zhang Xiao-Jian; Liu Zhi-Qiang; Zheng Yu-Guo. **ReToAd: simple method for the rapid replacement of promoters to improve protein production**. *Biotechnol. Lett.*, 2018, **40**, 957-964

[5] Cheng Feng; Xu Jian-Miao; **Xiang Chao**; Liu Zhi-Qiang; Zhao Li-Qing; Zheng Yu-Guo. Simple-MSSM: a simple and efficient method for simultaneous multi-site saturation mutagenesis. *Biotechnol. Lett.*, 2017, **39**, 567-575

Greifswald, Germany, July 2022

Chao Xiang

Acknowledgements

Autumn comes and goes, time flies like an arrow. I will be gone before the first snowfall of 2022. I look back on the past four years and my feelings come over me like a tidal wave. On the occasion of the completion of my graduation thesis, I would like to express my most sincere thanks and blessings to those who have guided me and helped me.

I would first like to thank my supervisor, Prof. Uwe Bornscheuer first, for his guidance through each stage of the process. Without his enlightening instruction, impressive kindness and patience, I could not have completed my thesis. His keen and vigorous academic observation enlightens me not only in this thesis but also for my future studies.

I would also like to thank Prof. Shuke Wu and Prof. Matthias Höhne. Their knowledge of science and passion for science really impressed me. Special thanks should go to all members of the Biotechnology and Enzyme Catalysis group for the wonderful time we had together.

I can't say enough to express it.

Articles

pubs.acs.org/acschemicalbiology

Articles

Creation of (*R*)-Amine Transaminase Activity within an α-Amino Acid Transaminase Scaffold

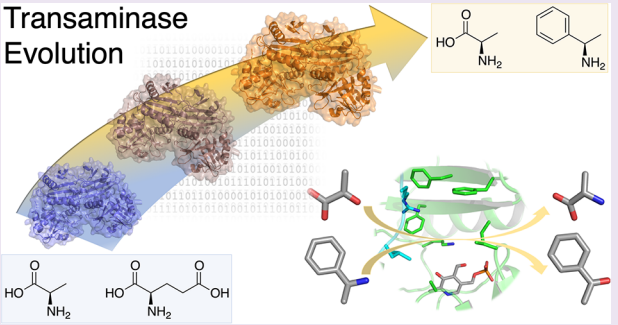
Moritz Voss, Chao Xiang, Jérémy Esque, Alberto Nobili, Marian J. Menke, Isabelle André, Matthias Höhne, and Uwe T. Bornscheuer*

Cite This: *ACS Chem. Biol.* 2020, 15, 416–424

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The enzymatic transamination of ketones into (*R*)-amines represents an important route for accessing a range of pharmaceuticals or building blocks. Although many publications have dealt with enzyme discovery, protein engineering, and the application of (*R*)-selective amine transaminases [(*R*)-ATA] in biocatalysis, little is known about the actual *in vivo* role and how these enzymes have evolved from the ubiquitous α-amino acid transaminases (α-AATs). Here, we show the successful introduction of an (*R*)-transaminase activity in an α-amino acid aminotransferase with one to six amino acid substitutions in the enzyme's active site. Bioinformatic analysis combined with computational redesign of the D-amino acid aminotransferase (DATA) led to the identification of a sextuple variant having a specific activity of 326 milliunits mg^{-1} in the conversion of (*R*)-phenylethylamine and pyruvate to acetophenone and D-alanine. This value is similar to those of natural (*R*)-ATAs, which typically are in the range of 250 milliunits mg^{-1}. These results demonstrate that (*R*)-ATAs can evolve from α-AAT as shown here for the DATA scaffold.



The family of pyridoxal 5′-phosphate (PLP)-dependent enzymes is a prime example for the natural diversity of enzyme activities that evolved within one structural scaffold. The superfamily of aspartate transaminases (fold type I) is the largest of seven known fold types. Besides enzymes catalyzing transamination, it contains activities that belong to EC classes 1−5, including racemization, decarboxylation, aldolase reactions, and a lipoamide-dependent oxidative decarboxylation and the C−C bond hydrolase kynureninase. Furthermore, different substrate specificities and stereopreferences are found among enzymes in these fold types.[1]

Transaminases are exclusively found in fold types I and IV and catalyze the reversible transfer of an amino group from an amino donor to an aldehyde, prochiral keto acid, or ketone, resulting in a (chiral) amine.[2] Transaminases are usually homodimers, with the active site located in the dimer interface. The transamination mechanism is a ping-pong bi-bi mechanism based on two half-reactions.[3] At the beginning of the catalytic cycle, the PLP is covalently bound as a Schiff base to the ε-group of the catalytic lysine in the enzyme's active site (called the internal aldimine). The first half of the reaction is initiated by the nucleophilic attack of the amino donor (e.g., alanine) where the amino donor replaces the lysine and forms the external aldimine. After abstraction of a proton from the Cα atom of the amino donor by the catalytic lysine, the planar quinonoid intermediate is formed. Via the transfer of the abstracted proton to C4′ of the cofactor, the chiral ketimine is created, which ultimately is hydrolyzed. This completes the first half-reaction with the release of the ketone while pyridoxamine 5′-phosphate (PMP) remains in the active site. In the second half-reaction, the reaction occurs in reverse order, starting with binding of the amino acceptor (e.g., a ketone) substrate.

Transaminases are highly enantioselective and substrate specific. α-Amino acid aminotransferases (AAT) are ubiquitous enzymes present in all organisms and exclusively convert α-amino acids or the respective α-keto acids, where the α-carboxylate function serves as an important recognition moiety for the substrate.[2] In contrast, the smaller subgroup of amine transaminases (ATAs) converts substrates that lack the α-carboxylic acid group, making ATAs suitable for the synthesis of industrially relevant chiral amines from prochiral ketones.[4,5] Interestingly, ATAs still accept alanine and pyruvate as co-substrates. The acceptance of both substrates with and without α-carboxylate in the same active site requires a dual-substrate recognition. In ATAs, this is facilitated by the side chain of an arginine residue, which stabilizes the negative charge of the α-carboxylate by a salt bridge, but also has the flexibility of "flipping out" of the active site in case an amine or ketone

Received: November 1, 2019
Accepted: January 28, 2020
Published: January 28, 2020

Downloaded via UNIV GREIFSWALD on June 27, 2022 at 20:09:37 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications

© 2020 American Chemical Society

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416–424

substrate lacking the α-carboxylate is binding as documented in the literature.[2,6,7]

Fold type IV of PLP-dependent enzymes is comprised of 4-amino-4-deoxychorismate lyases (ADCL) and three classes of transaminases: L-branched chain amino acid aminotransferases (BCAT), D-amino acid aminotransferases (DATA), and (*R*)-selective amine transaminases [(*R*)-ATA].[8] These different transaminase classes have a high degree of structural similarity in their overall fold but differ substantially in terms of their amino acid sequence, substrate scope, stereopreference, and activity.

The (*R*)-ATA is a group of enzymes discovered rather recently, since their activity was first reported in 2006 by Iwasaki et al., who described that the bacterial strain *Arthrobacter* sp. KNK168 can perform the whole-cell asymmetric synthesis of (*R*)-3,4-dimethoxyamphetamine.[9] In 2010, Savile et al. published the amino acid sequence of ATA117, a homologue of the transaminase that accounts for the (*R*)-ATA activity in *Arthrobacter* sp. KNK168, together with the extensively engineered variant ATA117-11Rd.[4] This variant harbors 27 mutations and was developed by Codexis and Merck & Co. for the asymmetric synthesis of the antidiabetic drug (*R*)-sitagliptin.[4] In parallel, Höhne et al. identified 17 novel (*R*)-ATAs by developing an algorithm for the sequence motif-based screening of protein data banks based on rational assignments.[8] At the time of that study, no protein or structural information for (*R*)-ATAs had been reported. Therefore, it was challenging to predict a sequence motif that correlates with (*R*)-ATA activity. The authors interrogated the protein sequences and structures of ADCLs, BCATs and DATAs, assuming that (*R*)-ATAs would also be a member of fold type IV. They based this assumption on the substrate coordination of DATA and BCAT, which share an overall similar active site architecture as assigned for the potential (*R*)-ATA.

Recently, α-AATs of fold type IV showing side activities toward (*R*)-amines were characterized.[10−14] Pavkov-Keller et al. described the DATA from *Curtobacterium pusillum* (*CPU*TA1),[13] and the thermostable BCAT from *Thermobaculum terrenum* (*TaTT*) was described by Bezsudnova et al.[11,14] Both of these transaminases exhibit the expected α-AAT activity, and significant activity toward (*R*)-amines was also observed, which distinguishes them from typical DATAs and BCATs. In addition, their sequence motifs differ from typical DATAs and BCATs, reflecting their unusual (*R*)-amine acceptance.[11]

Intrigued by the results of Höhne et al., we investigated whether it is possible to create (*R*)-ATAs from an α-transaminase by introducing key amino acid substitutions. Our major motivation was to develop a deeper understanding of how (*R*)-ATAs might have evolved by natural evolution within the α-AAT scaffold. Our overall engineering strategy was built upon the guidelines developed by Höhne et al. in combination with the now available crystal structures, which we and other groups have determined in the meantime [Protein Data Bank (PDB) entries 3WWH, 4CE5, 4CHI, 4CMD, and 6FTE].[15−19] These structures enable identification of key residues for (*R*)-amine coordination in the fold type IV class and eventually allow light to be shed on the evolutionary origin of (*R*)-ATAs. Key differences between (*R*)-ATAs and α-AATs were observed in the shape and polarity of the active site and the substrate coordination. On the basis of this structural information, we used a combination of sequence analysis and computational protein design studies to introduce (*R*)-ATA activity into the scaffold of an α-AAT from the fold type IV class of PLP-dependent enzymes. We chose the BCAT from *Escherichia coli* (*e*BCAT) and the DATA from *Bacillus subtilis* as α-AAT scaffolds, because these enzymes are well-studied, they do not show activity toward the typical ATA benchmark substrate 1-phenylethylamine (PEA), and their crystal structures had been determined (PDB entries 1IYE and 3DAA, respectively).[20,21]

RESULTS AND DISCUSSION

A Single Mutation Generates Initial (*R*)-ATA Activity on the *e*BCAT Scaffold. The *e*BCAT serves as a potential scaffold for the introduction of (*R*)-ATA activity, because the overall active site is composed like those in (*R*)-ATAs. As shown in Figure 1, the active site of both transaminases

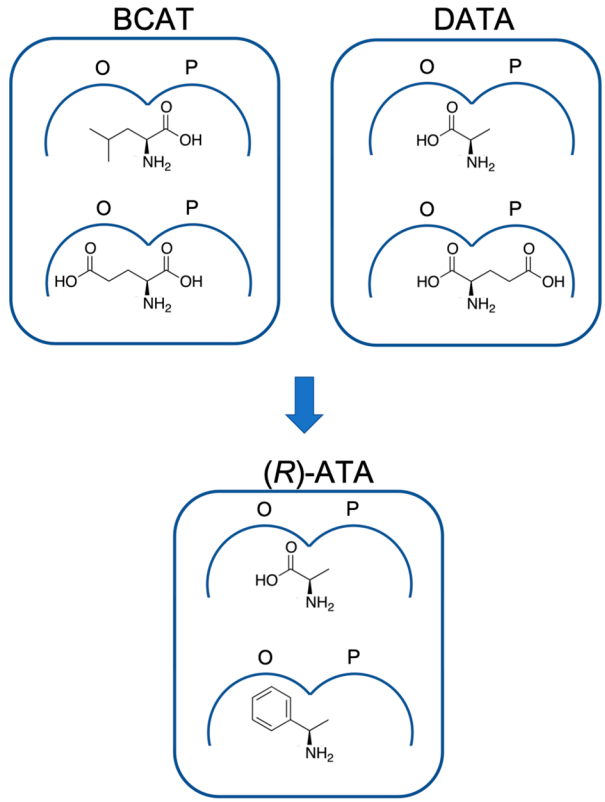


Figure 1. Pocket model of the active sites and the corresponding substrates of BCAT, DATA, and (*R*)-ATA.

consists of two binding pockets. The large binding pocket (O-site; active site region above PLP's hydroxyl group) of the *e*BCAT and (*R*)-ATAs can coordinate nonpolar alkyl and aryl groups, as well as carboxylic acids (the "dual-substrate recognition" mentioned above). Despite their similar architecture, the active site residues and properties differ according to their functions. The small binding pocket (P-site) of the *e*BCAT is restricted to accommodate the polar carboxylate function of L-amino acids, whereas the small binding pocket (P-site) of (*R*)-ATAs is limited to small nonpolar methyl or ethyl substituents.

To introduce the acceptance of (*R*)-PEA into the *e*BCAT scaffold, the mutation of the restricted P-site should be prioritized. Due to the reported activity toward L-phenylglycine, it can be assumed that the O-site can accept the phenyl

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

function of (*R*)-PEA.[22] Because the α-carboxylate in the P-site of *e*BCAT is coordinated without the direct coordination by basic amino acid side chains (further details are described in the Supporting Information), the acceptance of methyl substituents should be feasible by minor mutations in the hydrogen bond network. The substitution of the α-carboxylate function from an L-amino acid by a methyl group would already result in an (*R*)-amine as a switch in the designation of the absolute configuration takes place according to the CIP nomenclature.

Therefore, residues G38, R40, Y95, and R97 were chosen for mutagenesis in *e*BCAT, because these residues can account for the subtle α-carboxylate coordination in the P-site. To introduce the acceptance of (*R*)-PEA lacking the mandatory carboxylate function, the corresponding amino acids of the (*R*)-ATA from *Aspergillus fumigatus* were introduced, resulting in the G38V, R40S, Y95F, and R97E single mutants. Unfortunately, none of the single mutants showed detectable activity in the classical acetophenone assay with (*R*)-PEA as the amino donor and pyruvate as the acceptor.[23] Notably, PMP formation by the *e*BCAT R97E variant was observed, once the amino donor was added to the purified enzyme solution. This is evidence that, although no product formation was detected in the acetophenone assay, the first half-reaction was completed, and the amino donor is in principle accepted as it has reacted with the PLP cofactor to form the PMP. To enable catalysis, α-ketoisocaproate (the α-keto acid corresponding to the natural substrate leucine) was used as the amino acceptor instead of pyruvate. We were pleased to observe in this case an initial activity of 3.5 milliunits (mU) mg^{-1} in the spectrophotometric assay. Molecular modeling analysis revealed that the introduction of the R97E mutation created more space in the O-site, which could favor the binding of the phenyl ring in this pocket and the binding of the methyl group in the P-site (Figure S17). The R97E mutation caused the loss of activity in the native *e*BCAT reaction (with L-leucine and α-ketoglutarate in the glutamate dehydrogenase assay), because only 0.8% of WT activity (130 mU mg^{-1}) was detected. Unfortunately, further attempts to increase the activity of the R97E variant based on rational design or random mutagenesis did not lead to better variants (see the Supporting Information for details). Therefore, we considered using the D-amino acid transferases (DATA) as an alternative and complementary scaffold to the *e*BCAT.

Symmetrically with Respect to *e*BCAT, Y88 Is the Key Position for Allowing Initial (*R*)-ATA Activity in DATA. The D-amino acid transaminase (DATA, PDB entry 3DAA) from *B. subtilis* catalyzes the transfer between several D-amino acids and their corresponding α-keto acids. D-Alanine serves as the native amino donor for the synthesis of D-glutamate, both essential components for bacterial cell wall synthesis.[21] Therefore, the substrate specificity of only one of the two half-reactions needs to be altered. In comparison to our engineering efforts with BCAT, the challenge with DATA is that the coordination of the substrate's α-carboxylate must not be completely destroyed but needs to be modified in a way that both types of substrates, with or without an α-carboxylate, will be accepted.

The O-site of the DATA is exclusively coordinating the α-carboxylate function of the substrates by the polar side chains of residues Y31, R98, and H100. Therefore, Peisach et al. introduced the term "carboxylate trap" for this pocket.[21] The R98 side chain seems to have the strongest impact on D-alanine coordination, because the arginine residue forms a salt bridge with the α-carboxylate of alanine. Mutational studies by Kishimoto et al. showed that the R98M mutation yielded decreases in the specific activity (V_{max}) toward D-alanine and α-ketoglutarate of 4 orders of magnitude.[24] Besides R98, Y31 was also shown to drastically impact the D-alanine acceptance by Barber et al.[25] H100 also participates in the coordination of the α-carboxylate and completes the carboxylate trap of the DATA in the O-site.

The P-site harbors the methyl side chain of D-alanine and the γ-carboxylate function of D-glutamate. The P-site is formed by residues V33, S240, and T242.[26] These residues create a rather small binding pocket, but it is large enough to accommodate a wide variety of D-amino acid side chains.[21] Whether the glutamate's γ-carboxylate is coordinated by active site amino acids or remains in the solvent is currently not known.

To introduce (*R*)-PEA activity into the DATA scaffold, we chose the first coordination sphere of the active site as the target residues for mutation. In *e*BCAT, the R97E variant was the only beneficial one, leading to the introduction of (*R*)-PEA acceptance and hence activity. The corresponding position in the DATA scaffold is Y88, and we speculated that a substitution with glutamate could also have a beneficial effect here. Additionally, we sought to install a similar dual-substrate recognition mechanism as observed in ω-transaminases such as lysine-ε-aminotransferase (PDB entries 2CJD and 2CJH). In this enzyme, a positively charged guanidino group from R422 coordinates the α-carboxylate of α-ketoglutarate. To accept substrates lacking the α-carboxylate (e.g., the terminal amino group of lysine), a flexible glutamate residue E243 changes its orientation into a position where it neutralizes the positive charge of the arginine's guanidino group.[2,27] We speculated that a similar mechanism should also work in our DATA scaffold for the dual-substrate recognition (see also Figure S16). For this, the R98 residue could "switch" toward the introduced glutamate at position 88, coordinate with the γ-carboxylate for balancing the positive charge, and thus allow a nonpolar phenyl function to be coordinated in the O-site. Indeed, the glutamate substitution resulted in 12 mU mg^{-1} of activity in the acetophenone assay. Other single mutations or combinations did not further increase the (*R*)-PEA activity. Mutating residues F26 and Y31 also led to initial activities in the acetophenone assay (see Table S4).

On the basis of the gained knowledge that the DATA scaffold in principle can evolve toward (*R*)-PEA and its active site tolerates several mutations, we decided to follow a computer-aided enzyme design approach combined with amino acid co-evolution analysis. This should guide the selection of most beneficial amino acid mutations to broaden the substrate selectivity of the DATA. Here, the O-site was prioritized in our engineering strategy, because a more promiscuous binding pocket would be needed for the accommodation of the phenyl moiety of (*R*)-PEA.[28] This could be achieved by mutating the carboxylate trap forming residues Y31, R98, and H100 for (*R*)-PEA acceptance while keeping the D-alanine coordination intact.

The Combination of Computational Protein Design and Sequence Analysis Led to Wild Type-like (*R*)-ATA Activity in DATA. Starting from the crystal structure of D-amino acid aminotransferase determined in complex with the natural substrate D-alanine (PDB entry 3DAA),[21] we constructed a three-dimensional (3D) model of the enzyme

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

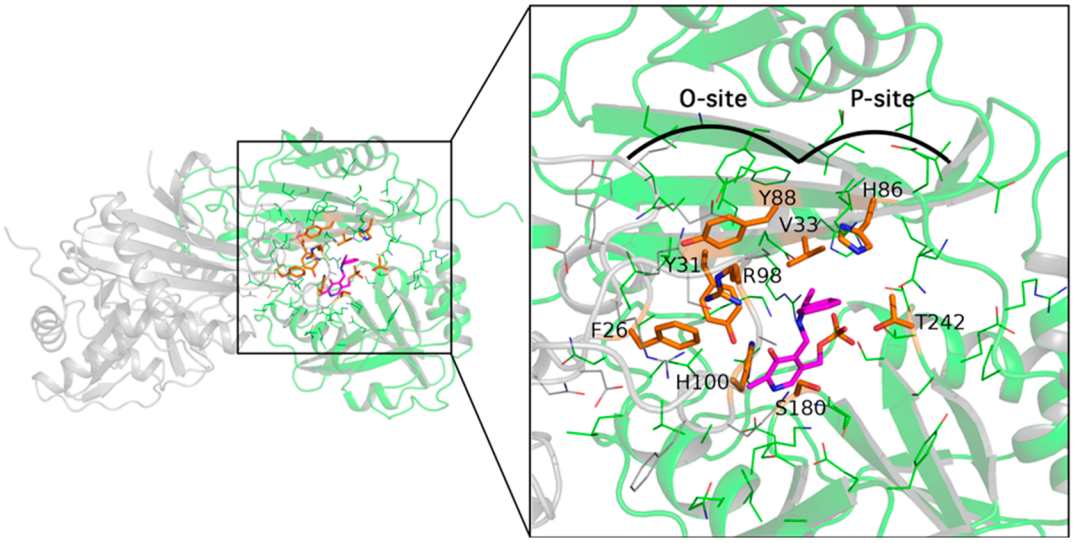


Figure 2. View of the overall three-dimensional organization of D-amino acid aminotransferase. On the left, the dimeric structural organization is shown with protein as a cartoon (green and gray). The right panel is a close-up of the catalytic site showing the mutable residues (orange sticks) and the flexible ones (green and gray lines) during RosettaDesign. In each panel, the PLP-(*R*)-phenylethylamine is represented as magenta sticks.

in complex with our substrate of interest, the PLP-(*R*)-phenylethylamine. Next, we undertook the computational redesign of the enzyme's active site using RosettaDesign[29] to adapt it to the recognition of PLP-(*R*)-phenylethylamine. Upon visual analysis of the amino acid residues surrounding PLP-(*R*)-phenylethylamine, we selected nine mutable (or designable) positions: Y31, V33, H86, Y88, S180, and T242 from the main subunit of the catalytic site (green cartoon in Figure 2) and F26, R98, and H100 from the adjacent subunit (gray cartoon in Figure 2). The remaining amino acid residues located within an 8 Å sphere centered on the ligand were defined as flexible through a rotamer library to allow side chain rearrangements during the design procedure. To explore combinatorial space, we then considered an enzyme design protocol based on the sampling of the nine positions by performing 20,000 independent runs of Rosetta *Enzyme_design*, which takes into account the backbone and side chain flexibility of "mutable" and "flexible" residues. The 3D models corresponding to the 20,000 designed sequences were subsequently rescored toward PLP-D-alanine derived from the crystallographic complex. For each complex, the total Rosetta score was plotted against the ligand energy contribution (ligand Rosetta score) for the PLP-(*R*)-phenylethylamine and the PLP-D-alanine ligands (Figure S15). The top-scoring sequences toward both ligands (corresponding to the bottom left quadrants) were further analyzed, leading to the identification of a small library of common sequences, of which the five top-scoring ones were selected for experimental testing. These sequences were found to contain seven or eight mutations, essentially composed of aromatic and aliphatic amino acid substitutions (Table S5).

Unfortunately, none of the proposed variants [M1−M5 (Table S5)] showed the desired (*R*)-PEA activity in the acetophenone assay with pyruvate as the amine acceptor. Interestingly, variant M2 exhibited some (*R*)-PEA acceptance when we assayed PMP formation after incubating the purified variant with (*R*)-PEA, which indicated a completed first half-reaction. The lack of activity in the acetophenone assay can be explained by the fact that the mutations might have had a detrimental effect on D-alanine coordination.

To identify mutations with a detrimental effect regarding the pyruvate acceptance from computational design results, we refined the variant M2 based on 3DM alignment statistics (Figures S3−S11). 3DM is a database that aligned sequences on the basis of their superfamily structure.[2,30] This allows the comparison of sequences with a low degree of sequence identity, because the alignment is based on the 3D position of the respective residue. The "D-aminotransferase PLP (2015)" database employed here includes 21,980 aligned sequences. Interestingly, only a few sequences containing W98 or W100 and no sequence harboring F88 were found. After these three mutations had been removed from the M2 variant, the resulting variant M2-4 (Y31F/H86F/S180A/T242I) resulted in an activity of 18 mU mg^{-1} (Table 1). Despite this still low activity, the variant did not show fast PMP formation upon incubation with (*R*)-PEA as observed with M2. However, when the Y88F substitution was included again [resulting in variant M2-5 (Y31F/H86F/Y88F/S180A/T242I)], rapid (*R*)-PEA-based PMP formation was restored and the activity increased to 185 mU mg^{-1} for M2-5. This clearly highlights the

Table 1. Activities of DATA Variants and Apparent Melting Temperatures

variant	(*R*)-amine activity[a] (mU mg^{-1})	native activity[b] (mU mg^{-1})	app. T_m (°C)
WT	0.2 ± 0.1	610 ± 30	70.2
Y31F	8 ± 1	208 ± 5	69.9
H86F	0.2 ± 0.1	nm	nm
Y88E	12 ± 1	nm	62.9
Y88F	0.6 ± 0.1	nm	nm
S180A	0 ± 1	nm	nm
M2-3	41 ± 4	210 ± 20	64.4
M2-4	18.3 ± 0.1	nm	69.9
M2-5	185 ± 7	17 ± 1	65.7
M2-6	326 ± 3	9 ± 4	69.1
M2	0.3 ± 0.1	6 ± 4	nm

[a]Initial activity measured in the acetophenone assay [(*R*)-PEA and pyruvate]. [b]Initial activity measured in the D-amino acid oxidase assay with D-glutamate and pyruvate. nm, not measured.

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

beneficial effect of Y88F on the (*R*)-PEA acceptance, similar to the results obtained for the *e*BCAT scaffold. Interestingly, the Y88F mutation had no activity as a single variant and is not present in any variant in the 3DM data set (Figure S7). A further increase in activity was achieved by combining the M2-5 variant with the H100L mutation, which is also present in 15% of the aligned sequences in the superfamily of the 3DM data set (Figure S9). The activity of this final variant is 326 mU mg^{-1} [M2-6 (Y31F/H86F/Y88F/H100L/S180A/T242I)] (Table 1). We were pleased to find that the incorporation of the six mutations of M2-6 had no detrimental influence on the apparent melting point of the DATA enzyme (Table 1).

Interestingly, the introduced mutations beneficial for the (*R*)-amine activity gradually decreased the activity toward the native substrate D-glutamate (Table 1). Whereas variant M2-3 (Y31F/H86F/Y88F) shows significant activity toward (*R*)-PEA and retains 34% of the native activity, the final variant M2-6 shows almost complete depletion of the native DATA activity but high activity toward (*R*)-PEA.

In all of these variants, residue R98 was kept unmutated, because this residue is probably the most important one in the "carboxylate trap" to enable D-alanine acceptance.[24] Additionally, the side chain of R98 lies in a position similar to that of R126 in *A. fumigatus*, which was shown to be an important residue for mediating the dual-substrate recognition and D-alanine acceptance.[7]

Overall, the mutations introduced into the M2-6 variant were of a nonpolar nature (aromatic and aliphatic amino acid side chains), leading to a notable increase in the active site volume with respect to that of the parental enzyme, favoring thus binding of the bulkier and more hydrophobic (*R*)-phenylethylamine compared to D-alanine.

To better understand the increased activity of variant M2-6 at molecular and energetic levels, binding free energy calculations were performed using the MM/PBSA method on the wild type enzyme and the M2-6 variant in complex with PLP-D-alanine or PLP-(*R*)-PEA (Figure 3 and Table S6). The contribution of the active site amino acid residues to the binding free energy is shown in Figure 3, whereas Table S6 lists all key residues (hot spots).

These calculations predicted that the PLP-D-alanine binding free energy is nearly of the same order of magnitude for both wild type and M2-6 enzymes (−50 ± 11 and −41 ± 9 kcal mol^{-1}, respectively). Consistent with the experimental results, the binding free energy value predicted for PLP-(*R*)-PEA in M2-6 was found to be −52 ± 10 kcal mol^{-1}, while it was predicted to be only −26 ± 7 kcal mol^{-1} in the wild type enzyme. Analysis of the amino acid contribution to the binding free energy revealed that most of designable residues (Y31, R98, H100, and T242) contribute in fact unfavorably to PLP-(*R*)-PEA binding in the wild type enzyme (Figure 3 and Table S6). Interestingly, analysis of MD simulation also suggested that the torsion of the ketimine (rectangle in Figure 3B) could be essential for favoring a catalytically productive orientation of the substrate. These results appear to be consistent with the reaction mechanism in which the ketimine is converted to the quinonoid (planar structure).

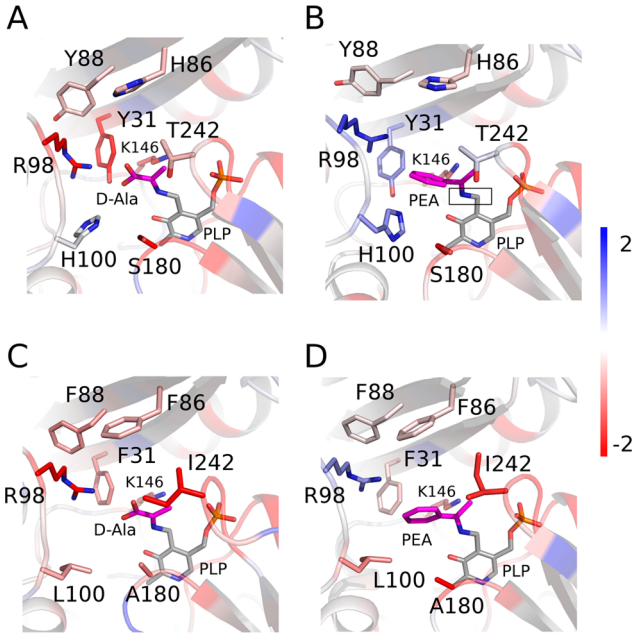


Figure 3. Binding free energy prediction using the MM/PBSA method. The contribution for each amino acid residue is projected on the 3D structures of the four complexes, i.e., wild type complexed with (A) PLP-D-alanine or (B) PLP-(*R*)-phenylethylamine and M2-6 complexed with (C) PLP-D-alanine or (D) PLP-(*R*)-phenylethylamine. The black rectangle in panel B highlights the torsion of the imine (ketimine form). Each snapshot was taken from the end of the MD simulation (5 ns), and the free energy values are given in kilocalories per mole.

CONCLUSIONS

The ubiquitous α-AATs are believed to be the evolutionary origin of ATAs. Therefore, we studied two members of the α-AAT of fold type IV of PLP-dependent enzymes, *e*BCAT and DATA, and their respective variants for the acceptance of the benchmark amine substrate (*R*)-PEA. The *e*BCAT was initially thought to be a suitable scaffold for the acceptance of (*R*)-PEA. We hypothesized that with a change in the P-site in the active site, the acceptance of a methyl function instead of the α-carboxylate (of the L-amino acid substrate) would lead to an (*R*)-ATA. However, the single mutation R97E was enabling only a very low level of (*R*)-PEA acceptance while depleting the native *e*BCAT activity. A further increase in activity by additional mutations could not be achieved. In contrast, we were able to evolve the DATA toward a moderately active (*R*)-ATA by mutating residues in the active site. By employing an *in silico* approach combining computational protein design and sequence co-evolution analysis, we finally identified a DATA variant harboring six mutations, yielding a total activity of 326 mU mg^{-1} toward (*R*)-PEA, while maintaining overall stability. Although from our experiments it is not possible to trace back the exact way in which ATAs have occurred during natural evolution, the shown DATA engineering deepens our understanding of how substrate specificity in α-AAT is affected and can be changed toward accepting amines. In addition, our results are in agreement with the evolutionary IAD model (innovation−amplification−divergence).[31] According to this theory, the original enzyme-encoding gene acquires mutations that introduce side activities (innovation) while maintaining its original function. This leads to a promiscuous enzyme, because it is catalyzing the original reaction (maybe with decreased activity) and the newly acquired side activity. In our study, variants like Y31F and M2-3 showed promiscuous activity. The divergence toward M2-6 is accompanied by the depletion of the native DATA activity.

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

METHODS

Plasmid Constructs. The *e*BCAT-encoding gene (GenBank entry YP_026247.1) was subcloned in the pET28b vector between the NdeI and XhoI restriction sites, resulting in a C-terminal His-tagged variant. The DATA gene (GenBank entry AAA22252.1) was cloned in the pGASTON vector between the NdeI and BamHI restriction sites with a C-terminal His tag. The D-amino acid oxidase from the *Trigonopsis variabilis* (*Tv*_DAAO; GenBank entry CAA90322.1) gene was ordered codon-optimized (for *E. coli* codon usage) from BioCat GmbH cloned into the pETDuet-1 vector between the EcoRI and HindIII restriction sites in the multiple cloning site 1 with an N-terminal His tag. The open reading frames and translated protein sequences of the three genes are given in the Supporting Information.

Site-Directed Mutagenesis. The creation of *e*BCAT and DATA mutants was performed using the QuikChange protocol (Agilent) using complementary primers with the desired codon and 3′ and 5′ homologue flanking regions. The primers were designed by using the standard settings of the web-based program PrimerX and are listed in Tables S1 and S2. For the polymerase chain reaction (PCR), 0.25 ng μL^{-1} template plasmid (carrying the DATA or eBCAT gene), 0.5 μM forward and reverse primers, 0.2 mM dNTPs, and 16 mU μL^{-1} *PfuPlus*! DNA polymerase (roboklon) in *Pfu* reaction buffer were used. The PCR was performed as follows: (a) 95 °C for 2 min, (b) 25 cycles at 95 °C for 30 s, 63 °C for 1 min, and 72 °C for 1 min kbp^{-1}, and (c) 72 °C for 10 min. The resulting PCR product was directly digested with the DpnI restriction enzyme (20 $\mu L\ mL^{-1}$; NEB) for 2 h at 37 °C followed by subsequent heat inactivation at 80 °C for 20 min. The digested PCR product was transformed in chemically competent *E. coli* TOP10 cells (Invitrogen), and single colonies were sequenced (Eurofins MWG GmbH) to verify the introduced mutation. The introduction of additional mutations was realized by iterating the described protocol.

Protein Expression and Purification of the Enzyme Variants. For the expression of the *e*BCAT and DATA variants, the plasmid constructs were transformed into *E. coli* BL21(DE3) cells, and these were incubated overnight at 37 °C in a 5 mL LB medium (Lysogeny Broth) preculture supplemented with the respective antibiotics (100 $\mu g\ mL^{-1}$ ampicillin or 50 $\mu g\ mL^{-1}$ kanamycin for the *e*BCAT and DATA, respectively). One milliliter of the preculture was used for the inoculation of 100 mL of TB medium (supplemented with the corresponding antibiotic) and incubated at 37 °C and 180 rpm. The expression of the *e*BCAT and DATA variants was induced at an optical density of approximately 0.6 at 600 nm with 0.2% L-rhamnose or 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG), respectively, and each was incubated at 20 °C overnight. The *Tv*_DAAO gene was expressed using the described protocol of Barber et al. at 17 °C overnight in TB medium.[25] The cells were harvested by centrifugation (20 min at 4,000 *g* and 4 °C).

For purification, the *e*BCAT and DATA variants were resuspended in 50 mM HEPES buffer (pH 7.5) containing 0.1 mM PLP, 0.3 M NaCl, and 0.01 M imidazole and lysed via ultrasonic treatment (50% pulse, 50% power, 2 × 5 min; Sonoplus HD2070, Bandelin Electronic GmbH). The lysate was clarified by centrifugation (1 h at 10,000 *g* and 4 °C) and purified by immobilized metal affinity chromatography with the following buffers: washing buffer [50 mM HEPES buffer (pH 7.5) containing 0.1 mM PLP, 0.3 M NaCl, and 0.02 M imidazole] and elution buffer [50 mM HEPES buffer (pH 7.5) containing 0.1 mM PLP, 0.3 M NaCl, and 0.3 M imidazole]. The protein was desalted in 50 mM HEPES buffer (pH 7.5) and 0.1 mM PLP using the PD-10 desalting column (GE Healthcare). The *Tv*_DAAO was lysed and purified in an analogous manner, but in 50 mM phosphate buffer (pH 8.0) with 20 μM FAD^+. The purified and desalted proteins were stored at −20 °C in 30% glycerol.

Acetophenone Assay. The activities of the purified *e*BCAT or DATA variants regarding the conversion of 1-phenylethylamine were determined photometrically by applying the acetophenone assay on an Infinite 200 PRO (TECAN) or FLUOstar Omega (BMG LABTECH GmbH) plate reader in UV-transparent microtiter plates (UV-Star, Greiner Bio-One GmbH).[23] The assay was performed with 2.5 mM (*S*)- or (*R*)-PEA as amine donors and 2.5 mM pyruvate or 2.5 mM α-ketoisocaproate as amine acceptors in 1.25−2.5% dimethyl sulfoxide (DMSO) and 50 mM CHES buffer (pH 9.0) at 30 °C. The formation of acetophenone was quantified by following the increase in absorption at 245 nm over time. One unit (U) was defined as the formation of 1 μmol of acetophenone per minute. All measurements were performed in triplicate.

Glutamate Dehydrogenase Assay. The activities of the purified *e*BCAT variants regarding the form of NADH were determined photometrically by applying the glutamate dehydrogenase assay on the FLUOstar Omega (BMG LABTECH GmbH) plate reader in microtiter plates (200 μL assay volume). The assay was performed with 2.5 mM amino donor (L-leucine), 2.5 mM α-ketoglutarate as the amine acceptor, 5 U mL^{-1} glutamate dehydrogenase (Sigma-Aldrich), 500 μM NAD^+, and 50 mM CHES buffer (pH 9.0). The formation of NADH was detected by following the increase in absorption at 340 nm over time.[32]

D-Amino Acid Oxidase Assay. The native activity of the purified DATA variants was measured using the D-amino acid oxidase assay described by Barber et al.[25] The reaction was performed in microtiter plates (200 μL assay volume) with 2.5 mM D-glutamate as the amine donor and 2.5 mM pyruvate as the amine acceptor, 0.7 U mL^{-1} D-amino acid oxidase from *T. variabilis* (*Tv*_DAAO), 22 U mL^{-1} HRP (Sigma-Aldrich), and 50 μM Ampliflu Red in 50 mM CHES buffer (pH 9.0) with 0.5% DMSO and 2.5% ethanol as co-solvents. The formation of Resorufin was quantified by following the increase in absorption at 560 nm over time on the FLUOstar Omega (BMG LABTECH GmbH) plate reader.

Analysis of the Half-Reaction by PMP Formation. The formation of PMP was determined photometrically on the FLUOstar Omega (BMG LABTECH GmbH) plate reader in microtiter plates (200 μL assay volume). The reaction was performed with 50 μL of purified transaminase, 2.5 mM (*R*)-PEA in 1.25% ethanol, and 50 mM CHES buffer (pH 9.0). The formation of PMP was detected by following the change in absorbance from 290 to 470 nm.

Melting Point Determination. The apparent melting points of most variants were measured with the nanoDSF instrument Prometheus NT.48 (NanoTemper) and represents the inflection point of the tryptophan fluorescence ratio (350 nm to 330 nm) with a 0.1 °C min^{-1} temperature ramp from 20 to 95 °C. For the analysis, the purified enzyme at a concentration of 1 mg mL^{-1} in 50 mM HEPES buffer (pH 7.5) with 0.1 mM PLP was used. The apparent melting points of the DATA variants are listed in Table 1 and Table S4.

Computational Enzyme Design. The crystal structure of the D-amino acid aminotransferase (DATA) from *Bacillus* sp. YM-1 (PDB entry 3DAA)[21] was used as a scaffold for the computational enzyme redesign. This crystal structure was determined in the dimeric form with a high resolution of 1.90 Å, containing 277 amino acids for each subunit. All missing residues, only the N- and C-termini, compared to the UniProt sequence (P19938), were modeled using Modeler 9.19.[33]

The crystal structure also formed a complex with PLP-D-alanine. The enzyme complexed with PLP-(*R*)-PEA was built in two steps. (i) PLP-(*R*)-PEA was built and geometrically optimized using the Avogadro software,[34] and (ii) PLP-(*R*)-PEA was superimposed on the enzyme−PLP-D-alanine complex. For both PLP-D-alanine and PLP-(*R*)-PEA, RESP charges were obtained after fitting partial charges (ESP) computed from Hartree−Fock theory (HF/6-31G*) by using Gaussian G09.[35] RESP charges were included in the ligand parameters for further calculations (Figure S14).

Only the PLP-(*R*)-PEA complex was used for design using Rosetta 3.9. A fast relaxation (energy minimization) was first carried out on the protein side chains and the ligand [PLP-(*R*)-phenylethylamine]. Then, the design was performed on nine selected mutable residues belonging to the first shell of ligand coordination: Y31, V33, H86, Y88, S180, and T242 from the main subunit (catalytic site) and F26, R98, and H100 from the adjacent subunit. All other residues surrounding the ligand in a sphere of 8 Å were considered flexible (through the use of a rotamer library), whereas the rest of the protein was fixed. Harmonic constraints were applied to maintain the nitrogen

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

of the catalytic lysine (K146) close (~3 Å) to the nitrogen bonding the phenylethylamine to PLP during the relaxation and the design. From these calculations, 20,000 amino acid sequences and their corresponding 3D structures were extracted as output. For each corresponding 3D structure (20,000), superimposition of PLP-D-alanine was performed to build complexes (enzyme−PLP-D-alanine). Fast relaxation was performed as previously described to rearrange the side chains around the new ligand, and the new complexes [enzyme−PLP-(*R*)-PEA] were scored with the same energy function as the enzyme−PLP-D-alanine complex.

MD Simulations and Molecular Mechanics Poisson−Boltzmann Surface Area (MM/PBSA) Calculations. The mutant models (M2-6) were built using Modeller9.19.[33] Models with the lowest DOPE score[36] were kept for further MD simulations. All complexes were first minimized in vacuum to release steric clashes. MD simulations were performed using the AMBER ff14SB[37,38] force field for enzymes and GAFF for the ligands using pmemd.CUDA[39] of AMBER16 software. The system was protonated using the propka web server to set the experimental pH to 9. The catalytic lysine was kept neutral as expected in the reaction mechanism with the ketimine. MD simulations were carried out at a constant temperature (300 K) and a constant pressure (1 bar) using the Berendsen algorithm.[40] The integration time step was 2 fs, and covalent bonds involving hydrogens were constrained using SHAKE.[41] The nonbonded pair list was updated heuristically. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME)[42] approach. Nonbonded interactions were treated with a 9 Å direct space cutoff. All enzyme systems were neutralized with Na^+ ions (minimal salt condition),[43] in explicit TIP3P water molecules.[44] Periodic boundary conditions (PBCs) were applied on primary cubic boxes having minimal distances from the dimers of 10 Å. The water molecules and counterions were energy-minimized and equilibrated at 100 K around the constrained solute for 100 ps in the *NVT* ensemble; the entire system was then heated incrementally over 100 ps from 100 to 300 K in 5 K steps with harmonic positional restraints of 25.0 kcal mol^{-1} $Å^{-2}$ on the solute atoms. The MD simulations were continued in the *NPT* ensemble. The positional restraints were gradually removed over 250 ps and followed by the production phase. In all stages, the imine bond (N−C) was kept close to the catalytic lysine by setting a distance constraint of 3.5 Å using a force constant of 50 kcal mol^{-1} $Å^{-2}$. MD snapshots were saved every 10 ps. The MD simulations were carried out for a total of 5 ns for all complexes.

MM/PBSA calculations were performed using MMPBSA.py software and default parameters for Poisson−Boltzmann as described by Miller et al.[45] The calculations were performed on 20 snapshots equitably extracted between 1 and 5 ns.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.9b00888.

A description of additional *e*BCAT protein engineering studies, the nucleotide and protein sequences of the *e*BCAT, DATA, and *Tv*_DAAO enzymes, lists of mutants, and energies from the *in silico* experiments, and figures of 3DM alignments and molecular modeling (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Uwe T. Bornscheuer − *Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, 17487 Greifswald, Germany;* orcid.org/0000-0003-0685-2696; Phone: (+49) 3834 420 4367; Email: uwe.bornscheuer@uni-greifswald.de

Authors

Moritz Voss − *Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, 17487 Greifswald, Germany*

Chao Xiang − *Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, 17487 Greifswald, Germany*

Jérémy Esque − *Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRA, INSA, F-31077 Toulouse, France*

Alberto Nobili − *Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, 17487 Greifswald, Germany*

Marian J. Menke − *Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, 17487 Greifswald, Germany*

Isabelle André − *Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRA, INSA, F-31077 Toulouse, France;* orcid.org/0000-0001-6280-4109

Matthias Höhne − *Protein Biochemistry, Institute of Biochemistry, Greifswald University, 17487 Greifswald, Germany;* orcid.org/0000-0002-2542-725X

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.9b00888

Author Contributions

M.V. and C.X. contributed equally to this work. U.T.B. and M.H. initiated and supervised the project. The computational design and molecular modeling of enzymes were performed by I.A. and J.E. Initial experiments were performed by A.N. M.V., C.X., A.N., and M.J.M. created and biochemically characterized all enzyme variants. U.T.B., M.H., I.A., M.V., C.X., and J.E. analyzed the results. U.T.B., M.H., and M.V. drafted the manuscript, and M.V., J.E., I.A., M.H., and U.T.B. wrote the manuscript. The manuscript was revised and approved by all authors.

Funding

The authors are grateful for funding by the DFG (Grants HO 4754/3-1 and BO 1862/16-1) and the China Scholarship council for a Ph.D. stipend to C.X. (Grant 201808330394). A.N. especially thanks the European Union (KBBE-2011-5, Grant 289350) for financial support within the European Union Seventh Framework Programme.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to Y. Tao and M. Stricker for their help in studying the *e*BCAT enzyme. This work was granted access to the HPC resources on the TGCC-Curie and Occigen supercomputers and the Computing mesocenter of Région Midi-Pyrénées (CALMIP, Toulouse, France).

■ ABBREVIATIONS

BCAT, L-branched chain amino acid aminotransferase; DATA, D-amino acid aminotransferase; PEA, 1-phenylethylamine

■ REFERENCES

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

(2) Steffen-Munsberg, F., Vickers, C., Kohls, H., Land, H., Mallin, H., Nobili, A., Skalden, L., van den Bergh, T., Joosten, H. J., Berglund, P., Höhne, M., and Bornscheuer, U. T. (2015) Bioinformatic Analysis of a PLP-Dependent Enzyme Superfamily Suitable for Biocatalytic Applications. *Biotechnol. Adv. 33*, 566−604.
(3) Henson, C. P., and Cleland, W. W. (1964) Kinetic Studies of Glutamic Oxaloacetic Transaminase Isozymes. *Biochemistry 3*, 338−345.
(4) Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. (2010) Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science 329*, 305−309.
(5) Gomm, A., and O'Reilly, E. (2018) Transaminases for Chiral Amine Synthesis. *Curr. Opin. Chem. Biol. 43*, 106−112.
(6) Manta, B., Cassimjee, K. E., and Himo, F. (2017) Quantum Chemical Study of Dual-Substrate Recognition in ω-Transaminase. *ACS Omega 2*, 890−898.
(7) Skalden, L., Thomsen, M., Höhne, M., Bornscheuer, U. T., and Hinrichs, W. (2015) Structural and Biochemical Characterization of the Dual Substrate Recognition of the (*R*)-Selective Amine Transaminase from *Aspergillus Fumigatus*. *FEBS J. 282*, 407−415.
(8) Höhne, M., Schätzle, S., Jochens, H., Robins, K., and Bornscheuer, U. T. (2010) Rational Assignment of Key Motifs for Function Guides in Silico Enzyme Identification. *Nat. Chem. Biol. 6*, 807−813.
(9) Iwasaki, A., Yamada, Y., Kizaki, N., Ikenaka, Y., and Hasegawa, J. (2006) Microbial Synthesis of Chiral Amines by (*R*)-Specific Transamination with *Arthrobacter* sp. KNK168. *Appl. Microbiol. Biotechnol. 69*, 499−505.
(10) Boyko, K. M., Stekhanova, T. N., Nikolaeva, A. Y., Mardanov, A. V., Rakitin, A. L., Ravin, N. V., Bezsudnova, E. Y., and Popov, V. O. (2016) First Structure of Archaeal Branched-Chain Amino Acid Aminotransferase from *Thermoproteus Uzoniensis* Specific for L-Amino Acids and *R*-Amines. *Extremophiles 20*, 215−225.
(11) Bezsudnova, E. Y., Boyko, K. M., Nikolaeva, A. Y., Zeifman, Y. S., Rakitina, T. V., Suplatov, D. A., and Popov, V. O. (2019) Biochemical and Structural Insights into PLP Fold Type IV Transaminase from *Thermobaculum Terrenum*. *Biochimie 158*, 130−138.
(12) Zeifman, Y. S., Boyko, K. M., Nikolaeva, A. Y., Timofeev, V. I., Rakitina, T. V., Popov, V. O., and Bezsudnova, E. Y. (2019) Functional Characterization of PLP Fold Type IV Transaminase with a Mixed Type of Activity from *Haliangium Ochraceum*. *Biochim. Biophys. Acta, Proteins Proteomics 1867*, 575−585.
(13) Pavkov-Keller, T., Strohmeier, G. A., Diepold, M., Peeters, W., Smeets, N., Schürmann, M., Gruber, K., Schwab, H., and Steiner, K. (2016) Discovery and Structural Characterisation of New Fold Type IV-Transaminases Exemplify the Diversity of This Enzyme Fold. *Sci. Rep. 6*, 38183.
(14) Bezsudnova, E. Y., Dibrova, D. V., Nikolaeva, A. Y., Rakitina, T. V., and Popov, V. O. (2018) Identification of Branched-Chain Amino Acid Aminotransferases Active towards (*R*)-(+)-1-Phenylethylamine among PLP Fold Type IV Transaminases. *J. Biotechnol. 271*, 26−28.
(15) Guan, L. J., Ohtsuka, J., Okai, M., Miyakawa, T., Mase, T., Zhi, Y., Hou, F., Ito, N., Iwasaki, A., Yasohara, Y., and Tanokura, M. (2015) A New Target Region for Changing the Substrate Specificity of Amine Transaminases. *Sci. Rep. 5*, 10753.
(16) Sayer, C., Martinez-Torres, R. J., Richter, N., Isupov, M. N., Hailes, H. C., Littlechild, J. A., and Ward, J. M. (2014) The Substrate Specificity, Enantioselectivity and Structure of the (*R*)-Selective Amine: Pyruvate Transaminase from *Nectria Haematococca*. *FEBS J. 281*, 2240−2253.
(17) Thomsen, M., Skalden, L., Palm, G. J., Höhne, M., Bornscheuer, U. T., and Hinrichs, W. (2014) Crystallographic Characterization of the (*R*)-Selective Amine Transaminase from *Aspergillus Fumigatus*. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 70*, 1086−1093.
(18) Łyskowski, A., Gruber, C., Steinkellner, G., Schürmann, M., Schwab, H., Gruber, K., and Steiner, K. (2014) Crystal Structure of an (*R*)-Selective ω-Transaminase from *Aspergillus Terreus*. *PLoS One 9*, No. e87350.
(19) Telzerow, A., Paris, J., Håkansson, M., González-Sabín, J., Ríos-Lombardía, N., Schürmann, M., Gröger, H., Morís, F., Kourist, R., Schwab, H., and Steiner, K. (2019) Amine Transaminase from *Exophiala Xenobiotica* - Crystal Structure and Engineering of a Fold IV Transaminase That Naturally Converts Biaryl Ketones. *ACS Catal. 9*, 1140−1148.
(20) Goto, M., Miyahara, I., Hayashi, H., Kagamiyama, H., and Hirotsu, K. (2003) Crystal Structures of Branched-Chain Amino Acid Aminotransferase Complexed with Glutamate and Glutarate: True Reaction Intermediate and Double Substrate Recognition of the Enzyme. *Biochemistry 42*, 3725−3733.
(21) Peisach, D., Chipman, D. M., Van Ophem, P. W., Manning, J. M., and Ringe, D. (1998) Crystallographic Study of Steps along the Reaction Pathway of D-Amino Acid Aminotransferase. *Biochemistry 37*, 4958−4967.
(22) Bommer, M., and Ward, J. M. (2013) A 1-Step Microplate Method for Assessing the Substrate Range of l-α-Amino Acid Aminotransferase. *Enzyme Microb. Technol. 52*, 218−225.
(23) Schätzle, S., Höhne, M., Redestad, E., Robins, K., and Bornscheuer, U. T. (2009) Rapid and Sensitive Kinetic Assay for Characterization of ω-Transaminases. *Anal. Chem. 81*, 8244−8248.
(24) Kishimoto, K., Yoshimura, T., Soda, K., and Esaki, N. (1997) Mutation of Arginine 98, Which Serves as a Substrate-Recognition Site of D-Amino Acid Aminotransferase, Can Be Partly Compensated for by Mutation of Tyrosine 88 to an Arginyl Residue. *J. Biochem. 122*, 1182−1189.
(25) Barber, J. E. B., Damry, A. M., Calderini, G. F., Walton, C. J. W., and Chica, R. A. (2014) Continuous Colorimetric Screening Assay for Detection of D-Amino Acid Aminotransferase Mutants Displaying Altered Substrate Specificity. *Anal. Biochem. 463*, 23−30.
(26) Walton, C. J. W., Parmeggiani, F., Barber, J. E. B., McCann, J. L., Turner, N. J., and Chica, R. A. (2018) Engineered Aminotransferase for the Production of D-Phenylalanine Derivatives Using Biocatalytic Cascades. *ChemCatChem 10*, 470−474.
(27) Mani Tripathi, S., and Ramachandran, R. (2006) Direct Evidence for a Glutamate Switch Necessary for Substrate Recognition: Crystal Structures of Lysine ε-Aminotransferase (Rv3290c) from *Mycobacterium Tuberculosis* H37Rv. *J. Mol. Biol. 362*, 877−886.
(28) Li, R., Wijma, H. J., Song, L., Cui, Y., Otzen, M., Tian, Y., Du, J., Li, T., Niu, D., Chen, Y., Feng, J., Han, J., Chen, H., Tao, Y., Janssen, D. B., and Wu, B. (2018) Computational Redesign of Enzymes for Regio- and Enantioselective Hydroamination Article. *Nat. Chem. Biol. 14*, 664−670.
(29) Liu, Y., and Kuhlman, B. (2006) Rosetta Design Server for Protein Design. *Nucleic Acids Res. 34*, W235−W238.
(30) Nobili, A., Gall, M. G., Pavlidis, I. V., Thompson, M. L., Schmidt, M., and Bornscheuer, U. T. (2013) Use of "small but Smart" Libraries to Enhance the Enantioselectivity of an Esterase from Bacillus Stearothermophilus towards Tetrahydrofuran-3-Yl Acetate. *FEBS J. 280*, 3084−3093.
(31) Bergthorsson, U., Andersson, D. I., and Roth, J. R. (2007) Ohno's Dilemma: Evolution of New Genes under Continuous Selection. *Proc. Natl. Acad. Sci. U. S. A. 104*, 17004−17009.
(32) Steffen-Munsberg, F., Matzel, P., Sowa, M. A., Berglund, P., Bornscheuer, U. T., and Höhne, M. (2016) *Bacillus Anthracis* ω-Amino Acid:Pyruvate Transaminase Employs a Different Mechanism for Dual Substrate Recognition than Other Amine Transaminases. *Appl. Microbiol. Biotechnol. 100*, 4511−4521.
(33) Webb, B., and Sali, A. (2016) Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf. 54*, 5.6.1−5.6.37.
(34) Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf. 4*, 17.

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

(35) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) *Gaussian 09*, Gaussian, Inc., Wallingford, CT.

(36) Shen, M., and Sali, A. (2006) Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci. 15*, 2507−2524.

(37) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. A. (1995) Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc. 117*, 5179−5197.

(38) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput. 11*, 3696−3713.

(39) Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput. 9*, 3878−3888.

(40) Berendsen, H. J. C. C., Postma, J. P. M. M., van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys. 81*, 3684−3690.

(41) van Gunsteren, W. F., and Berendsen, H. J. C. (1977) Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys. 34*, 1311−1327.

(42) Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys. 98*, 10089−10092.

(43) Åqvist, J. (1990) Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem. 94*, 8021−8024.

(44) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79*, 926−935.

(45) Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA. Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput. 8*, 3314−3321.

https://dx.doi.org/10.1021/acschembio.9b00888
ACS Chem. Biol. 2020, 15, 416−424

Supporting information for

Creation of (*R*)-amine transaminase activity within an α-amino acid transaminase scaffold

Moritz Voss, Chao Xiang, Jérémy Esque, Alberto Nobili, Marian J. Menke, Isabelle André, Matthias Höhne, Uwe T. Bornscheuer

Table of contents

Further attempts to improve (*R*)-ATA activity in the eBCAT R97E variant

The active site mutation R97E in the *e*BCAT scaffold led to an initial (*R*)-ATA acceptance, with germinal activity of 3.5 mU mg^{-1} in the adapted acetophenone assay with (*R*)-PEA as amino donor and α-ketoisocaproate as amino acceptor. The R97E mutation disrupts the hydrogen bond network in the active site, important for the coordination of the amino acids α-carboxylate functions.

In the *e*BCAT wild type's active site, the α-carboxylate of the amino acid substrate is not actively coordinated by the positively charged arginine residues, as typically observed in transaminases. Instead, the polarized OH function of Y95 and the polarized main-chain NH groups of T257 and A258 indirectly coordinate the α-carboxylate. Additionally, the hydroxyl function of Y95 is polarized by the positively charged R97 and the two main chain NH functions are polarized by R40. This enables the partially positive hydrogens to form a hydrogen-bond with the negatively charged α-carboxylic function of the substrate. In contrast, the O-site coordinates different types of functional groups and therefore enables the dual-substrate recognition.[1] In case of L-glutamate, the γ-carboxylate is directly coordinated in the O-site by the positively charged R97 and the polarized side-chain hydroxyl function of Y31. In addition, the main-chain NH group of V109 and the hydroxyl function of Y129 complete the hydrogen bond network. The side chains of L-leucine, L-isoleucine, and L-valine are also accepted in the O-site, but they do not form the above described hydrogen bond network. This allows the coordination of aryl- and alkyl moieties as well as carboxylic functions in the O-site.

After studying the active site residues G38, R40, Y95, and R97 and identifying the R97E substitution as beneficial, we aimed to increase the initial (*R*)-amine activity by combining this beneficial substitution with other mutations. We incorporated more amino acids variations derived from the *A. fumigatus* ATA as well as from the general amino acid motif identified earlier for (*R*)-ATAs.[2] Unfortunately, a further increase of the activity found for the R97E variant could not be achieved (Table S3). Additionally, random mutagenesis of the *e*BCAT scaffold carrying the beneficial R97E mutation was attempted, but screening of ~4,000 clones using the glycine oxidase solid-phase screening assay with (*R*)-PEA and glyoxylate as substrate pairs did not lead to the identification of additional beneficial mutations.[3] We thus speculated that for further increase of this newly achieved ATA activity not only the engineering of the P-site for the acceptance of methyl groups is required, but also the introduction of an α-carboxylate recognition site in the O-site in order to accept D-alanine. In (*R*)-ATAs like the ones from *A. fumigatus* or *A. terreus*, the guanidino function of the arginine side chain (R126 or R128, respectively) was shown to interact with the carboxylic function of pyruvate in the O-site.[4,5] This arginine residue is located on a flexible loop, shaping the O-site and having two distinct conformations. The crystal structure of *A. fumigatus* with the bound inhibitor gabaculine (PDB ID: 4UUG) shows the two confirmations of the loop (G121-N135), resulting in a closed and open state. While the closed state

shows that R126 interacts with the carboxylate function of the inhibitor *m*-carboxyphenyl-pyridoxamine phosphate (*m*CPP), the open state is shifted away and an alternative salt-bridge between the R126 and D132 was observed.[4] This rearrangement of the loop could therefore mediate the dual-substrate recognition in the (*R*)-ATA.[4,6] This carboxylate recognition mechanism is not present in the *e*BCAT, since the α-carboxylate of the accepted amino acids are bound in the P-site. In fact, Łyskowski *et al*. compared the B-factors of the flexible loop of the *A. terreus* (*R*)-ATA with the corresponding loop in the *e*BCAT and reported a much lower B-factor for the *e*BCAT, indicating that the corresponding loop region is less flexible.[5]

In order to introduce the carboxylate recognition in the O-site and therefore potentially enabling the dual-substrate recognition, we transferred the loop sequence from the (*R*)-ATA from *A. fumigatus* into the *e*BCAT (data not shown). Even though the *e*BCAT chimera enzyme was expressed as soluble protein, no (*R*)-ATA activity was detectable using the sensitive acetophenone assay. The introduction of the beneficial R97E mutation into this *e*BCAT loop variant also did not enable the (*R*)-PEA substrate acceptance. One might speculate that the loop was not positioned correctly due the adjacent residues of the *e*BCAT structure, but this would require further fine-tuning and extensive experiments.

Supporting methods and sequences

*e*BCAT (pET28b) DNA sequence (ORF)

ATGACGACGAAAAAAGCTGATTATATCTGGTTCAATGGCGAAATGGTCCGCTGGGAAGATGCGAAAGTTCACGTTATGTC
TCACGCTCTGCATTATGGCACCTCAGTGTTTGAAGGTATTCGTTGCTACGATTCGCACAAAGGTCCGGTGGTTTTCCGTC
ATCGCGAACACATGCAGCGTCTGCATGACAGTGCGAAAATTTATCGCTTTCCGGTCAGCCAATCTATCGATGAACTGATG
GAAGCATGTCGTGACGTGATCCGCAAAAACAATCTGACCTCTGCTTACATTCGTCCGCTGATCTTTGTCGGCGATGTGGG
TATGGGTGTGAACCCGCCGGCAGGTTATAGCACGGACGTTATTATCGCAGCATTCCCGTGGGGTGCATACCTGGGTGCAG
AAGCCCTGGAACAGGGCATTGATGCCATGGTTAGCTCTTGGAACCGCGCAGCTCCGAATACCATCCCGACGGCAGCAAAA
GCCGGCGGTAATTATCTGAGTTCCCTGCTGGTTGGTAGTGAAGCGCGTCGCCACGGCTATCAGGAAGGTATTGCCCTGGA
TGTCAACGGCTACATCTCCGAAGGCGCAGGTGAAAACCTGTTTGAAGTGAAAGACGGCGTTCTGTTTACCCCGCCGTTCA
CGTCATCGGCTCTGCCGGGTATTACCCGTGATGCAATTATCAAACTGGCTAAAGAACTGGGCATCGAAGTCCGTGAACAG
GTGCTGAGTCGCGAATCCCTGTACCTGGCAGATGAAGTGTTTATGAGCGGTACCGCAGCTGAAATTACGCCGGTTCGCTC
TGTCGACGGCATCCAAGTTGGCGAAGGTCGTTGCGGTCCGGTCACGAAACGCATTCAGCAAGCCTTCTTTGGTCTGTTCA
CGGGCGAAACGGAAGATAAATGGGGCTGGCTGGATCAAGTCAATCAGCTCGAGCACCACCACCACCACCACTGA

eBCAT protein sequence (ORF)

MTTKKADYIWFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLHDSAKIYRFPVSQSIDELM
EACRDVIRKNNLTSAYIRPLIFVGDVGMGVNPPAGYSTDVIIAAFPWGAYLGAEALEQGIDAMVSSWNRAAPNTIPTAAK
AGGNYLSSLLVGSEARRHGYQEGIALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQ
VLSRESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGLFTGETEDKWGWLDQVNQLEHHHHHH

DATA (pGASTON) DNA sequence (ORF)

ATGGGCTATACCCTGTGGAATGATCAGATTGTGAAAGATGAAGAAGTGAAAATTGATAAAGAAGATCGTGGCTATCAGTT
TGGCGATGGCGTGTATGAAGTGGTGAAAGTGTATAATGGCGAAATGTTTACCGTGAATGAACATATTGATCGTCTGTATG
CGAGCGCGGAAAAAATTCGTATTACCATTCCGTATACCAAAGATAAATTTCATCAGCTGCTGCATGAACTGGTGGAAAAA
AATGAACTGAATACCGGCCATATTTATTTTCAGGTGACCCGTGGCACCAGCCCGCGTGCGCATCAGTTTCCGGAAAATAC
CGTGAAACCGGTGATTATTGGCTATACCAAAGAAAATCCGCGTCCGCTGGAAAATCTGGAAAAAGGCGTGAAAGCGACCT
TTGTGGAAGATATTCGTTGGCTGCGTTGCGATATTAAAAGCCTGAATCTGCTGGGCGCGGTGCTGGCGAAACAGGAAGCG
CATGAAAAAGGCTGCTATGAAGCGATTCTGCATCGTAATAATACCGTGACCGAAGGCAGCAGCAGCAATGTGTTTGGCAT
TAAAGATGGCATTCTGTATACCCATCCGGCGAATAATATGATTCTGAAAGGCATTACCCGTGATGTGGTGATTGCGTGCG
CGAATGAAATTAATATGCCGGTGAAAGAAATTCCGTTTACCACCCATGAAGCGCTGAAAATGGATGAACTGTTTGTGACC
AGCACCACCAGCGAAATTACCCCGGTGATTGAAATTGATGGCAAACTGATTCGTGATGGCAAAGTGGGCGAATGGACCCG
TAAACTTCAGAAACAGTTTGAAACCAAAATTCCGAAACCGCTGCATATTTCAGGAGGATCCCATCATCATCATCATCATT
GA

DATA protein sequence (ORF)

MGYTLWNDQIVKDEEVKIDKEDRGYQFGDGVYEVVKVYNGEMFTVNEHIDRLYASAEKIRITIPYTKDKFHQLLHELVEK
NELNTGHIYFQVTRGTSPRAHQFPENTVKPVIIGYTKENPRPLENLEKGVKATFVEDIRWLRCDIKSLNLLGAVLAKQEA
HEKGCYEAILHRNNTVTEGSSSNVFGIKDGILYTHPANNMILKGITRDVVIACANEINMPVKEIPFTTHEALKMDELFVT
STTSEITPVIEIDGKLIRDGKVGEWTRKLQKQFETKIPKPLHISGGSHHHHHH

*Tv*_DAAO (pETDuet-1) DNA sequence (ORF)

ATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCGAATTCGATGGCAAAAATCGTTGTTATCGGTGCAGGCGT
TGCCGGTCTGACCACCGCACTGCAGCTGCTGCGTAAAGGCCATGAAGTGACCATTGTTAGTGAGTTTACTCCGGGCGATC
TGAGCATTGGTTATACCAGTCCGTGGGCAGGTGCCAATTGGCTGACCTTTTATGATGGTGGCAAACTGGCCGATTATGAT
GCAGTTAGTTATCCGATTCTGCGTGAACTGGCACGTAGTAGTCCGGAAGCAGGTATTCGCCTGATTAGTCAGCGTAGTCA
TGTGCTGAAACGCGATCTGCCGAAACTGGAAGTGGCCATGAGTGCCATTTGCCAGCGTAATCCGTGGTTTAAAAATACCG
TGGATAGTTTTGAAATCATCGAAGATCGCAGTCGTATTGTTCATGATGATGTTGCATATCTGGTGGAATTTCGTAGCGTT
TGCATTCATACCGGTGTTTATCTGAATTGGCTGATGAGTCAGTGCCTGAGCCTGGGCGCCACCGTTGTGAAACGTCGCGT
GAATCATATTAAGGATGCAAATCTGCTGCATAGTAGTGGCAGCCGTCCGGATGTGATTGTGAATTGCAGTGGCCTGTTTG
CACGTTTTCTGGGCGGTGTGGAAGATAAAAAGATGTATCCGATTCGTGGTCAGGTGGTTCTGGTGCGTAATAGTCTGCCG
TTTATGGCAAGTTTTAGTAGCACCCCGGAAAAAGAAAATGAAGATGAAGCACTGTATATCATGACCCGTTTTGATGGCAC
CAGTATTATTGGTGGCTGCTTTCAGCCGAATAATTGGAGCAGCGAACCGGATCCGAGTCTGACCCATCGTATTCTGAGTC
GCGCCCTGGATCGCTTTCCGGAACTGACCAAAGATGGCCCGCTGGATATTGTTCGTGAATGCGTTGGCCATCGTCCGGGT
CGCGAAGGTGGTCCGCGTGTGGAACTGGAAAAAATTCCGGGCGTTGGTTTTGTGGTTCATAATTATGGTGCCGCAGGCGC
AGGCTATCAGAGCAGTTATGGCATGGCAGATGAAGCCGTTAGCTATGTGGAACGTGCCCTGACCCGTCCGAATCTGTAA

*Tv*_DAAO protein sequence (ORF)

MGSSHHHHHHSQDPNSMAKIVVIGAGVAGLTTALQLLRKGHEVTIVSEFTPGDLSIGYTSPWAGANWLTFYDGGKLADYD
AVSYPILRELARSSPEAGIRLISQRSHVLKRDLPKLEVAMSAICQRNPWFKNTVDSFEIIEDRSRIVHDDVAYLVEFRSV
CIHTGVYLNWLMSQCLSLGATVVKRRVNHIKDANLLHSSGSRPDVIVNCSGLFARFLGGVEDKKMYPIRGQVVLVRNSLP
FMASFSSTPEKENEDEALYIMTRFDGTSIIGGCFQPNNWSSEPDPSLTHRILSRALDRFPELTKDGPLDIVRECVGHRPG
REGGPRVELEKIPGVGFVVHNYGAAGAGYQSSYGMADEAVSYVERALTRPNL

Supporting Tables

Table S1 List of primer sequences for the creation of the *e*BCAT variants.

Primer name	3' → 5' nucleotide sequence
eBCAT A160N fw	CCGCCGTTTTTTGCTGCCGTC
eBCAT A160N rv	GACGGCAGCAAAAAACGGCGG
eBCAT S166D fw	CCAACCAGCAGGGAATCCAGATAATTACCGC
eBCAT S166D rv	GCGGTAATTATCTGGATTCCCTGCTGGTTGG
eBCAT S166D_Y164W fw	CAACCAGCAGGGAATCCAGCCAATTACCGC
eBCAT S166D_Y164W rv	GCGGTAATTGGCTGGATTCCCTGCTGGTTG
eBCAT A259G fw	GGCGTAATTTCACCTGCGG
eBCAT A259G rv	CCGCAGGTGAAATTACGCC
eBCAT I96V_R97E fw	GACCTCTGCTTACGTTGAGCCGCTGATC
eBCAT I96V_R97E rv	GATCAGCGGCTCAACGTAAGCAGAGGTC
eBCAT V49R fw	GCACAAAGGTCCGCGGGTTTTCCGTCATC
eBCAT V49R rv	GATGACGGAAAACCCGCGGACCTTTGTGC
eBCAT V85M fw	GAAGCATGTCGTGACATGATCCGCAAAAAC
eBCAT V85M rv	GTTTTTGCGGATCATGTCACGACATGCTTC
eBCAT A195S fw	CATCTCCGAAGGCTCAGGTGAAAACCTG
eBCAT A195S rv	CAGGTTTTCACCTGAGCCTTCGGAGATG
eBCAT Y31H fw	CACGCTCTGCATCATGGCACCTCAG
eBCAT Y31H rv	CTGAGGTGCCATGATGCAGAGCGTG
eBCAT Y31R fw	GTCTCACGCTCTGCATAGAGGCACCTCAGTGTTTG
eBCAT Y31R rv	CAAACACTGAGGTGCCTCTATGCAGAGCGTGAGAC
eBCAT Y95F fw	CTGACCTCTGCTTTCGTTGAGCCGCTG
eBCAT Y95F rv	CAGCGGCTCAACGAAAGCAGAGGTCAG
eBCAT V109S fw	CGGCGATGTGGGTATGGGTAGTAACCCGCCGGCAGGTTATAG
eBCAT V109S rv	CTATAACCTGCCGGCGGGTTACTACCCATACCCACATCGCCG
eBCAT V109L fw	GATGTGGGTATGGGTTTAAACCCGCCGGCAGG
eBCAT V109L rv	CCTGCCGGCGGGTTTAAACCCATACCCACATC
eBCAT W126F fw	CGCAGCATTCCCGTTTGGTGCATACCTGG
eBCAT W126F rv	CCAGGTATGCACCAAACGGGAATGCTGCG
eBCAT Y129V fw	ATTCCCGTGGGGTGCAGTACTGGGTGCAGAAG
eBCAT Y129V rv	GGGCACCCCACGTCATGACCCACGTCTTCGGG
eBCAT Y266S fw	CCAAGCGAGACAGCTG
eBCAT Y266S rv	CAGCTGTCTCGCTTGG
eBCAT R97K fw	CCTCTGCTTACATTAAACCGCTGATCTTTGTCG
eBCAT R97K rv	CGACAAAGATCAGCGGTTTAATGTAAGCAGAGG
eBCAT R40S fw	CTCAGTGTTTGAAGGTATTAGCTGCTACGATTCGCACAAAG
eBCAT R40S rv	CTTTGTGCGAATCGTAGCAGCTAATACCTTCAAACACTGAG
eBCAT R40A fw	CTCAGTGTTTGAAGGTATTGCGTGCTACGATTCGCACAAAG
eBCAT R40A rv	CTTTGTGCGAATCGTAGCACGCAATACCTTCAAACACTGAG
eBCAT R97M fw	GACCTCTGCTTACATTATGCCGCTGATCTTTGTCGG
eBCAT R97M rv	CCGACAAAGATCAGCGGCATAATGTAAGCAGAGGTC
eBCAT A258H fw	GTGTTTATGAGCGGTACCCATGCTGAAATTACGCCGGTTC
eBCAT A258H rv	GAACCGGCGTAATTTCAGCATGGGTACCGCTCATAAACAC
eBCAT W126A fw	GTTATTATCGCAGCATTCCCGGCGGGTGCATACC
eBCAT W126A rv	GGTATGCACCCGCCGGGAATGCTGCGATAATAAC
eBCAT W126G fw	GTTATTATCGCAGCATTCCCGGGCGGTGCATACCTG
eBCAT W126G rv	CAGGTATGCACCGCCCGGGAATGCTGCGATAATAAC
eBCAT Y129I fw	CATTCCCGTGGGGTGCAATTCTGGGTGCAGAAG

eBCAT Y129I rv	CTTCTGCACCCAGAATTGCACCCCACGGGAATG
eBCAT Y129T fw	CATTCCCGTGGGGTGCAACCCTGGGTGCAGAAG
eBCAT Y129T rv	CTTCTGCACCCAGGGTTGCACCCCACGGGAATG
eBCAT V102R_R97L fw	CTTCCGCTGATCTTTCGTGGCGATGTGGGTATGG
eBCAT V102R_R97L rv	CCATACCCACATCGCCACGAAAGATCAGCGGAAG
eBCAT T33D fw	CACGCTCTGCATTATGGCGATTCAGTGTTTGAAGGTATTC
eBCAT T33D rv	GAATACCTTCAAACACTGAATCGCCATAATGCAGAGCGTG
eBCAT R175M fw	GTTGGTAGTGAAGCGATGCGCCACGGCTATCAG
eBCAT R175M rv	CTGATAGCCGTGGCGCATCGCTTCACTACCAAC
eBCAT E181Y fw	CACGGCTATCAGTATGGTATTGCCCTGGATG
eBCAT E181Y rv	CATCCAGGGCAATACCATACTGATAGCCGTG
eBCAT R175M_R176D fw	GGTTGGTAGTGAAGCGATGGATCACGGCTATCAG
eBCAT R175M_R176D rv	CTGATAGCCGTGATCCATCGCTTCACTACCAACC
eBCAT R176D fw	GTAGTGAAGCGCGTGACCACGGCTATCAG
eBCAT R176D rv	CTGATAGCCGTGGTCACGCGCTTCACTAC
eBCAT T221N fw	CGGCTCTGCCGGGTATTAATCGTGATGCAATTATCAAAC
eBCAT T221N rv	GTTTGATAATTGCATCACGATTAATACCCGGCAGAGCCG
eBCAT I234N fw	GCTAAAGAACTGGGCAATGAAGTCCGTGAACAGG
eBCAT I234N rv	CCTGTTCACGGACTTCATTGCCCAGTTCTTTAGC
eBCAT K282N fw	GTCCGGTCACGAATCGCATTCAGCAAG
eBCAT K282N rv	CTTGCTGAATGCGATTCGTGACCGGAC
eBCAT M107R fw	GCGATGTGGGTAGGGGTGTGAACCC
eBCAT M107R rv	GGGTTCACACCCCTACCCACATCGC

Table S2 List of primer sequences for the creation of the DATA variants.

Primer name	**3' → 5' nucleotide sequence**
DATA H86F fw	CTGAATACCGGCTTTATTTATTTTCAGGTGACCC
DATA H86F rv	GGGTCACCTGAAAATAAATAAAGCCGGTATTCAG
DATA H86F_I87V_Y88E fw	CTGAATACCGGCTTTGTGGAATTTCAGGTGACCCGTGGCAC
DATA H86F_I87V_Y88E rv	GTGCCACGGGTCACCTGAAATTCCACAAAGCCGGTATTCAG
DATA T242A fw	GTTTGTGACCAGCACCGCGAGCGAAATTACCCCG
DATA T242A rv	CGGGGTAATTTCGCTCGCGGTGCTGGTCACAAAC
DATA S180G fw	GTGACCGAAGGCAGCGGCAGCAATGTGTTTG
DATA S180G rv	CAAACACATTGCTGCCGCTGCCTTCGGTCAC
DATA Y165T fw	CGCATGAAAAAGGCTGCACCGAAGCGATTCTGCATC
DATA Y165T rv	GATGCAGAATCGCTTCGGTGCAGCCTTTTTCATGCG
DATA T242G fw	GTTTGTGACCAGCACCGGCAGCGAAATTACCCC
DATA T242G rv	GGGGTAATTTCGCTGCCGGTGCTGGTCACAAAC
DATA Y31F fw	GGCGATGGCGTGTTTGAAGTGGTGAAAG
DATA Y31F rv	CTTTCACCACTTCAAACACGCCATCGCC
DATA Y31M fw	GGCGATGGCGTGATGGAAGTGGTGAAAG
DATA Y31M rv	CTTTCACCACTTCCATCACGCCATCGCC
DATA Y31F_V33A fw	GATGGCGTGTTTGAAGCGGTGAAAGTGTATAATG
DATA Y31F_V33A rv	CATTATACACTTTCACCGCTTCAAACACGCCATC
DATA H86L_Y88F fw	CTGAATACCGGCCTGATTTTTTTTCAGGTGACCCG
DATA H86L_Y88F rv	CGGGTCACCTGAAAAAAAATCAGGCCGGTATTCAG
DATA H86F_Y88F fw	GAACTGAATACCGGCTTTATTTTTTTTCAGGTGACCCG
DATA H86F_Y88F rv	CGGGTCACCTGAAAAAAAATAAAGCCGGTATTCAGTTC
DATA H86L_Y88W fw	CTGAATACCGGCCTGATTTGGTTTCAGGTGACCCG
DATA H86L_Y88W rv	CGGGTCACCTGAAACCAAATCAGGCCGGTATTCAG
DATA H86I_Y88V fw	CTGAATACCGGCATTATTGTGTTTCAGGTGACCCG
DATA H86I_Y88V rv	CGGGTCACCTGAAACACAATAATGCCGGTATTCAG
DATA R98W_H100L fw	GCACCAGCCCGTGGGCGCTGCAGTTTCCGGAAAATAC
DATA R98W_H100L rv	GTATTTTCCGGAAACTGCAGCGCCCACGGGCTGGTGC
DATA R98W_H100W fw	GCACCAGCCCGTGGGCGTGGCAGTTTCCGGAAAATAC
DATA R98W_H100W rv	GTATTTTCCGGAAACTGCCACGCCCACGGGCTGGTGC

DATA R98Y_H100L fw	GCACCAGCCCGTATGCGCTGCAGTTTCCGGAAAATAC
DATA R98Y_H100L rv	GTATTTTCCGGAAACTGCAGCGCATACGGGCTGGTGC
DATA R98T_H100W fw	GCACCAGCCCGACCGCGTGGCAGTTTCCGGAAAATAC
DATA R98T_H100W rv	GTATTTTCCGGAAACTGCCACGCGGTCGGGCTGGTGC
DATA S180L fw	GTGACCGAAGGCAGCCTGAGCAATGTGTTTGGC
DATA S180L rv	GCCAAACACATTGCTCAGGCTGCCTTCGGTCAC
DATA S180V fw	GTGACCGAAGGCAGCGTGAGCAATGTGTTTGGC
DATA S180V rv	GCCAAACACATTGCTCACGCTGCCTTCGGTCAC
DATA T242I fw	GTTTGTGACCAGCACCATTAGCGAAATTACCCCGG
DATA T242I rv	CCGGGGTAATTTCGCTAATGGTGCTGGTCACAAAC
DATA T242W fw	GTTTGTGACCAGCACCTGGAGCGAAATTACCCCG
DATA T242W rv	CGGGGTAATTTCGCTCCAGGTGCTGGTCACAAAC
DATA T242L fw	GTTTGTGACCAGCACCCTGAGCGAAATTACCCCG
DATA T242L rv	CGGGGTAATTTCGCTCAGGGTGCTGGTCACAAAC
DATA H86A fw	CTGAATACCGGCGCGATTTATTTTCAGG
DATA H86A DATA rv	CCTGAAAATAAATCGCGCCGGTATTCAG
DATA S180A fw	GTGACCGAAGGCAGCGCGAGCAATGTGTTTGG
DATA S180A rv	CCAAACACATTGCTCGCGCTGCCTTCGGTCAC
DATA R98K fw	GTGGCACCAGCCCGAAAGCGCATCAGTTTC
DATA R98K rv	GAAACTGATGCGCTTTCGGGCTGGTGCCAC
DATA Y31F fw	GGCGATGGCGTGTTTGAAGTGGTGAAAG
DATA Y31F rv	CTTTCACCACTTCAAACACGCCATCGCC
DATA Y31M fw	GGCGATGGCGTGATGGAAGTGGTGAAAG
DATA Y31M rv	CTTTCACCACTTCCATCACGCCATCGCC
DATA Y31F_V33A fw	GATGGCGTGTTTGAAGCGGTGAAAGTGTATAATG
DATA Y31F_V33A rv	CATTATACACTTTCACCGCTTCAAACACGCCATC
DATA H86L_Y88F fw	CTGAATACCGGCCTGATTTTTTTTCAGGTGACCCG
DATA H86L_Y88F rv	CGGGTCACCTGAAAAAAAATCAGGCCGGTATTCAG
DATA H86F_Y88F fw	GAACTGAATACCGGCTTTATTTTTTTTCAGGTGACCCG
DATA H86F_Y88F rv	CGGGTCACCTGAAAAAAAATAAAGCCGGTATTCAGTTC
DATA H86L_Y88W fw	CTGAATACCGGCCTGATTTGGTTTCAGGTGACCCG
DATA H86L_Y88W rv	CGGGTCACCTGAAACCAAATCAGGCCGGTATTCAG
DATA H86I_Y88V fw	CTGAATACCGGCATTATTGTGTTTCAGGTGACCCG
DATA H86I_Y88V rv	CGGGTCACCTGAAACACAATAATGCCGGTATTCAG
DATA R98W_H100L fw	GCACCAGCCCGTGGGCGCTGCAGTTTCCGGAAAATAC
DATA R98W_H100L rv	GTATTTTCCGGAAACTGCAGCGCCCACGGGCTGGTGC
DATA R98W_H100W fw	GCACCAGCCCGTGGGCGTGGCAGTTTCCGGAAAATAC
DATA R98W_H100W rv	GTATTTTCCGGAAACTGCCACGCCCACGGGCTGGTGC
DATA R98Y_H100L fw	GCACCAGCCCGTATGCGCTGCAGTTTCCGGAAAATAC
DATA R98Y_H100L rv	GTATTTTCCGGAAACTGCAGCGCATACGGGCTGGTGC
DATA R98T_H100W fw	GCACCAGCCCGACCGCGTGGCAGTTTCCGGAAAATAC
DATA R98T_H100W rv	GTATTTTCCGGAAACTGCCACGCGGTCGGGCTGGTGC
DATA S180L fw	GTGACCGAAGGCAGCCTGAGCAATGTGTTTGGC
DATA S180L rv	GCCAAACACATTGCTCAGGCTGCCTTCGGTCAC
DATA S180V fw	GTGACCGAAGGCAGCGTGAGCAATGTGTTTGGC
DATA S180V rv	GCCAAACACATTGCTCACGCTGCCTTCGGTCAC
DATA T242I fw	GTTTGTGACCAGCACCATTAGCGAAATTACCCCGG
DATA T242I rv	CCGGGGTAATTTCGCTAATGGTGCTGGTCACAAAC
DATA T242W fw	GTTTGTGACCAGCACCTGGAGCGAAATTACCCCG
DATA T242W rv	CGGGGTAATTTCGCTCCAGGTGCTGGTCACAAAC
DATA T242L fw	GTTTGTGACCAGCACCCTGAGCGAAATTACCCCG
DATA T242L rv	CGGGGTAATTTCGCTCAGGGTGCTGGTCACAAAC
DATA H86Ffw	GAACTGAATACCGGCTTTATTTATTTTCAGGTGACCCG
DATA H86F rv	CGGGTCACCTGAAAATAAATAAAGCCGGTATTCAGTTC
DATA Y88Ffw	GAATACCGGCCATATTTTTTTTCAGGTGACCCGTG
DATA Y88F rv	CACGGGTCACCTGAAAAAAAATATGGCCGGTATTC
DATA H100A fw	GCCCGCGTGCGGCGCAGTTTCCGGAAAATAC
DATA H100A rv	GTATTTTCCGGAAACTGCGCCGCACGCGGGC

DATA Y31A fw	GTTTGGCGATGGCGTGGCGGAAGTGGTGAAAGTG
DATA Y31A rv	CACTTTCACCACTTCCGCCACGCCATCGCCAAAC
DATA R98A fw	TGGCACCAGCCCGGCGGCGCATCAGTTTC
DATA R98A rv	ACCGTGGTCGGGCCGCCGCGTAGTCAAAG
DATA F26A fw	GAAGATCGTGGCTATCAGGCGGGCGATGGCGTGTATGAAG
DATA F26A rv	CTTCATACACGCCATCGCCCGCCTGATAGCCACGATCTTC
DATA V33A fw	GATGGCGTGTATGAAGCGGTGAAAGTGTATAATG
DATA V33A rv	CATTATACACTTTCACCGCTTCATACACGCCATC
DATA Y31A_F26A fw	GGGCGATGGCGTGGCGGAAGTGGTGAAAGTG
DATA Y31A_F26A rv	CACTTTCACCACTTCCGCCACGCCATCGCCC
DATA F26H fw	GAAGATCGTGGCTATCAGCATGGCGATGGCGTGTATG
DATA F26H rv	CATACACGCCATCGCCATGCTGATAGCCACGATCTTC
DATA Y88A fw	GAATACCGGCCATATTGCTTTTCAGGTGACCCG
DATA Y88A rv	CGGGTCACCTGAAAAGCAATATGGCCGGTATTC
DATA Y31L fw	GTTTGGCGATGGCGTGCTGGAAGTGGTGAAAGTG
DATA Y31L rv	CACTTTCACCACTTCCAGCACGCCATCGCCAAAC
DATA Y31F fw	GGCGATGGCGTGTTTGAAGTGGTGAAAG
DATA Y31F rv	CTTTCACCACTTCAAACACGCCATCGCC
DATA Y88E fw	GAATACCGGCCATATTGAATTTCAGGTGACCCGTGG
DATA Y88E rv	CCACGGGTCACCTGAAATTCAATATGGCCGGTATTC
DATA Y31S fw	GTTTGGCGATGGCGTGAGCGAAGTGGTGAAAGTG
DATA Y31S rv	CACTTTCACCACTTCGCTCACGCCATCGCCAAAC

Table S3 Created *e*BCAT variants. None of the presented variants had an activity of >5 mU mg^{-1} in the acetophenone assay with (*R*)-PEA and α-ketoisocaproate.

#	mutation	#	mutation	#	mutation
1	Y31H	43	R68C/V252L	85	Y95F/R97E/M107R
2	Y31R	44	D76V/T221N	86	Y95F/R97E/Y164W
3	F36A	45	V85M/R97E	87	Y95F/R97E/G256T
4	F36Y	46	Y95F/R97E	88	Y95F/Y164W/G256T
5	G38A	47	Y95F/Y164W	89	R97E/V109L/W126F
6	G38V	48	Y95F/G256T	90	R97E/V109L/I234N
7	R40A	49	I96V/R97E	91	R97E/Y164W/G256T
8	R40S	50	R97E/V109S	92	R97E/G256T/A259G
9	Y95A	51	R97E/M107R	93	R97L/R175M/E181Y
10	Y95F	52	R97E/Y164W	94	V22I/V85M/R97E/I234N
11	R97A	53	R97E/A195S	95	V22I/R97E/V109L/I234N
12	R97D	54	R97E/G256T	96	F36Y/G38V/Y95F/R97E
13	R97E	55	R97E/A258H	97	F36Y/G38V/R97E/M107R
14	R97F	56	R97L/W126A	98	F36Y/Y95F/R97E/M107R
15	R97L	57	R97L/W126G	99	F36Y/Y95F/R97E/Y164W
16	R97M	58	R97L/Y129V	100	F36Y/R97E/M107R/V109H
17	M107A	59	R97L/Y129I	101	G38V/Y95F/R97E/M107R
18	M107R	60	R97L/Y129T	102	G38V/Y95F/R97E/Y164W
19	V109A	61	R97L/R176D	103	G38V/Y95F/R97E/G256T
20	V109H	62	R97M/A258H	104	G38V/Y95F/Y164W/G256T
21	G138D	63	M107R/V109H	105	G38V/R97E/Y164W/G256T
22	S144N	64	M107R/Y164W	106	V85M/R97E/V109L/I234N
23	Y164A	65	V109L/W126F	107	V85M/R97E/I234N/K282N
24	Y164W	66	Y164W/G256T	108	Y95F/R97E/M107R/Y164W
25	G256A	67	L199P/E238V	109	Y95F/R97E/Y164W/G256T
26	G256T	68	T209S/V252M	110	R97E/A160N/Y164W/S166D
27	A258H	69	T33D/R97L/V102R	111	R97L/R175M/R176D/E181Y
28	K282N	70	F36I/N163S/K230N	112	V22I/V85M/R97E/V109L/I234N
29	V22I/I234N	71	F36Y/G38V/R97E	113	V22I/V85M/R97E/I234N/K282N

30	Y31H/R97E	72	F36Y/Y95F/R97E	114	V22I/R97E/V109L/I234N/K282N
31	Y31R/R97E	73	F36Y/R97E/M107R	115	F36Y/Y95F/R97E/M107R/Y164W
32	F36Y/G38V	74	F36Y/M107R/V109H	116	G38V/Y95F/R97E/Y164W/G256T
33	F36Y/R97E	75	G38V/Y95F/R97E	117	V85M/R97E/V109L/T221N/I234N
34	F36Y/M107R	76	G38V/Y95F/Y164W	118	V85M/R97E/V109L/I234N/K282N
35	F36Y/Y164W	77	G38V/Y95F/G256T	119	V22I/V85M/R97E/V109L/T221N/I234N
36	G38V/Y95F	78	G38V/R97E/M107R	120	V22I/V85M/R97E/V109L/I234N/K282N
37	G38V/R97E	79	G38V/R97E/Y164W	121	F36Y/G38V/Y95F/R97E/Y164W/G256T
38	G38V/Y164W	80	G38V/R97E/G256T	122	F36Y/Y95F/R97E/M107R/V109H/Y164W
39	G38V/G256T	81	G38V/Y164W/G256T	123	G38V/Y95F/R97E/M107R/Y164W/G256T
40	R40A/R97E	82	V85M/R97E/V109L	124	V22I/V85M/R97E/V109L/T221N/I234N/K282N
41	R40S/R97E	83	V85M/R97E/I234N	125	F36Y/G38V/Y95F/R97E/M107R/Y164W/G256T
42	V49R/R97E	84	Y95F/I96V/R97E		

Table S4 Created DATA variants with activities and apparent melting points. The activity was measured with (R)-PEA and pyruvate in the acetophenone or DAAO assay by the detection of the formed acetophenone or H_2O_2 respectively.

Name	DATA variant	Activity / mU mg^{-1}	app. Tm / °C
	WT	0.4 ± 0.1	67.0
	F26A	6.1 ± 0.4	68.1
	F26H	2.7 ± 0.1	65.9
	F26L	0.9 ± 0.2	66.0
	F26T	3.1 ± 0.1	65.0
	Y31A	5.1 ± 0.1	63.6
	Y31F	7.3 ± 1.0	69.9
	Y31L	3.2 ± 0.2	64.7
	Y31T	3.8 ± 0.1	61.9
	V33A	0.8 ± 0.1	65.1
	H86A	0.5 ± 0.1	n.m.
	H86F	0.2 ± 0.1	n.m.
	Y88A	3.4 ± 0.7	64.3
	Y88E	11.9 ± 0.2	62.9
	Y88F	0.6 ± 0.0	n.m.
	R98A	1.8 ± 0.1	67.0
	R98K	0.8 ± 0.1	n.m.
	H100A	1.4 ± 0.1	69.3
	S180A	0.6 ± 0.0	67.0
	S180L	0.0 ± 0.0	n.m.
	S180V	0.0 ± 0.5	n.m.
	S180W	0.8 ± 0.0	67.1
	T242A	0.2 ± 0.1	n.m.
	F26A/Y31A	9.6 ± 0.1	63.2
	F26A/R98K	2.3 ± 0.1	n.m.
	F26H/Y88E	3.3 ± 0.2	n.m.
	Y31F/S180A	3.2 ± 0.1	n.m.
	Y31M/S180V	5.2 ± 0.1	62.9
	H86F/Y88F	0.9 ± 0.1	65.6
	Y88E/Y165T	6.4 ± 0.3	61.0
	Y88E/S180G	4.9 ± 0.3	62.0
	Y88E/T242A	1.4 ± 0.0	n.m.
	Y88E/T242G	0.7 ± 0.0	60.1
	R98A/A99R	1.1 ± 0.0	66.6

	F26A/Y31A/Y88A	4.8 ± 0.1	58.4
	F26A/Y31A/Y88E	0.1 ± 0.0	n.m.
M2-3	Y31F/H86F/Y88F	40.9 ± 3.7	64.4
	Y31F/S180A/T242I	9.1 ± 0.2	65.6
	Y31M/S180V/T242W	14.2 ± 0.3	57.4
	H86F/I87V/Y88E	2.4 ± 0.2	n.m.
	H86F/Y88F/S180A	2.3 ± 0.1	62.6
	H86F/Y88F/T242I	4.6 ± 0.2	64.5
	Y88E/R98A/A99R	0.6 ± 0.0	n.m.
	Y31F/H86F/Y88F/T242I	120.7 ± 12.8	66.4
M2-4	Y31F/H86F/S180A/T242I	18.3 ± 0.1	69.9
	Y31F/Y88F/S180A/T242I	121.4 ± 0.7	66.5
	H86F/Y88F/S180A/T242I	5.7 ± 0.3	63.7
M2-5	Y31F/H86F/Y88F/S180A/T242I	184.7 ± 6.9	65.7
	Y31M/H86L/Y88T/S180V/T242W	12.3 ± 2.2	n.m.
	Y31F/H86F/Y88F/R98K/S180A/T242I	94.8 ± 1.3	n.m.
	Y31F/H86F/Y88F/H100K/S180A/T242I	128.8 ± 1.2	n.m.
M2-6	Y31F/H86F/Y88F/H100L/S180A/T242I	325.8 ± 2.6	69.1
M2	Y31F/H86F/Y88F/R98W/H100W/S180A/T242I	0.3 ± 0.0	n.m.
M5	Y31M/H86I/Y88V/R98Y/H100L/S180L/T242L	0.0 ± 0.0	n.m.
M1	Y31M/H86L/Y88F/R98W/H100L/S180L/T242I	0.3 ± 0.0	n.m.
M4	Y31M/H86L/Y88W/R98T/H100W/S180V/T242W	0.0 ± 0.0	66.2
M3	Y31F/V33A/H86L/Y88F/R98Y/H100L/S180L/T242W	0.0 ± 0.0	n.m.

Table S5 Proposed DATA mutants (M1-M5) from the Rosetta analysis.

										Rosetta score			
Variant	**Amino acid residue number**									**PLP-(*R*)-phenylethylamine (Rosetta Unit)**		**PLP-D-alanine (Rosetta Unit)**	
	26	**31**	**33**	**86**	**88**	**98**	**100**	**180**	**242**	**Total**	**Ligand**	**Total**	**Ligand**
wt	F	Y	V	H	Y	R	H	S	T	-539.83	-10.03	-540.85	-5.80
M1	-	M	-	L	F	W	L	L	I	-585.44	-12.41	-569.78	-6.39
M2	-	F	-	F	F	W	W	A	I	-584.83	-13.15	-564.86	-6.27
M3	-	F	A	L	F	Y	L	L	W	-582.35	-12.04	-560.83	-6.32
M4	-	M	-	L	W	T	W	V	W	-581.82	-12.77	-563.38	-6.17
M5	-	M	-	I	V	Y	L	L	L	-581.72	-12.41	-566.20	-6.06

Table S6 MM/PBSA binding free energy decomposed by amino acid residue (ΔG). The list gives hot spot residues having a binding free energy value >1 kcal mol^{-1} in absolute value. σ: standard deviation in kcal mol^{-1}.

Wild type - PLP-D-alanine					Wild type - PLP-(*R*)-PEA					M2-6 - PLP-D-alanine					M2-6 - PLP-(*R*)-PEA				
AA	**Chain**	**#res**	**ΔG**	**σ**	**AA**	**Chain**	**#res**	**ΔG**	**σ**	**AA**	**Chain**	**#res**	**ΔG**	**σ**	**AA**	**Chain**	**#res**	**ΔG**	**σ**
ASP	A	21	1.19	0.20	GLU	A	32	7.69	0.96	ASP	A	28	1.18	0.35	GLU	A	32	8.55	1.36
ASP	A	28	1.26	0.25	HIS	A	47	-1.64	0.46	GLU	A	32	12.90	1.62	HIS	A	47	-2.62	0.74
TYR	A	31	-2.14	0.87	ARG	A	50	-25.05	1.41	HIS	A	47	-1.75	0.56	ARG	A	50	-25.28	2.15
GLU	A	32	12.32	1.06	ARG	A	138	-1.72	0.93	ARG	A	50	-31.28	2.33	ARG	A	138	-2.64	0.85
HIS	A	47	-2.00	0.69	LYS	A	156	-1.85	0.33	LYS	A	57	-1.16	0.34	LYS	A	156	-1.96	0.42
ARG	A	50	-29.91	1.71	GLU	A	166	3.24	0.65	ARG	A	120	-1.37	0.22	GLU	A	166	2.78	0.56
LYS	A	57	-1.10	0.26	GLU	A	177	5.98	2.20	ASP	A	136	1.88	0.71	GLU	A	177	4.93	2.65
LYS	A	116	-1.11	0.35	GLY	A	178	1.10	0.29	ARG	A	138	-5.87	2.17	GLY	A	178	1.06	0.33
ARG	A	120	-1.43	0.22	SER	A	180	-1.80	0.62	ASP	A	143	1.44	0.45	ALA	A	180	-2.08	0.53
ASP	A	136	2.95	0.45	LEU	A	201	-2.17	0.39	LYS	A	156	-3.92	0.72	SER	A	181	-1.49	0.53
ARG	A	138	-7.56	1.16	LYS	A	202	-1.58	0.37	GLU	A	166	5.26	0.94	LEU	A	201	-1.80	0.51
ASP	A	143	1.80	0.41	GLY	A	203	-6.28	0.66	ARG	A	171	-1.63	0.58	LYS	A	202	-1.07	0.46
LYS	A	145	-1.34	0.46	ILE	A	204	-7.01	0.45	GLU	A	177	3.71	1.63	GLY	A	203	-6.31	1.07
LYS	A	156	-4.94	0.52	THR	A	205	-9.19	0.97	SER	A	179	2.10	0.72	ILE	A	204	-7.66	1.14
GLU	A	166	4.90	0.59	ARG	A	206	-5.38	0.77	SER	A	181	-1.68	0.69	THR	A	205	-9.47	1.67
ARG	A	171	-2.02	0.43	ASP	A	207	1.69	0.54	LEU	A	201	-2.17	0.27	ARG	A	206	-4.32	1.24
GLU	A	177	10.46	2.06	THR	A	239	-1.20	0.46	LYS	A	202	-2.00	0.50	SER	A	240	-9.02	2.69
SER	A	179	1.15	0.45	SER	A	240	-4.11	0.72	GLY	A	203	-7.66	0.63	THR	A	241	-6.52	1.02
SER	A	180	-2.54	0.92	THR	A	241	-5.92	1.12	ILE	A	204	-7.40	1.16	ILE	A	242	-1.52	0.38
SER	A	181	-1.91	0.59	GLU	A	244	1.81	0.26	THR	A	205	-10.65	2.24	GLU	A	244	1.84	0.36
ILE	A	200	-1.25	0.21	ARG	B	98	1.37	0.66	ARG	A	206	-5.21	1.06					
LEU	A	201	-2.08	0.62						ASP	A	207	1.20	0.66					
LYS	A	202	-2.09	0.47						SER	A	240	-8.90	2.73					
GLY	A	203	-7.04	0.72						THR	A	241	-9.29	1.00					
ILE	A	204	-7.93	0.65						ILE	A	242	-2.04	0.42					
THR	A	205	-11.25	0.90						GLU	A	244	2.72	0.54					
ARG	A	206	-6.24	0.63						ASP	A	304	1.37	0.32					
ASP	A	207	1.90	0.50						ARG	A	305	-2.31	0.40					
SER	A	240	-8.56	1.34						ASP	A	311	2.00	0.45					
THR	A	241	-7.82	1.34						ARG	B	93	-1.23	0.32					
GLU	A	244	3.06	0.46						ARG	B	98	-16.19	2.91					
ASP	A	304	1.37	0.18															
ARG	A	305	-1.95	0.25															
ASP	A	311	1.76	0.23															
ARG	B	93	-1.17	0.19															
ARG	B	98	-14.01	2.29															

Supporting Figures

```
            N-terminal helix
DATA        -------------------------GYTLWNDQIVKDEEVKIDKEDRGYQFGDGVYEVV 34
R-ATA       ASMDKVFSGYYARQKLLERSDNPFSKGIAYVEGKLVLPSDARIPLLDEGFMHSDLTYDVI 60

                                                                    L1
DATA        KVYNGEMFTVNEHIDRLYASAEKIRITIPYTKDKFHQLLHELVEKNELNTGHIYFQVTRG 94
R-ATA       SVWDGRFFRLDDHLQRILESCDKMRLKFPLALSSVKNILAEMVAKSGIRDAFVEVIVTRG 120

            L1                      L2
DATA        TSPRAHQFPENTVKPVIIGYTKENPR--PLENLEKGVKATFV--EDIRWL---RCDIKSL 147
R-ATA       LTGVRGSKPEDLYNNNIYLLVLPYIWVMAPENQLHGGEAIITRTVRRTPPGAFDPTIKNL 180

DATA        NLLGAVLAKQEAHEKGCYEAILHR-NNTVTEGSSSNVFGIKDGILYTHPANNMILKGITR 206
R-ATA       QWGDLTKGLFEAMDRGATYPFLTDGDTNLTEGSGFNIVLVKNGIIYTPDR--GVLRGITR 238

DATA        DVVIACANEINMPVKEIPFTTHEALKMDELFVTSTTSEITPVIEIDGKLIRDGKVGEWTR 266
R-ATA       KSVIDVARANSIDIRLEVVPVEQAYHSDEIFMCTTAGGIMPITLLDGQPVNDGQVGPITK 298

DATA        KLQKQFETKIPKPLHI                                            282
R-ATA       KIWDGYWEMHYNPAYSFPVDYGSG                                    322
```

Figure S1 Protein sequence alignment of the DATA and the (*R*)-ATA from *A. fumigatus*. The alignment is based on the structural inspection of superimposition of the respective crystal structures (PDB ID: 3DAA and 4CHI) in PyMOL and the rcsb.org tool: *Sequence & Structural Alignment* using the FATCAT protocol. Green: DATA residues mutated in the Rosetta approach; purple: catalytic lysine.

```
            N-terminal helix
BCAT        ----------------------MTTKKADYIWFNGEMVRWEDAKVHVMSHALHYGTSVFEG 39
R-ATA       MASMDKVFSGYYARQKLLERSDNPFSKGIAYVEGKLVLPSDARIPLLDEGFMHSDLTYDV 60

BCAT        IRCYDSHKGPVVFRHREHMQRLHDSAKIYRFPVSQSIDELMEACRDVIRKNNLTSAYIRP 99
R-ATA       ISVWD----GRFFRLDDHLQRILESCDKMRLKFPLALSSVKNILAEMVAKSGIRDAFVEV 116

               L1                       L2
BCAT        LIFVGDVGMGVNPP-AGYSTDVIIAAFPWGAYLGAEALEQGIDAMVSS-WNRAAPNTIPT 157
R-ATA       IVTRGLTGVRGSKPEDLYNNNIYLLVLPYIWVMAPENQLHGGEAIITRTVRRTPPGAFDP 176

BCAT        AAKAGGNYLSSLLVGSEARRHGYQEGIALDVNGYISEGAGENLFEVKDGVLFTPPFTSSA 217
R-ATA       TIKNLQ-WGDLTKGLFEAMDRGATYPFLTDGDTNLTEGSGFNIVLVKNGIIYTPDR--GV 233

BCAT        LPGITRDAIIKLAKELGIEVREQVLSRESLYLADEVFMSGTAAEITPVRSVDGIQVGEGR 277
R-ATA       LRGITRKSVIDVARANSIDIRLEVVPVEQAYHSDEIFMCTTAGGIMPITLLDGQPVNDGQ 293

BCAT        CGPVTKRIQQAFFGLFTGETEDKWGWLDQVNQ---                         309
R-ATA       VGPITKKIWDGYWEMHYN-----PAYSFPVDYGSG                         323
```

Figure S2 Protein sequence alignment of the *e*BCAT and the (*R*)-ATA from *A. fumigatus*. The alignment is based on the structural inspection of superimposition of the respective crystal structures (PDB ID: 3DAA and 4CHI) in PyMOL and the rcsb.org tool: *Sequence & Structural Alignment* using the FATCAT protocol. Purple: catalytic lysine.

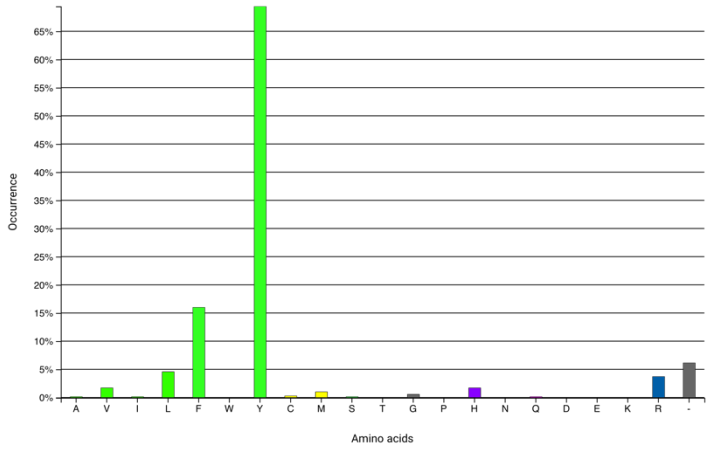


Figure S3 Amino acid distribution on the 3D-position 26 (DATA: F26) in the aligned core region in the 3DM-dataset.

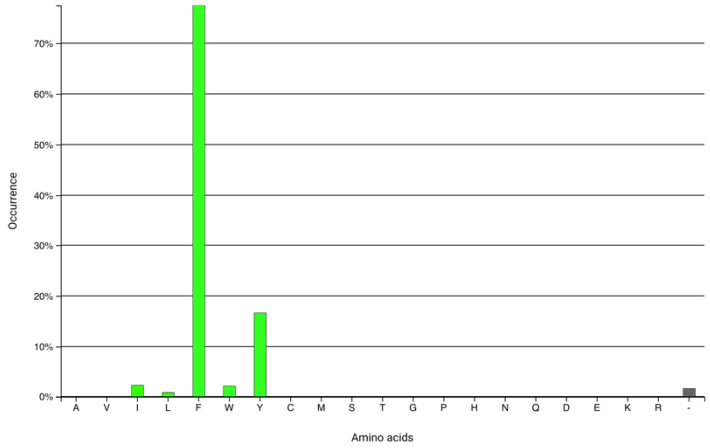


Figure S4 Amino acid distribution on the 3D-position 31 (DATA: Y31) in the aligned core region in the 3DM-dataset.

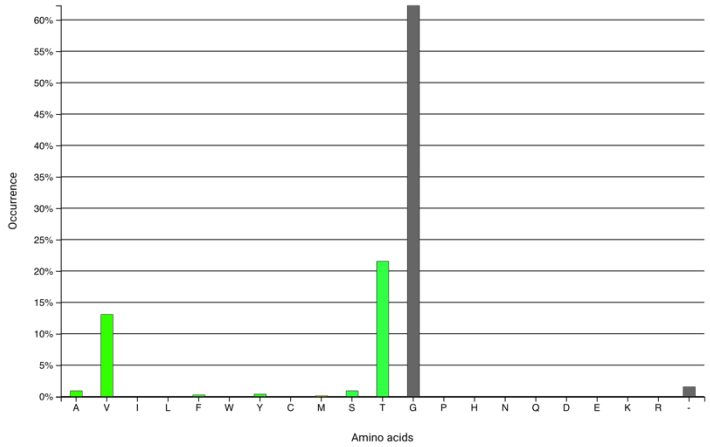


Figure S5 Amino acid distribution on the 3D-position 33 (DATA: V33) in the aligned core region in the 3DM-dataset.

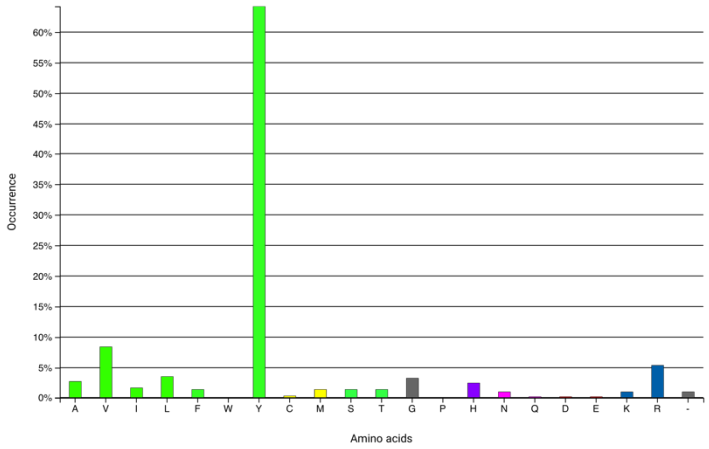


Figure S6 Amino acid distribution on the 3D-position 86 (DATA: H86) in the aligned core region in the 3DM-dataset.

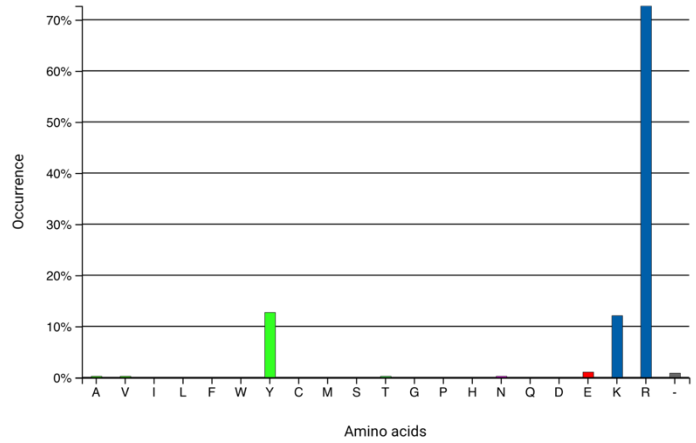


Figure S7 Amino acid distribution on the 3D-position 88 (DATA: Y88) in the aligned core region in the 3DM-dataset.

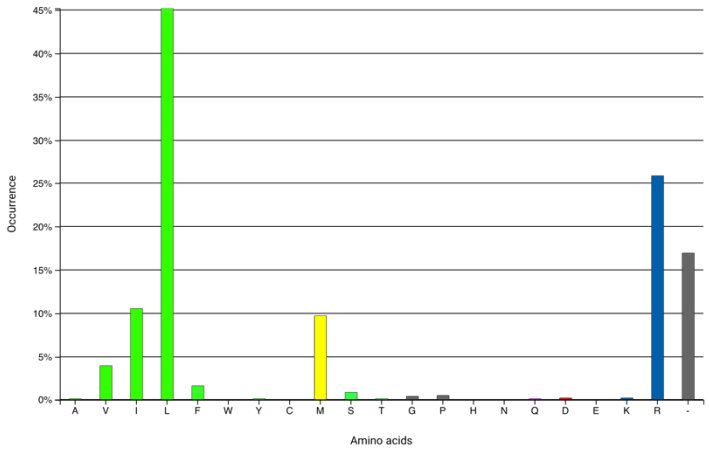


Figure S8 Amino acid distribution on the 3D-position 98 (DATA: R98) in the aligned core region in the 3DM-dataset.

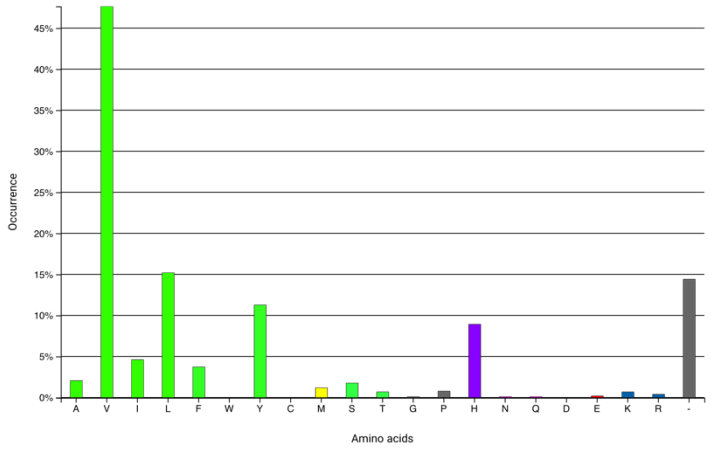


Figure S9 Amino acid distribution on the 3D-position 100 (DATA: H100) in the aligned core region in the 3DM-dataset.

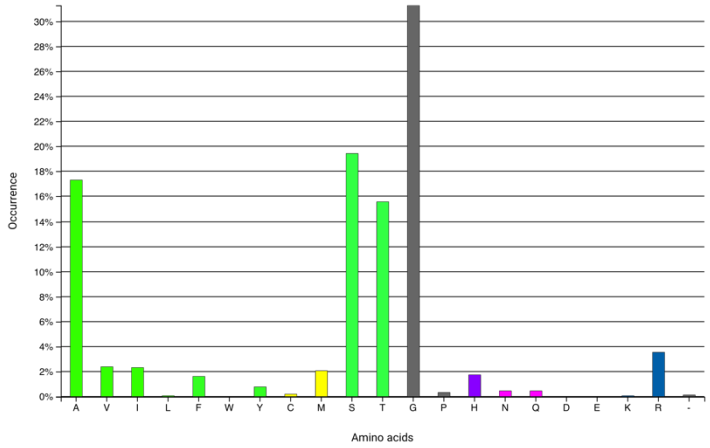


Figure S10 Amino acid distribution on the 3D-position 179 (DATA: S180) in the aligned core region in the 3DM-dataset.

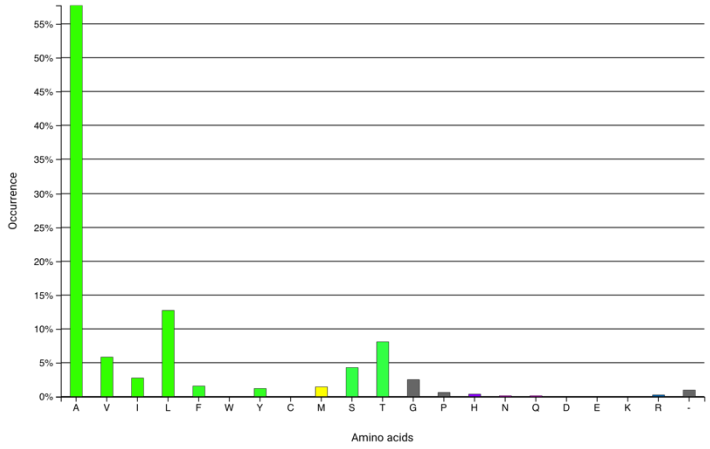


Figure S11 Amino acid distribution on the 3D-position 241 (DATA: T242) in the aligned core region in the 3DM-dataset.

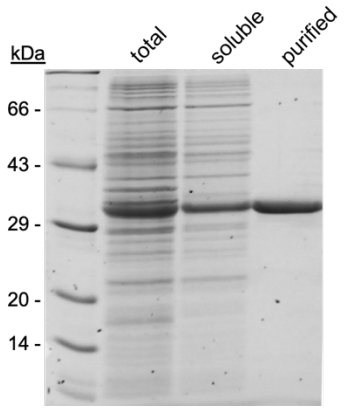


Figure S12 SDS-PAGE analysis of the expression and purification of the DATA wild type. Theoretical size: 33.5 kDa.

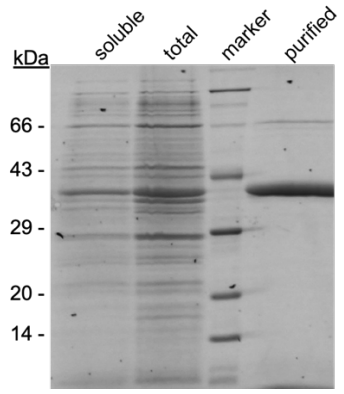


Figure S13 SDS-PAGE analysis of the *Tv*_DAAO expression and purification. The theoretical MW of the *Tv*_DAAO protein is 41.1 kDa.

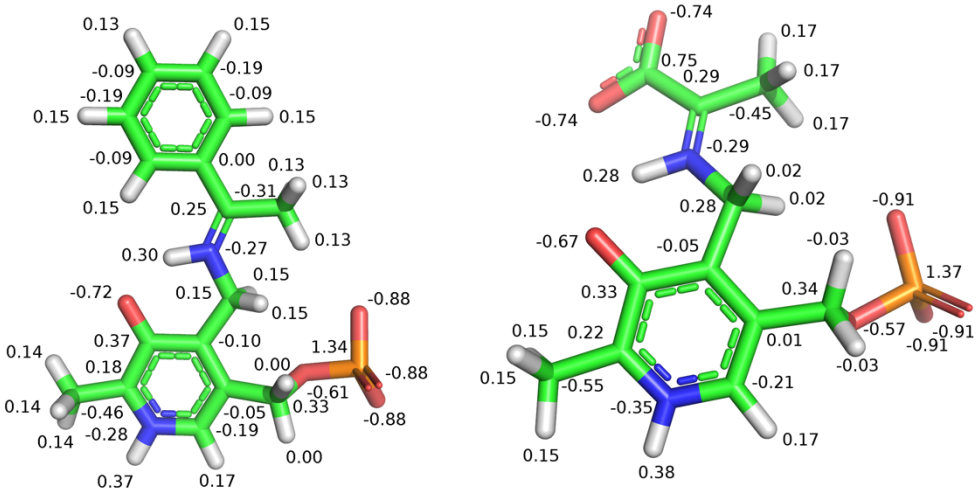


Figure S14 Charge parametrization of the ligand (ketimine). The ketimine was preferred to quinoid structure for a parametrization facility. The ligands correspond to the PLP-(*R*)-phenylethylamine (on the left) or PLP-D-alanine (on the right).

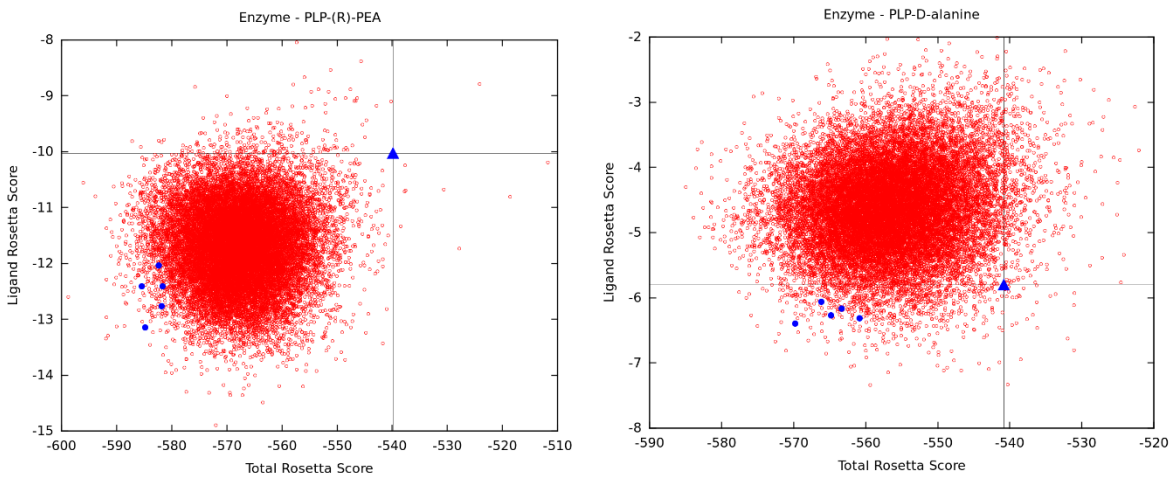


Figure S15 Graphics representing the total score (X-axis) against the enzyme:ligand (PLP-(*R*)-phenylethylamine or PLP-D-alanine) interface score (y-axis). All 20,000 sequences were plotted on the scatter plot and compared to the wild-type sequence (blue triangles crossed by vertical and horizontal lines).

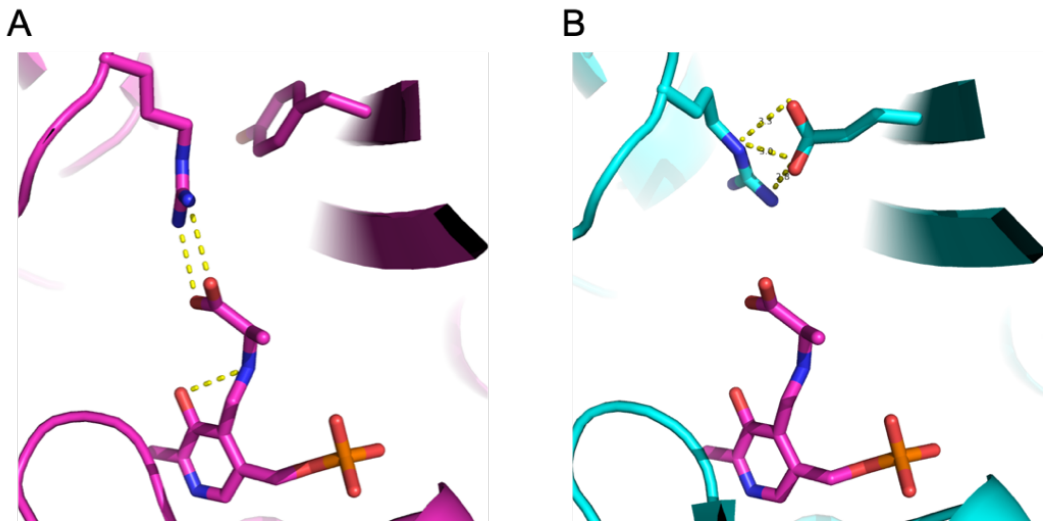


Figure S16 Model of the potential dual-substrate recognition in the DATA Y88E variant. A: Coordination of the α-carboxylate function of the D-alanine-PLP ligand with the R98 guanidino function as shown in the crystal structure PDB ID: 3DAA. B: Model of the Y88E mutation and the potential coordination of the guanidino function of R98.

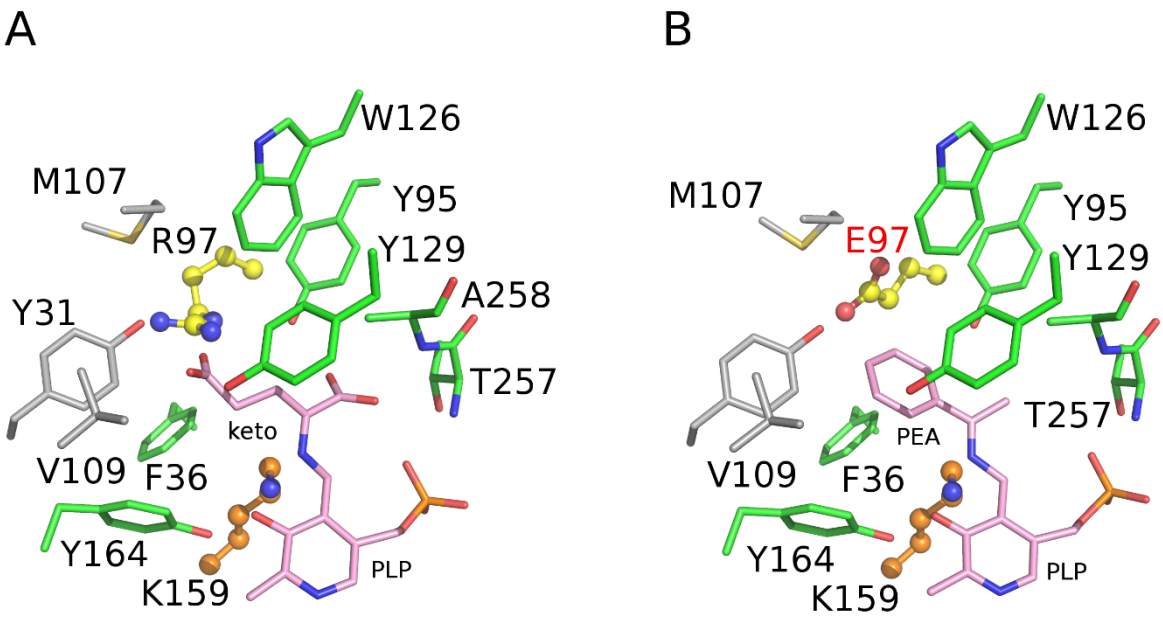


Figure S17 Binding pocket of the *e*BCAT WT and the R97E mutant. The wild type (PDB ID: 1IYE) is shown in A, whereas the mutant R97E is shown in B. The residues close to the ketoglutarate (keto) or (*R*)-PEA regions are displayed in sticks (green for chain A and grey for chain B). The mutated residue (R97) is represented in yellow ball and sticks, whereas the catalytic lysine is colored in orange. The mutant model was built by superimposition of PLP-PEA on the crystallographic ligand, whereas the mutation was performed using Modeller 9.19.[7]

References

(1) Hirotsu, K.; Goto, M.; Okamoto, A.; Miyahara, I. Dual substrate recognition of aminotransferases. *Chem. Rec.* **2005**, *5*, 160–172.

(2) Höhne, M.; Schätzle, S.; Jochens, H.; Robins, K.; Bornscheuer, U. T. Rational assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.* **2010**, *6*, 807–813.

(3) Weiß, M. S.; Pavlidis, I. V.; Vickers, C.; Hohne, M.; Bornscheuer, U. T. Glycine oxidase based high-throughput solid-phase assay for substrate profiling and directed evolution of (*R*)- and (*S*)-selective amine transaminases. *Anal. Chem.* **2014**, *86*, 11847–11853.

(4) Skalden, L.; Thomsen, M.; Höhne, M.; Bornscheuer, U. T.; Hinrichs, W. Structural and biochemical characterization of the dual substrate recognition of the (*R*)-selective amine transaminase from *Aspergillus fumigatus*. *FEBS J.* **2015**, *282*, 407–415.

(5) Łyskowski, A.; Gruber, C.; Steinkellner, G.; Schürmann, M.; Schwab, H.; Gruber, K.; Steiner, K. Crystal structure of an (*R*)-selective ω-transaminase from *Aspergillus terreus*. *PLoS One* **2014**, *9*.

(6) Guan, L. J.; Ohtsuka, J.; Okai, M.; Miyakawa, T.; Mase, T.; Zhi, Y.; Hou, F.; Ito, N.; Iwasaki, A.; Yasohara, Y.; et al. A new target region for changing the substrate specificity of amine transaminases. *Sci. Rep.* **2015**, *5*, 1–8.

(7) Webb, B.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **2016**, *2016*, 5.6.1-5.6.37.

Bioorg. Med. Chem. 43 (2021) 116271

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Directed evolution of an amine transaminase for the synthesis of an Apremilast intermediate via kinetic resolution

Chao Xiang [a], Shuke Wu [a,b], Uwe T. Bornscheuer [a,*]

[a] Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, Felix Hausdorff-Str. 4, 17487 Greifswald, Germany
[b] State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Street, Wuhan 430070, PR China

ARTICLE INFO

Keywords:
Biocatalysis
Chiral amine
Directed evolution
Kinetic resolution
Transaminase

ABSTRACT

Apremilast is an important active pharmaceutical ingredient that relies on a resolution to produce the key chiral amine intermediate. To provide a new catalytic and enzymatic process for Apremilast, we performed the directed evolution of the amine transaminase from *Vibrio fluvialis*. Six rounds of evolution resulted in the VF-8M-E variant with > 400-fold increase specific activity over the wildtype enzyme. A homology model of VF-8M-E was built and a molecular docking study was performed to explain the increase in activity. The purified VF-8M-E was successfully applied to produce the key chiral amine intermediate in enantiopure form and 49% conversion via a kinetic resolution, representing a new enzymatic access towards Apremilast.

1. Introduction

The chiral amine group is a prestigious motif widely found in bioactive natural products and man-made compounds, such as pharmaceuticals and agrochemicals. Many optically pure amines are key chiral building blocks for these bioactive chemicals and are also useful chiral ligands, or chiral reagents for resolution. The common chemical methods to access enantiopure chiral amines include traditional resolution using chiral acids and enantioselective catalysis with metal catalysts and chiral ligands.[1] However, these methods suffer from the use of stoichiometric chiral resolution reagents, or toxic transition metals and costly complex ligands. On the other hand, nature's catalysts, enzymes, provide a catalytic, highly selective, and environmentally friendly alternative method for the production of various chiral chemicals.[2] For the synthesis of chiral amines, the enzyme toolbox includes lipases,[3] transaminases,[4] amine dehydrogenases,[5] monoamine oxidases,[6] and imine reductases.[7] Despite the availability of these enzymes, due to the relatively high substrate specificity, natural enzymes usually require extensive directed evolution[8] for the production of a specific target, especially for the active pharmaceutical ingredients (APIs).[9]

A particularly interesting API containing a chiral amine group is Apremilast, (*S*)-*N*-{2-[1-(3-Ethoxy-4-methoxyphenyl)-2-methanesulfonylethyl]-1,3-dioxo-2,3-dihydro-1*H*-isoindol-4-yl}acetamide.[10] It is an analog of thalidomide and a potent inhibitor of phosphodiesterase 4 and tumor necrosis factor-α.[11] Apremilast was developed by Celgene for the treatment of autoimmune and inflammatory diseases (plaque psoriasis and psoriatic arthritis) under the name of Otezla®. The (*S*)-enantiomer of Apremilast is much more potent than the (*R*)-enantiomer,[10] thus the (*S*)-form was further developed into the API. The original synthetic route to Apremilast involves the resolution of racemic 1-(3-ethoxy-4-methoxyphenyl)-2-(methylsulfonyl)ethanamine (*rac*-**1**) to (*S*)-**1** with stoichiometric *N*-Ac-L-leucine (Scheme 1a).[10] The asymmetric hydrogenation approach had been explored to produce (*S*)-**1** from the corresponding enamine or ketone with precious and toxic rhodium or ruthenium complexes.[12] To our knowledge, there is only one report of an enzyme-mediated synthesis of (*S*)-**1**: lipase (Novozym 435) had been employed for the resolution of *rac*-**1** via acylation,[13] yet the *ee* of acylated product was not perfect.

As part of our research focus on amine transaminases (ATAs),[14–16] we were interested in developing a novel ATA-based method to produce (*S*)-**1** in enantiopure form (Scheme 1b), and thus to provide a catalytic, environmentally friendly, and economic process for Apremilast.

2. Results and discussion

2.1. Identification of a suitable ATA

We screened a range of wild-type (*S*)-selective ATAs and some of

* Corresponding author.
E-mail address: uwe.bornscheuer@uni-greifswald.de (U.T. Bornscheuer).

https://doi.org/10.1016/j.bmc.2021.116271
Received 29 March 2021; Received in revised form 30 May 2021; Accepted 4 June 2021
Available online 10 June 2021
0968-0896/© 2021 Elsevier Ltd. All rights reserved.

their mutants available in our laboratory for the conversion of *rac*-**1** with pyruvate, and analyzed the reaction products by HPLC. Although several wild-type and engineered (*S*)-selective ATAs displayed some activities towards bulky substrates,[14–16] most of them had no detectable activity towards *rac*-**1**. These results indicate the unique structural requirement of **1** in comparison with the bulky amines previously investigated. The only ATA with detectable activity was a double mutant of the ATA from *Vibrio fluvialis* (F85L/V153A, abbreviated as VF-2M), which was previously engineered for (*S*)-phenylbutylamine with a specific activity of 4.99 U/mg.[16a] VF-2M was purified, and its specific activity towards *rac*-**1** was determined by the arylketone assay[17] to be 0.62 mU/mg with pyruvate (Table 1, entry 2). The wildtype VF-ATA had no detectable activity (<0.1 mU/mg), demonstrating the importance of the F85L/V153A mutations for the initial activity.

2.2. Directed evolution of the Vibrio fluvialis ATA

To further improve the activity towards *rac*-**1**, we analyzed the crystal structure of VF-ATA (PDB: 4E3Q)[18] and focused on the residues forming the substrate binding pocket (Figure 1), including 14 residues of the small pocket (F19, L56, F85, F86, R88, R146, W147, V153, K163, Y165, T268, E315, E316), and 11 residues of the large pocket (W57, W150, A228, V258, I259, N286, R415, L417, C424, P426, F427). The residue Y249 at the entry of the substrate binding site was also selected. Using VF-2M as the starting template, single site-saturated mutagenesis of these 26 residues was performed with the Q5 mutagenesis method with NNK codons. For each site, 94 clones were picked and cultured in deep 96-well plates. The activity of VF-ATA variants in the cell lysate was screened for conversion of *rac*-**1** with pyruvate. The characteristic absorbance ($\lambda = 310$ nm) of the corresponding ketone **2**, was measured by a microtiter plate reader.

The iterative saturation mutagenesis (ISM)[19] strategy was applied to improve the activity of VF-ATA. The first round of directed evolution led to the triple mutant VF-3M (bearing an additional W57L mutation) with an activity of 4.38 mU/mg, representing a 7-fold improvement compared to the template VF-2M (Table 1). Using VF-3M as the next template, the second round of directed evolution gave two variants, 4M−A (additional K163A mutation) and 4M−E (additional K163E mutation), with slightly improved activities (Table 1). Further directed evolution pursued two different routes with 4M−A and 4M−E as the templates, respectively (Table 1). The third round of evolution led to 5M−A and 5M−E with the same additional N286C mutation (Table 1). Continuing these two routes gave 6M−A (12.45 mU/mg) and 6M−E (15.40 mU/mg) with the same additional R415K mutation (Table 1). The fifth round of evolution with 6M−A and 6M−E was performed, however, only the 6M−E direction gave a variant, 7M−E (additional F19H mutation), with an improved activity of 32.34 mU/mg (Table 1). The sixth round of evolution of 7M−E led to the best variant, 8M-E (additional Y249F mutation), with an activity of 39.43 mU/mg (Table 1). The VF-8M-E variant thus contains eight mutations: F19H/W57L/F85L/V153A/K163E/Y249F/N286C/R415K. In comparison with the wildtype VF-wt (<0.1 mU/mg) and starting template VF-2M (0.62 mU/mg), the final VF-8M-E variant displays > 400-fold and > 60-fold increased activity, respectively.

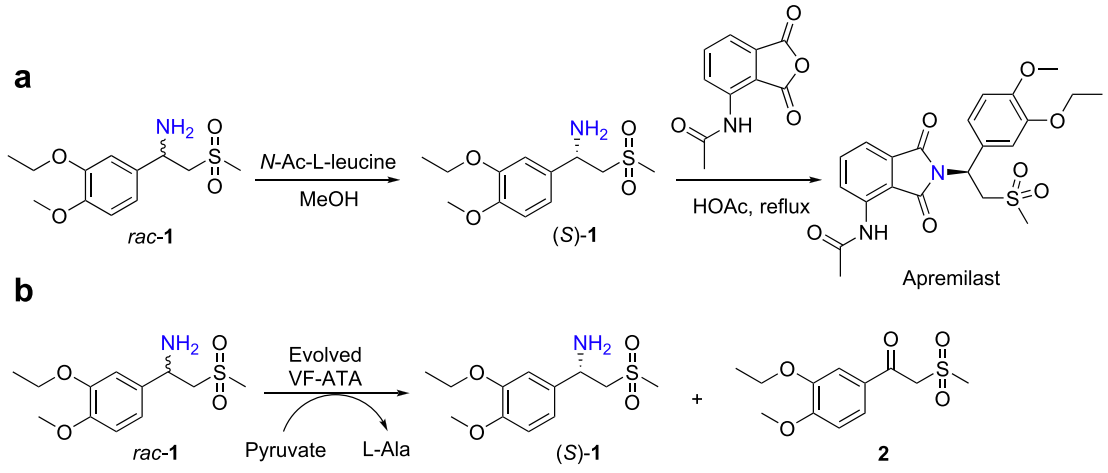
Table 1
VF-ATA variants and their corresponding activity towards *rac*-**1**/pyruvate, *rac*-**1**/pentanal[a].

Variant	Mutation	Specific Activity [mU/mg]	
		Pyruvate	Pentanal
VF-wt	–	< 0.1	< 0.1
2M	**F85L/V153A**	0.62 ± 0.07	0.49 ± 0.34
3M	**W57L**/F85L/ V153A	4.38 ± 0.12	0.96 ± 0.09
4M−A	W57L/F85L/V153A/**K163A**	5.43 ± 0.03	2.26 ± 0.11
4M−E	W57L/F85L/V153A/**K163E**	6.01 ± 0.07	4.42 ± 0.12
5M−A	W57L/F85L/V153A/K163A/**N286C**	8.06 ± 1.41	4.20 ± 0.33
5M−E	W57L/F85L/V153A/K163E/**N286C**	8.65 ± 0.23	6.03 ± 0.33
6M−A	W57L/F85L/V153A/K163A/N286C/**R415K**	12.45 ± 0.16	13.38 ± 0.23
6M−E	W57L/F85L/V153A/K163E/N286C/**R415K**	15.40 ± 0.55	20.95 ± 0.36
7M−E	**F19H**/W57L/F85L/V153A/K163E/N286C/R415K	32.34 ± 1.82	14.86 ± 0.48
8M-E	F19H/W57L/F85L/V153A/K163E/**Y249F**/N286C/R415K	39.43 ± 3.68	n.m.[b]

[a] Arylketone assay conditions: The specific activities of the purified VF-ATA variants for the conversion of *rac*-**1** were determined by using the arylketone assay in 96-well microtiter plates, and measured on the Infinite® 200 PRO (TECAN) plate reader. The assay was performed with *rac*-**1** (5 mM) as amine donor and pyruvate (2.5 mM) or pentanal (2.5 mM) as amine acceptors in DMSO (1.25–2.5%), HEPES buffer (50 mM, pH 6.5) at 30 °C. The formation of **2** was quantified by following the increase of absorption at 310 nm over time. One unit (U) was defined as the formation of 1 μmol **2** per minute. All measurements were performed in triplicates.
[b] n.m. = not measured.

2.3. Characterization of the best three variants of VF-ATA

The best three variants, VF-6M−E, VF-7M−E, and VF-8M-E, were selected for further characterization. Because the protein engineering rounds were performed with *rac*-**1**, we further examined the enantioselectivity of the three variants for converting (*S*)-**1**. Fortunately, all the VF-ATA variants did not show any detectable activity towards (*S*)-**1** (<0.1 mU/mg). This proved the excellent enantioselectivity of VF-ATA



Scheme 1. **a** The original synthetic route to Apremilast involves resolution of *rac*-**1** to (*S*)-**1**. **b** Kinetic resolution of *rac*-**1** to (*S*)-**1** by the evolved VF-ATA.

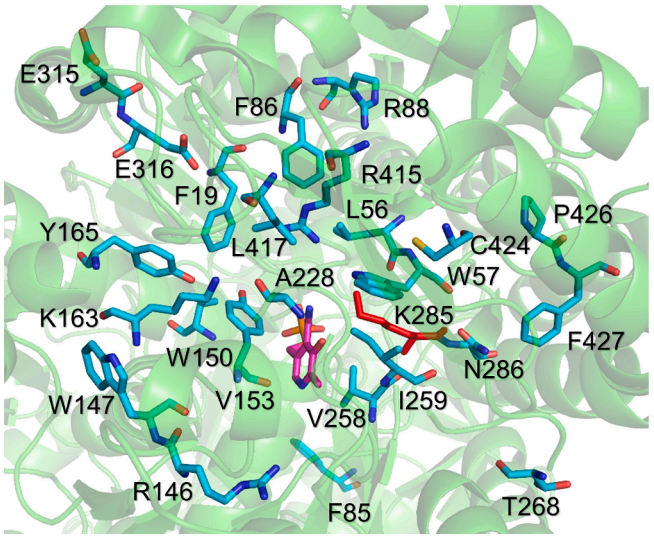


Figure 1. Structure of VF-ATA (PDB: 4E3Q) with the key residues labeled. Key residues are shown as sticks: pyridoxamine-5′-phosphate (PMP) in pink, the catalytic lysine (K285) is highlighted in red, other sites for mutation are in blue.

variants. The three variants were purified and tested in different reaction buffer (Figure 2). The maximum specific activity was achieved at the optimal pH of 6.0 or 6.5. Then, the best three variants were applied for the kinetic resolution of *rac*-**1** under the optimal conditions. As shown in Figure 3, 20 mM of *rac*-**1** was resolved by VF-8M-E in 24 h to reach 51% conversion, while VF-7M−E and VF-6M−E took 48 h to reach > 50% conversion. The optical purity of the remaining (*S*)-**1** at 48 h was determined to be > 99% *ee*. During the reaction, a small percentage (~10%) of byproduct 1-(3-ethoxy-4-methoxyphenyl)ethenone was observed, probably due to the α,β-elimination of (*R*)-**1**. Nevertheless, these results demonstrated that the highly enantioselective VF-ATA variants could be applied to produce the enantiopure amine intermediate for Apremilast. The unpurified cell lysates were also proven as catalysts for the kinetic resolution (see Figure S2 in SI). To demonstrate the synthetic application, the kinetic resolution was performed on 20 mL scale (100 mg) using the stable variant (VF-6M−E). The reaction took 48 h to reach a conversion of 50%. Simple workup (filtration, extraction, and evaporation) afforded (*S*)-**1** (>99% *ee*) in 41% isolated yield, and ketone **2** in 43% isolated yield.

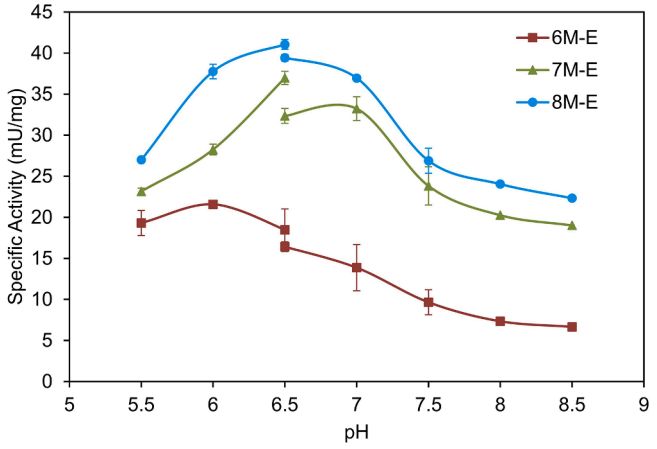


Figure 2. Specific activity of the best three VF-ATA variants for conversion of *rac*-**1** in different buffers. Assay conditions: 5 mM *rac*-**1**, 50 mM potassium phosphate buffer (pH 5.5–6.5) or HEPES buffer (pH 6.5–8.5), 0.1 mM PLP, 2.5 mM pyruvate, 1% (vol/vol) DMSO, 30 °C.

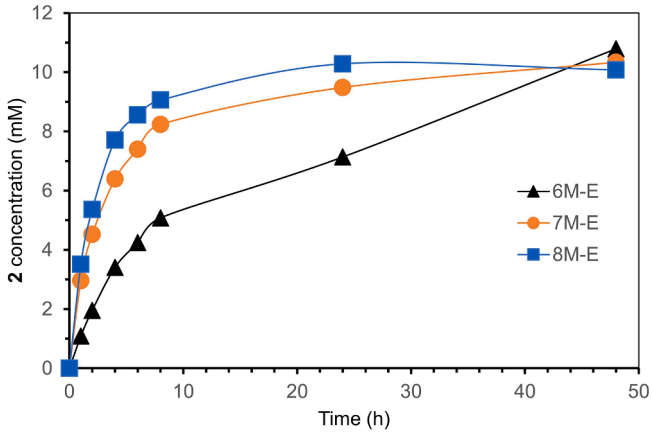


Figure 3. Kinetic resolution of *rac*-**1** with the best three VF-ATA variants. Assay conditions: 20 mM *rac*-**1**, 50 mM HEPES buffer pH 6.5, 0.1 mM PLP, 200 mM pyruvate, 10% (vol/vol) DMSO, 0.3–0.8 mg/ml purified enzyme, 37 °C and 180 rpm.

2.4. Homology modeling and docking study of VF-8M-E

To elucidate the possible mechanism and the molecular rational for the observed improved activity towards *rac*-**1**, a homology model of the VF-8M-E was built with YASARA. As shown in Figure 4a and 4b, in comparison to the wildtype VF-wt, the substrate binding pocket was significantly expanded in the VF-8M-E variant, which hence can now accommodate the bulky substrate (*R*)-**1** much better. The molecular docking of the reaction intermediate, PLP-(*R*)-**1** complex, to the model of VF-8M-E was simulated with YASARA. As depicted in Figure 4c, the position of PLP in VF-8M-E was very similar to that of PMP in VF-wt, which validated that PLP-(*R*)-**1** complex was docked in a reasonable conformation. It is clear that, the W57L and R415K mutations significantly expanded the substrate binding pocket to accommodate the challenging methylsulfonyl group. On the other side of the binding pocket, F19H, F85L, and V153A contributed to host the 3-ethoxy-4-methoxyphenyl group. The contribution of the other three mutations is minor.

3. Conclusion

To provide a new enzymatic access to the chiral amine intermediate (*S*)-**1** for the synthesis of Apremilast, we identified a double mutant of the ATA from *Vibrio fluvialis*, which had initial activity for the conversion of *rac*-**1**. Subsequently, six rounds of directed evolution resulted in the VF-8M-E variant with > 60-fold increase specific activity. A homology model of VF-8M-E was built and molecular docking study was performed to explain the increase of activity. The purified VF-8M-E was successfully applied in the kinetic resolution of *rac*-**1** to produce enantiopure (*S*)-**1** in 49% conversion. The preparative kinetic resolution was performed with the stable variant VF-6M−E to afford enantiopure (*S*)-**1** in 41% isolated yield. The application of this ATA variant represents an alternative method for the synthesis of the key chiral amine intermediate for the manufacturing of Apremilast.

4. Experimental section

4.1. Site-directed mutagenesis

All variants were prepared using Q5® site-directed mutagenesis kit from New England BioLabs. The degenerated primers were designed non-overlapping by using the standard setting of NEBaseChanger. For the PCR, template plasmid (0.25 ng/μL, carrying the VF-2M gene), forward and reverse primers (0.5 μM each), Q5® hot start high-fidelity

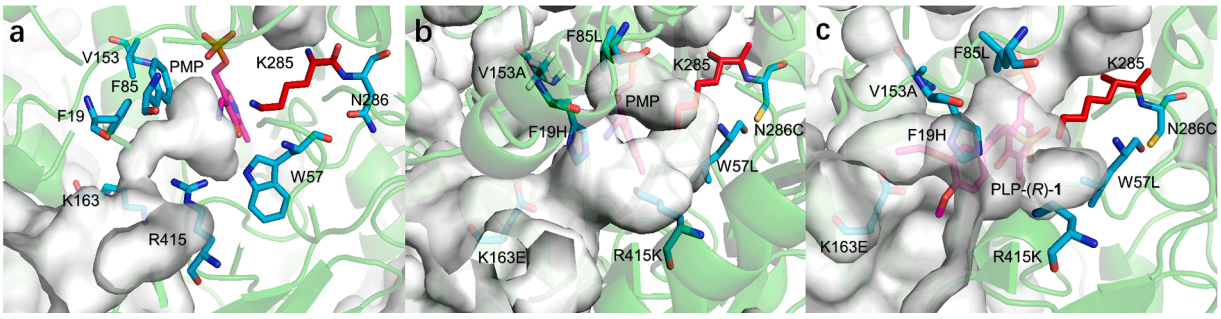


Figure 4. Structural study of VF-wt and VF-8M-E. **a** The substrate binding pocket of VF-wt; **b** the substrate binding pocket of the homology model of VF-8M-E; **c** docking of PLP-(*R*)-**1** complex to the model of VF-8M-E. The PMP and PLP-(*R*)-**1** are in pink, the catalytic lysine (K285) is highlighted in red, and the mutated residues are highlighted in blue.

2X master mix were used. The PCR was performed as follows: (i) 98 °C, 30 s; (ii) 30 cycles: 98 °C, 10 s; 50–72 °C, 30 s; 72 °C, 0.5 min/kbp; (iii) 72 °C, 2 min. The resulting PCR product was directly treated with the kinase, ligase & *Dpn*I (KLD enzyme mix; NEB) at room temperature for 30 min and then transformed in chemically competent *E. coli* TOP10 cells. After confirmation of the codon distribution in the case of site-saturation mutagenesis, the PCR products were transformed in chemical competent *E. coli* BL21(DE3) cells for expression.

4.2. Protein expression and purification of the enzyme variants

The plasmid constructs containing the genes of VF-ATA variants were transformed into *E. coli* BL21(DE3) cells, and the resulted cells were incubated in an LB-medium (Lysogeny Broth, 5 mL) preculture with kanamycin (50 μg/mL) at 37 °C overnight. The preculture was transferred into TB-medium (Terrific Broth, 100 mL) with kanamycin (50 μg/mL) and incubated at 37 °C, 180 rpm. The expression of the Vf-ATA variants was induced with IPTG (0.5 mM, isopropyl β-D-thio-galactopyranoside) at an optical density of approx. 0.6 at 600 nm, and incubated at 26 °C overnight. The cells were harvested by centrifugation (20 min, 4000 g).

For purification of the VF-ATA variants, the cell pellets were resuspended in HEPES buffer (50 mM, pH 7.5) containing NaCl (300 mM), PLP (0.1 mM), imidazole (10 mM), and then lysed by ultrasonication (50% power, 50% pulse, 2 × 5 min). The lysate was clarified by centrifugation (1 h, 10,000 g, 4 °C) and purified by affinity chromatography (Ni-NTA agarose) with the following buffers: washing buffer containing HEPES (50 mM, pH 7.5) PLP (0.1 mM), NaCl (300 mM), and imidazole (20 mM), and elution buffer containing HEPES (50 mM, pH 7.5) PLP (0.1 mM), NaCl (300 mM), and imidazole (300 mM). The VF-ATAs were desalted in HEPES buffer (50 mM pH 7.5) with PLP (0.1 mM) using the PD-10 desalting column (GE Healthcare). The purified VF-ATAs were stored in 30% glycerol at −20 °C.

4.3. Determination of activity of VF-ATA variants

The specific activities of the purified VF-ATA variants for the conversion of *rac*-**1** were determined by using the arylketone assay in 96-well microtiter plates, and measured on the Infinite® 200 PRO (TECAN) plate reader. The assay was performed with *rac*-**1** (5 mM) as amine donor and pyruvate (2.5 mM) or pentanal (2.5 mM) as amine acceptors in DMSO (1.25–2.5%), HEPES buffer (50 mM, pH 6.5) at 30 °C. The formation of **2** was quantified by following the increase of absorption at 310 nm over time.

4.4. Kinetic resolution of rac-1 with VF-ATA variants

To a 100-mL flask, the following components were added to form a 10-mL system: HEPES buffer (50 mM, pH 6.5) with PLP (0.1 mM), *rac*-**1** (20 mM), pyruvate (200 mM), DMSO (10%), purified VF-ATA (0.3–0.8 mg/mL). The reaction was incubated in a shaking incubator at 37 °C and 180 rpm. At 1, 2, 4, 6, 8, 24, and 48 h of the reaction, samples (100 μL) were taken and mixed with acetonitrile (500 μL) and trifluoracetic acid solution (0.1%, 400 μL) for HPLC analysis of the conversion of **1**. The samples (125 μL) were basified with NaOH (100 mM, 125 μL) and extracted with ethyl acetate (0.5 mL). The organic phase was dried over Na_2SO_4, filtered and the supernatant was subjected to evaporation. The dried residues were diluted in a mixture of hexane/ethanol/isopropanol (50:40:10) for chiral HPLC analysis of *ee*.

4.5. Preparation of (S)-1 with VF-ATA variants

To a 100-mL flask, HEPES buffer (50 mM, pH 6.5, including 0.1 mM PLP and 100 mM pyruvate), *rac*-**1** (100 mg), and purified VF-6 M−E (20 mg) were added to form a 20 mL reaction volume system. The reaction was maintained at pH 6.5 and 30 °C until reaching 50% conversion analyzed by HPLC. Afterwards, the reaction was quenched with HCl (100 mM) to pH 2.0 and the mixture was filtered. The filtrate was washed with CH_2Cl_2 (2 × 5 mL) and then NaOH solution (1 M) was added to adjust the pH to 12. The basified solution was extracted with CH_2Cl_2 (2 × 25 mL), and the CH_2Cl_2 was evaporated to afford (*S*)-**1** (41 mg, 41% yield) as yellow solid. ^{1}H NMR (DMSO-d_6) δ: 7.02 (m, 1H), 6.89 (m, 2H), 4.28–4.23 (m, 1H), 4.03–3.98 (q, J = 6.9 Hz, 2H), 3.73 (s, 3H), 3.45–3.20 (m, 2H), 2.96 (s, 3H), 2.07 (brs, 2H), 1.35–1.30 (t, J = 6.9 Hz, 3H). To isolate **2**, the solid of the initial filtration was washed with HCl (100 mM), dissolved in acetonitrile, and filtered again, the filtrate was evaporated to give ketone **2** (43 mg, 43% yield) as white solid. ^{1}H NMR (DMSO-d_6) δ: 7.73–7.72 (m, 1H), 7.50–7.49 (m, 1H), 7.13–7.09 (m, 1H), 5.04 (s, 2H), 4.13–4.05 (q, J = 7.2 Hz, 2H), 3.87 (s, 3H), 3.13 (s, 3H), 1.38–1.32 (t, J = 7.2 Hz, 3H).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

X. C. thanks the China Scholarship Council for financial support of a PhD thesis project (File No.: 201808330394). S. W. thanks the Alexander von Humboldt-Stiftung for a Humboldt Research Fellowship. The authors thank Ina Menyes for support with HPLC analysis.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmc.2021.116271.

References

1 (a) Breuer M, Ditrich K, Habicher T, et al. *Angew Chem Int Ed.* 2004;43:788–824. https://doi.org/10.1002/anie.200300599;
(b) Nugent TC, El-Shazly M. *Adv Synth Catal.* 2010;352:753–819. https://doi.org/10.1002/adsc.200900719.

2 (a) Sheldon RA, Brady D, Bode ML. *Chem Sci.* 2020;11:2587–2605. https://doi.org/10.1039/C9SC05746C;
(b) Hauer B. *ACS Catal.* 2020;10:8418–8427. https://doi.org/10.1021/acscatal.0c01708;
(c) Hollmann F, Opperman DJ, Paul CE. *Angew Chem Int Ed.* 2021;60:5644–5665. https://doi.org/10.1002/anie.202001876;
(d) Winkler CK, Schrittwieser JH, Kroutil W. *ACS Cent Sci.* 2021;7:55–71. https://doi.org/10.1021/acscentsci.0c01496;
(e) Wu S, Snajdrova R, Moore JC, Baldenius K, Bornscheuer UT. *Angew Chem Int Ed.* 2021;60:88–119. https://doi.org/10.1002/anie.202006648;
(f) Marshall JR, Mangas-Sanchez J, Turner NJ. *Tetrahedron.* 2021;82. https://doi.org/10.1016/j.tet.2021.131926.

3 (a) Ghislieri D, Turner NJ. *Top Catal.* 2014;57:284–300. https://doi.org/10.1007/s11244-013-0184-1;
(b) Kohls H, Steffen-Munsberg F, Höhne M. *Curr Opin Chem Biol.* 2014;19:180–192. https://doi.org/10.1016/j.cbpa.2014.02.021.

4 (a) Hohne M, Bornscheuer UT. *ChemCatChem.* 2009;1:42–51. https://doi.org/10.1002/cctc.200900110;
(b) Koszelewski D, Tauber K, Faber K, Kroutil W. *Trends Biotechnol.* 2010;28:324–332. https://doi.org/10.1016/j.tibtech.2010.03.003;
(c) Tufvesson P, Lima-Ramos L, Jensen JS, Al-Haque N, Neto W, Woodley JM. *Biotechnol Bioeng.* 2011;108:1479–1493. https://doi.org/10.1002/bit.23154;
(d) Mathew S, Yun H. *ACS Catal.* 2012;2:993–1001. https://doi.org/10.1021/cs300116n;
(e) Fuchs M, Farnberger JE, Eur KW. *J Org Chem.* 2015;6965–6982. https://doi.org/10.1002/ejoc.201500852;
(f) Steffen-Munsberg F, Vickers C, Kohls H, et al. *Biotechnol Adv.* 2015;33:566–604. https://doi.org/10.1016/j.biotechadv.2014.12.012;
(g) Guo F, Berglund P. *Green Chem.* 2017;19:333–360. https://doi.org/10.1039/C6GC02328B;
(h) Slabu I, Galman JL, Lloyd RC, Turner NJ. *ACS Catal.* 2017;7:8263–8284. https://doi.org/10.1021/acscatal.7b02686;
(i) Kelly SA, Pohle S, Wharry S, et al. *Chem Rev.* 2018;118:349–367. https://doi.org/10.1021/acs.chemrev.7b00437;
(j) Gomm A, O'Reilly E. *Curr Opin Chem Biol.* 2018;43:106–112. https://doi.org/10.1016/j.cbpa.2017.12.007;
(k) Patil MD, Grogan G, Bommarius A, Yun H. *Catalysts.* 2018;8:254. https://doi.org/10.3390/catal8070254;
(l) Rocha JF, Pina AF, Sousa SF, Cerqueira NMFSA. *Catal Sci Technol.* 2019;9:4864–4876. https://doi.org/10.1039/C9CY01210A.

5 Grogan G. *Curr. Opin. Chem. Biol.* 2018;43:15–22.

6 Batista VF, Galman JL, Pinto DCGA, Silva AMS, Turner NJ. *ACS Catal.* 2018;8:11889–11907.

7 Mangas-Sanchez J, Montgomery SL, Aleku GA, et al. *Curr. Opin. Chem. Biol.* 2017;37:19–25.

8 (a) Zeymer C, Hilvert D. *Ann Rev Biochem.* 2018;87:131–157. https://doi.org/10.1146/annurev-biochem-062917-012034;
(b) Angew AFH. *Chem Int Ed.* 2019;58:14420–14426. https://doi.org/10.1002/anie.201907729;
(c) Bornscheuer UT, Hauer B, Jaeger KE, Schwaneberg U. *Angew Chem Int Ed.* 2019;58:36–40. https://doi.org/10.1002/anie.201812717;
(d) Qu G, Li A, Acevedo-Rocha CG, Sun Z, Reetz MT. *Angew Chem Int Ed.* 2020;59:13204–13231. https://doi.org/10.1002/anie.201901491.

9 (a) Patel RN. *Bioorg Med Chem.* 2018;26:1252–1274. https://doi.org/10.1016/j.bmc.2017.05.023;
(b) Sun H, Zhang H, Ang EL, Zhao H. *Bioorg. Med. Chem.* 2018;26:1275–1284. https://doi.org/10.1016/j.bmc.2017.06.043;
(c) Li G, Wang J, Reetz MT. *Bioorg. Med. Chem.* 2018;26:1241–1251. https://doi.org/10.1016/j.bmc.2017.05.021;
(d) Adams JP, Brown MJB, Diaz-Rodriguez A, Lloyd RC, Roiban G. *Adv. Synth. Catal.* 2019;361:2421–2432. https://doi.org/10.1002/adsc.201900424;
(e) Albarrán-Velo J, González-Martínez D, Gotor-Fernández V. *Biocatal. Biotransformat.*. 2018;36:102–130. https://doi.org/10.1080/10242422.2017.1340457;
(f) Fryszkowska A, Devine PN. *Curr. Opin. Chem. Biol.* 2020;55:151–160. https://doi.org/10.1016/j.cbpa.2020.01.012.

10 Man H-W, Schafer P, Wong LM, et al. *J. Med. Chem.* 2009;52:1522–1524.

11 (a) Schett G, Sloan VS, Stevens RM, Schafer P. *Ther. Adv. Musculoskel. Dis.* 2010;2:271–278. https://doi.org/10.1177/1759720X10381432;
(b) Papp K, Reich K, Leonardi CL, et al. *J. Am. Acad. Dermatol.* 2015;73:37–49. https://doi.org/10.1016/j.jaad.2015.03.049.

12 Ruchelman AL, Connolly TJ. *Tetrahedron Asymmetry*. 2015;26:553–559.

13 Doubsky, J.; Klvana, R.; Richter, J.; Lehnert, P. WO2016192694, 2016.

14 Pavlidis IV, Weiß MS, Genz M, et al. *Nat. Chem.* 2016;8:1076–1082.

15 Voss M, Das D, Genz M, et al. *ACS Catal.* 2018;8:11524–11533.

16 (a) Nobili A, Steffen-Munsberg F, Kohls H, et al. *ChemCatChem.* 2015;7:757–760. https://doi.org/10.1002/cctc.201403010;
(b) Genz M, Vickers C, van denBergh T, et al. *Int. J. Mol. Sci.* 2015;16:26953–26963. https://doi.org/10.3390/ijms161126007;
(c) Genz M, Melse O, Schmidt S, et al. *ChemCatChem.* 2016;8:3199–3202. https://doi.org/10.1002/cctc.201601007.

17 Schätzle S, Höhne M, Redestad E, Robins K, Bornscheuer UT. *Anal. Chem.* 2009;81:8244–8248.

18 Midelfort KS, Kumar R, Han S, et al. *Protein Eng. Des. Sel.* 2013;26:25–33.

19 Reetz MT, Carballeira JD. *Nat. Prot.* 2007;2:891–903.

Supporting Information for

Directed evolution of an amine transaminase for the synthesis of an Apremilast intermediate via kinetic resolution

Chao Xiang[a], Shuke Wu[a,b] and Uwe T. Bornscheuer[a*]

[a] *Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, Felix Hausdorff-Str. 4, 17487 Greifswald, Germany*

[b] *State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Street, Wuhan, 430070 P. R. China*

Table of Contents

* Corresponding author. E-mail: uwe.bornscheuer@uni-greifswald.de

1. Materials

All chemicals were either purchased from commercial suppliers (Acros Organics, Sigma-Aldrich, Alfa Aesar, Combiblocks, TCI Europe, Fluorochem, Merck, Fisher, Enamine). The key chemicals and their commercial suppliers are listed below: *rac*-**1** 1-(3-Ethoxy-4-methoxyphenyl)-2-(methylsulfonyl)ethanamine (CAS 253168-94-4, TCI), (*S*)-**1** (*S*)-1-(3-Ethoxy-4-methoxyphenyl)-2-(methylsulfonyl)ethanamine (CAS 608141-42-0, BLDPharm), **2** 1-(3-Ethoxy-4-methoxyphenyl)-2-(methylsulfonyl)ethanone (CAS 1450657-28-9, Toronto Research Chemicals).

Q5® hot start high-fidelity 2X master mix and kinase, ligase & DpnI (KLD enzyme mix) was purchased from New England Biolabs. Plasmid miniprep-kit and gel extraction-kit were purchased from Macherey-Nagel. DNA oligos were purchased from Thermo Fisher.

LB (Lysogeny Broth) medium was used for routine culturing of *E. coli* for genetic engineering and seed culturing. TB (Terrific Broth) medium was used for large-scale expression of VF-ATA.

2. Protein and DNA sequences of VF-ATA

>Protein sequence of VF-ATA wildtype:

MNKPQSWEARAETYSLYGFTDMPSLHQRGTVVVTHGEGPYIVDVNGRRYLDANSGLWNMVA
GFDHKGLIDAAKAQYERFPGYHAFFGRMSDQTVMLSEKLVEVSPFDSGRVFYTNSGSEANDT
MVKMLWFLHAAEGKPQKRKILTRWNAYHGVTAVSASMTGKPYNSVFGLPLPGFVHLTCPHYW
RYGEEGETEEQFVARLARELEETIQREGADTIAGFFAEPVMGAGGVIPPAKGYFQAILPILRKYD
IPVISDEVICGFGRTGNTWGCVTYDFTPDAIISSKNLTAGFFPMGAVILGPELSKRLETAIEAIEEF
PHGFTASGHPVGCAIALKAIDVVMNEGLAENVRRLAPRFEERLKHIAERPNIGEYRGIGFMWAL
EAVKDKASKTPFDGNLSVSERIANTCTDLGLICRPLGQSVVLCPPFILTEAQMDEMFDKLEKALD
KVFAEVATDPNSSSVDKLAAALEHHHHHH

>DNA sequence of VF-ATA wildtype:

ATGAACAAACCGCAAAGCTGGGAAGCCCGGGCCGAGACCTATTCGCTCTATGGTTTCACC
GACATGCCTTCGCTGCATCAGCGCGGCACGGTCGTCGTGACCCATGGCGAGGGACCCTA
TATCGTCGATGTGAATGGCCGGCGTTATCTGGACGCCAACTCGGGCCTGTGGAACATGGT
CGCGGGCTTTGACCACAAGGGGCTGATCGACGCCGCCAAGGCCCAATACGAGCGTTTTC
CCGGTTATCACGCCTTTTTCGGCCGCATGTCCGATCAGACGGTAATGCTGTCGGAAAAGCT
GGTCGAGGTGTCGCCCTTTGATTCGGGCCGGGTGTTCTATACAAACTCGGGGTCCGAGGC
GAATGACACCATGGTCAAGATGCTATGGTTCCTGCATGCAGCCGAGGGCAAACCGCAAAA
GCGCAAGATCCTGACCCGCTGGAACGCCTATCACGGCGTGACCGCCGTTTCGGCCAGCA
TGACCGGCAAGCCCTATAATTCGGTCTTTGGCCTGCCGCTGCCGGGCTTTGTGCATCTGA

CCTGCCCGCATTACTGGCGCTATGGCGAAGAGGGCGAAACCGAAGAGCAGTTCGTCGCC
CGCCTCGCCCGCGAGCTGGAGGAAACGATCCAGCGCGAGGGCGCCGACACCATCGCCG
GTTTCTTTGCCGAACCGGTGATGGGCGCGGGCGGCGTGATTCCCCCGGCCAAGGGGTAT
TTCCAGGCGATCCTGCCAATCCTGCGCAAATATGACATCCCGGTCATCTCGGACGAGGTG
ATCTGCGGTTTCGGACGCACCGGTAACACCTGGGGCTGCGTGACCTATGACTTTACACCC
GATGCAATCATCTCGTCCAAGAATCTTACAGCGGGCTTTTTCCCCATGGGGGCGGTGATCC
TTGGCCCGGAACTTTCCAAACGGCTGGAAACCGCAATCGAGGCGATCGAGGAATTCCCCC
ATGGCTTTACCGCCTCGGGCCATCCGGTCGGCTGTGCTATTGCGCTGAAAGCAATCGACG
TGGTGATGAATGAAGGGCTGGCTGAGAACGTCCGCCGCCTTGCCCCCCGTTTCGAGGAAA
GGCTGAAACATATCGCCGAGCGCCCGAACATCGGTGAATATCGCGGCATCGGCTTCATGT
GGGCGCTGGAGGCTGTCAAGGACAAGGCAAGCAAGACGCCGTTCGACGGCAACCTGTCG
GTCAGCGAGCGTATCGCCAATACCTGCACCGATCTGGGGCTGATTTGCCGGCCGCTTGGT
CAGTCCGTCGTCCTTTGTCCGCCCTTTATCCTGACCGAGGCGCAGATGGATGAGATGTTC
GATAAACTCGAAAAAGCCCTTGATAAGGTCTTTGCCGAGGTTGCCACGGATCCGAATTCGA
GCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA

>Protein sequence of VF-ATA-8M-E (the mutation sites were highlighted):

MNKPQSWEARAETYSLYGHTDMPSLHQRGTVVVTHGEGPYIVDVNGRRYLDANSGLLNMVA
GFDHKGLIDAAKAQYERFPGYHALFGRMSDQTVMLSEKLVEVSPFDSGRVFYTNSGSEANDT
MVKMLWFLHAAEGKPQKRKILTRWNAYHGATAVSASMTGEPYNSVFGLPLPGFVHLTCPHYW
RYGEEGETEEQFVARLARELEETIQREGADTIAGFFAEPVMGAGGVIPPAKGYFQAILPILRKFD
IPVISDEVICGFGRTGNTWGCVTYDFTPDAIISSKCLTAGFFPMGAVILGPELSKRLETAIEAIEEF
PHGFTASGHPVGCAIALKAIDVVMNEGLAENVRRLAPRFEERLKHIAERPNIGEYRGIGFMWAL
EAVKDKASKTPFDGNLSVSERIANTCTDLGLICKPLGQSVVLCPPFILTEAQMDEMFDKLEKALD
KVFAEVATDPNSSSVDKLAAALEHHHHHH

>DNA sequence of VF-ATA-8M-E (the mutation sites were highlighted):

ATGAACAAACCGCAAAGCTGGGAAGCCCGGGCCGAGACCTATTCGCTCTATGGTCATACC
GACATGCCTTCGCTGCATCAGCGCGGCACGGTCGTCGTGACCCATGGCGAGGGACCCTA
TATCGTCGATGTGAATGGCCGGCGTTATCTGGACGCCAACTCGGGCCTGCTTAACATGGT
CGCGGGCTTTGACCACAAGGGGCTGATCGACGCCGCCAAGGCCCAATACGAGCGTTTTC
CCGGTTATCACGCCCTCTTCGGCCGCATGTCCGATCAGACGGTAATGCTGTCGGAAAAGC
TGGTCGAGGTGTCGCCCTTTGATTCGGGCCGGGTGTTCTATACAAACTCGGGGTCCGAGG
CGAATGACACCATGGTCAAGATGCTATGGTTCCTGCATGCAGCCGAGGGCAAACCGCAAA
AGCGCAAGATCCTGACCCGCTGGAACGCCTATCACGGCGCGACCGCCGTTTCGGCCAGC
ATGACCGGCGAGCCCTATAATTCGGTCTTTGGCCTGCCGCTGCCGGGCTTTGTGCATCTG
ACCTGCCCGCATTACTGGCGCTATGGCGAAGAGGGCGAAACCGAAGAGCAGTTCGTCGC
CCGCCTCGCCCGCGAGCTGGAGGAAACGATCCAGCGCGAGGGCGCCGACACCATCGCC
GGTTTCTTTGCCGAACCGGTGATGGGCGCGGGCGGCGTGATTCCCCCGGCCAAGGGGTA
TTTCCAGGCGATCCTGCCAATCCTGCGCAAATTTGACATCCCGGTCATCTCGGACGAGGT
GATCTGCGGTTTCGGACGCACCGGTAACACCTGGGGCTGCGTGACCTATGACTTTACACC
CGATGCAATCATCTCGTCCAAGTGTCTTACAGCGGGCTTTTTCCCCATGGGGGCGGTGAT
CCTTGGCCCGGAACTTTCCAAACGGCTGGAAACCGCAATCGAGGCGATCGAGGAATTCCC
CCATGGCTTTACCGCCTCGGGCCATCCGGTCGGCTGTGCTATTGCGCTGAAAGCAATCGA

CGTGGTGATGAATGAAGGGCTGGCTGAGAACGTCCGCCGCCTTGCCCCCCGTTTCGAGG
AAAGGCTGAAACATATCGCCGAGCGCCCGAACATCGGTGAATATCGCGGCATCGGCTTCA
TGTGGGCGCTGGAGGCTGTCAAGGACAAGGCAAGCAAGACGCCGTTCGACGGCAACCTG
TCGGTCAGCGAGCGTATCGCCAATACCTGCACCGATCTGGGGCTGATTTGCAAGCCGCTT
GGTCAGTCCGTCGTCCTTTGTCCGCCCTTTATCCTGACCGAGGCGCAGATGGATGAGATG
TTCGATAAACTCGAAAAAGCCCTTGATAAGGTCTTTGCCGAGGTTGCCACGGATCCGAATT
CGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA

3. Genetic engineering of VF-ATA plasmids

The gene of VF-ATA was amplified with Q5 DNA polymerase using primers VfTA-F (taataaggagatataccATGAACAAACCGCAAAGCTG) and VfTA-R (gtttctttaccagactcgag TCAGTGGTGG TGGTGGTG). The pRSFduet-1 vector (Novagen) was amplified with Q5 DNA polymerase using primers pRSF-R (CATGGTATATCTCCTTATTAAAGTTAAAC) and pRSF-F (CTCGAGTCTGGTAAAGAA AC). The two PCR products were assembled using SLiCE method and then transformed into chemically competent *E. coli* TOP10 cells to afford the plasmid pRSF-VF-ATA. It was used for expression of VF-ATA and as the template for directed evolution.

4. List of primers for directed evolution of VF-ATA

F19NNK Q5 fw: NNKACCGACATGCCTTCGCTGCATC

F19NNK Q5 rv: ACCATAGAGCGAATAGGTCTCGGCC

L56MNN Q5 Rv: MNNGCCCGAGTTGGCGTCCAGATAAC

L56MNN Q5 fw: TGGAACATGGTCGCGGGCTTTG

W57NNK Q5 fw: NNKAACATGGTCGCGGGCTTTGACC

W57NNK Q5 Rv: CAGGCCCGAGTTGGCGTCC

L85MNN Q5 Rv: MNNGGCGTGATAACCGGGAAAACGCTC

L85MNN Q5 fw: TTCGGCCGCATGTCCGATCAG

F86NNK Q5 fw: NNKGGCCGCATGTCCGATCAGACG

F86NNK Q5 Rv: GAAGGCGTGATAACCGGGAAAACG

R88NNK Q5 Fw: TTCGGCNNKATGTCCGATCAGACG

R88NNK Q5 Rv: GAGGGCGTGATAACCGGGAAAAC

W147NNK Q5 Fw: CCTGACCCGCNNKAACGCCTATC

W147NNK Q5 Rv: ATCTTGCGCTTTTGCGGT

V153NNK Q5 fw：CTATCACGGCNNKACCGCCGTTTCGG

V153NNK Q5 rv：GCGTTCCAGCGGGTCAGG

Y150NNK Q5 fw：CNNKCACGGCGCCACCGCCGTTTCGG

Y150NNK Q5 rv：GCGTTCCAGCGGGTCAGG

K163NNK Q5 Fw: CATGACCGGCNNKCCCTATAATTCGG

K163NNKQ5 Rv: CTGGCCGAAACGGCGGTG

Y165NNK Q5 fw: NNKAATTCGGTCTTTGGCCTGCCGC

Y165NNK Q5 rv: GGGCTTGCCGGTCATGCTGG

E233NNK Q5 Fw: CGGCGTGATTNNKCCGGCCAAGGGGTATTTCC

E233NNK Q5 Rv: CCCGCGCCCATCACCGGT

Y249NNK Q5 Fw: CGCAAANNKGACATCCCGG

Y249NNK Q5 Rv: CAGGATTGGCAGGATC

V258NNK Q5 Fw: CTCGGACGAGNNKATCTGCGGTTTCG

V258NNK Q5 Rv: ATGACCGGGATGTCATATTTGC

I259NNK Q5 Fw: CTCGGACGAGGTGNNKTGCGGTTTCG

I259NNK Q5 Rv: ATGACCGGGATGTCATATTTGC

N286NNK Q5 Fw: CTCGTCCAAGNNKCTTACAGCGG

N286NNK Q5 Rv: ATGATTGCATCGGGTGTAAAG

V297NNKQ5 Fw: CATGGGGGCGNNKATCCTTGGCC

V297NNKQ5 Rv: GGGAAAAAGCCCGCTGTAAGATTCTTG

E316NNK Q5 Fw: NNKTTCCCCCATGGCTTTACC

E316NNK Q5 Rv: CTCGATCGCCTCGATTGC

T322NNK Q5 Fw: CCATGGCTTTNNKGCCTCGGGCC

T322NNK Q5 Rv: GGGAATTCCTCGATCGCC

A383NNK Q5 Fw: GGCGCTGGAGNNKGTCAAGGACAAG

A383NNK Q5 Rv: CACATGAAGCCGATGCCG

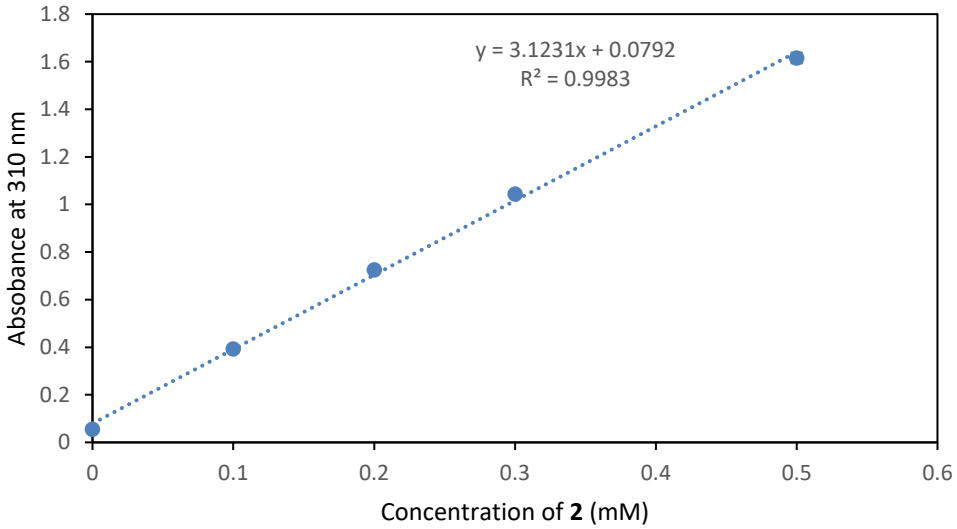
R415MNN Q5 Rv: CGGMNNGCAAATCAGCCCCAGATCG

R415MNN Q5 Fw: CTTGGTCAGTCCGTCGTCCTTTGTC

L417NNK Q5 fw: NNKGGTCAGTCCGTCGTCCTTTGTCC

L417NNK Q5 Rv: CGGCCGGCAAATCAGCCC

5. Standard curve for acetophenone assay



6. Other VF-ATA variants and their corresponding activities[a]

Table S1. Specific activity of other VF-ATA variants.

Nr.	Amino acid position and mutation	activity [mU mg-1]	
		Pyruvate	Pentanone
VF-wt	--	n.m.	n.m.
VF-2M	F85L/V153A	0.62 ± 0.07	0.49 ± 0.34
2M/W57L	F85L /W57L/V153A	4.38 ± 0.12	0.96 ± 0.09
2M/K163E	F85L/V153A/K163E	1.48 ± 0.05	0.83 ± 0.06
2M/N286A	F85L/V153A/N286A	2.11 ± 0.23	0.59 ± 0.08
2M/N286V	F85L/V153A/N286V	1.30 ± 0.04	0.84 ± 0.06
2M/R415K	F85L/V153A /R415K	2.58 ± 0.32	2.95 ± 0.04
2M/W57L/K163A	F85L /W57L/V153A/K163A	5.43 ± 0.03	2.26 ± 0.11
2M/W57L/K163E	F85L/V153A /W57L/K163E	6.01 ± 0.07	4.42 ± 0.12
2M/W57L/N286V	F85L/V153A /W57L/N286V	2.52 ± 0.16	3.94 ± 0.31
2M/W57L/N286A	F85L/V153A /W57L/N286A	5.44 ± 0.40	3.98 ± 0.10
2M/W57L/R415K	F85L/V153A /W57L/R415K	2.2±.09	1.95 ± 0.13
2M/W57L/R415K/K163A	F85L/V153A /W57L/R415K/K163A	8.16 ± 0.23	8.57 ± 0.23
4M-AA	F85L/V153A /W57L/R415K/K163A/N286C	11.88 ± 0.19	20.09 ± 0.23

[a] Acetophenone assay conditions: The specific activities of the purified VF-ATA variants for the conversion of *rac*-**1** were determined by using the acetophenone assay in 96-well microtiter plates, and measured on the Infinite® 200 PRO (TECAN) plate reader. The assay was performed with *rac*-**1** (5 mM) as amine donors and pyruvate (2.5 mM) or pentanal (2.5 mM) as amine acceptors in DMSO (1.25-2.5%), HEPES buffer (50 mM, pH 6.5) at 30°C. The formation of **2** was quantified by following the increase of absorption at 310 nm over time. One unit (U) was defined as the formation of 1 µmol **2** per minute. All measurements were performed in triplicates.

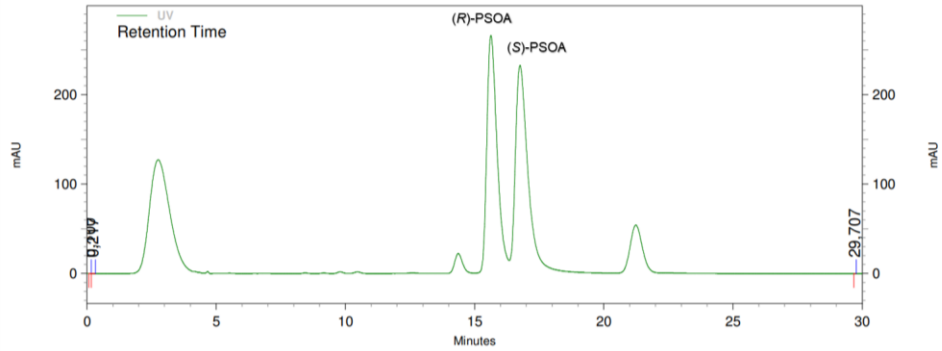
7. Analytic methods

Analysis of concentration of **1** and **2** was performed on an HPLC system with a Luna Omega 5 um Polar C18 (150 × 4.6 mm, Phenomenex): eluent H_2O/CH_3CN/TFA 50:50:0.1, flow rate at 1 mL/min, detection at 210 nm, column temperature at 25°C. Authentic standards were analyzed before the analysis of the reaction mixtures. Retention times: 1.89 min for **1**, 3.19 min for **2**.

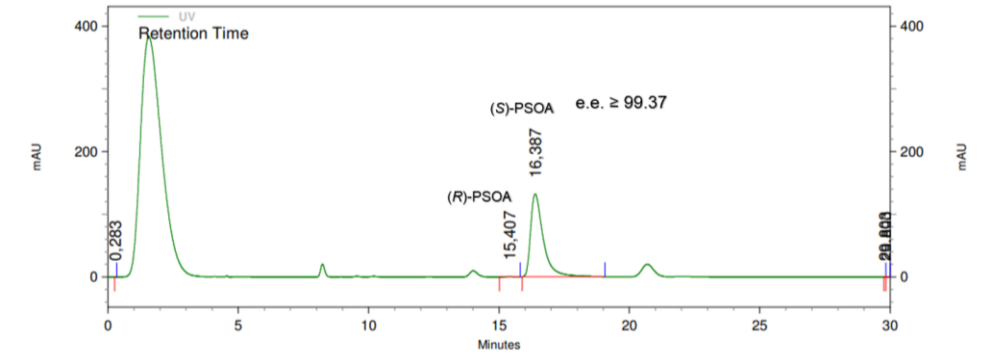
Analysis of the enantiomeric excesses of **1** was performed on an HPLC system with a Chiralpak AD-H column (5 µm, 4.6 mm × 250 mm, Daicel): eluent: hexane/ethanol/isopropanol 50:40:10, flow rate 0.8 mL/min, detection at 280 nm, column temperature at 25°C. The authentic *rac*-**1** and (*S*)-**1** were used as the reference. Retention times: 15.5 min for (*R*)-**1**, 16.6 min for (*S*)-**1**.

8. Chiral HPLC chromatograms

rac-**1** standard



(*S*)-**1** produced by kinetic resolution of *rac*-**1** by VF-ATA-8M-E



A growth selection for the directed evolution of amine-forming/converting enzymes

Shuke Wu[1,2,6]*, Chao Xiang[1,6], Yi Zhou[2], Mohammad Saiful Hasan Khan[1], Weidong Liu[1,3], Christian G. Feiler[4], Ren Wei[1], Gert Weber[4], Matthias Höhne[5] and Uwe T. Bornscheuer[1]*

[1] Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, Felix Hausdorff-Str. 4, D-17489 Greifswald, Germany

[2] State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Street, Wuhan 430070, P. R. China

[3] Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, P. R. China

[4] Macromolecular Crystallography, Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Straße 15, 12489 Berlin, Germany

[5] Protein Biochemistry, Institute of Biochemistry, University of Greifswald, Felix Hausdorff-Str. 4, D-17489 Greifswald, Germany

[6] These authors contributed equally: Shuke Wu, Chao Xiang

*e-mail: shukewu@mail.hzau.edu.cn; uwe.bornscheuer@uni-greifswald.de

Abstract

Fast screening of enzyme variants is crucial for tailoring biocatalysts for asymmetric synthesis of non-natural chiral chemicals, such as amines. However, most existing screening methods are either limited by the throughput or require specialized equipment. Herein, we report an ultra-high throughput, simple, low-equipment dependent, and generally applicable growth selection system for engineering amine-forming/converting enzymes and applied it to improve biocatalysts belonging to three different enzyme classes. This resulted in (i) an amine transaminase variant with 110-fold increased specific activity for the asymmetric synthesis of the chiral amine intermediate of Linagliptin; (ii) a 270-fold improved monoamine oxidase to prepare the chiral amine intermediate of Cinacalcet by deracemization; and (iii) an ammonia lyase variant with a 26-fold increased activity in the asymmetric synthesis of a non-natural amino acid. Our growth selection method is adaptable to different enzyme classes, varying levels of enzyme activities, and thus a flexible tool for various stages of an engineering campaign.

Chiral amines are indispensable building blocks for a large variety of bioactive pharmaceuticals and agrochemicals, and they are also widely employed as chiral auxiliaries and resolving agents. Consequently, catalytic synthesis of optically pure amines has been the key subject in chemical and pharmaceutical industries.[1-4] In this context, biocatalytic approaches are very attractive for highly selective and green synthesis of chiral chemicals.[5-10] Recent discoveries and developments of transaminases,[11-15] monoamine oxidases,[16] amine dehydrogenases,[17] imine reductases,[18] and ammonia lyases[19] provided a versatile biocatalyst toolbox to access chiral amines. Since many synthetically useful chiral intermediates are new to these natural enzymes, extensive and time-consuming directed evolution[20-23] is usually required to tailor natural enzymes for the production of specific non-natural chiral amine targets.[24-26] Therefore, high-throughput screening methods have been developed for specific classes of enzymes.[27,28] For example, specialized amine donors that elicit colored responses were exploited for assaying transaminases;[29] monoamine oxidase activities are visualized with H_2O_2-coupled assays;[30] and $NAD(P)^+$/NAD(P)H-coupled assays facilitate screening of dehydrogenases/reductases variants.[31] For challenging lyase-catalyzed reactions without easily detectable co-products, mutant design mostly had to rely on computational redesign[32, 33] or expensive high-throughput mass spectrometry.[34] However, most of these methods are either still limited by the throughput (when applied in multi-well plates) or require specialized and expensive equipment (e.g., robotic platforms[35] or microfluidic devices[30,36]), thus preventing a more widespread implementation of these enzymes. Furthermore, none of these methods is applicable for different classes of enzymes. Therefore, an ultra-high throughput, low-equipment dependent, and generally applicable method is highly desired for the directed evolution of enzymes for the synthesis of non-natural chiral amines.

In contrast to screening with the methods mentioned above, growth selection is intrinsically connected with ultra-high throughput and simplicity because only the desired highly active variants (instead of every variant) of the library are detectable and generate cell growth as an easily measured output signal via the optical density (OD) in liquid culture or colony formation/size in solid culture.[37-39] However, selection requires a direct link between the activity of the enzyme and the survival of the host, which is

often highly specialized and difficult to be established for many synthetically useful enzymes. Pioneering studies in selection-based enzyme evolution often focused on enzymes conferring antibiotic resistance (e.g., β-lactamases)[40,41] and the proteins linked to the expression of antibiotic resistance genes[42] or phage coat protein pIII (essential for phage propagation).[43] Yet, it is difficult to apply this concept to synthetically useful enzymes. Another approach is to exploit the auxotrophy generated by deleting essential genes for certain metabolites (e.g., natural amino acids)[44-46] and then to complement them by directed evolution of related enzymes, which are usually limited to those cases where natural metabolites or biochemicals are produced as exemplified for a chorismate mutase[45] or natural amino acid racemases.[46] Very recently, redox cofactor ($NAD(P)^+$/NAD(P)H) auxotrophs were developed via extensive genome engineering, and employed for selection-based evolution of several NAD(P)H-dependent enzymes.[47-50] This approach has proven useful in switching the cofactor preference of several enzymes (e.g., formate dehydrogenase),[47,48] yet with limited success in improving activity (~10-fold) for non-native substrates.[49,50] Despite the fact that growth selection is theoretically a very powerful strategy, its great potential of evolving enzymes for asymmetric synthesis/resolution is underrepresented, except for early reports on selection-based evolution of hydrolases.[51,52] To the best of our knowledge, there is no report of selection-based improvement of enzyme activities for the synthesis of non-natural chiral amines.

Herein, we report an ultra-high throughput, simple, low-equipment dependent, and generally applicable growth selection system for the directed evolution of amine-forming/converting enzymes (Fig. 1). This methodology is demonstrated by evolution of three different classes of enzymes for the synthesis of three important non-natural chiral amines in optically pure forms: 1) an amine transaminase (TA, enzyme class EC 2) for (*R*)-1-*Boc*-3-aminopiperidine, the key chiral intermediate for the antidiabetic drugs Linagliptin,[53] Trelagliptin and Alogliptin;[54] 2) a monoamine oxidase (MAO, enzyme class EC 1) for (*R*)-1-(1-naphthyl)ethylamine, the key intermediate for the calcimimetic drug Cinacalcet;[55] and 3) an ammonia lyase (enzyme class EC 4) for (*S*)-2-amino-3-(naphthalen-1-yl)propanoic acid, the key chiral synthon for a highly potent and long-acting Gα12 agonist.[56] The rapid improvements of catalytic activities (26-270 fold)

in only one or two rounds of evolution without specialized equipment showcase the great potential of growth selection for wider application in creation of suitable biocatalysts to meet the speed requirement in industry.[57]

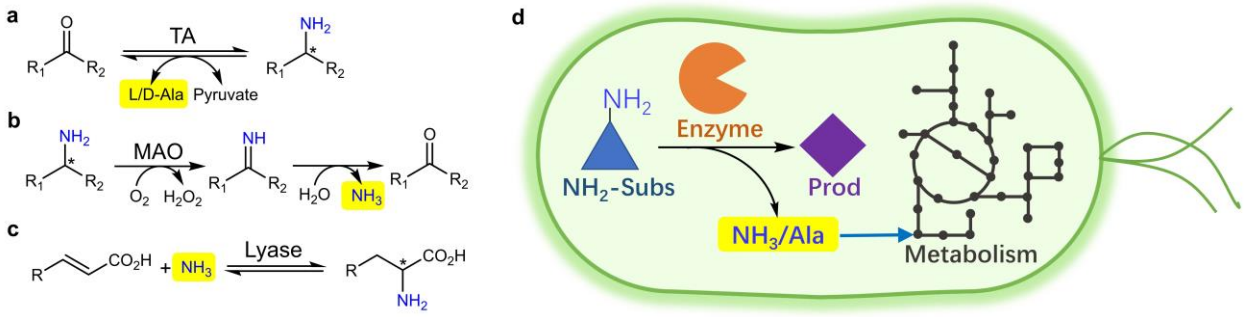


Fig 1. Concept of growth selection method for directed evolution of transaminases (TAs), monoamine oxidases (MAOs), and ammonia lyases. a, TA catalyzes reversible transamination of a targeted amine and releases easily usable L- or D-alanine (highlighted). **b**, MAO catalyzes oxidation of a targeted amine to its imine which is auto-hydrolyzed to release ammonia (highlighted). **c**, lyase catalyzes reversible conversion of an α,β-unsaturated acid and ammonia (highlighted) to a targeted amino acid. **d**, alanine or ammonia generated from a targeted amine by the active enzymes is then utilized *in situ* for cell growth serving as an output signal.

Results and Discussion

Design and general procedure of the growth selection. To design a generally applicable growth selection system for amine-forming/converting enzymes, we exploited growth based on alanine or ammonia generated from the targeted amine by amine-forming/converting enzymes: reversible transamination of the targeted amine by TA with intracellular pyruvate releases L- or D-alanine (Fig. 1a); oxidation of the targeted amine by MAO produces the imine, which is auto-hydrolyzed to release ammonia (Fig. 1b); reversible conversion of the targeted amino acid by lyase producing an α,β-unsaturated acid, and ammonia (Fig. 1c). Alanine (L- or D-) or ammonia could be easily utilized as the nitrogen source for cell growth (Fig. 1d). Hence, when the targeted amine is supplied as the only nitrogen source in a chemically-defined medium,

only the cells containing active enzyme variants (producing alanine or ammonia) could survive and grow. The precondition is that the host cannot directly utilize the targeted amines. Fortunately, this is often valid for many synthetically useful non-natural amines, because *E. coli* BL21(DE3) has a rather limited scope of nitrogen sources (Supplementary Fig. 1) and its endogenous transaminases are mainly acting on amino acids. Indeed, the use of non-natural amines as the only nitrogen source was often applied to isolate wild-type microorganisms containing amine-converting enzymes. However, to the best of our knowledge, this strategy has not been demonstrated for the directed evolution of enzymes for the synthesis of non-natural chiral amines.

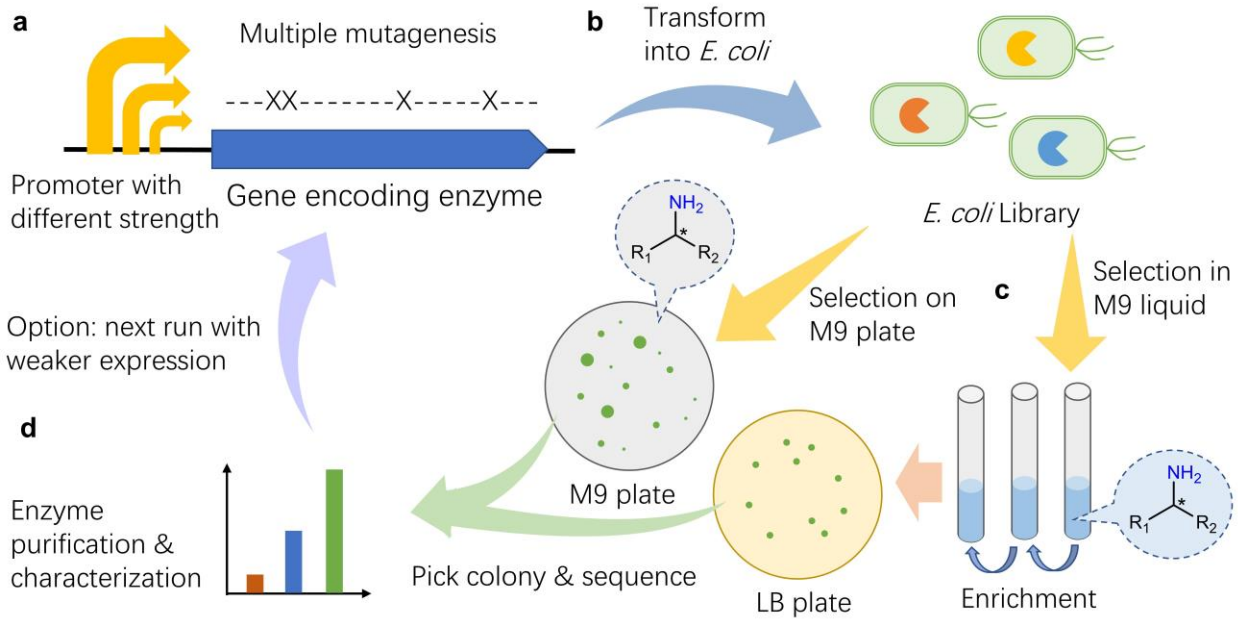


Fig 2. General procedure of the growth-based directed evolution of amine-forming/converting enzymes. a, mutagenesis of the gene encoding the enzyme of interest cloned under a promoter with suitable strength **b**, *E. coli* host cells are transformed with the gene library. **c**, selection with M9 plates or liquid medium with the targeted amine as the only nitrogen source. **d**, isolation of single colonies, sequencing of the genes, and characterization of the enzyme variants.

The general procedure of the growth-based directed evolution of amine-forming/converting enzymes is proposed as follows (Fig. 2): 1) genetic construction of a vector harboring the gene encoding the enzyme of interest with a suitable promoter, and mutagenesis by PCR (e.g., multiple site-directed mutagenesis, error-prone PCR or DNA shuffling, Fig. 2a); 2) highly efficient transformation of *E. coli* with the constructed gene library by electroporation (Fig. 2b); 3) selection of improved variants within the *E. coli* library on the M9 plate or liquid medium (enrichment followed with spreading on LB plate) with the targeted amine as the only nitrogen source (Fig. 2c); 4) isolation of the single colonies, sequencing of genes, purification, and characterization of the enzyme variants (Fig. 2d). By this, a genotype-phenotype linkage is ensured. A common issue in the growth selection is that the cell growth rate may not be proportional to the specific activity of the enzyme due to the complexity of the cell growth. To address this, we employed four small constitutive promoters with different strengths (Supplementary Fig. 2) on the vectors to modulate the expression levels of the enzyme to fine-tune the selection pressure. A strong promoter resulting in high expression will facilitate the conversion of sufficient amounts of amine to promote growth when the starting activity is low. A medium or low expression, on the contrary, ensures growth of cells bearing only the most active variants if a moderate or decent template activity already exists. Other strategies, such as inducible promoter, 5'-untranslated region,[41] or protein degradation tags[45] could also be used to tune the expression levels of the enzyme.

Growth selection-based evolution of an amine transaminase for (*R*)-1. To demonstrate growth selection for evolution of TAs, we selected the (*R*)-selective AtTA from *Aspergillus terreus* for the synthesis of (*R*)-**1**, the key chiral synthon for Linagliptin,[53] Trelagliptin, and Alogliptin (Fig. 3a).[54] Although the synthesis of (*R*)-**1** with commercial TAs was reported recently,[58] the sequence of the enzyme is unknown. We previously discovered that AtTA(wt) could be applied for synthesis of (*R*)-**1**,[59] however, the conversion was only 11%. We found the specific activity of AtTA(wt) towards (*R*)-**1** is only 0.038 U/mg, which is almost two orders of magnitude lower than for the benchmark substrate (*R*)-1-phenylethylamine (2.9 U/mg).

The efficient synthesis of (*R*)-**1** by AtTA is thus not possible until protein engineering significantly improves the specific activity. To verify the feasibility of growth selection, we cloned the AtTA(wt) gene under the control of four promoters with different strengths, and assayed the growth of the resulting transformed *E. coli* cells on M9 agar plates with D-alanine (positive control), (*R*)-1-phenylethylamine (positive control), and (*R*)-**1** as the only nitrogen source (Supplementary Fig. 3). The results showed that 1) all *E. coli* could easily use D-alanine for growth; 2) expression of AtTA(wt) at strong to weak levels enabled *E. coli* to grow on (*R*)-1-phenylethylamine, while *E. coli* was unable to grow with very weak expression of AtTA(wt) or without it (empty vector); 3) all *E. coli* were unable to grow on (*R*)-**1**. Therefore, if the activity of AtTA towards (*R*)-**1** is remarkably improved, the *E. coli* cell containing this variant should stand out by forming a colony on the plate (Fig. 3b).

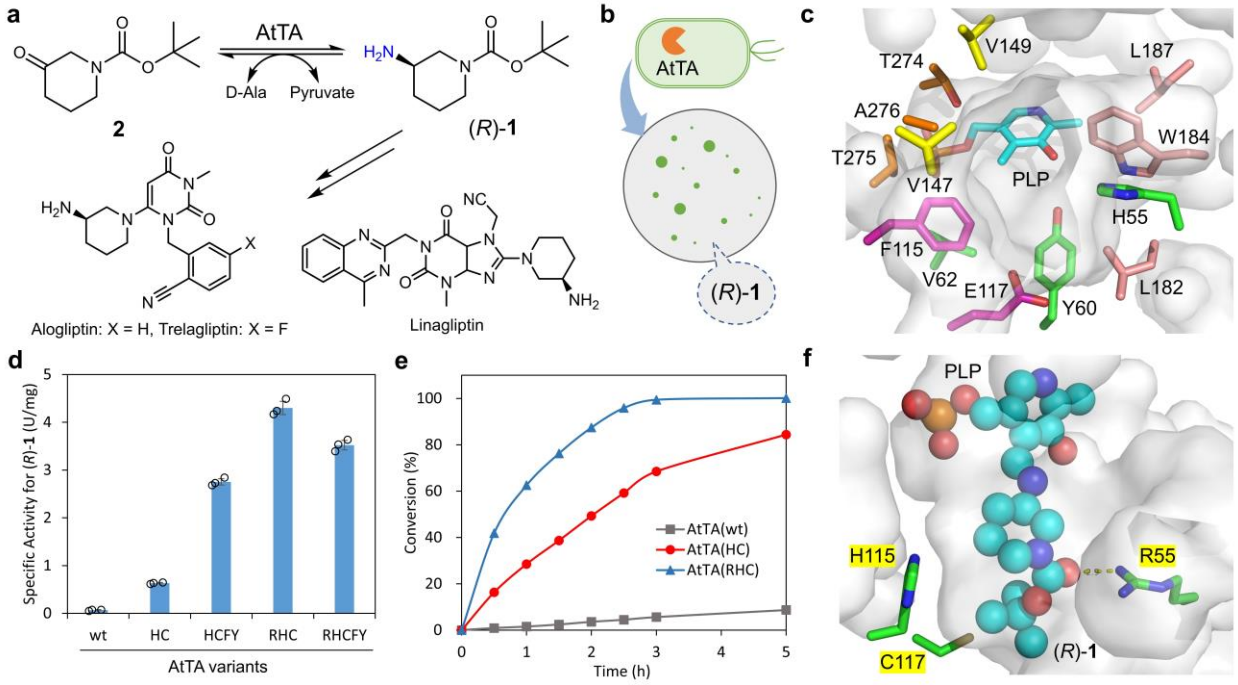


Fig 3. Growth-based directed evolution of AtTA for the asymmetric synthesis of (*R*)-1. a, targeted reaction converting **2** to (*R*)-**1** in an asymmetric synthesis. **b**, the principle of the growth-selection is based on the M9 agar plate where amine (*R*)-**1** serves as the only nitrogen source. **c**, the active site of AtTA(wt) (PDB: 4CE5) showing the target sites (grouped in different colors) for mutagenesis. **d**, specific activity of purified AtTA variants in the conversion of (*R*)-**1** with pyruvate. **e**, time course of the asymmetric synthesis

of (*R*)-**1** from **2** by purified AtTA variants. **f**, docking of the PLP-(*R*)-**1** complex (represented as cyan spheres) in the active site of the structure of AtTA(RHC) containing the mutations H55R/F115H/E117C. Data in **d** are mean values of triplicate experiments with error bars indicating the s. d. (n = 3). Data in **e** are from one independent experiment.

To maximize the chance of success in the directed evolution rounds, we first analyzed the structure of AtTA(wt) (PDB: 4CE5) and selected 13 residues in the active site region (Fig. 3c). These residues were grouped into five clusters for the ease of simultaneous mutagenesis: H55-Y60-V62, F115-E117, V147-V149, L182-W184-L187, and T274-T275-A276. Saturation mutagenesis of each group was performed with two or three NNK codons, thus leading to two libraries of 1,024 and three libraries of 32,768 unique combinations. The AtTA libraries were constructed on the vectors bearing a strong or medium promoter. These were introduced into *E. coli* BL21(DE3) cells by electroporation, and selected on M9 agar plates with (*R*)-**1** as the sole nitrogen source. Many colonies appeared on all five plates of AtTA libraries with a strong promoter (Supplementary Fig. 4), indicating that already a slight improvement of activity or expression may allow *E. coli* to grow. On the other hand, medium expressed AtTA allows a more stringent selection: only around 20 colonies were found in the F115-E117 library (Supplementary Fig. 5). The plasmids of 12 representative colonies were isolated to yield eight unique variants, which turned out to have much higher specific activity than the wild-type (0.11-0.53 vs 0.038 U/mg, Supplementary Fig. 6). The best three variants (0.40-0.53 U/mg) were chosen as templates for further mutagenesis. To increase and fine-tune the selection pressure, they were subcloned to the vectors with a weak or very weak promoter. Growth experiments confirmed (Supplementary Fig. 7) that the very weak promoter yields the best condition for the three AtTA variants for further selection and thus we started the second round of evolution by introducing saturation mutagenesis at the other four residue groups. Only the AtTA(YHC) with D5Y/F115H/E117C mutations gave some colonies for all four groups (Supplementary Fig. 8). A total of 96 colonies were sequenced and led to 15 different variants, which were evaluated to yield two significantly more active variants (additional H55R and V147F/V149Y, 1.7-1.9 U/mg, Supplementary Fig. 9). We further combined the beneficial mutants and removed the accidentally introduced D5Y mutation from the

first round of evolution. The resulting variants were evaluated (Fig. 3d) and AtTA(RHC) with H55R/F115H/E117C gave the highest activity of 4.2 U/mg (110-fold over the wildtype) while maintaining their very high enantioselectivity. The best AtTA variants were applied for asymmetric synthesis of (*R*)-**1** (Fig. 3e): AtTA(RHC) produced (*R*)-**1** (98% *ee*) in quantitative conversion in 5 h, while AtTA(wt) only gave (*R*)-**1** in 8.7% conversion under the same conditions. The preparative scale synthesis was performed with purified AtTA(RHC) to obtain (*R*)-**1** (98% *ee*) in 98% isolated yield (Supplementary Fig. 10).

To elucidate the possible molecular basis of the improved activity, we solved the crystal structure of AtTA(RHC) (Supplementary Fig. 11, Supplementary Table 1) in apo form (7XG5) and with bound PLP (7XG6) and performed a docking experiment with the reaction intermediate, the PLP-(*R*)-**1** complex (external aldimine). As depicted in Figure 3f and Supplementary Fig. 12, the PLP part binds to AtTA(RHC) in a very similar position to PLP in the AtTA(wt) and AtTA(RHC), and the (*R*)-**1** part fits into the active site of AtTA(RHC) very well. According to the docking results, particularly the F115H mutation precisely carved out additional space in the substrate-binding pocket to accommodate the piperidine ring of (*R*)-**1**. E117C expanded the pocket and may facilitate the binding of the *tert*-butyl group of (*R*)-**1**. Furthermore, the H55R variation offers a positively charged side chain which provides a hydrogen bond (2.9 Å) between the guanidinium group and the carbonyl group of (*R*)-**1**. Therefore, only three precise mutations have been sufficient to significantly (>100-fold) improve the activity of AtTA towards (*R*)-**1**, enabling practical asymmetric synthesis of this key chiral amine intermediate of several drugs.

Growth selection-based evolution of a monoamine oxidase for (*R*)-3**.** To prove the applicability of growth selection for MAOs, we selected the cyclohexylamine oxidase (CHAO) from *Brevibacterium oxydans* for deracemization of **3** via simultaneous biocatalytic oxidation and chemical reduction to produce (*R*)-**3**, the key chiral amine intermediate for the calcimimetic drug Cinacalcet (Fig. 4a).[55] Although (*R*)-**3** was prepared by biocatalysis before (mainly via TAs),[61] there are no reports dealing with a biocatalytic deracemization. The specific activities of CHAO(wt) towards (*S*)- and (*R*)-**3** were 0.009 and <0.001 U/mg,

respectively (Fig. 4d). This indicated that the CHAO(wt) has the desired selectivity towards **3**, but the specific activity towards (*S*)-**3** is too low for synthetic applications. To degrade the highly toxic H_2O_2 generated by CHAO, catalase was co-expressed with CHAO during the selection. Co-expressing of CHAO(wt) at different levels and catalase enabled *E. coli* to grow in M9 liquid culture with ammonia, cyclohexylamine (activity of 5.6 U/mg),[62] or cyclopentylamine (activity of 0.33 U/mg)[62] as the only nitrogen source (Supplementary Fig. 13) Over-expressing catalase is necessary for growth on substrates, such as cyclohexylamine, which are converted with high activity (Supplementary Fig. 14). However, (*S*)-**3** and the corresponding ketone are toxic to *E. coli* cells (Supplementary Fig. 15). At a lowered concentration (1 mM) of (*S*)-**3**, no toxicity effects are observed, but colony formation on agar plate was difficult to evaluate. Therefore, we performed enrichment of *E. coli* cells in liquid cultures and added methyl laurate as a growth-compatible second phase which acts as a reservoir of (*S*)-**3** and also extracts the ketone *in situ*. At the reduced concentration of (*S*)-**3** in the medium, *E. coli* cells bearing active variants of CHAO could be enriched and isolated (Fig. 4b).

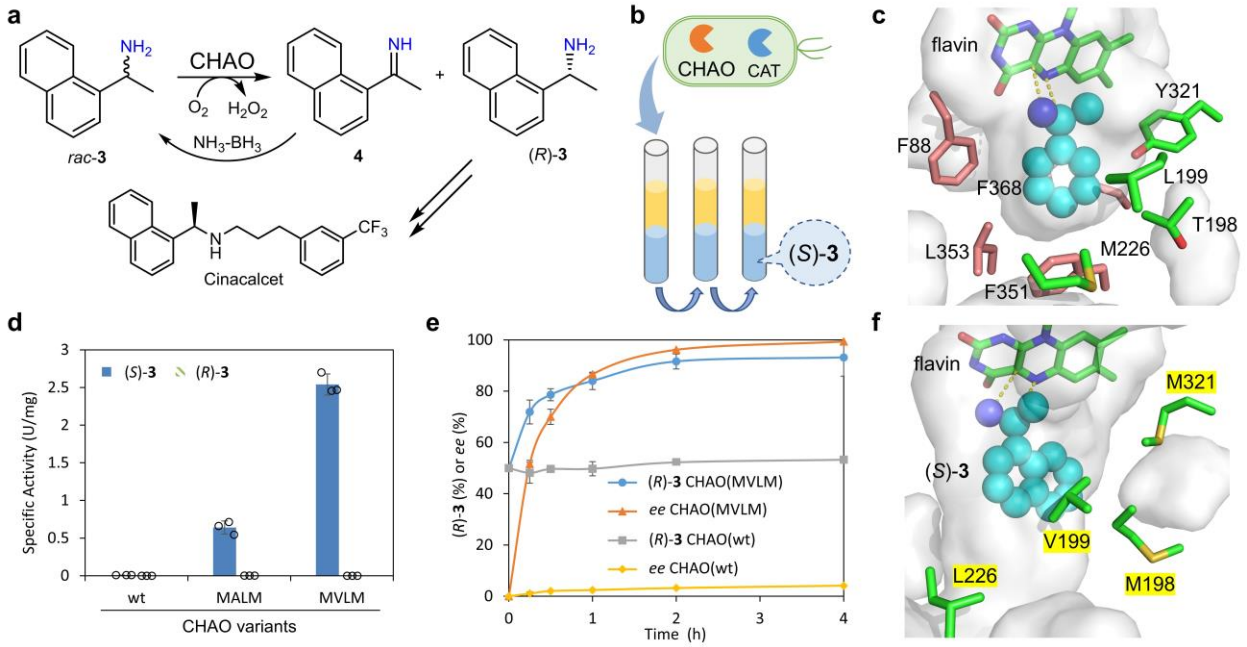


Fig 4. Growth-based directed evolution of CHAO for deracemization to produce (*R*)-3. a, targeted reaction for deracemization of *rac*-**3** to (*R*)-**3**. **b**, the principle of the growth-selection is based on *E. coli*

cells co-expressing CHAO variants and catalase (CAT) in the two-phase system (methyl laurate and M9 liquid medium) where amine (*S*)-**3** (1 mM in the aqueous phase) serves as the only nitrogen source. This ensures a non-toxic concentration but sufficient supply of (*S*)-**3** to the cells and simultaneous removal of the toxic ketone product. **c**, identification of two groups of hot spots by docking of (*S*)-1-phenylethylamine (truncated substrate, represented as cyan spheres) into the active site of CHAO(wt) (PDB: 4I59). **d**, specific activities of purified CHAO variants for converting (*S*)- and (*R*)-**3**. **e**, time course of the deracemization of *rac*-**3** by *E. coli* co-expressing CHAO variants and catalase and NH_3-BH_3. **f**, docking of (*S*)-**3** (represented as cyan spheres) in the active site of the structure model of CHAO(MVLM) with T198M/L199V/M226L/Y321M. Data in **d** and **e** are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).

To accelerate the developing timelines,[57] we envisioned that the growth selection could allow dramatic improvements of activity via a single round of evolution. CHAO(wt) showed a high activity of 3.5 U/mg towards (*S*)-1-phenylethylamine,[62] a similar but less bulky aromatic amine. Thus, it was docked into the active site of CHAO(wt) (PDB: 4I59)[63] to identify possible key residues for mutagenesis (Fig. 4c). We speculated that (*S*)-**3** may bind in a similar pose but the additional ring of (*S*)-**3** may occupy either side (F88-F351-L353-F368 or T198-L199-M226-Y321 labeled in salmon or green, Fig. 4c). These two groups of hot spots were subjected to simultaneous mutagenesis using the DBS codon, which is standing for 18 different codons and encoding 12 hydrophobic and small amino acids: A, R, C, G, I, L, M, F, S, T, W and V. Simultaneously, a four-site mutagenesis of CHAO(wt) was performed with the GoldenGate method,[64] leading to two libraries of 104,976 unique combinations. These libraries were constructed on the plasmids with medium to very weak promoters, transformed into *E. coli* BL21(DE3) cells (expressing catalase) by electroporation, and then underwent enrichment in M9 medium with (*S*)-**3** in the presence of methyl laurate. The culture was diluted every 24 hours 5 times, and significant growth was only observed for the medium expressed library of T198-L199-M226-Y321 (Supplementary Fig. 16). The resulting enriched culture was spread on an LB plate, and 12 colonies were isolated to yield two unique variants, CHAO(MALM) with T198M/L199A/M226L/Y321M and CHAO(MVLM) with T198M/L199V/M226L/Y321M mutations. The specific activities of the purified CHAO(MALM) and CHAO(MVLM) towards (*S*)-**3** were 0.64 and 2.54 U/mg, respectively (Fig. 4d). This demonstrated a 70–270-fold increase in specific activity, while the high

enantioselectivity was preserved (⩽0.001 U/mg towards (*R*)-**3** for both variants). The enzyme kinetics of CHAO(MVLM) towards (*S*)-**3** were k_{cat} = 240 min^{-1}, K_m = 2.2 mM, K_i = 8.4 mM (Supplementary Fig. 17). In comparison with CHAO(wt) (k_{cat} = 0.91 min^{-1}, K_m = 4.5 mM, K_i = 25 mM), the quadruple mutant mainly improved the k_{cat}, reduced the K_m but also has some substrate inhibition. *E. coli* cells co-expressing CHAO(MVLM) and catalase were employed for deracemization of *rac*-**3**, leading to enantiopure (*R*)-**3** (>99% *ee*) in 93% conversion, while the same reaction with *E. coli* cells expressing CHAO(wt) gave almost racemic **3** (4% *ee*) (Fig. 4e, Supplementary Fig. 18). The preparative scale synthesis was also performed to obtain (*R*)-**3** (>99% *ee*) in 69% isolated yield (Supplementary Fig. 19).

Next, (*S*)-**3** was docked to a homology model of CHAO(MVLM) to elucidate the possible rationale for the observed improved activity (Fig. 4f). According to the docking results, (*S*)-**3** binds to the active site CHAO(MVLM) in a similar pose but at a slightly different angle compared to (*S*)-1-phenylethylamine in the CHAO(wt). The Y321M mutation slightly increased the hydrophobicity and size of the pocket. T198M also slightly expanded the space by pointing the methionine side chain away from the active site. L199V may provide a suitable hydrophobic interaction with the naphthalene ring. Besides the change in the active site, the M226L mutation flips the side chain away from F351 and opens the active site to the substrate channel. With proper structure-guided selection of mutation residues, a single round of evolution with growth selection was sufficient to significantly (>200-fold) improve the activity of CHAO towards (*S*)-**3**.

Growth selection-based evolution of an ammonia lyase for (*S*)-5**.** To further demonstrate the applicability of growth selection, we have chosen the phenylalanine ammonia lyase from *Petroselinum crispum* (PcPAL)[63] for synthesis of non-natural amino acid, (*S*)-**5**, the key chiral intermediate for a highly potent Gα12 agonist (Fig. 5a).[56] The specific activities of PcPAL(wt) towards L-phenylalanine and the target (*S*)-**5** were 0.147 and 0.0027 U/mg, respectively (Fig. 5d), suggesting a large room for activity improvement. However, the solubility of (*S*)-**5** in the culture medium is very low (i.e., < 1.5 mM). Furthermore, the growth selection using amino acids is complicated by indigenous enzymes in the host (*e.g.*, amino acid

transaminases). Thus, we tailored the *E. coli* host by expressing the aromatic amino acid transporter (AroP, to increase the availability of (*S*)-**5**) and deleting the gene encoding the aromatic amino acid transaminase (*tyrB*, to minimize the interference of indigenous enzymes). We tested *E. coli* host (with or without deleting *tyrB*) for the expression of PcPAL(wt) at different levels (with or without co-expressing AroP) on M9 agar plates supplemented with ammonia, L-Phe, or (*S*)-**5** as the only nitrogen sources (Supplementary Fig. 20). As expected, deleting *tyrB* significantly reduced background growth on L-Phe, while expressing AroP contributed to the fast utilization of L-Phe. With these two modifications, the growth depends on the expression level of PcPAL(wt). Although no obvious cell growth on (*S*)-**5** was observed, deleting *tyrB* and co-expressing AroP may help to identify PcPAL variants with improved activity on (*S*)-**5**. Thus, the selection was performed using this engineered *E. coli* host in liquid cultures with 1 mM (*S*)-**5** (Fig. 5b).

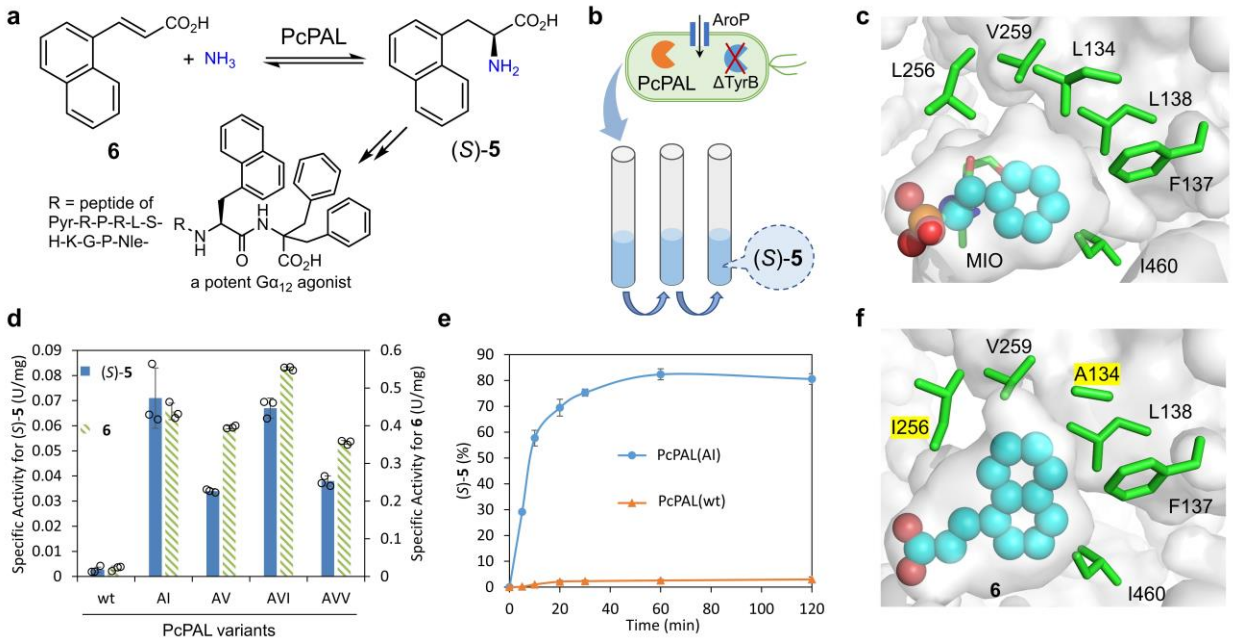


Fig 5. Growth-based directed evolution of PcPAL for asymmetric synthesis of (*S*)-5. a, targeted reaction to convert **6** to (*S*)-**5**. **b**, the principle of the growth-selection is based on *E. coli* Δ*tyrB* co-expressing PcPAL variants and a transporter (AroP) in the M9 liquid medium where the amino acid (*S*)-**5** serves as the only nitrogen source. **c**, identification of six key residues by analyzing the active site of PcPAL(wt) (PDB: 6HQF) with (*R*)-(1-amino-2-phenylethyl)phosphonic acid (inhibitor, represented as cyan spheres). **d**, specific activities of purified PcPAL variants for converting (*S*)-**5** and **6**. **e**, time course of the asymmetric

synthesis of (*S*)-**5** from **6** by *E. coli* expressing PcPAL(AI) or PcPAL(wt). **f**, docking of (*S*)-**5** (represented as cyan spheres) in the active site of the structure model of PcPAL(AI) with L134A/L256I. Data in **d** and **e** are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).

Taking advantage of previous studies of PcPAL,[19] we aimed to improve activity via a single round of evolution. Guided by the structure of PcPAL(wt) with the inhibitor (*R*)-(1-amino-2-phenylethyl)phosphonic acid (PDB: 6HQF),[65] six hydrophobic residues lining the phenyl ring (L134, F137, L138, L256, V259, I460) were selected for simultaneous multi-site mutagenesis (Fig. 5c). Instead of using NNK or DBS codons, we only focused on small hydrophobic side chains: the DYA codon (encodes L, I, V, T, S, and A) was used for L134, L138, and L256; DYC codon (encodes F, I, V, T, S, and A) was used for F137; RYC codon (encodes I, V, T, and A) was used for V259 and I460. Six-site directed mutageneses of PcPAL(wt) were achieved with sequentially performed Q5 mutagenesis rounds, leading to a library of 20,736 unique combinations. The libraries were constructed on the plasmids with strong to weak promoters, then introduced into *E. coli* BL21(DE3)Δ*tyrB* cells (expressing the transporter AroP) and subjected to enrichment in M9 medium supplemented with (*S*)-**5**. The culture was diluted every 24 hours 5 times, and significant growth was only observed for the library with strong expression level (Supplementary Fig. 21). The resulting enriched culture was spread on a LB plate and 24 colonies were isolated to yield four unique variants: PcPAL(AI) with L134A/L256I, PcPAL(AV) with L134A/F137V, PcPAL(AVI) with L134A/F137V/L256I, and PcPAL(AVV) with L134A/F137V/I460V. These variants showed specific activities of 0.034-0.071 U/mg for converting (*S*)-**5** and specific activities of 0.36-0.55 U/mg for converting **6** (Fig. 5d). These activities are 13-26 times higher than that of PcPAL(wt). *E. coli* whole cells expressing PcPAL variants were evaluated for the asymmetric synthesis of (*S*)-**5** (Supplementary Fig. 22). The best PcPAL(AI) afforded enantiopure (*S*)-**5** in 82% conversion in 1 h, whereas the wild-type PcPAL only gave (*S*)-**5** in 3% conversion under the same conditions (Fig. 5e, Supplementary Fig. 23).

To provide molecular insights for the improved activity, compound **6** was docked to the homology model of PcPAL(AI) (Fig. 5f). Clearly, **6** binds to the active site of PcPAL(AI) in the same pose as the inhibitor in the PcPAL(wt) (Fig. 5c) as shown from our docking models. The L134A mutation (most important, reflected in all variants) clearly caved out additional space in the pocket to accommodate the naphthyl group. The L256I mutation (found in the two most active variants) slightly and precisely enlarged the pocket to accommodate **6**. Another important mutation, F137V, also provides more space but in a direction that may not directly contact **6**. With these modifications, the specific activity of PcPAL towards **6** increased 26-fold to 0.55 U/mg, enabling the biocatalytic synthesis of the key chiral non-natural amino acid, though further engineering of PcPAL may be required for practical applications.

Discussion

Fast identification of active enzyme mutants is crucial for directed evolution of biocatalysts for the efficient synthesis of chemicals. Instead of widely adopted high-throughput screening methods, herein, we developed a facile growth selection system – fine-tuned by a set of suitable promoters – generally applicable for different classes of amine-forming/converting enzymes. The principle of the methodology was based on the targeted chiral amine as the sole nitrogen source for cell growth in a chemically-defined medium. Our selection method is fast and efficient: for carrying out the protocol on agar plates, one week is sufficient to proceed from a cloned library to identified and sequenced hits. Two weeks are necessary if the selection is done via growth enrichment in liquid medium. A clear advantage of this approach is that an organic phase can be used as a substrate/product reservoir if a hydrophobic substrate (or product) has to be used that shows a low solubility or toxicity at a higher concentration. The largest library in this study included 10^5 clones for the selection of a MAO variant, but the upper limit is only given by the transformation efficiency and might easily be increased to 10^6-10^7 variants. This makes this strategy especially appealing for simultaneous site-saturation of multiple positions to harness potentially synergistic effects and thus to identify variants

with a significant improvement in one round of evolution. A potentially higher activity gain on the other hand also facilitates a clear growth advantage and the reduction of false-positive hits.

In comparison with previously developed screening methods using medium-throughput analytic instruments or high-throughput microfluidic devices, the growth selection methods presented here can interrogate very large enzyme libraries by circumventing the analysis of every variant and the method works also without specialized instruments/devices. This is particularly useful for the scenario with limited access to these instruments/devices, such as small companies or labs in developing countries. Another advantage of the growth selection method is its general applicability: the growth selection procedures could be easily transferable to evolve other classes of amine-converting enzymes. The use of fine-tuned promoters and expression systems enables this growth selection applies to other enzyme classes and growth-facilitating metabolites produced. This general applicability also reduces the barriers for researchers to adopt this method.

In conclusion, the growth selection was demonstrated for significant improvement of catalytic activity (26-270 fold) of a TA, a MAO, and a lyase in only one or two rounds of evolution without expensive and specialized equipment. We believe that the concept and method described here could be generally applicable to develop other enzymes (amine dehydrogenases, imine reductases, etc.) for the synthesis of non-natural chiral amines, amino acids, amides, and other nitrogen-containing molecules. Considering its outstanding features of high throughput, simple and low-equipment dependent, we envision that growth selection can be widely adopted in academia and industry to develop suitable biocatalysts for various applications.

Methods

Growth selection of AtTA libraries for converting (*R*)-1**.** The AtTA libraries with strong and medium constitutive promoters were transformed into electrocompetent cells of *E. coli* BL21 (DE3) by using an

electroporation cuvette (1 mm gap) and default *E. coli* settings (1.8 kV) for the MicroPulser Electroporator (Biorad). Immediately after the electroporation shock, prewarmed LB medium (2 ml, 37 °C) was added to the cuvette and the cells were cultured in an incubator for 2 h at 30 °C, 200 rpm. Then, a small portion of the cells (20 μl) was isolated to test the electroporation efficiency. Half of the cells were added into LB medium (5 ml) containing kanamycin (50 mg l^{-1}) for isolation of the plasmids from the library. The other half of the cells were centrifuged (5 min, 4000 g), and the supernatant of the LB medium was removed. The cell pellets were resuspended in M9 medium (300 μl) without nitrogen source, and then spread on the agar plates of M9 medium containing kanamycin (25 mg l^{-1}) and (*R*)-**1** (10 mM) as the sole nitrogen source. The plates were kept in an incubator (30 °C) for 5 days. Photos of the plates were taken every 24 h. Colonies were usually observed in the 2nd or 3rd day, if there are positive hits. Some big and representative colonies were picked and transferred into LB liquid medium (5 ml) containing kanamycin (50 mg l^{-1}) for preparation of cell stocks and plasmids for sequencing. The growth selection for the second round of evolution of AtTA was performed following the same procedures above.

Asymmetric synthesis of (*R*)-1 by purified AtTA variants. The reaction was run in a vessel fitted with a magnetic stirrer, temperature probe, pH probe, and base addition. To a 100-ml flask, the following components were added to form a 40-ml system: CHES buffer (100 mM, pH 8.5) including PLP (1 mM), D-alanine (500 mM), oxidized nicotinamide adenine dinucleotide (NAD^+, 2–3 mM), glucose dehydrogenase (3 U ml^{-1}), D-glucose (550 mM) and lactate dehydrogenase (6 U ml^{-1}); 8 ml stock solution of **2** (250 mM in DMSO); purified protein (1-2 mg ml^{-1}). The addition of the respective ketone as a DMSO solution started the reaction and the reactor was then stirred at 30 °C and with the pH maintained between 8.4-8.6 for 12 h by automatic addition of NaOH solution (2 M). To monitor the progression of the reaction, samples (50 μl) were taken at different times of the reaction (0.5, 1, 1.5, 2, 2.5, 3, 5, and 12 h), and mixed with acetonitrile (500 μl) and hydrochloric acid solution (100 mM, 450 μl) for HPLC analysis of the conversion of **2**. Afterwards, the reaction was quenched with HCl to pH 2.0. The products were extracted with ethyl acetate (EtOAc) at pH 10 (NaOH addition), delivering (after drying and evaporation) the crude amines as oils.

These were characterized (HPLC and achiral GC). The *ee* values of (*R*)-**1** were analyzed by chiral GC analysis. After extraction of (*R*)-**1** with EtOAc, the compound was dissolved in CH_2Cl_2, then derivatization to the trifluoroacetamide was performed by adding a 20-fold excess of trifluoroacetic anhydride. After purging with nitrogen to remove excess anhydride and residual trifluoroacetic acid, the derivatized compound was dissolved in CH_2Cl_2 and analyzed.

Growth selection of CHAO libraries for converting (*S*)-3**.** The CHAO libraries with medium, weak, and very weak constitutive promoters were transformed into electrocompetent cells of *E. coli* BL21 (DE3) containing the plasmid SCm-KatE (constitutively expressing the catalase from *E. coli*) by the same electroporation procedures above. A small portion of the cells (20 μl) were used to test the electroporation efficiency and half of the cells were used to prepare the plasmids of the library. The other half of the cells were centrifuged (5 min, 4000 g), and the supernatant of the LB medium was removed. The cell pellets were resuspended in M9 liquid medium (2 ml) containing appropriate antibiotics (25 mg l^{-1} kanamycin and 12.5 mg l^{-1} ampicillin) and (*S*)-**3** (1 mM) as the sole nitrogen source. Methyl laurate (0.5 ml) was added into the aqueous M9 medium as a growth-compatible second phase. The two-phase cultures were kept in an incubator (30 °C, 200 rpm), and diluted (10x) in fresh M9 medium every 24 h 5 times (5 days in total). The dilution was performed by adding existing aqueous culture (200 μl) into fresh M9 medium (1.8 ml) containing appropriate antibiotics (25 mg l^{-1} kanamycin and 12.5 mg l^{-1} ampicillin) and (*S*)-**3** (1 mM) and fresh methyl laurate (0.5 ml). Photos of the cultures were taken and the optical densities were measured every 24 h. An aliquot (50 μl) of the enriched culture of the CHAO library with medium expressing levels was spread on an LB agar plate containing kanamycin (50 mg l^{-1}) for isolation of single colonies (37 °C, for 24 h). The representative colonies were picked and transferred into LB liquid medium (5 ml) containing kanamycin (50 mg l^{-1}) for preparation of cell stocks and plasmids for sequencing.

Deracemization of *rac*-3 by *E. coli* co-expressing CHAO variants and catalase and NH_3-BH_3. Fresh *E. coli* cells containing pRSF-CHAO(MVLM) or pRSF-CHAO(wt) and SCm-KatE were employed as whole-cell catalysts for the reaction. The cell pellets of a 50-ml culture were first washed with Tris-HCl buffer

(100 mM, pH 8.0) and subjected to centrifugation again (4000 g, 15 min). The supernatant was discarded and the cell pellets were resuspended in new Tris-HCl buffer (100 mM, pH 8.0) and the optical density of the cell suspensions was measured. To a reaction vial (20 ml) with screwcap, a stock solution of *rac*-**3** (500 mM HCl salt in water), a stock solution of NH_3-BH_3 (1 M in Tris-HCl buffer), Tris-HCl buffer (100 mM, pH 8.0), and *n*-dodecane (1 ml) were added to form a two-phase catalytic system composed of the aqueous buffer (1 ml) containing *rac*-**3** (25 mM), NH_3-BH_3 (500 mM) and the organic phase (1 ml). The cell suspensions were added to start the reaction (final density of cells in aqueous phase was 10 g l^{-1} dcw, dry cell weight). 36 reaction vials were incubated at 30 °C, 200 rpm for 4 h (the caps were opened for venting at 1 and 2 h for 1 min). To monitor the progression of the reaction, NaOH solution (5 M, 100 μl) was added to the reaction vials at different times (0, 0.25, 0.5, 1, 2, and 4 h) to stop the reaction. Then, *n*-hexane (1 ml) was added to the vial for extraction. The organic phase (*n*-hexane and *n*-dodecane) was centrifuged (12000 g, 3 min) and an aliquot of the supernatant (100 μl) was mixed with *n*-hexane (900 μl, containing 2 mM ethylbenzene as internal standard) for simultaneous chiral HPLC analysis of the conversion and *ee*.

Growth selection of PcPAL libraries for converting (*S*)-5. The PcPAL libraries with strong, medium, and weak constitutive promoters were transformed into electrocompetent cells of *E. coli* BL21 (DE3) Δ*tyrB* containing the plasmid SCm-AroP (constitutively expressing the aromatic amino acid transporter from *E. coli*) by the same electroporation procedures above. A small portion of the cells (20 μl) was used to test the electroporation efficiency and half of the cells were used to prepare the plasmids of the library. The other half of the cells were centrifuged (5 min, 4000 g), and the supernatant of the LB medium was removed. The cell pellets were resuspended in M9 liquid medium (2 ml) containing appropriate antibiotics (25 mg l^{-1} kanamycin and 12.5 mg l^{-1} ampicillin) and (*S*)-**5** (1 mM) as the sole nitrogen source. The cultures were kept in an incubator (30 °C, 200 rpm), and diluted (10x) in fresh M9 medium each 24 h for 5 times (5 days in total). The dilution was performed by adding the existing aqueous culture (200 μl) into fresh M9 medium (1.8 ml) containing appropriate antibiotics (25 mg l^{-1} kanamycin and 12.5 mg l^{-1} ampicillin) and (*S*)-**5** (1 mM). The optical densities of the cultures were measured for the last three days, and a photo was taken at

the end of enrichment. The enriched culture of PcPAL library with strong expressing levels was spread on an LB agar plate containing kanamycin (50 mg l^{-1}) for isolation of single colonies (37 ^{o}C, for 24 h). The representative colonies were picked and transferred into LB liquid medium (5 ml) containing kanamycin (50 mg l^{-1}) for preparation of cell stocks and plasmids for sequencing.

Asymmetric synthesis of (*S*)-5 by *E. coli* expressing PcPAL variants. Fresh *E. coli* cells containing pRSF-PcPAL(AI) or pRSF-PcPAL(wt) were employed as whole-cell catalysts for the reaction. The cell pellets from a 50-ml culture were resuspended in NH_3/NH_4Cl buffer (6 M, pH 10.0) and the optical density of the cell suspensions was measured. To a reaction vial (20 ml) with a screwcap, a stock solution of **6** (1 M in DMSO) and NH_3/NH_4Cl buffer (6 M, pH 10.0) were added to form a catalytic system (1 ml) containing **6** (22 mM). The cell suspensions were added to start the reaction (final density of cells was 5 g dcw l^{-1}). 42 reaction vials were incubated at 30 ^{o}C, 200 rpm for 2 h. To monitor the progression of the reaction, HCl solution (6 M, 1 ml) was added to the reaction vials at different times (0, 5, 10, 20, 30, 60, and 120 min) to stop the reaction. To determine the conversion, methanol (2 ml, containing 2 mM acetophenone as the internal standard) was added to the vial. An aliquot of the mixture (1 ml) was centrifuged (12000 g, 10 min), and an aliquot of the supernatant (700 μl) was used for HPLC analysis of the conversion. To determine the *ee* of **5**, methanol (300 μl) was added to the vial. The mixture was centrifuged (12000 g, 10 min), and an aliquot of the supernatant (700 μl) was used for chiral HPLC analysis.

Data availability statement

Additional data supporting the findings of this study are available as Supplementary Information. All other are available from the corresponding authors upon reasonable request.

Code availability statement

No custom code or mathematical algorithm was used in this study.

References

1. Nugent, T. C. & El-Shazly, M. Chiral amine synthesis-recent developments and trends for enamide reduction, reductive amination, and imine reduction. *Adv. Synth. Catal.* **352**, 753–819 (2010).
2. Höhne, M. & Bornscheuer, U. T. Biocatalytic routes to optically active amines. *ChemCatChem* **1**, 42–51 (2009).
3. Kroutil, W. et al. Asymmetric preparation of prim-, sec-, and tert-amines employing selected biocatalysts. *Org. Process Res. Dev.* **17**, 751–759 (2013).
4. Ghislieri, D. & Turner, N. J. Biocatalytic approaches to the synthesis of enantiomerically pure chiral amines. *Top. Catal.* **57**, 284 (2014).
5. Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).
6. Sheldon, R. A. & Woodley, J. M. Role of biocatalysis in sustainable chemistry. *Chem. Rev.* **118**, 801–838 (2018).
7. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: enzymatic synthesis for industrial applications. *Angew. Chem. Int. Ed.* **60**, 88–119 (2021).
8. Bell, E. L. et al. Biocatalysis. *Nat. Rev. Methods Primers* **1**, 46 (2021).
9. Yi, D. et al. Recent trends in biocatalysis. *Chem. Soc. Rev.* **50**, 8003–8049 (2021).
10. Simić, S. et al. Shortening synthetic routes to small molecule active pharmaceutical ingredients employing biocatalytic methods. *Chem. Rev.* **122**, 1052−1126 (2022).
11. Koszelewski, D., Tauber, K., Faber, K. & Kroutil W. ω-Transaminases for the synthesis of non-racemic α-chiral primary amines. *Trends Biotechnol.* **28**, 324–332 (2010).
12. Steffen-Munsberg, F. et al. Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. *Biotechnol Adv.* **33**, 566–604 (2015).

13. Slabu, I., Galman, J. L., Lloyd, R. C. & Turner, N. J. Discovery, engineering, and synthetic application of transaminase biocatalysts. *ACS Catal.*, **7**, 8263–8284 (2017).
14. Gomm, A. & O'Reilly, E. Transaminases for chiral amine synthesis. *Curr. Opin. Chem. Biol.* **43**, 106–112 (2018).
15. Patil, M. D., Grogan, G., Bommarius, A. & Yun, H. Recent advances in ω-transaminase-mediated biocatalysis for the enantioselective synthesis of chiral amines. *Catalysts* **8**, 254 (2018).
16. Batista, V. F., Galman, J. L., Pinto, D. C. G. A., Silva, A. M. S. & Turner, N. J. Monoamine oxidase: tunable activity for amine resolution and functionalization. *ACS Catal.* **8**, 11889–11907 (2018).
17. Ducrot, L., Bennett, M., Grogan, G. & Vergne-Vaxelaire, C. NAD(P)H-dependent enzymes for reductive amination: active site description and carbonyl-containing compound spectrum. *Adv. Synth. Catal.* **363**, 328–351 (2021).
18. Mangas-Sanchez, J. et al. Imine reductases (IREDs). *Curr. Opin. Chem. Biol.* **37**, 19–25 (2017).
19. Parmeggiani, F., Weise, N. J., Ahmed, S. T. & Turner, N. J. Synthetic and therapeutic applications of ammonia-lyases and aminomutases. *Chem. Rev.* **118**, 73–118 (2018).
20. Arnold, F. H. Innovation by evolution: bringing new chemistry to life (Nobel lecture). *Angew. Chem. Int. Ed.* **58**, 14420–14426 (2019).
21. Bornscheuer, U. T., Hauer, B., Jaeger, K. E. & Schwaneberg, U. Directed evolution empowered redesign of natural proteins for the sustainable production of chemicals and pharmaceuticals. *Angew. Chem. Int. Ed.* **58**, 36–40 (2019).
22. Qu, G., Li, A., Acevedo-Rocha, C. G., Sun, Z. & Reetz M. T. *Angew. Chem. Int. Ed.* **59**, 13204–13231 (2020).
23. Wang, Y. et al. Directed evolution: methodologies and applications. *Chem. Rev.* **121**, 12384−12444 (2021).
24. Savile, C. K. et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).

25. Schober, M. et al. Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2**, 909–915 (2019).
26. Kumar, R. et al. Biocatalytic reductive amination from discovery to commercial manufacturing applied to abrocitinib JAK1 inhibitor. *Nat. Catal.* **4**, 775–782 (2021).
27. Markel, U. et al. Advances in ultrahigh-throughput screening for directed enzyme evolution. *Chem. Soc. Rev.* **49**, 233–262 (2020).
28. Sheludko, Y. V. & Fessner, W.-D. Winning the numbers game in enzyme evolution-fast screening methods for improved biotechnology proteins. *Curr. Opin. Struct. Biol.* 63, 123–133 (2020).
29. Green, A. P., Turner, N. J. & O'Reilly, E. Chiral amine synthesis using ω-transaminases: an amine donor that displaces equilibria and enables high-throughput screening. *Angew. Chem. Int. Ed.* **53**, 10714−10717 (2014).
30. Debon, A. et al. Ultrahigh-throughput screening enables efficient single-round oxidase remodelling. *Nat. Catal.* **2**, 740–747 (2019).
31. Marshall, J. R. et al. Screening and characterization of a diverse panel of metagenomic imine reductases for biocatalytic reductive amination. *Nat. Chem.* **13**, 140–140 (2021).
32. Li, R. et al. Computational redesign of enzymes for regio- and enantioselective hydroamination. *Nat. Chem. Biol.* **14**, 664–670 (2018).
33. Cui, Y. et al. Development of a versatile and efficient C–N lyase platform for asymmetric hydroamination via computational enzyme redesign. *Nat. Catal.* **4**, 364–367 (2021).
34. Kempa, E. E. et al. Rapid screening of diverse biotransformations for enzyme evolution. *JACS Au* **1**, 508−516 (2021).
35. Dörr, M. et al. Fully automatized high-throughput enzyme library screening using a robotic platform. *Biotechnol. Bioeng.* **113**, 1421–1432 (2016).
36. Stucki, A., Vallapurackal, J., Ward, T. R. & Dittrich, P. S. Droplet microfluidics and directed evolution of enzymes: an intertwined journey. *Angew. Chem. Int. Ed.* **60**, 24368–24387 (2021).

37. Taylor, S. V., Kast, P. & Hilvert, D. Investigating and engineering enzymes by genetic selection. *Angew. Chem. Int. Ed.* **40**, 3310–3335 (2001).

38. Acevedo-Rocha, C. G., Agudo, R. & Reetz, M. T. Directed evolution of stereoselective enzymes based on genetic selection as opposed to screening systems. *J. Biotechnol.* **191**, 3–10 (2014).

39. Xiao, H., Bao, Z. & Zhao, H. High throughput screening and selection methods for directed enzyme evolution. *Ind. Eng. Chem. Res.* **54**, 4011–4020 (2015).

40. Crameri, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).

41. Nearmnala, P., Thanaburakorn, M., Panbangred, W., Chaiyen, P. & Hongdilokkul, N. An in vivo selection system with tightly regulated gene expression enables directed evolution of highly efficient enzymes. *Sci. Rep.* **11**, 11669 (2021).

42. Collins, C. H., Leadbetter, J. R. & Arnold, F. H. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **24**, 708–712 (2006).

43. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).

44. Yano, T., Oue, S. & Kagamiyama, H. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl Acad. Sci. USA* **95**, 5511–5515 (1998).

45. Neuenschwander, M., Butz, M., Heintz, C., Kast, P. & Hilvert, D. A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol.* **25**, 1145–1147 (2007).

46. Baxter, S. et al. An improved racemase/acylase biotransformation for the preparation of enantiomerically pure amino acids. *J. Am. Chem. Soc.* **134**, 19310−19313 (2012).

47. Calzadiaz-Ramirez, L. et al. *In vivo* selection for formate dehydrogenases with high efficiency and specificity toward NADP+. *ACS Catal.* **10**, 7512−7525 (2020).

48. Sellés Vidal, L., Murray, J.W. & Heap, J. T. Versatile selective evolutionary pressure using synthetic defect in universal metabolism. *Nat. Commun.* **12**, 6859 (2021).

49. Maxel, S. et al. A growth-based, high-throughput selection platform enables remodeling of 4-hydroxybenzoate hydroxylase active site. *ACS Catal.* **10**, 6969−6974 (2020).

50. Kramer, L. et al. Engineering carboxylic acid reductase (CAR) through a whole-cell growth-coupled NADPH recycling strategy. *ACS Synth. Biol.* **9**, 1632−1637 (2020).

51. Reetz, M. T., Hobenreich, H., Soni, P. & Fernandez, L. A genetic selection system for evolving enantioselectivity of enzymes. *Chem. Commun.* 5502–5504 (2008).

52. Fernández-Álvaro, E. et al. A combination of *in vivo* selection and cell sorting for the identification of enantioselective biocatalysts. *Angew. Chem. Int. Ed.* **50**, 8584–8587 (2011).

53. Feng, J. et al. Discovery of alogliptin: a potent, selective, bioavailable, and efficacious inhibitor of dipeptidyl peptidase IV. *J. Med. Chem.* **50**, 2297–2300 (2007).

54. Eckhardt, M. et al. 8-(3-(R)-aminopiperidin-1-yl)-7-but-2-ynyl-3-methyl-1-(4-methyl-quinazolin-2-yl methyl)-3,7-dihydropurine-2,6-dione (BI 1356), a highly potent, selective, long-acting, and orally bioavailable DPP-4 inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* **50**, 6450–6453 (2007).

55. Block, G. A. et al. Cinacalcet for secondary hyperparathyroidism in patients receiving hemodialysis. *N. Engl. J. Med.* **350**, 1516–1525 (2004).

56. Trân, K. et al. Constraining the side chain of C-terminal amino acids in apelin-13 greatly increases affinity, modulates signaling, and improves the pharmacokinetic profile. *J. Med. Chem.* **64**, 5345−5364 (2021).

57. Truppo, M. D. Biocatalysis in the pharmaceutical industry: the need for speed. *ACS Med. Chem. Lett.* **8**, 476–480 (2017).

58. Petri, A., Colonna, V. & Piccolo, O. Asymmetric synthesis of a high added value chiral amine using immobilized ω-transaminases. *Beilstein J. Org. Chem.* **15**, 60–66 (2019).

59. Höhne, M., Schätzle, S., Jochens, H., Robins, K. & Bornscheuer, U. T. Rational assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.* **6**, 807–813 (2010).

60. Łyskowski, A. et al. Crystal structure of an (*R*)-selective ω-transaminase from *Aspergillus terreus*. *PLoS ONE* **9**, e87350 (2014).

61. Marx, L. et al. Chemoenzymatic approaches to the synthesis of the calcimimetic agent Cinacalcet employing transaminases and ketoreductases. *Adv. Synth. Catal.* **360**, 2157–2165 (2018).
62. Li, G. et al. Substrate profiling of cyclohexylamine oxidase and its mutants reveals new biocatalytic potential in deracemization of racemic amines. *Appl. Microbiol. Biotechnol.* **98**, 1681–1689 (2014).
63. Mirza, I. A. et al. Structural analysis of a novel cyclohexylamine oxidase from *Brevibacterium oxydans* IH-35A. *PLoS ONE* **8**, e60072 (2013).
64. Püllmann, P. et al. Golden mutagenesis: an efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **9**, 10932 (2019).
65. Bata, Z. et al. Substrate tunnel engineering aided by X-ray crystallography and functional dynamics swaps the function of MIO-enzymes. *ACS Catal.* **11**, 4538–4539 (2021).

Acknowledgements

We acknowledge the financial supports from the National Natural Science Foundation of China (No. 32101229 to S.W.), the Alexander von Humboldt-Stiftung (to S.W.), and the China Scholarship Council (a PhD thesis project to X.C., File No.: 201808330394). W.L. thanks the financial supports from Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-PTJJ-008, TSBICIP-IJCP-003, TSBICIP-KJGG-009-01, TSBICIP-KJGG-002-06), Youth Innovation Promotion Association CAS and China Scholarship Council. We thank Martin Woicke for preliminary experiments on CHAO, Ina Menyes for support with chiral GC analysis, Prof. Michael Lammers for providing reagents and facilities for crystallization, Prof. Huai-Long Teng for support with chiral HPLC analysis, and the staffs from Beamline 14.1 at BESSY for assistance during data collection.

Author contributions

S.W. and U.T.B. initiated the project. S.W. designed the study and selected the enzymes and targets. X.C. performed the evolution of AtTA; S.W. and M.S.H.K. performed the evolution of CHAO; S.W. and Y.Z. performed the evolution of PcPAL. W.L., C.G.F., R.W. and G.W. solved the crystal structure of AtTA(RHC). S.W., X.C., M.H. and U.T.B. discussed the findings and wrote the manuscript with inputs from all authors.

Competing interests

The authors declare no competing interests.

Supplementary Information for:

A growth selection for the directed evolution of amine-forming/converting enzymes

Shuke Wu[1,2,6]*, Chao Xiang[1,6], Yi Zhou[2], Mohammad Saiful Hasan Khan[1], Weidong Liu[1,3], Christian G. Feiler[4], Ren Wei[1], Gert Weber[4], Matthias Höhne[5] and Uwe T. Bornscheuer[1]*

[1] Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, Felix Hausdorff-Str. 4, D-17489 Greifswald, Germany

[2] State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, No. 1 Shizishan Street, Wuhan 430070, P. R. China

[3] Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, P. R. China

[4] Macromolecular Crystallography, Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Straße 15, 12489 Berlin, Germany

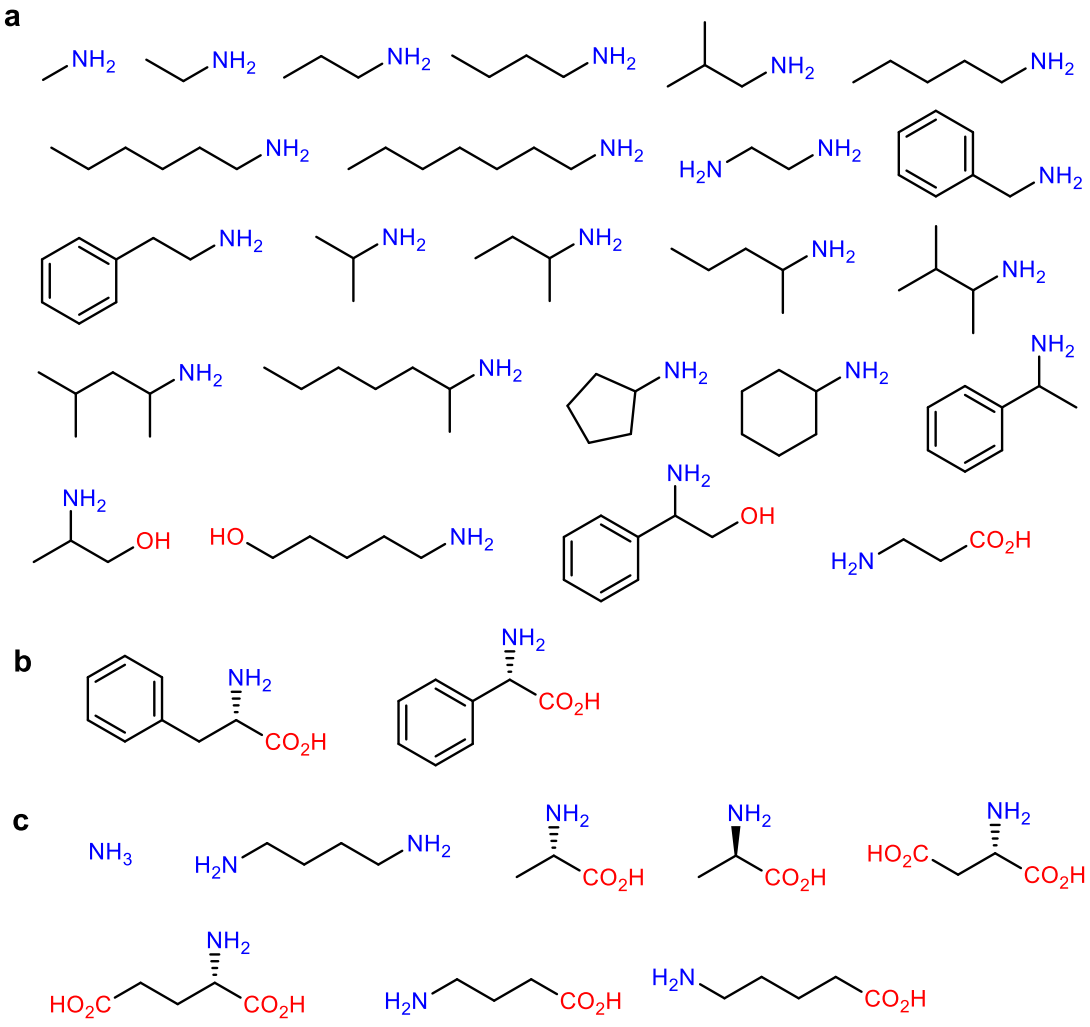
[5] Protein Biochemistry, Institute of Biochemistry, University of Greifswald, Felix Hausdorff-Str. 4, D-17489 Greifswald, Germany

[6] These authors contributed equally: Shuke Wu, Chao Xiang

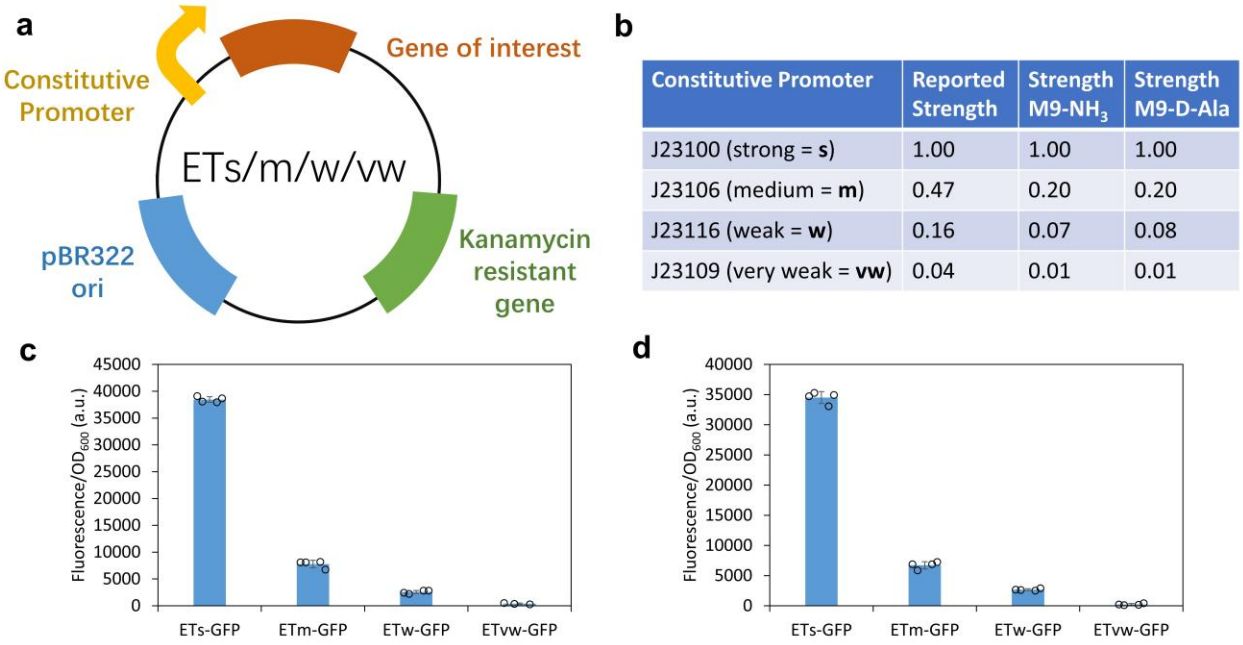
*e-mail: shukewu@mail.hzau.edu.cn; uwe.bornscheuer@uni-greifswald.de

Table of Contents

Supplementary Figures

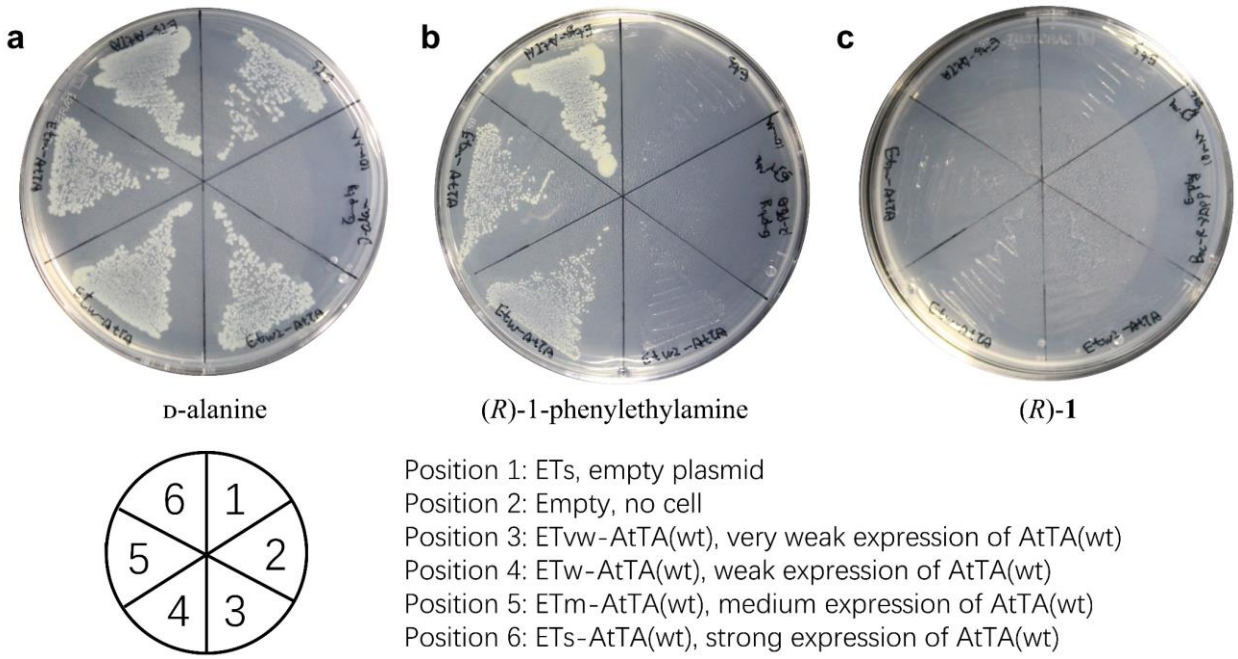


Supplementary Figure 1. Growth of *E. coli* BL21(DE3) with different amines and amino acids as the only nitrogen source. **a**, no growth on these amines (OD_{600} < 0.1 in 72 h). **b**, moderate growth on these amino acids (0.1 < OD_{600} < 0.5 in 72 h). **c**, growth on these amines and amino acids (OD_{600} > 0.5 in 72 h). Conditions: An aliquot (20 µl) of overnight culture of *E. coli* BL21(DE3) containing an empty plasmid in LB medium was inoculated into M9 medium (2 ml) with glucose (4 g l^{-1}) and the amine or amino acid (10 mM) as the sole nitrogen source and incubated at 30 ºC, 200 rpm for 72 h. The growth test did not consider the toxicity of the amine and amino acids.

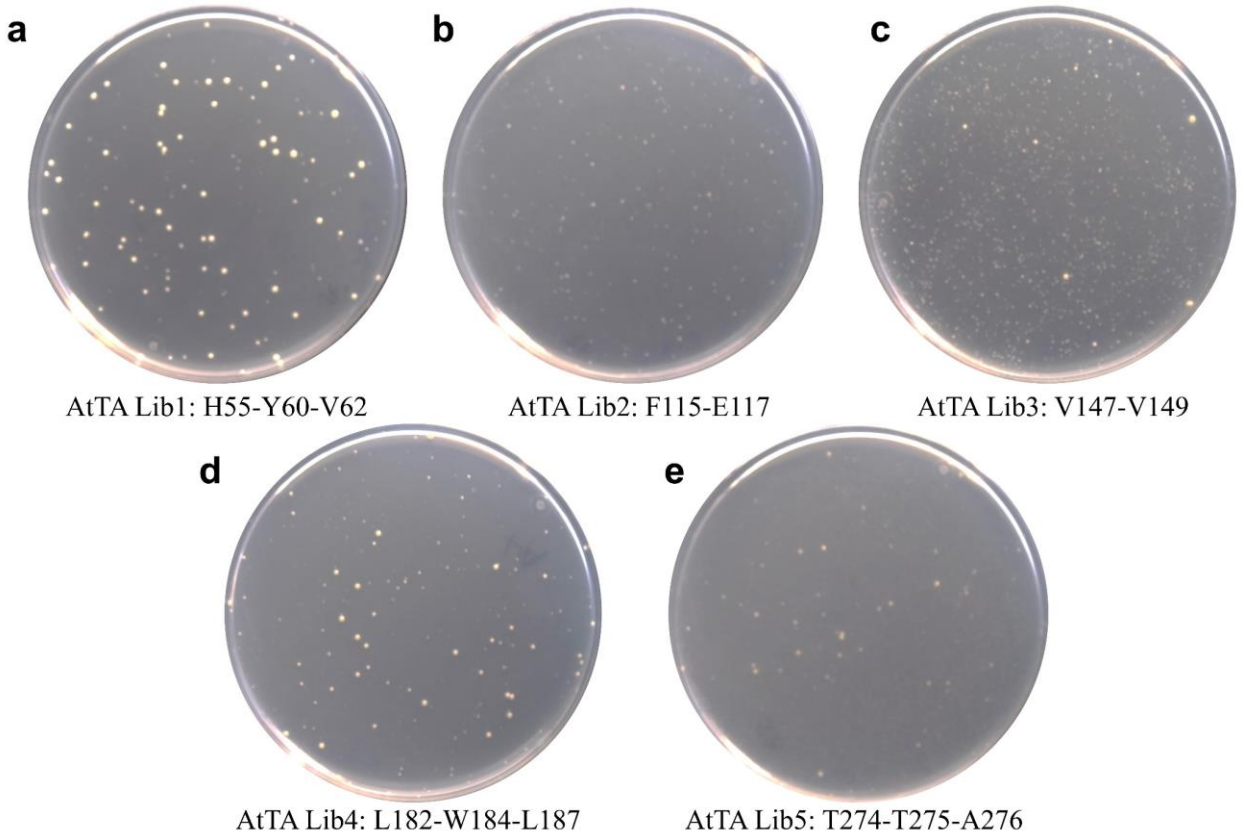


Constitutive Promoter	Reported Strength	Strength M9-NH_3	Strength M9-D-Ala
J23100 (strong = **s**)	1.00	1.00	1.00
J23106 (medium = **m**)	0.47	0.20	0.20
J23116 (weak = **w**)	0.16	0.07	0.08
J23109 (very weak = **vw**)	0.04	0.01	0.01

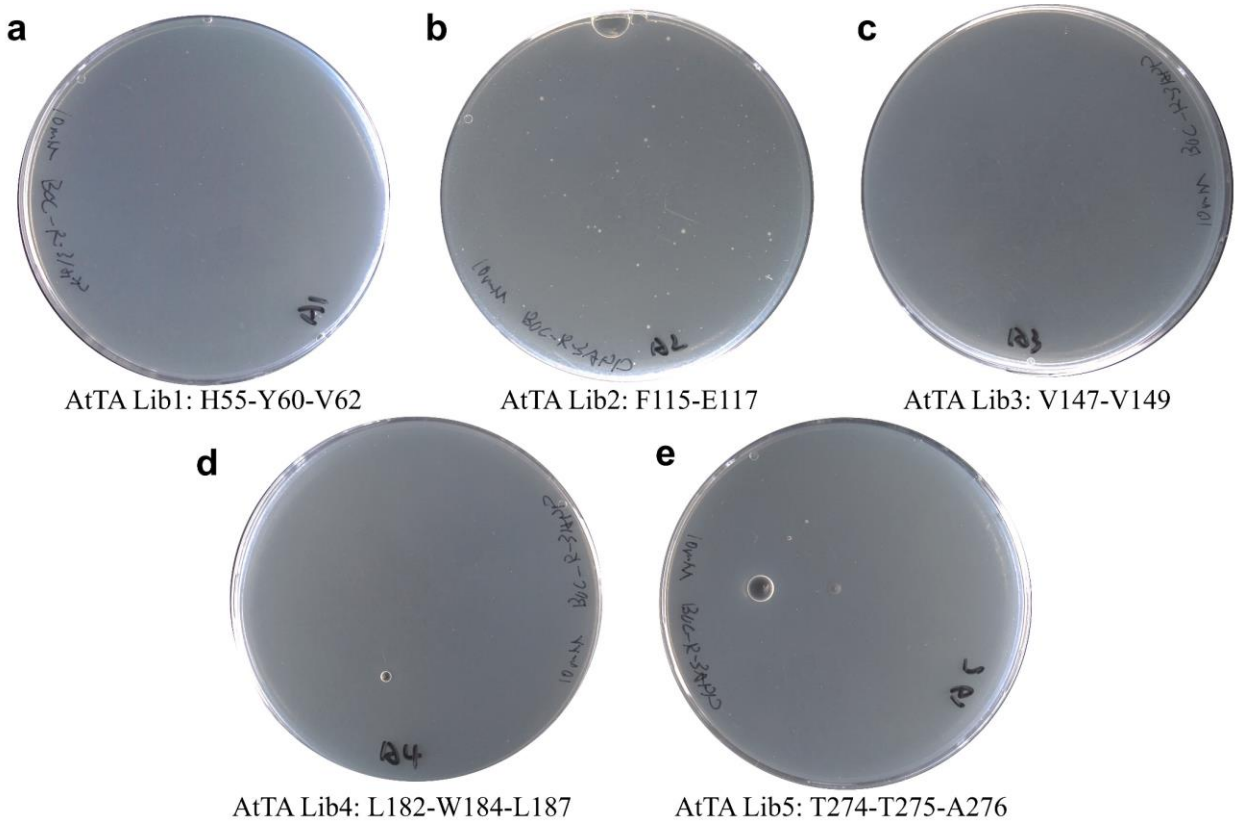
Supplementary Figure 2. Vectors and constitutive promoters used for growth selection in this study. **a**, the ETs/m/w/vw plasmids used in the selection consist of a pBR322 ori, a kanamycin resistance gene, and the gene of interest under the control of a constitutive promoter (strong, medium, weak, and very weak strength). The details of genetic engineering of the plasmids are provided in the method part. **b**, the four constitutive promoters used in this study. The reported strength is according to a previous study (https://parts.igem.org/Part:BBa_J23100). The strength in M9-NH_3 and M9-D-Ala was measured by using the fluorescence of superfolder GFP. **c**, the strength of the promoters was measured by the GFP fluorescence of the cells in M9-NH_3 medium. *E. coli* BL21(DE3) cells containing ETs/m/w/vw-GFP plasmid were inoculated (1%) in M9 medium (1 ml) with NH_3 (10 mM) as the sole nitrogen source, and incubated in a deep 96-well plate at 30 °C and 800 rpm for 24 h. **d**, the strength of the promoters was measured by the GFP fluorescence of the cells in M9-D-Ala medium. *E. coli* BL21(DE3) cells containing ETs/m/w/vw-GFP plasmid were inoculated (1%) in M9 medium (1 ml) with D-Ala (10 mM) as the sole nitrogen source, and incubated in a deep 96-well plate at 30 °C and 800 rpm for 24 h. Fluorescence and OD_{600} data were measured at the end of the cultivation. Data in **c** and **d** are mean values of quadruplicate experiments with error bars indicating the s. d. (n = 4).



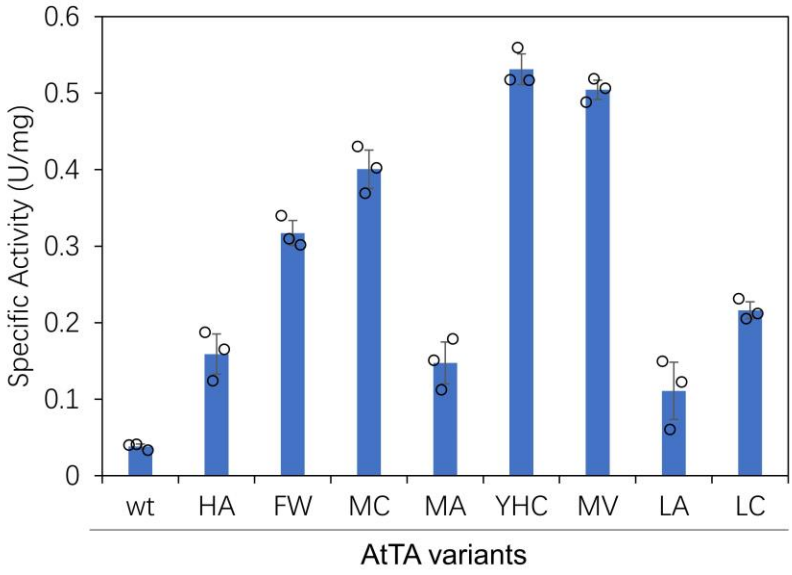
Supplementary Figure 3. Growth of *E. coli* cells expressing AtTA(wt) on M9 agar plates with different nitrogen sources. a, growth of the cells on an M9 agar plate with D-alanine (10 mM) as the only nitrogen source. **b**, growth of the cells on an M9 agar plate with (*R*)-1-phenylethylamine (10 mM) as the only nitrogen source. **c**, growth of the cells on an M9 agar plate with (*R*)-**1** (10 mM) as the only nitrogen source. *E. coli* BL21(DE3) cells containing ETs/m/w/vw-AtTA(wt) or ETs plasmid were streaked on the M9 agar plates with different nitrogen sources and incubated at 30 °C for 72 h, and photos were taken.



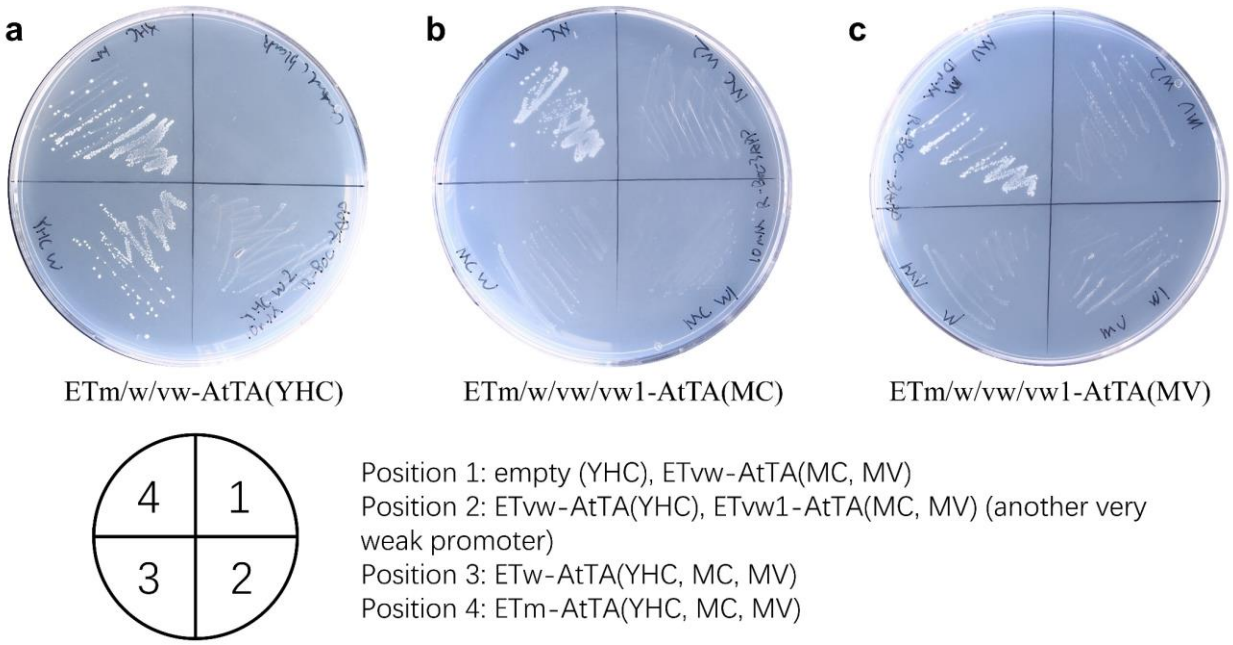
Supplementary Figure 4. Growth of *E. coli* cells containing ETs-AtTA libraries (strong expression levels) on M9 agar plates with (*R*)-1 (10 mM) as the sole nitrogen source. a, growth of *E. coli* cells containing ETs-AtTA-Lib1 (H55-Y60-V62). **b**, growth of *E. coli* cells containing the ETs-AtTA-Lib2 (F115-E117). **c**, growth of *E. coli* cells containing ETs-AtTA-Lib3 (V147-V149). **d**, growth of *E. coli* cells containing ETs-AtTA-Lib4 (L182-W184-L187). **e**, growth of *E. coli* cells containing ETs-AtTA-Lib5 (T274-T275-A276). Due to the close of the university during the outbreak of COVID-19 (no incubator was available), the plates were kept at room temperature (~15 °C) for 3 weeks and photos were taken then.



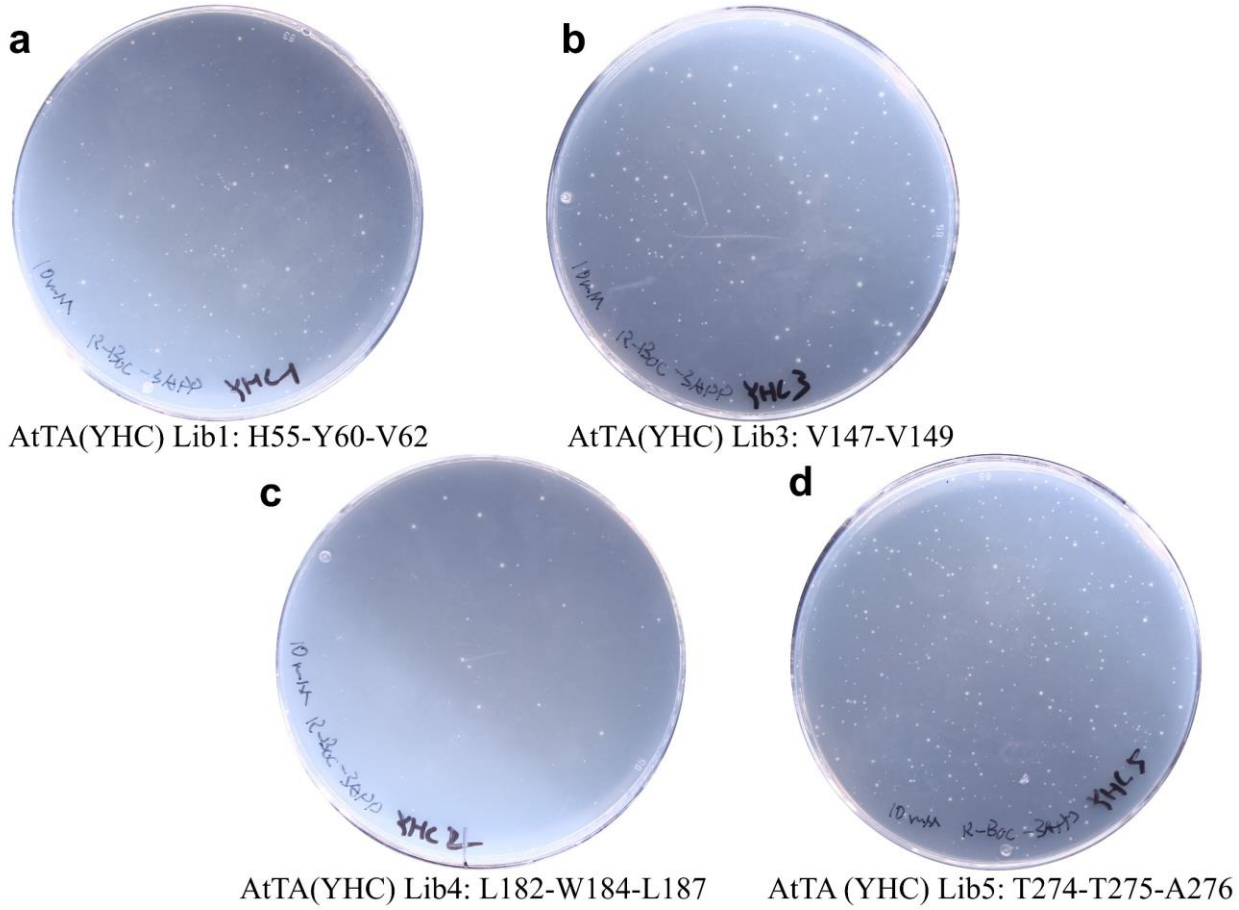
Supplementary Figure 5. Growth of *E. coli* cells containing ETm-AtTA libraries (medium expression level) on M9 agar plates with (*R*)-1 (10 mM) as the sole nitrogen source. a, growth of *E. coli* cells containing ETm-AtTA-Lib1 (H55-Y60-V62). **b**, growth of *E. coli* cells containing **ETm-AtTA-Lib2 (F115-E117)**. **c**, growth of *E. coli* cells containing ETm-AtTA-Lib3 (V147-V149). **d**, growth of *E. coli* cells containing ETm-AtTA-Lib4 (L182-W184-L187). **e**, growth of *E. coli* cells containing ETm-AtTA-Lib5 (T274-T275-A276). The plates were incubated at 30 °C for 72 h and photos were taken then.



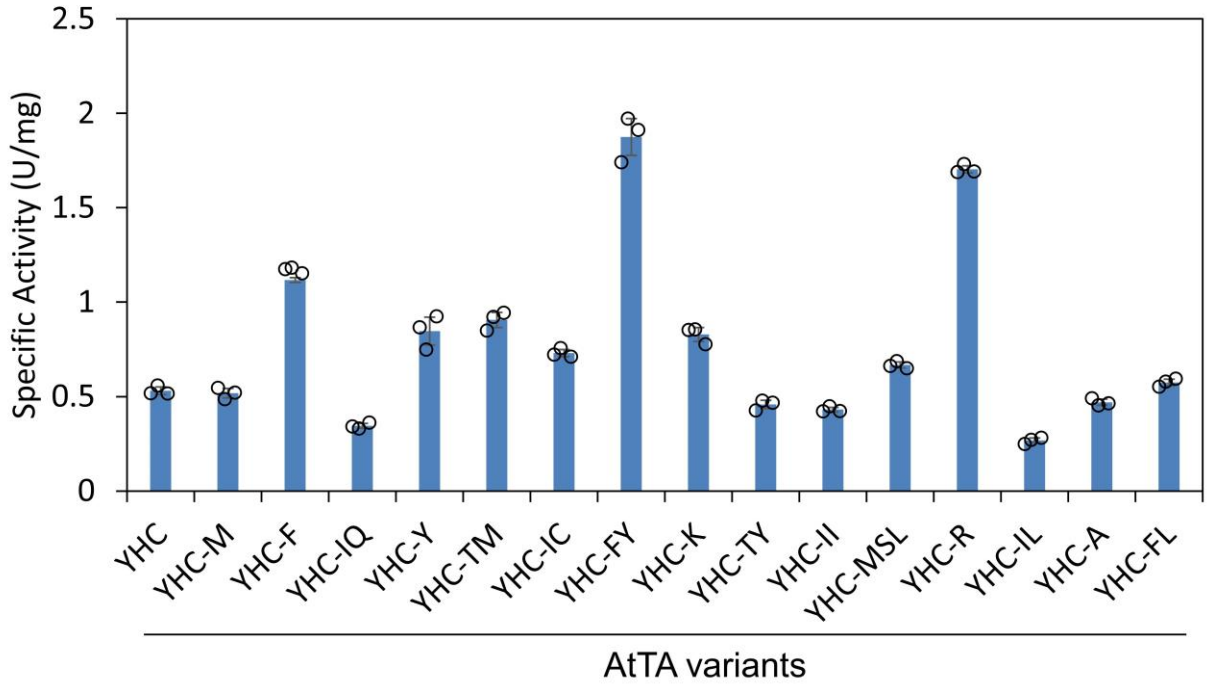
Supplementary Figure 6. Specific activities of purified AtTA variants (1st round from the F115-E117 library) towards (*R*)-1. AtTA variants and the corresponding mutations: HA, F115H/E117A; FW, E117W; **MC, F115M/E117C**; MA, F115M/E117A; **YHC, D5Y/F115H/E117C**; **MV, F115M/E117V**; LA, F115L/E117A; LA, F115L/E117C. The D5Y in the YHC variant was accidentally introduced during the mutagenesis. The specific activities of the purified AtTA variants were determined by the DAAO (D-alanine oxidase) assay. The detailed procedures of the activity assay are provided in the supplementary methods. Data are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



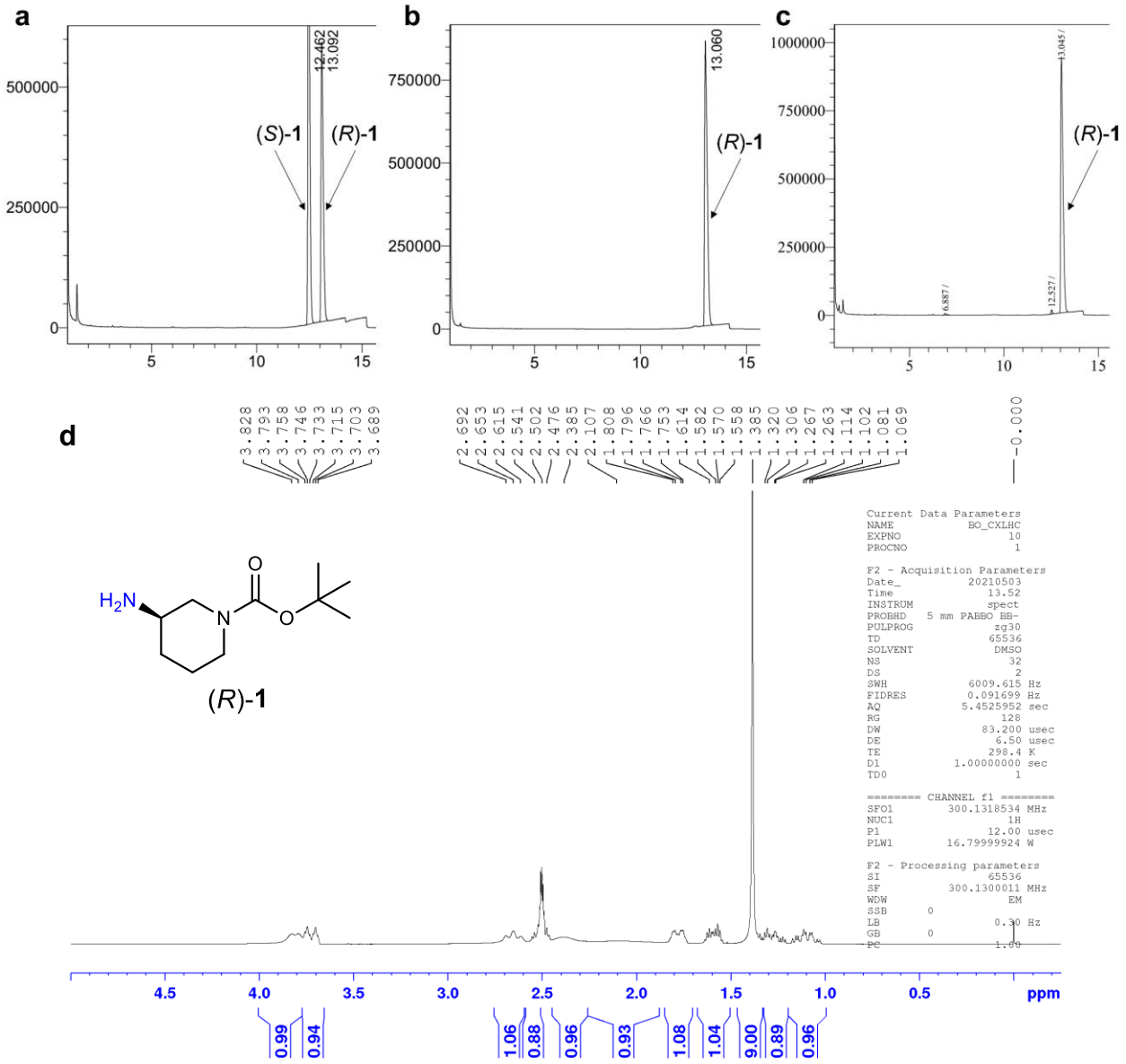
Supplementary Figure 7. Growth of *E. coli* cells containing ETm/w/vw-AtTA(MC/MV/YHC) (medium to very weak expression levels) on M9 agar plates with (*R*)-1 (10 mM) as the sole nitrogen source. a, growth of *E. coli* cells containing ETm/w/vw-AtTA(YHC) on an M9 agar plate with (*R*)-**1** (10 mM). **b**, growth of *E. coli* cells containing ETm/w/vw/vw1-AtTA(MC) on an M9 agar plate with (*R*)-**1** (10 mM). **c**, growth of *E. coli* cells containing ETm/w/vw/vw1-AtTA(MV) on an M9 agar plate with (*R*)-**1** (10 mM). For AtTA(MC) and AtTA(MV), another very weak promoter (vw1, J23117) was also used in the test. The plates were incubated at 30 °C for 72 h and photos were taken then.



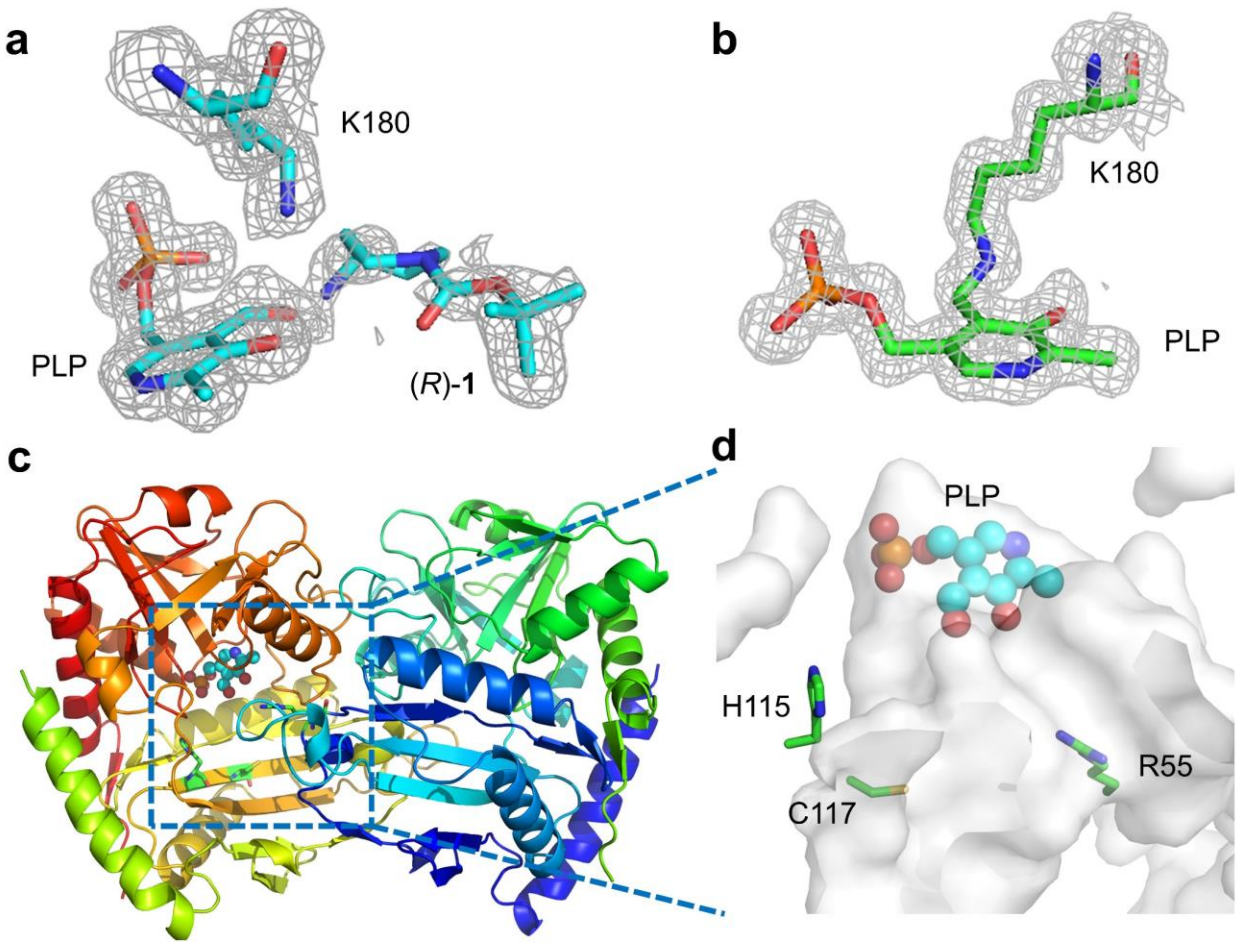
Supplementary Figure 8. Growth of *E. coli* cells containing ETvw-AtTA(YHC) libraries (very weak expression levels) on M9 agar plates with (*R*)-1 **(10 mM) as the sole nitrogen source. a**, growth of *E. coli* cells containing ETvw-AtTA(YHC)-Lib1 (H55-Y60-V62). **b**, growth of *E. coli* cells containing ETvw-AtTA(YHC)-Lib3 (V147-V149). **c**, growth of *E. coli* cells containing ETvw-AtTA(YHC)-Lib4 (L182-W184-L187). **d**, growth of *E. coli* cells containing ETvw-AtTA(YHC)-Lib5 (T274-T275-A276). The plates were incubated at 30 °C for 72 h and photos were taken then.



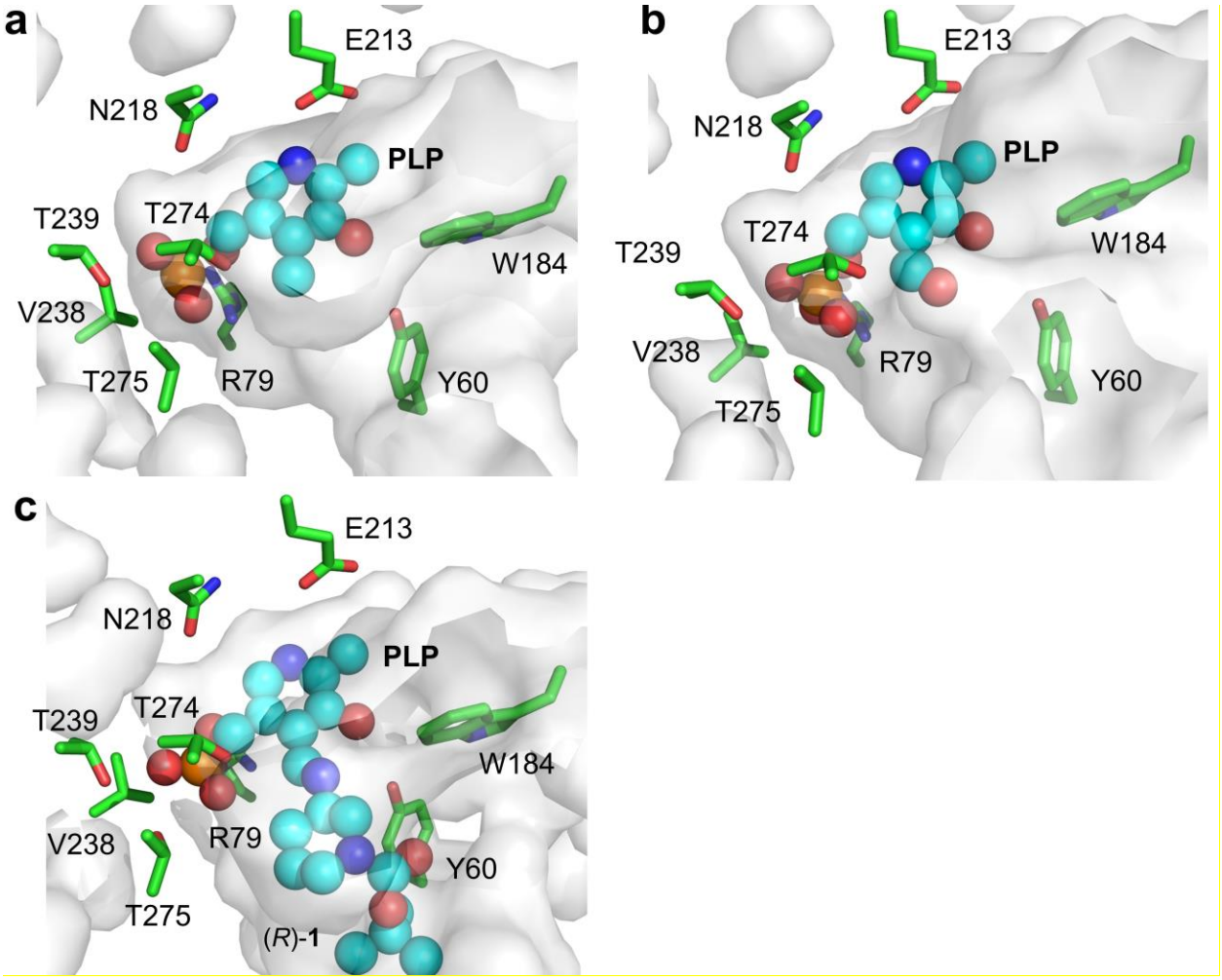
Supplementary Figure 9. Specific activities of purified AtTA variants (2nd round with YHC as the template) towards (*R*)-1. AtTA variants and the corresponding additional mutations (besides D5Y/F115H/E117C): YHC-M, L187M; YHC-F, V147F; YHC-IQ, V147I/V149Q; YHC-Y, V147Y; YHC-TM, V147T/V149M; YHC-IC, V147I/V149C; **YHC-FY, V147F/V149Y**; YHC-K, H55K; YHC-TY, V147T/V149Y; YHC-II, V147I/V149I; YHC-MSL, L73M/V147S/V149L; **YHC-R, H55R**; YHC-IL, V147I/V149L; YHC-A, T275A; YHC-FL, V147F/V149L. The specific activities of the purified AtTA variants were determined by the DAAO (D-alanine oxidase) assay. The detailed procedures of the activity assays are provided in the supplementary methods. Data are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



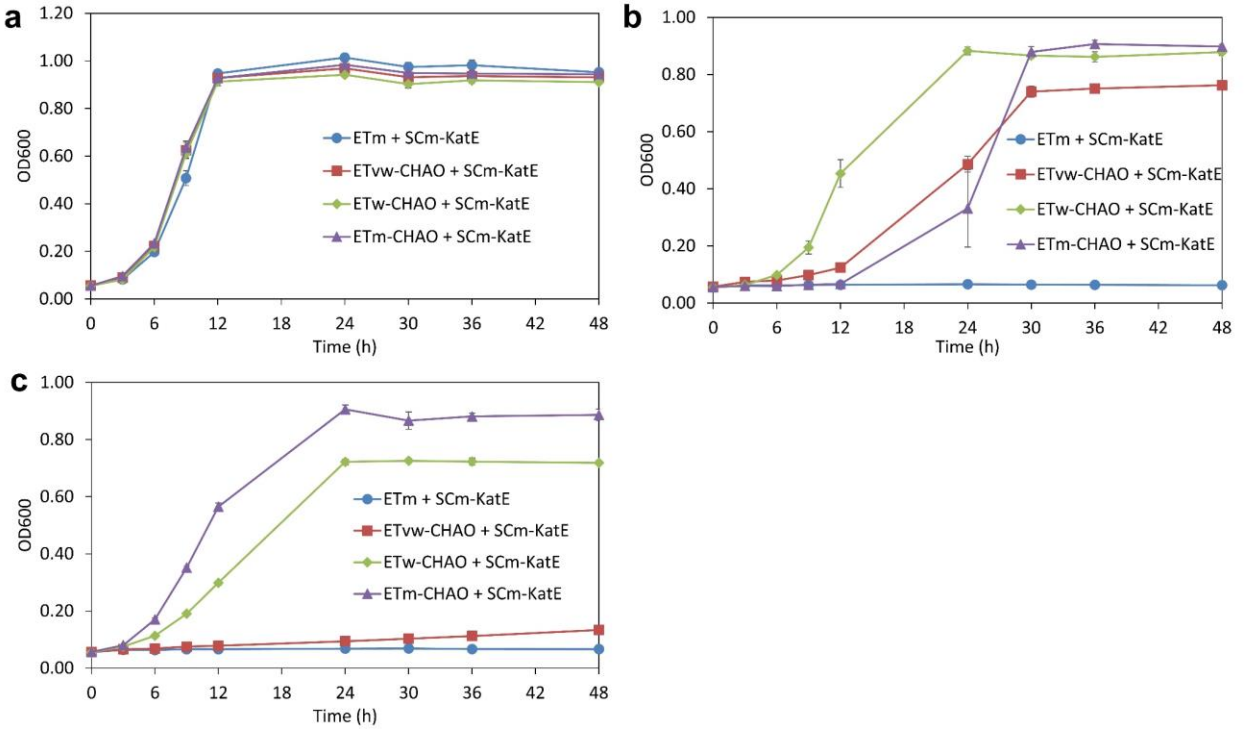
Supplementary Figure 10. Preparative scale synthesis of (*R*)-1 from 2 with purified AtTA(RHC). a, chiral GC chromatogram of the commercial standard of *rac*-**1**. **b**, chiral GC chromatogram of the commercial standard of (*R*)-**1**. **c**, chiral GC chromatogram of the synthesized (*R*)-**1**. **d**, ^{1}H-NMR spectrum of the synthesized (*R*)-**1**. The detailed procedures of the preparative scale synthesis are provided in the supplementary methods.



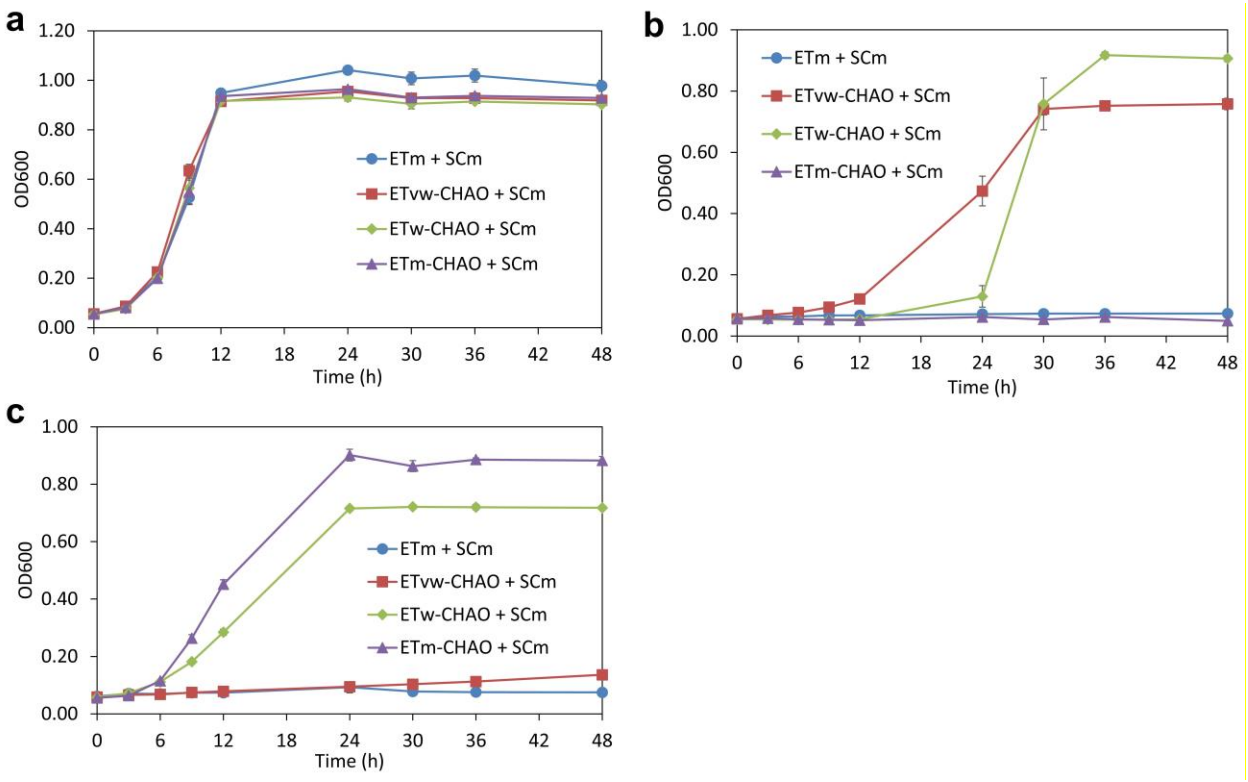
Supplementary Figure 11. Crystal structures of AtTA(RHC). a, 2m|Fo|-D|Fc| map (contoured at 1.0 σ) of the AtTA(RHC) structure (PDB: 7XG5) with PLP and (*R*)-**1**. Only partial density was observed on the (*R*)-**1** part. **b**, 2m|Fo|-D|Fc| map (contoured at 1.0 σ) of the AtTA(RHC) structure (PDB: 7XG6) with PLP bound to K180. **c**, overall structure of AtTA(RHC) showed a typical aminotransferase class IV fold. **d**, substrate-binding pocket of AtTA(RHC) with the mutations R55, H115, and C117. The detailed procedures of crystallization, data collection, and structure determination are provided in the supplementary methods. Data collection and refinement statistics are summarized in Supplementary Table 1.



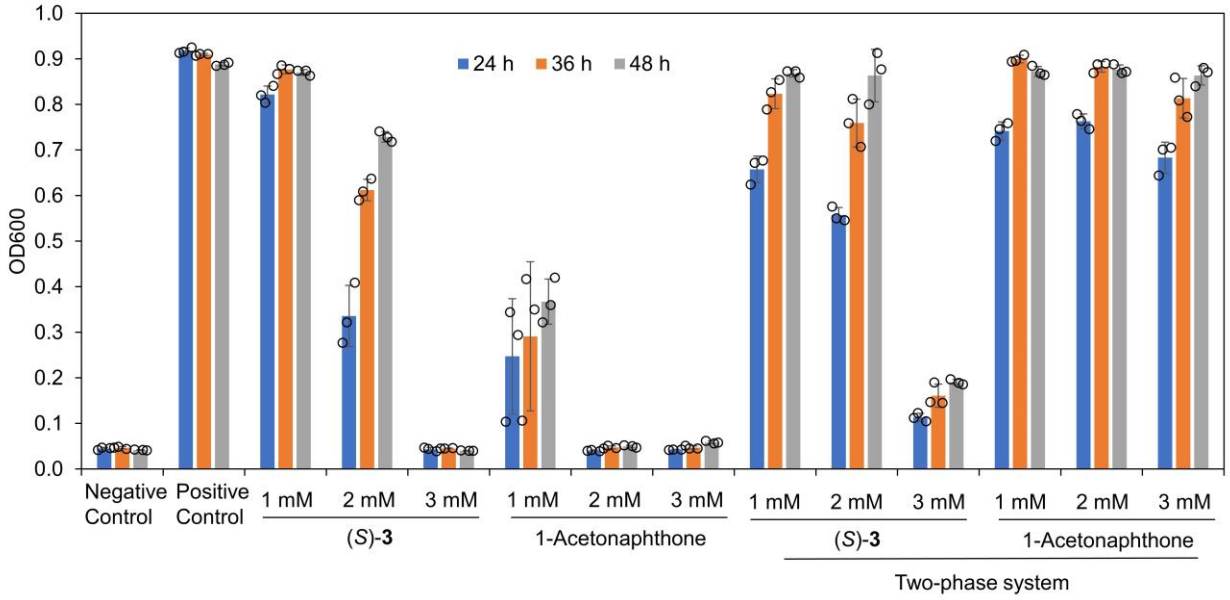
Supplementary Figure 12. PLP-binding pockets and surrounding residues. a, PLP-binding pocket of AtTA(wt) (PDB: 4CE5). b, PLP-binding pocket of AtTA(RHC) (PDB: 7XG6). c, binding pocket of the PLP part of the docked PLP-(*R*)-**1** complex.



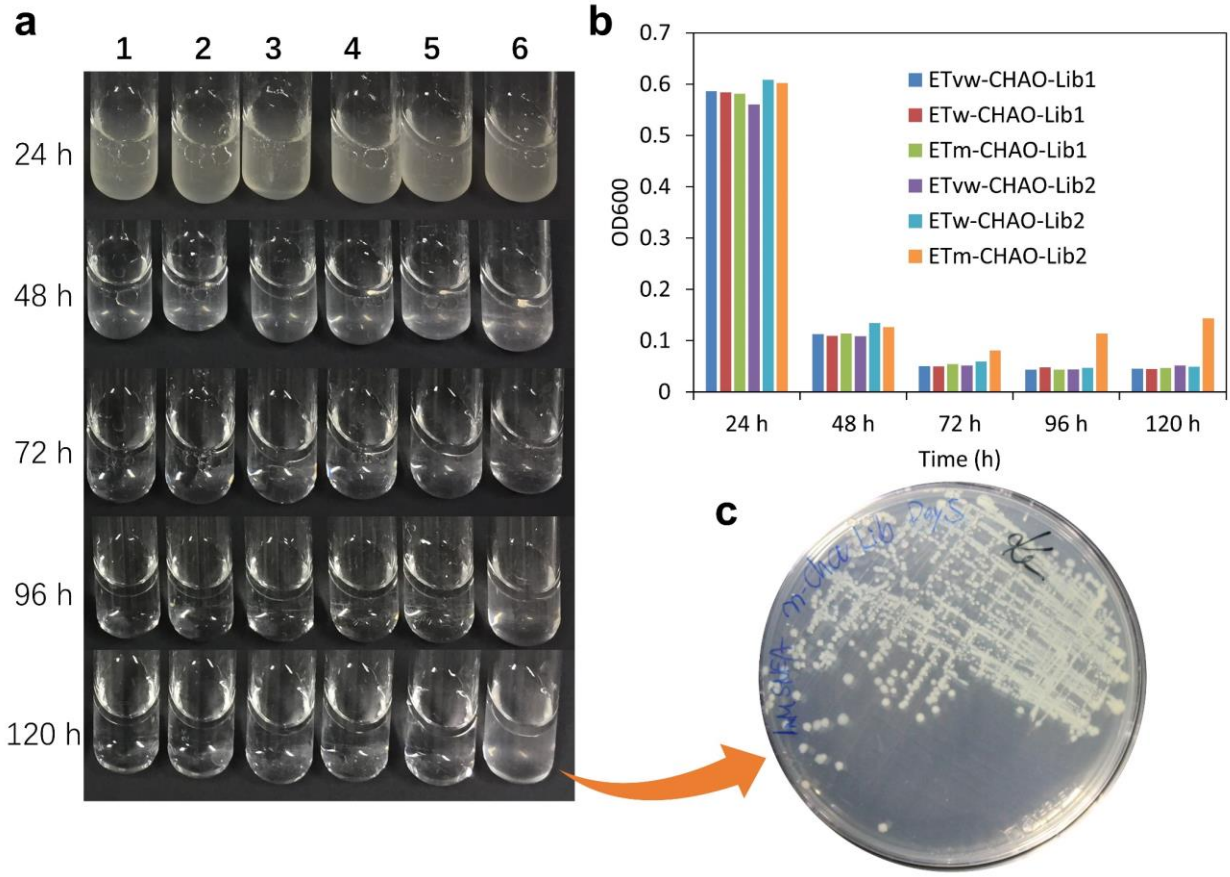
Supplementary Figure 13. Growth of *E. coli* cells co-expressing CHAO(wt) and catalase in M9 liquid medium with different nitrogen sources. a, growth of the cells in M9 liquid medium with ammonia (10 mM) as the only nitrogen source. **b**, growth of the cells in M9 liquid medium with cyclohexylamine (10 mM, specific activity of 5.6 U mg^{-1}) as the only nitrogen source. **c**, growth of the cells in M9 liquid medium with cyclopentylamine (10 mM, specific activity of 0.33 U mg^{-1}) as the only nitrogen source. *E. coli* BL21(DE3) cells containing ETm/w/vw-CHAO(wt) or ETm plasmid together with SCm-KatE plasmid (constitutive expressing of the catalase from *E. coli*) were used. ETs-CHAO(wt) with a strong promoter significantly inhibited the growth of *E. coli* even in LB medium (probably due to the burden of very strong expression), thus this was not included in this study. The cells (overnight cultured in M9 medium with ammonia) were inoculated (1%) in M9 medium (1 ml) with different nitrogen sources, and incubated in a deep 96-well plate at 30 °C and 800 rpm for 48 h. Data are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



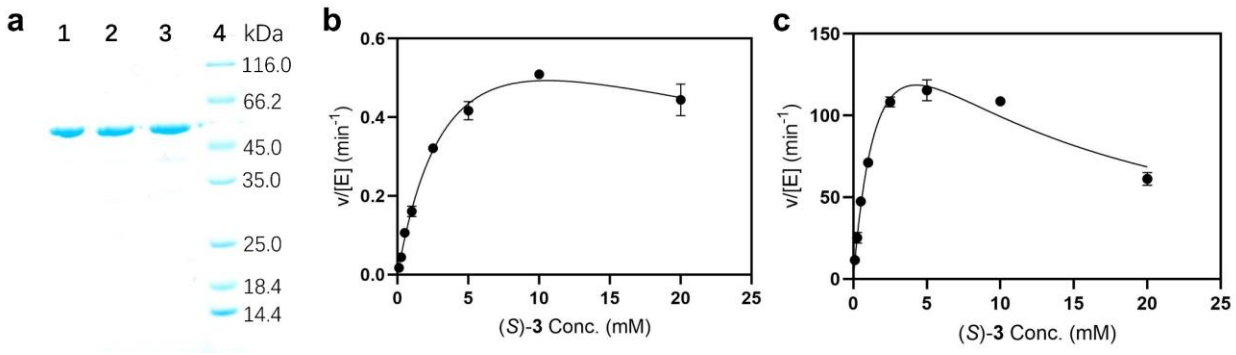
Supplementary Figure 14. Growth of *E. coli* cells expressing CHAO(wt) but without catalase in M9 liquid medium with different nitrogen sources. a, growth of the cells in M9 liquid medium with ammonia (10 mM) as the only nitrogen source. **b**, growth of the cells in M9 liquid medium with cyclohexylamine (10 mM, specific activity of 5.6 U mg^{-1}) as the only nitrogen source. **c**, growth of the cells in M9 liquid medium with cyclopentylamine (10 mM, specific activity of 0.33 U mg^{-1}) as the only nitrogen source. *E. coli* BL21(DE3) cells containing ETm/w/vw-CHAO(wt) or ETm plasmid together with SCm plasmid (empty, without catalase) were used. The cells (overnight cultured in M9 medium with ammonia) were inoculated (1%) in M9 medium (1 ml) with different nitrogen sources, and incubated in a deep 96-well plate at 30 °C and 800 rpm for 48 h. Data are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



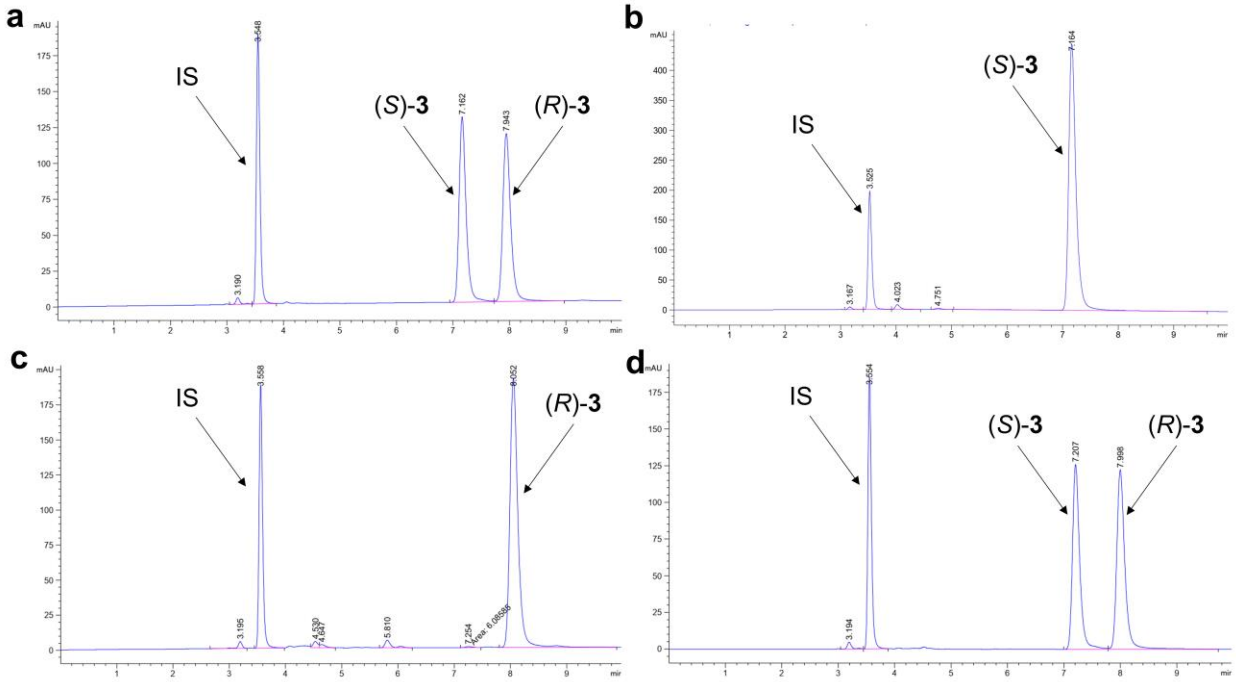
Supplementary Figure 15. Growth of *E. coli* cells in M9 liquid medium with ammonia (10 mM) and (*S*)-3 or the corresponding ketone at different concentrations. The toxicity of (*S*)-**3** or the corresponding ketone (1-acetonaphthone) was studied with *E. coli* BL21(DE3) cells containing ETm plasmid in M9 liquid medium with ammonia (10 mM). Negative control: in M9 liquid medium without nitrogen source. Positive control: in M9 liquid medium with ammonia (10 mM) without additional (*S*)-**3** or 1-acetonaphthone. Two-phase system: in M9 liquid medium with ammonia (10 mM) and methyl laurate (v/v: 4:1). The cells (overnight cultured in LB medium) were inoculated (1%) in M9 medium (2 ml) with ammonia (10 mM) and different amounts of (*S*)-**3** or 1-acetonaphthone, and incubated in cultural tubes at 30 °C and 200 rpm for 48 h. Data are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



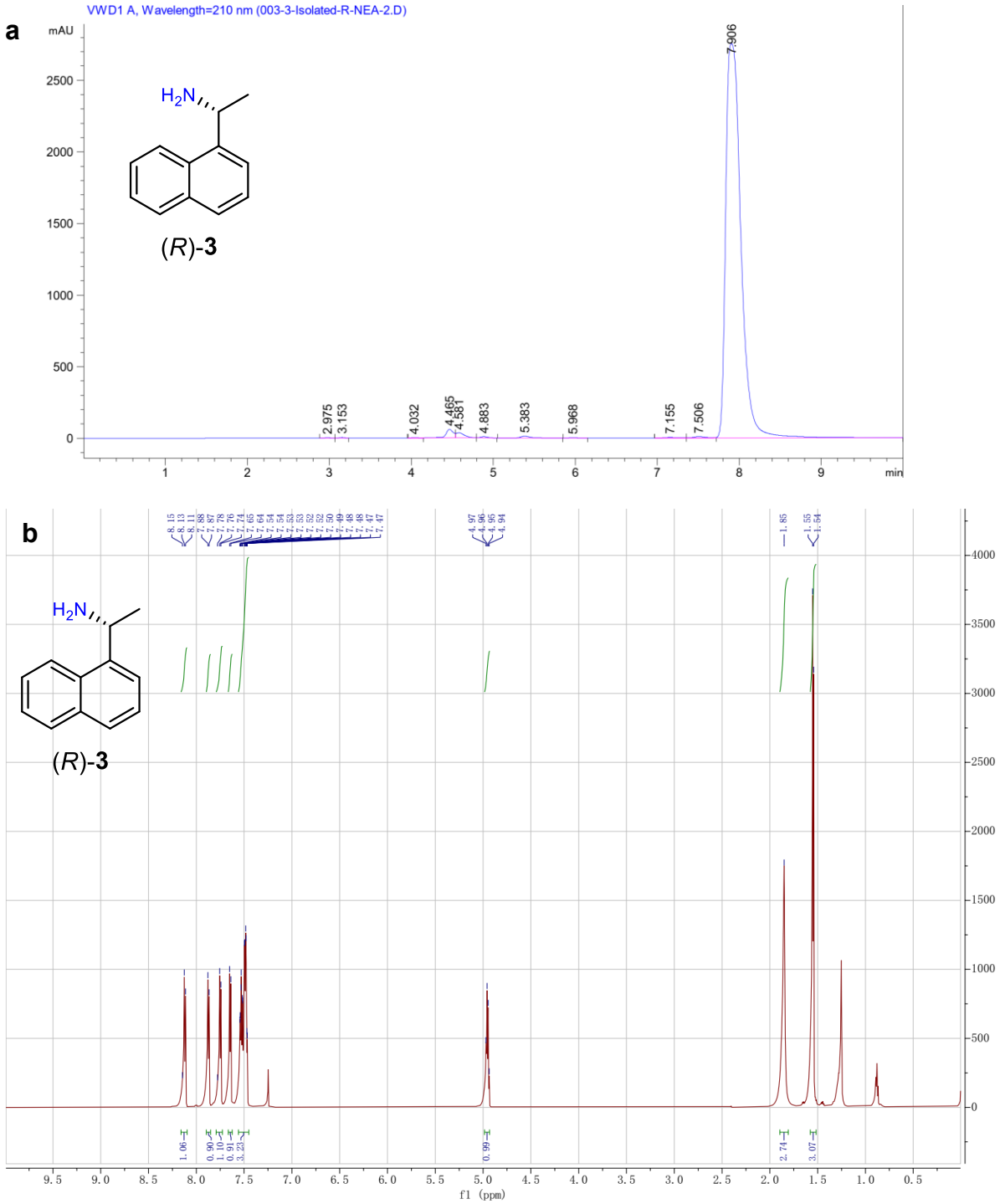
Supplementary Figure 16. Growth of *E. coli* cells co-expressing catalase and CHAO libraries in M9 liquid medium with (*S*)-3 **(1 mM) and methyl laurate (v/v: 4:1).** Lib1: simultaneous mutagenesis (DBS) of F88-F351-L353-F368. Lib2: simultaneous mutagenesis (DBS) of T198-L199-M226-Y321. **a**, Photos of the growth of *E. coli* cells containing different CHAO libraries. Line 1: *E. coli* with ETvw-CHAO-Lib1; Line 2: *E. coli* with ETw-CHAO-Lib1; Line 3: *E. coli* with ETm-CHAO-Lib1; Line 4: *E. coli* with ETvw-CHAO-Lib2; Line 5: *E. coli* with ETw-CHAO-Lib2; Line 6: *E. coli* with ETm-CHAO-Lib2. **b**, Optical density at 600 nm (OD600) of the *E. coli* cells containing different CHAO libraries. **c**, Isolation of single colonies from the enriched culture of *E. coli* cells containing ETm-CHAO-Lib2 (T198-L199-M226-Y321). The detailed procedures of the growth selection are provided in the methods of the main text. Data in **b** are from one independent experiment.



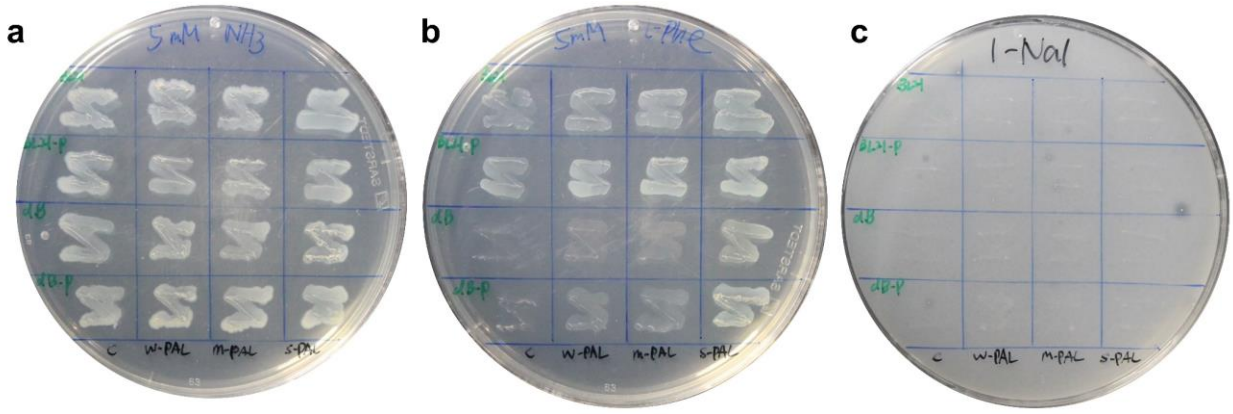
Supplementary Figure 17. Characterization of purified CHAO variants. a, SDS-PAGE analysis of purified CHAO variants (with a C-terminal His-tag). Lane 1: CHAO(wt); lane 2: CHAO(MALM); lane 3: CHAO(MVLM); lane 4: protein marker. **b**, enzyme kinetics of CHAO(wt) for oxidation of (*S*)-**3**. The solid line corresponds to the fitting to the Haldane equation. **c**, enzyme kinetics of CHAO(MVLM) for oxidation of (*S*)-**3**. The solid line corresponds to the fitting to the Haldane equation. The detailed procedures of enzyme purification and kinetic determination are provided in the supplementary methods. Data in **b** and **c** are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



Supplementary Figure 18. Chiral HPLC chromatograms for deracemization of *rac*-3 by *E. coli* co-expressing CHAO variants and catalase and NH_3-BH_3. **a**, *rac*-**3** in the reaction system without *E. coli* catalyst. **b**, commercial standard of (*S*)-**3**. **c**, (*R*)-**3** produced by *E. coli* co-expressing CHAO(MVLM) and catalase and NH_3-BH_3 for 4 h. **d**, *rac*-**3** left by *E. coli* co-expressing CHAO(wt) and catalase and NH_3-BH_3 for 4 h. Chiral HPLC analysis conditions: Daicel Chiralcel OD-3 column (3 μm, 4.6 mm × 250 mm), a constant flow (1.0 ml min^{-1}) of *n*-hexane (80%) and isopropanol (20%), detection at 210 nm. The detailed procedures of the reaction are provided in the methods of the main text.



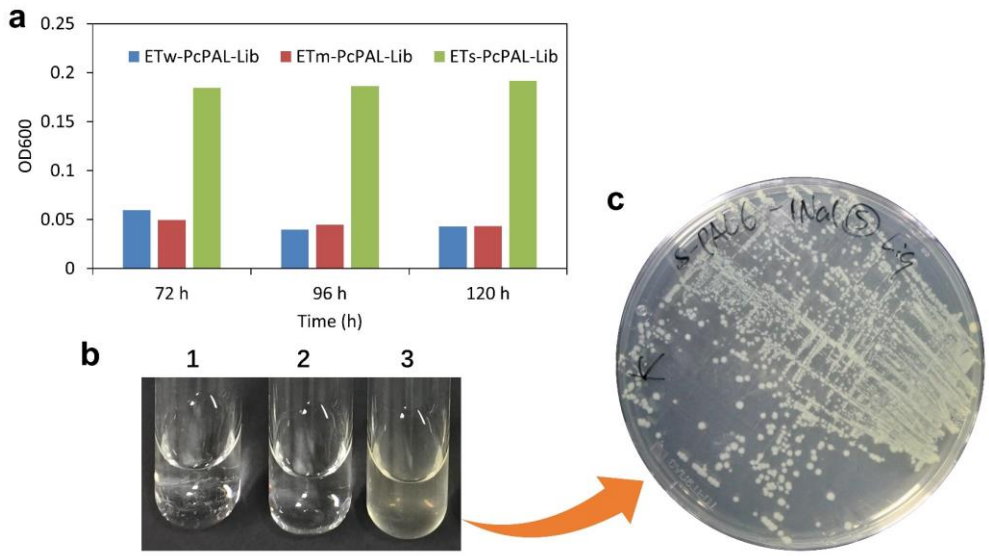
Supplementary Figure 19. Preparative scale synthesis of (*R*)-3 by deracemization of *rac*-3 by *E. coli* co-expressing CHAO(MVLM) and catalase and NH_3-BH_3. a, chiral HPLC chromatogram of the synthesized (*R*)-**3**. **b**, ^{1}H-NMR spectrum of the synthesized (*R*)-**3**. The detailed procedures for the preparative scale synthesis are provided in the supplementary methods.



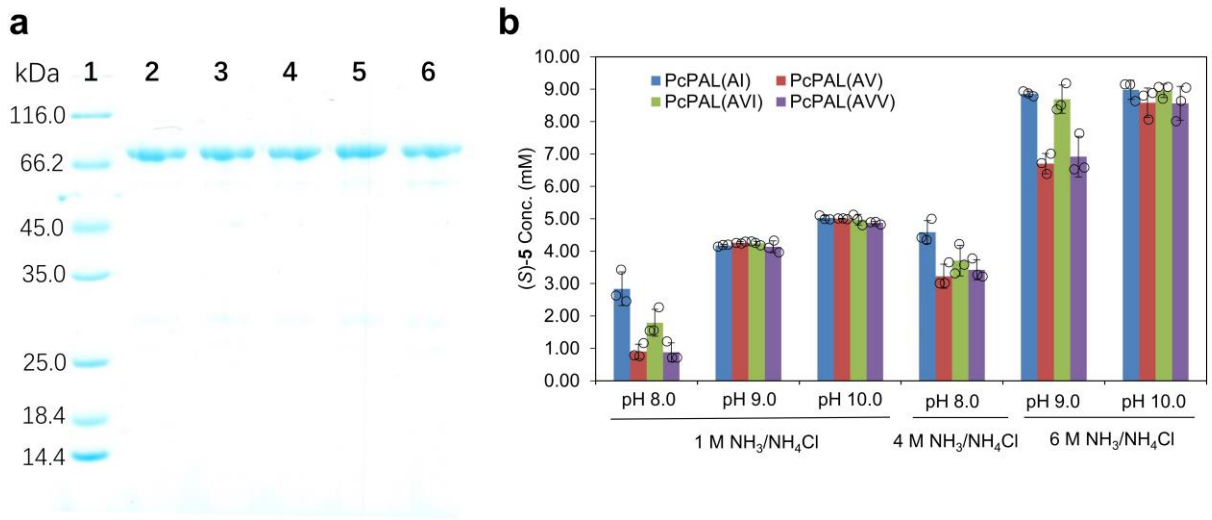
NH_3 | L-phenylalanine | (*S*)-**5**

Strain	Main Plasmid: ETm	ETw-PcPAL(wt)	ETm-PcPAL(wt)	ETs-PcPAL(wt)
E. coli BL21(DE3)	1	2	3	4
E. coli BL21(DE3) SCm-AroP	5	6	7	8
E. coli BL21(DE3) Δ*tyrB*	9	10	11	12
E. coli BL21(DE3) Δ*tyrB* SCm-AroP	13	14	15	16

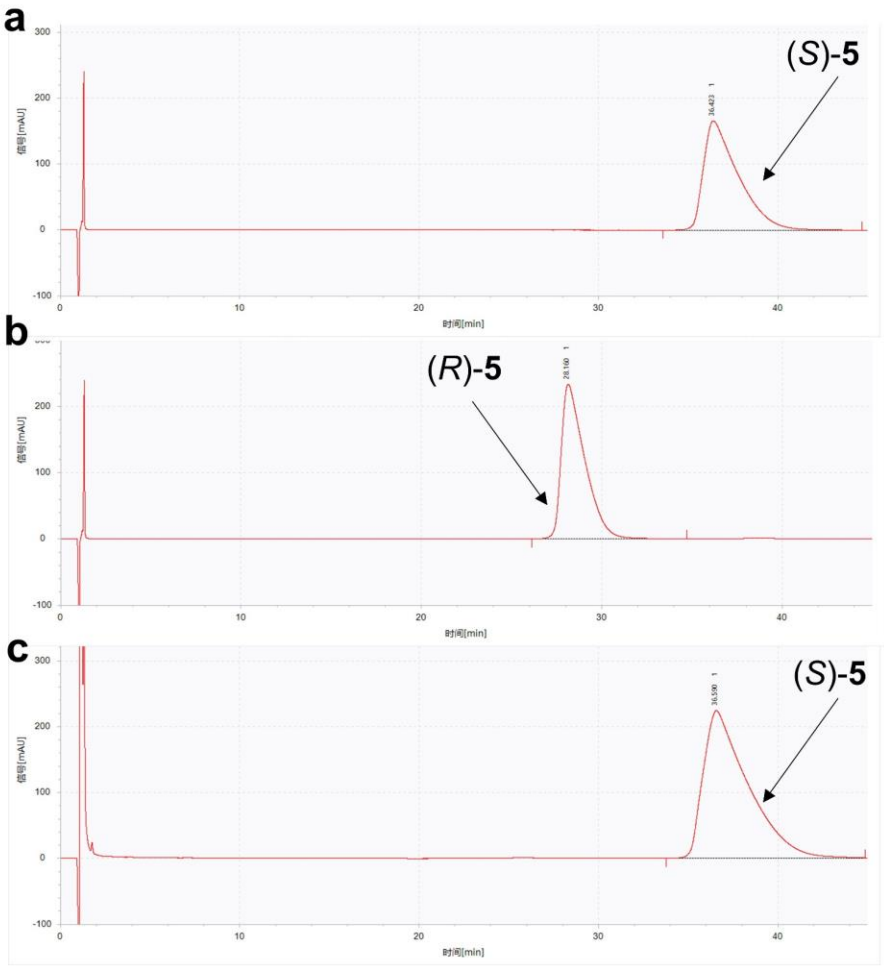
Supplementary Figure 20. Growth of different *E. coli* strains expressing PcPAL(wt) on M9 agar plates with different nitrogen sources. The strains differed in co-expressing the aromatic amino acid transporter (SCm-AroP) and the gene encoding the aromatic amino acid transaminase (Δ*tyrB*) was deleted. The main plasmids differed in expressing PcPAL(wt) at different levels. **a**, growth of the cells on an M9 agar plate with ammonia (5 mM) as the only nitrogen source. **b**, growth of the cells on an M9 agar plate with L-phenylalanine (5 mM) as the only nitrogen source. **c**, growth of the cells on an M9 agar plate with (*S*)-**5** (5 mM) as the only nitrogen source. Due to the limited solubility of (*S*)-**5**, the agar plate became opaque (probably because of precipitation of (*S*)-**5**) when cooling to room temperature. The *E. coli* cells containing different plasmids were streaked on the M9 agar plates with different nitrogen sources and incubated at 30 °C for 72 h, and photos were taken.



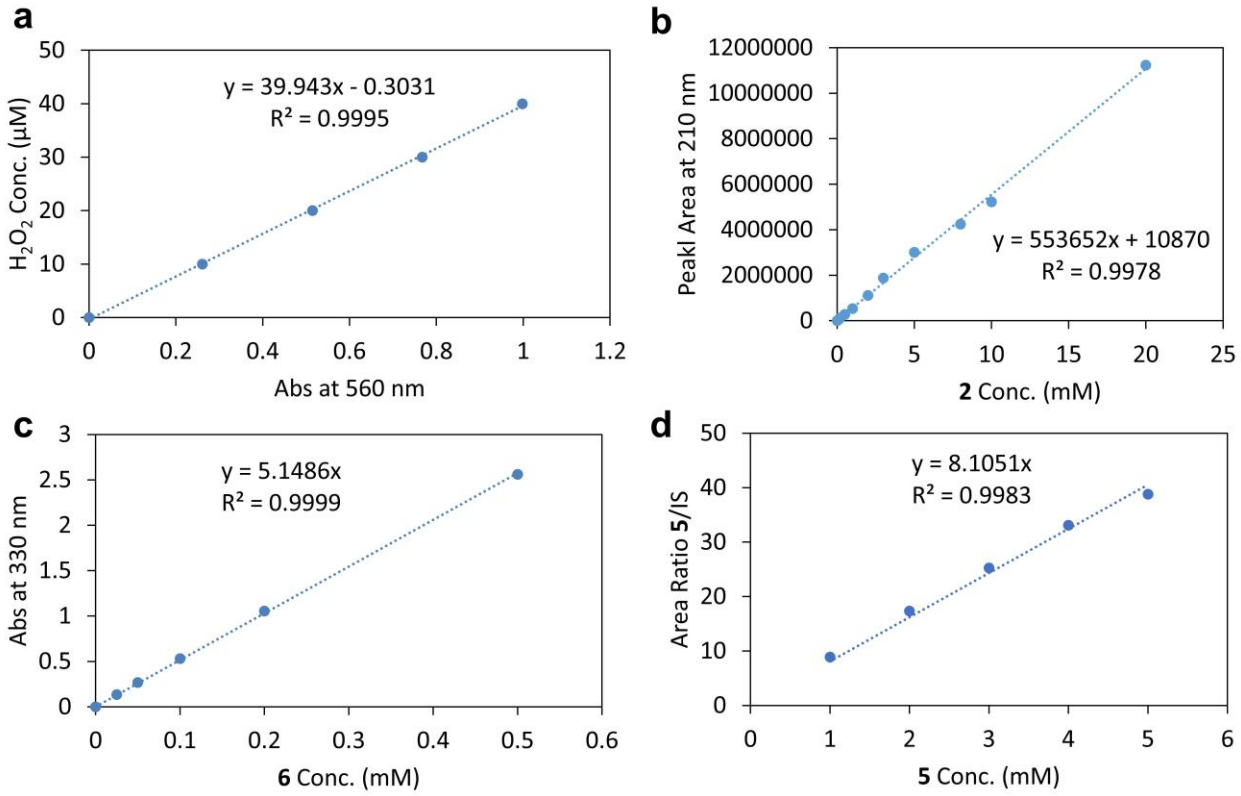
Supplementary Figure 21. Growth of *E. coli* BL21(DE3)Δ*tyrB* cells co-expressing an amino acid transporter and PcPAL libraries in M9 liquid medium with (*S*)-5 **(1 mM). a**, optical density at 600 nm (OD600) of the *E. coli* cells containing different PcPAL libraries. **b**, photo of the growth of *E. coli* cells containing different PcPAL libraries at 120 h. Line 1: *E. coli* with the ETw-PcPAL-Lib; Line 2: *E. coli* with the ETm-PcPAL-Lib; Line 3: *E. coli* with the ETs-PcPAL-Lib. **c**, Isolation of single colonies from the enriched culture of *E. coli* cells containing the ETs-PcPAL-Lib (L134-F137-L138-L256-V259-I460). The detailed procedures of the growth selection are provided in the methods of the main text. Data in **a** are from one independent experiment.



Supplementary Figure 22. Characterization of purified PcPAL variants. a, SDS-PAGE analysis of purified PcPAL variants (with a C-terminal His-tag). Lane 1: protein marker; lane 2: PcPAL(wt); lane 3: PcPAL(AI); lane 4: PcPAL(AV); lane 5: PcPAL(AVI); lane 6: PcPAL(AVV). **b**, Asymmetric synthesis of (*S*)-**5** from **6** by *E. coli* cells expressing PcPAL variants in different reaction buffers. Reaction conditions: **6** (10 mM), *E. coli* cells (2 g dcw l^{-1}), various buffer, 30 °C, 200 rpm, 6 h. Data in **b** are mean values of triplicate experiments with error bars indicating the s. d. (n = 3).



Supplementary Figure 23. Chiral HPLC chromatograms for production of (*S*)-5 from 6 by *E. coli* cells expressing PcPAL(AI). **a**, commercial standard of (*S*)-**5**. **b**, commercial standard of (*R*)-**5**. **c**, (*S*)-**5** produced by *E. coli* cells expressing PcPAL(AI) for 2 h. Chiral HPLC analysis conditions: Daicel Crownpak CR(+) column (5 μm, 4.6 mm × 150 mm), a constant flow (1.0 ml min^{-1}) of H_2O (85%) containing trifluoracetic acid (0.1%) and methanol (15%), detection at 210 nm. The detailed procedures of the reaction are provided in the methods of the main text.



Supplementary Figure 24. Calibration curves. a, calibration curve of H_2O_2 (Ampliflu Red assay) measured using a microplate reader for the determination of the activity of AtTA and CHAO. **b**, calibration curve of **2** by HPLC. **c**, calibration curve of **6** by microplate reader. **d**, calibration curve of **5** by HPLC.

Supplementary Tables

Supplementary Table 1. Data collection and refinement statistics of AtTA(RHC) crystals.

	AtTA(RHC)-PLP[1]	AtTA(RHC)-LLP[2]
PDB code	7XG5	7XG6
Data quality		
space group	C 2 2 2_1	C 2 2 2_1
unit cell		
a/b/c [Å]	106.40/137.82/117.42	105.76/136.29/116.57
α /β /γ (°)	90/90/90	90/90/90
resolution [Å]	30.00 - 1.76 (1.80 - 1.76)	30.00 - 1.32 (1.34 - 1.32)
unique reflections	83896 (3657)	195796 (9406)
redundancy	13.2 (9.4)	7.4 (7.5)
completeness [%]	98.9 (93.3)	99.7 (98.9)
average $I/\sigma(I)$	11.1 (2.5)	15.9 (2.8)
CC 1/2	0.99 (0.86)	0.99 (0.84)
Refinement		
R_{work}[a] (95 % of data)	0.159 (0.237)	0.154 (0.208)
R_{free}[a] (5 % of data)	0.181 (0.270)	0.163 (0.231)
r.m.s.d. bonds [Å]	0.019	0.016
r.m.s.d. angles [º]	1.73	1.62
dihedral angles		
most favored [%]	98.12	98.26
allowed [%]	1.88	1.74
disallowed [%]	0	0
Average B-factor/ Number of non-hydrogen atoms		
Protein	21.6/5010	14.0/5030
Ion/ligands	16.4/32	11.8/48
solvent	34.4/899	28.0/1031

[1] AtTA(RHC) with PLP in cavity (not covalent bound Lys180) was prepared under soaking conditions.

[2] AtTA(RHC) with covalent bound LLP (PLP-Lys180) were prepared under whole structure conditions.

Values in parentheses are for the highest resolution shell.

Supplementary Table 2. List of key primers used in this study.

Name	Sequence
pRSF1623-R	TCAAATATGTATCCGCTCATGAGAC
pRSF2637-F	TCGTGCCAGCTGCATTAATGAAT
pET2022-F	GTCTCATGAGCGGATACATATTTGA
pET3983-R	ATTCATTAATGCAGCTGGCACGA
J23100S-F	cctaggtacagtgctagcTAATTTTGTTTAACTTTAATAAGGAG
J23100S-R	actgagctagccgtcaaATTCATTAATGCAGCTGG
J23106M-F	cctaggtatagtgctagcTAATTTTGTTTAACTTTAATAAGGAG
J23106M-R	actgagctagccgtaaaATTCATTAATGCAGCTGGCA
J23116W-F	ggactatgctagcTAATTTTGTTTAACTTTAATAAGGAG
J23116W-R	ctaggactgagctagctgtcATTCATTAATGCAGCTGGC
J23109VW-F	cctagggactgtgctagcTAATTTTGTTTAACTTTAATAAG
J23109VW-R	actgagctagctgtaaaATTCATTAATGCAGCTGGC
AtTAopt-F	taataaggagatatattATGGCATCAATGGATAAGGTTTTTGCCGGTTATG
AtTA-R	gtttctttaccagactcgagTCAATGATGATGATGATGATGGG
Duet-TT1stATG-F	GTTTAACTTTAATAAGGAGATATATTATG
Duet-TT1stATG-R	CATAATATATCTCCTTATTAAAGTTAAAC
Duet-1stATG-F	GTTTAACTTTAATAAGGAGATATACCATG
Duet-1stATG-R	CATGGTATATCTCCTTATTAAAGTTAAAC
Duet-XhoI-F	CTCGAGTCTGGTAAAGAAACCG
Duet-XhoI-R	CGGTTTCTTTACCAGACTCGAG
AtTA-6062X-F	tgaccnnkgatnnkCCGAGCGTTTGGGATG
AtTA-55X-R	gatcgctmnnCATAAAACCCTGATCCAGCAG
AtTA-117X-F	nnkCTGATTGTTACCCGTGGTC
AtTA-115X-R	cacmnnTGCATCACGAATACCGC
AtTA-149X-F	tggnnkATGGAACCGGATATGCAG
AtTA-147X-R	mnnATACGGCTGCACAAACATATAC
AtTA-1847X-F	nnkggtgatnnkGTTCGTGGTATGTTTGAAGC
AtTA-182X-R	ctgmnnATTTTTAACGGTCGGATCAATTG
AtTA-2756X-F	nnknnkGGCGGTATTATGCCGATTAC
AtTA-274X-R	mnnACACATAAAAATTTCATCACAGCG
CHAO-F	taataaggagatataccATGACTCACCTGAATACCTATGAG
CHAO-R	gtttctttaccagactcgagTTAGTGGTGGTGATGGTGATG
ET-dBsaI-F	ttGGTCTCagactCCGCAAGTGGCACTTTTCG
ET-dBsaI-R	ttGGTCTCaagtcCCGGTCGTCAGCTTGTC
CHAO-88-R	ttGGTCTCaCATACCACCGTAATCAATTGGAC
CHAO-88-F	ttGGTCTCaTATGdbsATCGGTGAGACGCATACCC
CHAO-3513-R	ttGGTCTCaCACTCCCACATCCTCGGTATC
CHAO-3513-F	ttGGTCTCaAGTGdbsTTAdbsGACGGAACTAAGCCCACCG
CHAO-368-R	ttGGTCTCaGCCGATGAGAGTGGCGAG
CHAO-368-F	ttGGTCTCaCGGCdbsATTGGCGGCAGCAACTAC
CHAO-1989-R	ttGGTCTCaATTGACGATTACGGTGTGAAG
CHAO-1989-F	ttGGTCTCaCAATdbsdbsTTAGGCGCTGATCCTTACG
CHAO-226-R	ttGGTCTCaGAGTGACTGGATGCCCTCAC

CHAO-226-F	ttGGTCTCaACTCdbsGGGACGCGCGACGGTG
CHAO-321-R	ttGGTCTCaACGACCCATCGGGGCGCG
CHAO-321-F	ttGGTCTCaTCGTdbsTATAAAGTCCAAGCACGCTACC
PcPAL-F	taataaggagatataccATGGAAAACGGTAACGGCGC
PcPAL-His-R	ttagtggtggtgatggtgatgGCAGATCGGCAGCGGAG
Duet-His-XhoI-F	catcaccatcaccaccactaaCTCGAGTCTGGTAAAGAAAC
PcPAL-704S-R	ccattccaggattccagggATTCCAGCAGCGGGTCGA
PcPAL-716S-F	tgctccgctgccgatctcCCATCACCATCACCACCAC
PcPAL-1378-F	GCdycdyaAACGCTGGTATCTTCGGTAAC
PcPAL-134-R	GGATtrhTTCTTTCTGCAGTGCACCC
PcPAL-259-F	CTGrycAACGGTACCGCTGTTGG
PcPAL-256-R	TGCtrhGCCCTCTTTCGGTTGCAG
PcPAL-460-F	CTATGGCTTCCTACTGTTCTGAAC
PcPAL-460-R	CgryTTCTGCACCTTTGAAACCATAGTC
EcKatE-F	taataaggagatataccATGTCGCAACATAACGAAAAGAAC
EcKatE-R	gtttctttaccagactcgagTTAGGCAGGAATTTTGTCAATCTTAG
EcAroP-F	taataaggagatataccATGATGGAAGGTCAACAGCAC
EcAroP-R	tctttaccagactcgagTTAATGCGCTTTTACGGCTTTG
tyrB-Up-F	gagtcgacctgcagaagcttGTAAGCGAACGTGATACCCGTC
tyrB-Up-R	ctttcctgcaGCGATGGTTCTCCAGGTTTACG
tyrB-Down-F	gaaccatcgcTGCAGGAAAGCAGGCTGGAG
tyrB-Down-R	gagctgcacatgaactcgagCACGCTTTGCTGTTTTGCCGAG
pTarget-F	CTCGAGTTCATGTGCAGCTC
pTarget-R	AAGCTTCTGCAGGTCGACTC
dTyrB-gRNA-F	ggcggctccgGTTTTAGAGCTAGAAATAGCAAGTT
dTyrB-gRNA-R	aagggtttgaACTAGTATTATACCTAGGACTGAGC

Supplementary Methods

Chemicals and Materials

All chemicals were purchased from commercial suppliers and used without further purification. The key chemicals are listed below.

Chemicals from BLD pharma purchased from ChemPur: (*R*)-1-*Boc*-3-aminopiperidine (*R*)-**1** (97%, CAS 188111-79-7), (*S*)-1-*Boc*-3-aminopiperidine (*S*)-**1** (97%, CAS 625471-18-3), (*R*)-1-(1-naphthyl)ethylamine (*R*)-**3** (98%, CAS 3886-70-2), (*S*)-1-(1-naphthyl)ethylamine (*S*)-**3** (98%, CAS 10420-89-0), 3-(1-naphthyl)-L-alanine (*S*)-**5** (97%, CAS 55516-54-6), 3-(1-naphthyl)-D-alanine (*R*)-**5** (97%, CAS 78306-92-0), D-alanine (97%, CAS 338-69-2), borane-ammonia complex (98%, CAS 13774-81-7).

Chemicals purchased from Enamine: 1-*Boc*-3-piperidinone **2** (95%, CAS 98977-36-7), (2*E*)-3-(naphthalen-1-yl)prop-2-enoic acid **6** (95%, CAS 2006-14-6).

Chemicals from Acros Organics: 1-(1-naphthyl)ethylamine *rac*-**3** (98%, CAS 42882-31-5), cyclohexylamine (98%, CAS 108-91-8), cyclopentylamine (98%, CAS 1003-03-8).

Chemicals from ABCR: (*R*)-1-phenylethylamine (98%, CAS 3886-69-9).

Chemicals from Sigma-Aldrich: 1-acetonaphthone (97%, CAS 941-98-0), methyl laurate (99%, CAS 111-82-0), L-phenylalanine (99%, CAS 63-91-2), Ampliflu™ Red (98%, CAS 119171-73-2), horseradish peroxidase (HRP).

T4 DNA ligase and DNA primers were purchased from Thermo Fisher.

Q5 high fidelity DNA polymerase, KLD Enzyme mix, and BsaI-HF v2 (for Golden Gate mutagenesis), were purchased from New England Biolabs.

The plasmid miniprep kit and the gel extraction kit were bought from Macherey-Nagel.

DNA sequencings were performed by Eurofins.

Culture media

Medium for routine molecular biology works of *E. coli*: LB (Lysogeny broth) medium containing tryptone (10 g l^{-1}), NaCl (5 g l^{-1}) and yeast extract (5 g l^{-1}).

Medium for expression of enzymes in *E. coli* for purification of enzymes and whole-cell reactions: TB (Terrific broth) medium containing tryptone (12 g l^{-1}), yeast extract (24 g l^{-1}), glycerol (4 ml l^{-1}), K_2HPO_4 (12.5 g l^{-1}), and KH_2PO_4 (2.3 g l^{-1}).

Medium for growth selection of active amine enzymes in *E. coli* cells: modified M9 medium containing Na_2HPO_4 (6 g l^{-1}), KH_2PO_4 (3 g l^{-1}), NaCl (0.5 g l^{-1}), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), biotin (1 mg l^{-1}), thiamine (1 mg l^{-1}), trace metal solution (1 ml l^{-1} of 1000× stock solution, see below), glucose (4 g l^{-1}, carbon source), and the specific nitrogen source (1-10 mM).

Trace metal solution (1000× stock solution) containing $FeCl_3$ (0.5 g), $ZnSO_4 \cdot 7H_2O$ (0.15 g), $MnCl_2 \cdot 4H_2O$ (50 mg), $CuCl_2 \cdot 2H_2O$ (20 mg), $CoCl_2 \cdot 6H_2O$ (30 mg), $Na_2MoO_4 \cdot 2H_2O$ (30 mg), and H_3BO_3 (30 mg) in HCl solution (0.1 M, 50 ml).

Additional agar (15 g l^{-1}) was included for the preparation of solid medium for agar plates.

Appropriate amount of antibiotics was used: kanamycin (50 mg l^{-1}) and ampicillin (25 mg l^{-1}) for cultures in LB or TB medium. The concentration was reduced to half during the selection in M9 medium.

DNA Sequences

The AtTA gene was reported in our previous report.[1] It was engineered to add a C-terminal His-tag.

DNA sequence:

>AtTA

ATGGCATCAATGGATAAGGTTTTTGCCGGTTATGCAGCACGTCAGGCAATTCTGGAAAGCACCGAAACCACCAATC
CGTTTGCAAAAGGTATTGCATGGGTTGAAGGTGAACTGGTTCCGCTGGCAGAAGCACGTATTCCGCTGCTGGATC
AGGGTTTTATGCATAGCGATCTGACCTATGATGTTCCGAGCGTTTGGGATGGTCGTTTTTTTCGTCTGGATGATCAT
ATTACCCGTCTGGAAGCCAGCTGTACCAAACTGCGTCTGCGTCTGCCGCTGCCTCGTGATCAGGTTAAACAAATTC
TGGTTGAAATGGTTGCCAAAAGCGGTATTCGTGATGCATTTGTGGAACTGATTGTTACCCGTGGTCTGAAAGGTGT
TCGTGGCACCCGTCCGGAAGATATCGTGAATAATCTGTATATGTTTGTGCAGCCGTATGTTTGGGTTATGGAACCG
GATATGCAGCGTGTTGGTGGTAGCGCAGTTGTTGCACGTACCGTTCGTCGTGTTCCGCCTGGTGCAATTGATCCGA

CCGTTAAAAATCTGCAGTGGGGTGATCTGGTTCGTGGTATGTTTGAAGCAGCAGATCGTGGTGCAACCTATCCGTT
TCTGACCGATGGTGATGCACATCTGACCGAAGGTAGCGGTTTTAACATTGTGCTGGTGAAAGATGGTGTTCTGTAT
ACACCGGATCGTGGTGTTCTGCAGGGTGTTACACGTAAAAGCGTGATTAATGCAGCAGAAGCCTTTGGTATTGAA
GTGCGTGTTGAATTTGTTCCGGTTGAACTGGCATATCGCTGTGATGAAATTTTTATGTGTACCACCGCAGGCGGTA
TTATGCCGATTACCACCCTGGATGGTATGCCGGTTAATGGTGGTCAGATTGGTCCGATTACCAAAAAAATTTGGGA
TGGCTATTGGGCAATGCATTATGATGCAGCCTATAGCTTTGAAATTGATTATAATGAACGCAATTCAGGATCCCAT
CATCATCATCATCATTGA

The CHAO gene originating from *Brevibacterium oxydans* IH-35A[2] was codon-optimized for *E. coli*, a C-terminal His-tag was added, and it was synthesized by Twist Bioscience. Optimized DNA sequence:

>CHAO

ATGACTCACCTGAATACCTATGAGAGTGTGACCCCGGACCCGGATGTAGATGTAATTATTATTGGCGCAGGCATCT
CCGGTAGCGCAGCGGCTAAGGCGCTGCACGACCAGGGAGCCAGCGTCCTGGTTGTTGAGGCGAATGACCGTATT
GGCGGACGTACGTGGACCGAGCAAGAGGGAGCACCCGGTGGTCCAATTGATTACGGTGGTATGTTCATCGGTGA
GACGCATACCCATCTGATTGAGCTTGGCACGTCTCTTGGTCTGGAGATGACGCCATCGGGGAAGCCAGGAGATGA
TACTTATATTGTTGCAGGGAATGTTCTGCGTGCACCAGACGACCAACTCGACCCTAATCTCCCTTTTGTGCCGGAGT
TTCTTAGCTCGTTAAAGGCCCTGGACGAGCTTGCGGATAGTGTAGGCTGGGACCAACCGTGGGCCAGTCCAAACG
CAGCCGCGTTAGACAGTAAGACCGTGGCCACCTGGCTCGCGGAAACCATCGAAAGCGAAGAGGTCCGTCGCCTTC
ACACCGTAATCGTCAATACCCTCTTAGGCGCTGATCCTTACGAGGTATCGCTCCTGTATTGGGCTTATTATGTGAGC
GAGTGTGAGGGCATCCAGTCACTCATGGGGACGCGCGACGGTGCACAATGGGCATGGTGGTTCGGTGGTGCCGC
GCAAGTGAGTTGGCGTATTGCTGACGCAATTGGGCGCGATAAGTTCTTACTGGAGTGGCCGGTAGACCGCATCGA
GCACGACGAGAGCGGGGTGACATTGTTCTCAGGGCAACGCTCACTGCGCGCTCGTCACATTGTGATTGCAATGAG
CCCTCTTGCGGCGAACCAGATCCGCTTCGAGCCGGCCCTGCCGACTTCTCGTGCCCAGCTGCAAGCACGCGCCCCG
ATGGGTCGTTATTATAAAGTCCAAGCACGCTACCCATCATCCTTTTGGGTAGAGCAAGGTTATTCGGGCGCGTTGC
TGGATACCGAGGATGTGGGAGTGTTCTTACTGGACGGAACTAAGCCCACCGACACCCTCGCCACTCTCATCGGCTT
TATTGGCGGCAGCAACTACGATCGTTGGGCGGCTCATACACCTCAAGAGCGTGAGCGCGCATTTCTTGACTTGTTA
GTGAAGGCATTCGGGCCACAAGCGGCCGACCCTTCATACTTTCACGAAACCGACTGGACTCAGCAAGAGTGGGCC
AAGGGCGGTCCAGTCACTTATATGCCGCCTGGAGTTTTGGCGAATTTCGGTGCCGCGCTCCGCGATCCAGTCGGC
AAGGTCCATTTTGCGGGTACTGAGGCATCATTCCAATGGTCCGGCTACATGGAAGGTGGCGTCCGCGCCGGCCAA
AAGGCCGCCGCGGCGATTGCGGAAGAACTCGAACGCACGGCGAACAAGGGAGCGTTAGTTCATCACCATCACCA
CCACTAA

The PcPAL gene was described in our previous report.[3] A C-terminal His-tag was added. DNA sequence:

>PcPAL

ATGGAAAACGGTAACGGCGCTACCACTAACGGTCACGTGAACGGCAACGGTATGGACTTCTGCATGAAAACCGAA
GATCCTCTGTACTGGGGCATCGCTGCGGAGGCTATGACTGGTTCCCACCTGGACGAAGTTAAAAAGATGGTTGCT
GAATATCGTAAACCGGTTGTTAAACTGGGTGGCGAAACTCTGACCATCTCCCAAGTTGCTGCAATCTCTGCTCGTG
ACGGTTCCGGTGTTACTGTTGAACTGTCCGAAGCTGCGCGTGCTGGTGTTAAAGCGTCCTCTGACTGGGTTATGGA
CTCCATGAACAAAGGTACCGACTCCTATGGCGTTACCACTGGTTTCGGCGCTACCTCCCATCGTCGTACCAAACAG
GGGGGTGCACTGCAGAAAGAACTGATCCGCTTCCTGAACGCTGGTATCTTCGGTAACGGTTCTGACAATACGCTG
CCGCATTCCGCTACCCGTGCTGCTATGCTGGTTCGTATCAACACCCTGCTGCAAGGTTACTCTGGTATCCGTTTCGA
AATCCTGGAGGCTATCACGAAATTCCTGAACCAGAACATCACCCCGTGCCTGCCGCTGCGTGGTACCATCACTGCT
TCCGGCGACCTGGTTCCACTGTCCTACATCGCTGGTCTGCTGACTGGTCGTCCGAACTCTAAAGCTGTTGGTCCGAC
TGGTGTTATCCTGTCCCCGGAAGAAGCGTTCAAACTGGCTGGTGTGGAAGGTGGTTTCTTTGAACTGCAACCGAA
AGAGGGCCTGGCACTGGTTAACGGTACCGCTGTTGGTTCTGGTATGGCGTCCATGGTTCTGTTCGAAGCTAACATC
CTGGCTGTTCTGGCGGAAGTGATGTCTGCTATCTTCGCTGAAGTTATGCAGGGTAAACCAGAGTTCACCGACCACC
TGACTCACAAACTGAAACACCACCCGGGTCAGATCGAAGCTGCTGCTATCATGGAACACATCCTGGACGGTTCTGC
CTACGTTAAAGCTGCTCAGAAACTGCACGAAATGGACCCGCTGCAAAAACCGAAACAGGACCGTTATGCTCTGCG
TACCTCTCCACAGTGGCTGGGCCCGCAAATCGAAGTTATCCGCTCCTCTACCAAGATGATCGAACGTGAAATCAAC
TCTGTTAACGACAACCCGCTGATCGACGTTTCCCGCAACAAAGCTATCCACGGTGGTAACTTCCAGGGGACCCCGA
TCGGCGTTTCCATGGACAACACCCGTCTGGCTATCGCAGCTATCGGTAAACTGATGTTCGCTCAATTCTCTGAACTG
GTTAACGACTTCTACAACAACGGTCTGCCATCTAACCTGTCTGGTGGTCGTAACCCGTCCCTGGACTATGGTTTCAA
AGGTGCAGAAATCGCTATGGCTTCCTACTGTTCTGAACTGCAATTCCTGGCTAACCCGGTTACCAACCACGTTCAGT
CCGCAGAACAGCACAACCAAGACGTTAACTCTCTGGGTCTGATCTCTTCTCGTAAAACCTCTGAAGCTGTTGAAAT
CCTGAAACTGATGTCCACTACCTTCCTGGTTGGTCTGTGTCAAGCTATCGACCTGCGTCACCTGGAAGAAAACCTG
AAATCCACCGTTAAAAACACCGTGTCTTCCGTGGCTAAACGTGTTCTGACGATGGGTGTTAATGGAGAACTGCACC
CGTCCCGTTTCTGCGAAAAAGACCTGCTGCGTGTTGTCGACCGTGAATACATCTTTGCTTACATCGACGACCCGTGC
TCCGCTACCTACCCACTGATGCAGAAACTGCGTCAGACCCTGGTTGAACATGCTCTGAAAAACGGTGACAACGAAC
GTAACCTGTCTACCTCCATCTTCCAGAAAATTGCAACCTTCGAAGATGAACTGAAAGCTCTGCTGCCGAAAGAAGT
TGAATCCGCTCGTGCAGCACTGGAATCTGGTAACCCTGCTATCCCAAACCGTATCGAAGAATGCCGTTCCTACCCG
CTGTACAAATTCGTTCGTAAAGAACTGGGCACTGAATACCTGACCGGTGAAAAAGTTACCTCCCCAGGTGAAGAG
TTCGAAAAAGTTTTCATCGCTATGTCCAAAGGTGAAATCATCGACCCGCTGCTGGAATGCCTGGAATCCTGGAATG
GTGCTCCGCTGCCGATCTGC CATCACCATCACCACCACTAA

General Methods for Molecular Cloning, Transformation, and Engineering of Vectors

In general, most of the molecular cloning was performed by using the SLiCE method[4] (summarized below) and the mutagenesis was performed by using the Q5 mutagenesis method with KLD treatment or Golden Gate Mutagenesis[5] (see specific sections for details). For the SLiCE method, the SLiCE extract of *E. coli* TOP10 strain was prepared by the following protocol. The *E. coli* TOP10 cells were grown in 2× YT medium (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl) until OD_{600} reached 3.0. Then, the cells were harvested (5000 g, 10 min) and washed with ice-cold milli-Q water (5000 g, 15 min). The washed cell pellet was weighed and resuspended in CelLytic B Cell lysis reagent (Sigma, 0.3 ml per 0.25 g of pellet). It was lysed at room temperature for 10 min. Then, the clarified lysates (16000 g, 5 min) were mixed with glycerol (1:1, v/v) and stored at -80 ºC. The SLiCE buffer (10×) was prepared by dissolving $MgCl_2$ (0.1 M), ATP (10 mM), and DTT (10 mM) in Tri-HCl buffer (0.5 M, pH 7.5), and stored at -20 ºC. The SLiCE reaction (total volume 5 μl) was performed with two (or more) purified DNA fragments with 20 bp overlapping sequences (4 μl), SLiCE buffer (10×, 0.5 μl), and SLiCE extract (0.5 μl) at 37 ºC for 1 h. The SLiCE product was introduced into competent cells of *E. coli* by heat shock.

All PCRs were performed with Q5 DNA polymerase (New England Biolabs) according to the instruction. Purification of DNA fragments was done with a Gel Extraction Kit (Macherey-Nagel) according to the instruction. The sequences of primers are provided in Supplementary Table 2.

Transformation methods. The standard heat-shock transformation was used for most of the cloning works and for building small libraries (e.g., for the first 2 rounds of Q5 mutagenesis of PcPAL). Large libraries were introduced into *E. coli* cells by electroporation with an electroporation cuvette (1 mm gap) and default *E. coli* settings (1.8 kV) for the MicroPulser Electroporator (Biorad). Immediately after the electroporation shock, prewarmed LB medium (2 ml, 37 ºC) was added to the cuvette and the cells were cultured in an incubator for 2 h at 30 ºC, 200 rpm. A small portion of the cells (20 μl) was mixed with fresh LB medium (980 μl) and one tenth of the mixture (100 μl, corresponds to 1/1000 of the whole electroporation) was spread on an LB agar plate to test the electroporation efficiency. Half of the cells were added into LB medium (5 ml) containing kanamycin (50 mg l^{-1}) for isolation of the plasmids from the library. The other half of the cells were used for growth selection.

Genetic engineering of the vectors ETs/m/w/vw. The vectors ETs/m/w/vw were constructed by amplifying the plasmid backbone of pRSFduet-1 (Novagen) by using primers pRSF1623-R and pRSF2637-F and amplifying the pBR322ori of pETduet-1 (Novagen) by using primers pET2022-F and pET3983-R. These two DNA fragments were purified and assembled by the SLiCE method to get an intermediate plasmid with pBR322ori, a kanamycin-resistant gene, and the expressing cassette of pRSFduet-1. To get

the ETs vector, the strong constitutive promoter J23100 was introduced by amplifying the intermediate plasmid by using primers J23100S-F and J23100S-R (which replaced the 1st T7 promoter with J23100), and the PCR product was treated with KLD Enzyme mix (New England Biolabs) and introduced into *E. coli* cells. Similarly, ETm/w/vw vectors were engineered by PCR with J23106M-F/J23106M-R, J23116W-F/J23116W-R, and J23109VW-F/J23109VW-R, respectively, and followed with KLD treatment and transformation.

Genetic Engineering of AtTA

Genetic engineering of ETs/m/w/vw-AtTA. The AtTA gene was amplified from pGASTON-AtTA[1] by using primers AtTAopt-F and AtTA-R. The vectors ETs/m/w/vw were amplified by using the primers Duet-TT1stATG-R and Duet-XhoI-F. The purified AtTA fragment was assembled with the purified vector fragments by the SLiCE method to get the plasmids ETs/m/w/vw-AtTA.

Genetic construction of AtTA libraries. To construct the ETs-AtTA-Lib1 (H55-Y60-V62), the plasmid ETs-AtTA was amplified by using primers AtTA-6062X-F and AtTA-55X-R, and the PCR product was treated with the KLD Enzyme mix according to the instruction, purified and then introduced into the electrocompetent cells of *E. coli* BL21(DE3) by electroporation. The efficiency of electroporation was checked to ensure a full coverage of the library size. Half of the resulting cells were cultured for preparation of the plasmids of the library (its quality was checked by sequencing), and another half of the cells were subjected to growth selection (see Methods in the main text). The other AtTA libraries were constructed similarly: Lib2 (F115-E117) using primers AtTA-117X-F and AtTA-115X-R, Lib3 (V147-V149) using primers AtTA-149X-F and AtTA-147X-R, Lib4 (L182-W184-L187) using primers AtTA-1847X-F and AtTA-182X-R, Lib5 (T274-T275-A276) using primers AtTA-2756X-F and AtTA-274X-R.

Subcloning of AtTA variants. The genes encoding the AtTA variants (such as YHC) were subcloned to ETw and ETvw by the following procedures: the gene of AtTA(YHC) was amplified by using the primers Duet-TT1stATG-F and Duet-XhoI-R, and the targeted vector ETw was amplified by using primers Duet-TT1stATG-R and Duet-XhoI-F. The purified AtTA(YHC) fragment was assembled with the purified vector fragment by the SLiCE method to get the plasmid ETw-AtTA(YHC). Other subcloning of AtTA variants was performed similarly.

Genetic Engineering of CHAO

Genetic engineering of ETs/m/w/vw-CHAO. The gene encoding CHAO was amplified from the synthesized gene fragment by using the primers CHAO-F and CHAO-R. The vectors ETs/m/w/vw were amplified by using the primers Duet-1stATG-R and Duet-XhoI-F. The CHAO fragment was assembled with the vector fragments by the SLiCE method to get the plasmids ETs/m/w/vw-CHAO.

Genetic construction of CHAO libraries by Golden Gate Mutagenesis. The procedure of Golden Gate Mutagenesis follows a previous report[5] with some modifications. The BsaI restriction site inside the vectors ETm/w/vw-CHAO was first removed (termed domestication) by amplification of the plasmids by using the primers ET-dBsaI-F and ET-dBsaI-R. The PCR products were purified and subjected to a 40-cycle of digestion by BsaI-HFv2 (37 ºC, 2 min) and ligation by T4 ligase (20 ºC, 5 min) in a 15-μl system of T4 ligase buffer containing BsaI-HFv2 (0.5 μl) and T4 ligase (1 μl). The mixture was subjected to enzyme inactivation at 80 ºC for 20 min and then introduced into the competent cells of *E. coli* BL21(DE3) to get the ETm/w/vw-CHAO without BsaI restriction site. These plasmids were used as the templates to construct the CHAO libraries. To construct ETm-CHAO-Lib1 (F88-F351-L353-F368), three DNA fragments were amplified from ETm-CHAO by using three pairs of primers, CHAO-88-F & CHAO-3513-R (fragment 1), CHAO-3513-F & CHAO-368-R (fragment 2), and CHAO-368-F & CHAO-88-R (fragment 3). The three fragments were purified and subjected to a 100-cycle of digestion by BsaI-HFv2 (37 ºC, 2 min) and ligation by T4 ligase (20 ºC, 5 min) in a 20-μl system of T4 ligase buffer containing BsaI-HFv2 (0.5 μl) and T4 ligase (1 μl). The mixture was subjected to enzyme inactivation at 80 ºC for 20 min and then introduced into the electrocompetent cells of *E. coli* BL21(DE3) SCm-KatE by electroporation. The efficiency of electroporation was checked to ensure a full coverage of the library size. Half of the resulting cells were cultured for preparation of the plasmids of the library (its quality was checked by sequencing), and another half of the cells were subjected to growth selection (see Methods in the main text). The other CHAO library was constructed similarly: Lib2 (T198-L199-M226-Y321) using primers CHAO-1989-F & CHAO-226-R (fragment 1), CHAO-226-F & CHAO-321-R (fragment 2), and CHAO-321-F & CHAO-1989-R (fragment 3).

Subcloning of CHAO variants. The genes of CHAO variants (such as MVLM) were subcloned to pRSFduet-1 by the following procedures: the gene encoding CHAO(MVLM) was amplified by using the primers Duet-1stATG-F and Duet-XhoI-R, and the targeted vector pRSFduet-1 was amplified by using the primers Duet-1stATG-R and Duet-XhoI-F. The purified CHAO(MVLM) fragment was assembled with the purified vector fragment by the SLiCE method to get the plasmid pRSF-CHAO(MVLM). Other subcloning of CHAO variants was performed similarly.

Genetic Engineering of PcPAL

Genetic engineering of ETs/m/w-PcPAL. The gene encoding PcPAL was amplified from the PcPAL plasmid[3] by using the primers PcPAL-F and PcPAL-His-R. The vectors ETs/m/w were amplified by using the primers Duet-1stATG-R and Duet-His-XhoI-F. The purified PcPAL fragment was assembled with the purified vector fragments by the SLiCE method to get the plasmids ETs/m/w-PcPAL. A previous study[6] showed that the C704S and C716S mutations increased the stability of PcPAL, thus C704S/C716S were introduced by Q5 mutagenesis using the primers PcPAL-704S-R and PcPAL-716S-F. The resulting ETs/m/w-PcPAL(C704S/C716S) were used for all the following studies.

Genetic construction of PcPAL libraries. The ETs-PcPAL-Lib (L134-F137-L138-L256-V259-I460) was constructed by sequentially performed Q5 mutagenesis rounds. The I460 was firstly introduced by amplifying the plasmid ETs-PcPAL using the primers PcPAL-460-F and PcPAL-460-R, and the PCR product was treated with the KLD Enzyme mix and then introduced into the competent cells of *E. coli* BL21(DE3). Using the resulting library ETs-PcPAL(I460) as the template, L256-V259 was then introduced by Q5 mutagenesis using the primers PcPAL-259-F and PcPAL-256-R. The efficiency of the heat-shock transformation was checked to ensure a full coverage of the library size. The resulting library of ETs-PcPAL(L256-V259-I460) was used as the template to further incorporate L134-F137-L138 by Q5 mutagenesis using the primers PcPAL-1378-F and PcPAL-134-R. The resulted library was introduced into the electrocompetent cells of *E. coli* BL21(DE3) Δ*tyrB* SCm-AroP by electroporation. The efficiency of electroporation was checked was checked to ensure a full coverage of the library size. Half of the resulting cells were cultured for preparation of the plasmids of the library (its quality was checked by sequencing), and another half of the cells were subjected to growth selection (see Methods in the main text). The other PcPAL libraries were constructed similarly.

Subcloning of PcPAL variants. The genes of PcPAL variants (such as AI) were subcloned to pRSFduet-1 by the following procedures: the gene of PcPAL(AI) was amplified by using the primers Duet-1stATG-F and Duet-XhoI-R, and the targeted vector pRSFduet-1 was amplified by using the primers Duet-1stATG-R and Duet-XhoI-F. The purified PcPAL(AI) fragment was assembled with the purified vector fragment by the SLiCE method to get the plasmid pRSF-PcPAL(AI). Other subcloning of PcPAL variants was performed similarly.

Genetic Engineering of Assisting Plasmids

The assisting plasmid SCm-KatE (expressing *E. coli* catalase) was constructed by the following procedures. The KatE gene was amplified from the genome of *E. coli* BL21(DE3) by using primers EcKatE-F and EcKatE-R. The vector SCm (with SC101ori, ampicillin resistant gene, and the constitutive promoter with medium strength) was amplified by using the primers Duet-1stATG-R and Duet-XhoI-F. The purified KatE fragment was assembled with the purified vector fragment by the SLiCE method to get the plasmid SCm-KatE.

The assisting plasmid SCm-AroP (expressing *E. coli* aromatic amino acid transporter) was constructed by the following procedures. The AroP gene was amplified from the genome of *E. coli* BL21(DE3) by using the primers EcAroP-F and EcAroP-R. The vector SCm (with SC101ori, ampicillin-resistant gene, and the constitutive promoter with medium strength) was amplified by using the primers Duet-1stATG-R and Duet-XhoI-F. The purified AroP fragment was assembled with the purified vector fragment by the SLiCE method to get the plasmid SCm-AroP.

Genome Engineering of *E. coli* BL21(DE3) Δ*tyrB*

The *E. coli* BL21(DE3)Δ*tyrB* strain was engineered following a previously reported CRISPR-Cas9 method[7] with two key plasmids pCas (Addgene #62225) and pTargetF (Addgene #62226). pTarget-dTryB plasmid was engineered by first engineering the flanking sequences of *tyrB* gene by assembling three DNA fragments: tyrB-Up fragment (amplified from *E. coli* genome by using the primers tyrB-Up-F and tyrB-Up-R), tyrB-Down fragment (amplified from *E. coli* genome by using the primers tyrB-Down-F and tyrB-Down-R), and pTarget fragment (amplified by using the primers pTarget-R and pTarget-F). Then, the gRNA was introduced by using Q5 mutagenesis with the primers dTyrB-gRNA-F and dTyrB-gRNA-R. The resulting pTarget-dTyrB plasmid was sequenced to confirm the existence of gRNA and flanking sequences of *tyrB* gene. *E. coli* BL21(DE3)Δ*tyrB* strain was engineered by using pCas and pTarget-dTyrB according to the previously reported procedure,[7] briefly outlined: 1) pCas was introduced into the *E. coli* BL21(DE3) cells, 2) electrocompetent cells of *E. coli* BL21(DE3) pCas were prepared with an arabinose-induced λ-Red system, 3) pTarget-dTyrB was introduced by electroporation, 4) selection was performed on LB plates with kanamycin and streptomycin, 5) perform colony PCR was performed for genotyping the colonies, 6) the plasmid pTarget-dTyrB (by using IPTG) and pCas (by culturing at 37 °C) were sequentially cured. The resulting *E. coli* BL21(DE3)*ΔtyrB* strain was confirmed by PCR amplification and sequencing of the genome sequence flanking the *tyrB* gene.

Expression of Enzymes and Preparation of Whole-cell Catalysts

The plasmids (ETs/ETm/pRSFduet-1) containing the genes of AtTA/CHAO/PcPAL variants were transformed into *E. coli* BL21(DE3) cells, and the resulting cells were incubated in a culture tube containing LB medium (5 ml) with kanamycin (50 mg ml^{-1}) at 37 ºC, 180 rpm for 6-8 h. The culture was transferred into a flask (500 ml) containing TB medium (100 ml) with kanamycin (50 mg ml^{-1}) and incubated at 37 ºC, 180 rpm. When the OD_{600} of the culture reached 0.6, the culture was incubated at 20 ºC (AtTA) or 22 ºC (CHAO and PcPAL), 180 rpm for another 12-14 h. For the pRSFduet-1 plasmid, IPTG (0.5 mM, isopropyl β-D-thiogalactopyranoside) was added (at OD_{600} = 0.6) to induce the expression of enzymes. For the ETs or ETm plasmid, no IPTG is needed. The cells were harvested by centrifugation (4000 g, 20 min).

Purification of Enzymes

For purification of AtTA variants, the cell pellets were resuspended in HEPES buffer (50 mM, pH 7.5) containing NaCl (300 mM), PLP (0.1 mM), imidazole (10 mM), and then lysed by ultrasonication (50% power, 50% pulse, 5 min for two times). The lysate was clarified by centrifugation (10000 g, 4 ºC, 1 h) and purified by affinity chromatography (Ni-NTA agarose) with the following buffers: washing buffer containing HEPES (50 mM, pH 7.5), PLP (0.1 mM), NaCl (300 mM), and imidazole (20 mM), and elution buffer containing HEPES (50 mM, pH 7.5), PLP (0.1 mM), NaCl (300 mM), and imidazole (300 mM). The AtTA variants were desalted in HEPES buffer (50 mM pH 7.5) with PLP (0.1 mM) using the PD-10 desalting column (GE Healthcare). The purified AtTA variants were stored in 30% glycerol at -20 ºC.

For purification of CHAO and PcPAL variants, the same procedure was employed with some difference in the buffer: sodium phosphate buffer (50 mM, pH 7.5) without PLP was used instead of HEPES buffer (50 mM, pH 7.5) with PLP (0.1 mM).

For purification of AtTA(RHC) for crystallization, the cells were resuspended in lysis buffer containing HEPES (25 mM, pH 7.0), NaCl (150 mM), PLP (0.1 mM), imidazole (20 mM), then disrupted by a French Press, cell debris was removed by centrifugation (17000 g, 4 ºC, 1 h). The supernatant was then applied onto a Ni-NTA column FPLC system (GE Health care) equilibrated with buffer (20 mM HEPES, pH 7.0, 150 mM NaCl, 0.1 mM PLP, 20 mM imidazole), target proteins were eluted at ~100 mM imidazole. Proteins were concentrated and applied to a size exclusion column (Hiload superdex 200 16/600, 120 ml, GE Healthcare) equilibrated with buffer (20 mM HEPES, pH 7.0, 0.1 mM PLP, 150 mM NaCl) at a flow rate of 1 ml min^{-1}. The purity of each purification step was checked by SDS-PAGE gel. The purified proteins

were concentrated in buffer (20 mM HEPES, pH 7.0, 0.1 mM PLP, 150 mM NaCl) for crystallization screening.

Activity Assays

The specific activities of purified AtTA variants were determined by the D-amino acid oxidase assay reported previously. The reaction was performed in microtiter plates (200 µl assay volume) with (*R*)-**1** (2.5 mM) as amine donor and pyruvate (2.5 mM) as amine acceptor, D-amino acid oxidase (0.7 U ml^{-1}) from *T. variabilis* (TvDAAO), horseradish peroxidase (HRP, 22 U ml^{-1}), Ampliflu Red (50 µM) in CHES buffer (50 mM, pH 9.0), and an appropriate amount of AtTA variants (10 µl of 1-1000× dilution). The formation of Resorufin was quantified by following the increase of absorption at 560 nm over time by a Tecan infinite 200pro plate reader.

The specific activities of purified CHAO variants were determined by the Ampliflu Red assay. The reaction was performed in microtiter plates (200 µl assay volume) with substrate (5 mM), horseradish peroxidase (HRP, 2 U per well), Ampliflu Red (50 µM) in potassium phosphate buffer (50 mM, pH 7.0), and an appropriate amount of CHAO (10 µl of 1-1000× dilution). The formation of Resorufin was quantified by following the increase of absorption at 560 nm over time by a Tecan infinite 200pro plate reader.

The specific activities of purified PcPAL variants were determined by measuring the absorbance of **6** at 330 nm. The amination reaction (**6→5**) was performed in microtiter plates (200 µl assay volume) with **6** (0.5 mM) in NH_3/NH_4Cl buffer (6 M, pH 9.5), and an appropriate amount of PcPAL (20 µl). The deamination reaction (**5→6**) was performed in microtiter plates (200 µl assay volume) with (*S*)-**5** (1 mM) in potassium phosphate buffer (10 mM, pH 7.0), and an appropriate amount of PcPAL (10 µl). The decrease/formation of **6** was quantified by following the change of absorption at 330 nm over time by a Tecan infinite 200pro plate reader.

Analytical Methods

Analysis of **1** and **2** was performed using a Hitachi Chromaster HPLC system with a Luna Omega 5 µm PS C18 100A LC column (150 mm × 4.6 mm) and the following program: 20% acetonitrile and 80% water (contains 0.1% TEA) at a constant flow rate (1.0 ml min^{-1}) and the absorbance was detected at 210 nm. Retention times: 3.4 min for **1**, 12.6 min for **2**. The compounds were quantified based on the peak areas (absorbance at 210 nm, external standard).

The *ee* of **1** was determined by chiral GC analysis using a Shimadzu GC-2010 Plus system with an FID detector. The samples of **1** (in CH_2Cl_2) were firstly derivatized to the trifluoroacetamide by adding a 20-fold excess of trifluoroacetic anhydride. The excess anhydride and residual trifluoroacetic acid were removed by nitrogen purging, and the derivatized compound was dissolved in CH_2Cl_2 and analyzed by the following program. Column: heptakis-(2,3-di-*O*-acetyl-6-*O*-*tert*-butyldimethylsilyl)-cyclodextrin (25 m × 0.25 mm). Temperature program: start at 160 °C for 10.5 min, then increase to 180 °C at 10 °C min^{-1}, hold at 180 °C for 20 min. Retention times: 12.5 min for the trifluoroacetamide of (*S*)-**1**, 13.1 min for the trifluoroacetamide of (*R*)-**1**.

The concentration and *ee* of **3** were simultaneously determined by chiral HPLC analysis on an Agilent 1200 infinity system with a Chiralcel OD-3 column (3 μm, 4.6 mm × 250 mm, Daicel) and the following program: 80% *n*-hexane and 20% isopropanol at a constant flow rate (1.0 ml min^{-1}) and the absorbance was detected at 210 nm. Retention times: 3.5 min for internal standard (ethylbenzene), 7.2 min for (*S*)-**3**, and 8.0 min for (*R*)-**3**. The compounds were quantified based on the peak areas (absorbance at 210 nm) using the internal standard.

The concentration of **5** was determined by HPLC analysis on an EasySep®-3030 HPLC system (Unimicro Technologies, Shanghai) with an Agilent Poroshell 120 SB-C18 column (2.7 μm, 4.6 mm × 100 mm) and the following program: 40% H_2O (contains 0.1% trifluoracetic acid) and 60% methanol at a constant flow rate (0.5 ml min^{-1}) and the absorbance was detected at 210 nm. Retention times: 3.5 min for **5**, 4.5 min for internal standard (acetophenone), and 12.0 min for **6**. The compounds were quantified based on the peak areas (absorbance at 210 nm) using the internal standard.

The *ee* of **5** was determined by chiral HPLC analysis on a Wooking K2025 HPLC system with a Crownpak CR(+) column (5 μm, 4.6 mm × 150 mm, Daicel) and the following program: 85% H_2O (contains 0.1% trifluoracetic acid) and 15% methanol at a constant flow rate (1.0 ml min^{-1}) and the absorbance was detected at 210 nm. Retention times: 28.1 min for (*R*)-**5**, and 36.4 min for (*S*)-**5**.

Crystallization, Data Collection, Structure Determination of AtTA(RHC)

All crystallization experiments were conducted at 25°C using the sitting-drop vapor-diffusion method. In general, 1 μl AtTA(RHC) (20 mg ml^{-1} in HEPES buffer (20 mM, pH 7.0, with 0.1 mM PLP, 150 mM NaCl)) was mixed with 1 μl of reservoir solution in 48-well Cryschem Plates, then equilibrated against 100 μl of the reservoir at 25 ˚C. The crystals of AtTA(RHC) were obtained under the following crystallization condition: 0.1M HEPES pH 7.5, 10% PEG8000, 15% ethylglycerol. Within one week, the crystals reached

dimensions suitable for X-ray diffraction. 0.1 mM PLP + 1 mM (*R*)-**1** were used in co-crystallization and soaking experiments.

All of the X-ray diffraction data were tested and collected at Beamline 14.1 at BESSY. The crystals were mounted in a cryoloop, soaked with cryoprotectant solution (0.1M HEPES pH 7.5, 10% PEG8000, 20% ethylglycerol) prior to data collection at 100 K. The diffraction images were processed by using XDS.[8] Both of the crystal structures were solved by molecular replacement (MR) method with Phaser[9] from the Phenix[10] suite using the structure of AtTA(wt) (PDB code 4CE5)[11] as the search model. The further refinement was carried out using programs phenix.refine[12] and medel building with Coot.[13] Prior to structural refinements, 5% randomly selected reflections were set aside for calculating R_{free} as a monitor. Data collection and refinement statistics are summarized in Supplementary Table 1. All figures were prepared by using the PyMOL program (https://pymol.org/2/).

Homology Modeling and Docking Experiment

If not stated otherwise, the modeling and docking experiment were mainly performed with YASARA Structure version 20.4.24.

For the *in silico* study of AtTA(RHC), the PLP-(*R*)-**1** complex (external aldimine) was created with ChemDraw and underwent energy minimization (force field: Yasara2) and the bonds and chirality of the complex were double-checked. The existing PLP in the crystal structure of AtTA(RHC) (PDB: 7XG5) was removed to prepare the structure for docking. The complex was docked into the active site of AtTA(RHC) using the AutoDock VINA implemented in YASARA Structure (force field: AMBER03). The VINA docking runs generated 25 poses (5 distinct complex conformations). The pose with the lowest binding energy shows a very similar binding of the PLP part compared to the binding of PLP in the AtTA(wt) and exhibits a hydrogen bond between the guanidinium group of H55R and the carbonyl group of (*R*)-**1**. Thus, this pose was used to create the figure with the PyMOL program.

For the *in silico* study of CHAO(MVLM), the homology model of CHAO(MVLM) was built using the structure of CHAO(wt) (PDB: 4I59)[14] as the template by YASARA. The model was refined to the optimal conformation by performing a 500 ps MD simulation. The snapshot with the lowest energy was used as the structure for docking. (*S*)-**3** was created with ChemDraw and underwent energy minimization (force field: Yasara2) and the bonds and chirality were double-checked. (*S*)-**3** was docked into the active site of CHAO(MVLM) using the AutoDock VINA implemented in YASARA Structure (force field: AMBER03). (*S*)-**3** was docked 32 times with VINA against each of the 5 receptors in the ensemble, yielding

the 160 results. A pose was found in the results with a suitable distance (3-4 Å) between the C-atom α to the amine of (*S*)-**3** and the N5 of the flavin. The pose was used to create the figure with the PyMOL program.

For the *in silico* study of PcPAL(AI), the homology model of PcPAL(AI) was built using the structure of PcPAL(wt) (PDB: 6F6T)[15] as the template by SWISS-MODEL. The GMQE score of the model was 0.89 and the QMEANDisCo Global score is 0.85 ± 0.05. **6** was created with ChemDraw and underwent energy minimization (force field: Yasara2), and the atoms and bonds were double-checked. **6** was docked into the active site of PcPAL(AI) using the AutoDock VINA implemented in YASARA Structure (force field: AMBER03). The VINA docking runs generated 25 poses (3 distinct complex conformations). The pose with the lowest binding energy shows a very similar binding of **6** compared to the binding of (*R*)-(1-amino-2-phenylethyl)phosphonic acid (inhibitor) in the PcPAL(wt). Thus, this pose was used to create the figure with the PyMOL program.

Preparative Scale Syntheses

The preparation of (*R*)-**1** followed the same procedure as the Methods of the main text. (*R*)-**1** was isolated as yellow oil (1.96 g, 98% yield, 98% *ee*). ^{1}H NMR (300 MHz, DMSO-d_6) δ: 3.83–3.79 (d, J = 10.5 Hz, 1H), 3.76–3.69 (m, 1H), 2.69–2.62 (t, J = 11.5 Hz, 1H), 2.54–2.47 (m, 1H), 2.39 (br, 1H), 1.81–1.75 (m, 1H), 1.61–1.55 (m, 1H), 1.39 (s, 9H), 2.07 (brs, 2H), 1.32–1.26 (m, 1H), 1.11–1.07 (m, 1H).

For the preparative deracemization to produce (*R*)-**3** fresh *E. coli* cells containing pRSF-CHAO(MVLM) and SCm-KatE were employed as whole-cell catalysts. The cell pellets of a 200-ml culture were first washed with Tris-HCl buffer (100 mM, pH 8.0) and subjected to centrifugation again (4000 g, 15 min). The supernatant was discarded and the cell pellets were resuspended in new Tris-HCl buffer (100 mM, pH 8.0) and the optical density of the cell suspensions was measured. To an Erlenmeyer flask (1000 ml), the cell suspensions, a stock solution of *rac*-**3** (428 mg, 500 mM HCl salt in water), a stock solution of NH_3-BH_3 (1 M in Tris-HCl buffer), Tris-HCl buffer (100 mM, pH 8.0), and *n*-dodecane (100 ml) were added to form a two-phase catalytic system composed of the aqueous buffer (100 ml) containing *rac*-**3** (25 mM), cells (10 g dcw l^{-1}) and NH_3-BH_3 (500 mM) and organic phase (100 ml). The flask was incubated at 30 °C, 200 rpm for 4 h. A sample of the reaction (100 μl) was taken, mixed with NaOH solution (5 M, 100 μl), and extracted with *n*-hexane (1 ml) to check the *ee* by chiral HPLC analysis. Then, a new batch of cell suspensions (~10 ml, from 200-ml culture) was added to the reaction mixture, and continued reaction for another 4 h (8 h in total). The *ee* reached > 99% by chiral HPLC analysis. The reaction was quenched by adding HCl solution to reach pH 1. The mixture was centrifuged (4000 g, 5 min) to separate the organic phase (*n*-dodecane), and then the aqueous phase was extracted by *n*-hexane (50 ml) three times. The

aqueous phase was basified by NaOH solution to pH 13, and extracted with *n*-hexane (100 ml) five times. The organic phase was combined and dried with Na_2SO_4. Further evaporation and drying under vacuum overnight offered (*R*)-**3** as yellowish oil (295 mg, 69% yield, > 99% *ee*). ^{1}H NMR (600 MHz, $CDCl_3$) δ: 8.15–8.13 (d, J = 8.4 Hz, 1H), 7.88–7.87 (d, J = 8.1 Hz, 1H), 7.76–7.75 (d, J = 8.1 Hz, 1H), 7.65–7.64 (d, J = 7.2 Hz, 1H), 7.54–7.47 (m, 3H), 4.97–4.94 (q, J = 6.8 Hz, 1H), 1.85 (br, 2H), 1.55–1.54 (d, J = 6.6 Hz, 3H). The NMR data are consistent with those in a previous report.[16]

Supplementary References

1. Höhne, M., Schätzle, S., Jochens, H., Robins, K. & Bornscheuer, U. T. Rational assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.* **6**, 807–813 (2010).
2. Li, G. et al. Substrate profiling of cyclohexylamine oxidase and its mutants reveals new biocatalytic potential in deracemization of racemic amines. *Appl. Microbiol. Biotechnol.* **98**, 1681–1689 (2014).
3. Bartsch, S. & Bornscheuer, U. T. A single residue influences the reaction mechanism of ammonia lyases and mutases. *Angew. Chem. Int. Ed.* **48**, 3362–3365 (2009).
4. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012).
5. Püllmann, P. et al. Golden mutagenesis: an efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **9**, 10932 (2019).
6. Nagy, E. M. Z. et al. Mapping the hydrophobic substrate binding site of phenylalanine ammonia-lyase from *Petroselinum crispum*. *ACS Catal.* **9**, 8825–8834 (2019).
7. Jiang, Y. et al. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **82**, 3693 (2016).
8. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
9. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
10. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
11. Łyskowski, A. et al. Crystal structure of an (*R*)-selective ω-transaminase from *Aspergillus terreus*. *PLoS ONE* **9**, e87350 (2014).
12. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
13. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

14. Mirza, I. A. et al. Structural analysis of a novel cyclohexylamine oxidase from *Brevibacterium oxydans* IH-35A. *PLoS ONE* **8**, e60072 (2013).
15. Bata, Z. et al. Substrate tunnel engineering aided by X-ray crystallography and functional dynamics swaps the function of MIO-enzymes. *ACS Catal.* **11**, 4538–4539 (2021).
16. Marx, L. et al. Chemoenzymatic approaches to the synthesis of the calcimimetic agent Cinacalcet employing transaminases and ketoreductases. *Adv. Synth. Catal.* **360**, 2157–2165 (2018).