ACS Catalysis

Research Article

Cite This: ACS Catal. 2018, 8, 11524−11533

pubs.acs.org/acscatalysis

In Silico Based Engineering Approach to Improve Transaminases for the Conversion of Bulky Substrates

Moritz Voss,[†] Devashish Das,[§] Maika Genz,[†] Anurag Kumar,[§] Naveen Kulkarni,[§] Jakub Kustosz,[†] Pravin Kumar,*[,§] Uwe T. Bornscheuer,*[,†] and Matthias Höhne*[,‡]

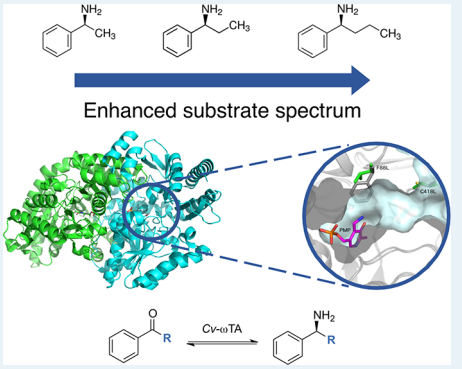
[†]Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany

[‡]Protein Biochemistry, Institute of Biochemistry, Greifswald University, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany

[§]Quantumzyme, LLP, No. 110/8, Krishnappa Layout, Lalbagh Road, Bangalore 560027, India

Supporting Information

ABSTRACT: Protein engineering is often applied to tailor substrate specificity, enantioselectivity, or stability of enzymes according to the needs of a process. In rational engineering approaches, molecular docking and molecular dynamics simulations are often used to compare transition states of wild-type and enzyme variants. Besides affecting the transition state energies by mutations, the entry of the substrate and its positioning in the active site (Michaelis complex) is also often studied, and mutagenesis of residues forming the substrate entry tunnel can have a profound impact on activity and selectivity. In this study, we combine the strengths of such a tunnel approach with MD followed by semiempirical QM calculations that allow the identification of beneficial positions and an in silico screening of possible variants. We exemplify this strategy in the expansion of the substrate scope of *Chromobacterium violaceum* amine transaminase toward sterically demanding substrates. Two double mutants (F88L/C418(G/L)) proposed by the modeling showed >200-fold improved activities in the conversion of 1-phenylbutylamine and enabled the asymmetric synthesis of this amine from the corresponding ketone, which was not possible with the wild-type. The correlation of interaction energies and geometrical parameters (distance of the substrate's carbonyl carbon to the cofactor's amino group) as obtained in the simulations suggests that this strategy can be used for in silico prediction of variants facilitating an efficient entry and placement of a desired substrate as a first requirement for catalysis. However, when choosing amino acid positions for substitution and modeling, additional knowledge of the enzymatic reaction mechanism is required, as residues that are involved in the catalytic machinery or that guarantee the structural integrity of the enzyme will not be recognized by the developed algorithm and should be excluded manually.



KEYWORDS: *protein engineering, asymmetric synthesis, biocatalysis, transaminase, semi-empirical quantum mechanics, molecular dynamics, alanine scanning*

INTRODUCTION

Transaminases are pyridoxal-5′-phosphate-(PLP)-dependent enzymes (Scheme 1) that catalyze the reversible transfer of an amine group from an amino donor to prochiral keto acids, ketones, or aldehydes.[1] Transaminases are highly enantioselective toward their natural substrates, which makes them of key interest in biocatalysis for industrial application. Beside α-amino acid transaminases, which are ubiquitous enzymes in all organisms, a smaller subgroup of amine transaminases (ATA) exists. These transaminases convert substrates without a α-carboxylic acid moiety.

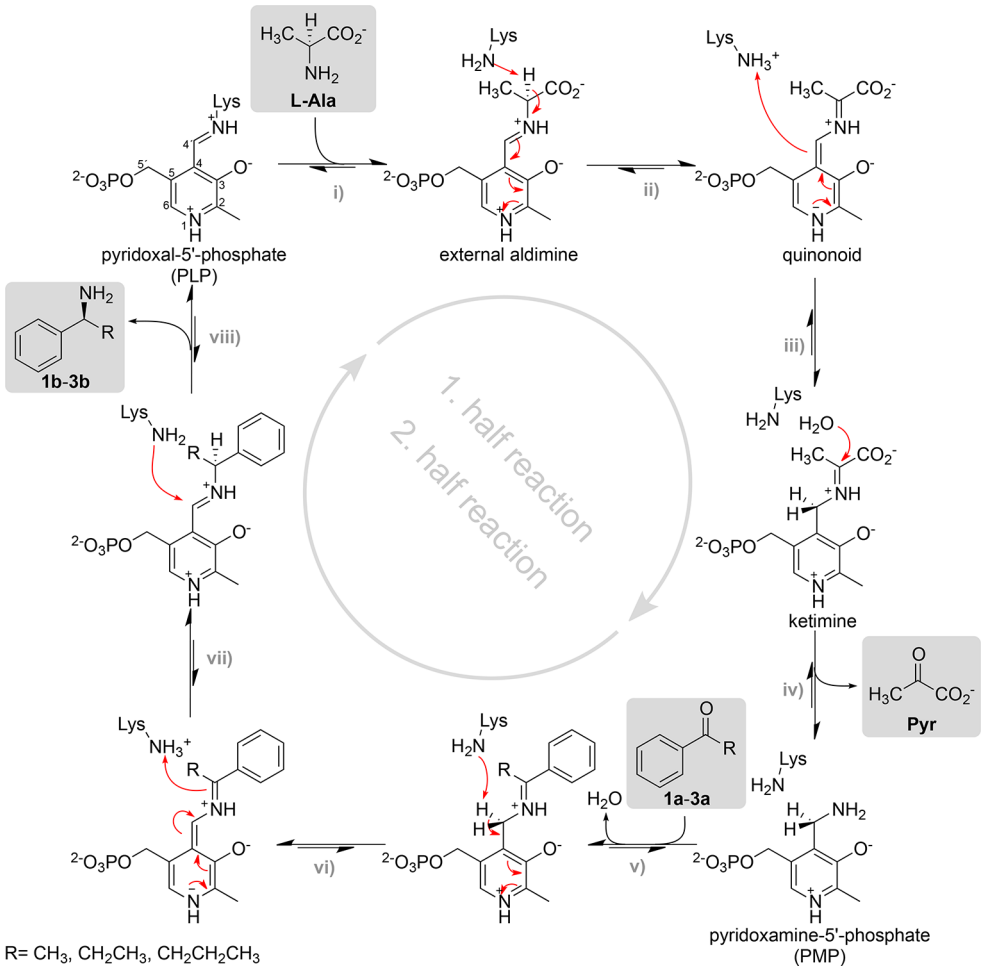
Due to their different overall folding, amine transaminases are found in the PLP fold type I and fold type IV. Type IV amine transaminases accept D-alanine with a general preference toward the (*R*)-enantiomer, whereas the structurally different transaminases of fold type I use L-alanine (L-Ala) and form the (*S*)-enantiomer as amine product.[2] Nevertheless, the substrate scope of these enzymes is limited due to the natural geometry of the transaminases' active site, which is located at the interface of the dimerized monomers and consists of a large (LBP) and a small binding pocket (SBP). If a ketone or α-keto acid binds to the enzyme, then the LBP can accommodate a bulky substituent such as an alkyl or aryl group adjacent to the ketone function (or the carboxylic acid group in an α-keto acid). The SBP is restricted to accommodate a methyl or in rare cases an ethyl substituent.[3−5]

The well-studied (*S*)-selective ω-transaminase from *Chromobacterium violaceum* (*Cv*-ωTA) is consistent with the above-mentioned substrate limitations. This transaminase was characterized in 2007 by Kaulmann et al. and its high potential

Received: September 27, 2018
Revised: October 23, 2018
Published: October 25, 2018

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

Scheme 1. Simplified Mechanism of the Transaminase-Catalyzed Reaction[a]



[a]The reaction is exemplified with alanine as amino donor substrate and acetophenone (**1a**), propiophenone (**2a**), or butyrophenone (**3a**) as ketone substrate. The catalytic lysine (K288) of *Cv*-ωTA is shown. The first half reaction starts with PLP bound as a Schiff's base to a lysine ε-amino group of the enzyme ("internal aldimine"). Nucleophilic attack of the substrate's amino group at C4' of PLP initiates a transamination, where the substrate (alanine) replaces the lysine residue and forms the external aldimine. The catalytic lysine abstracts the proton of the Cα atom of the bound substrate and transfers it to C4′ of the cofactor, whereby generating the ketimine intermediate. A quinonoid intermediate is formed in between these steps. Hydrolysis of the ketimine yields PMP and the ketone product. In the second half reaction, this sequence of reactions is passed through in reverse order.

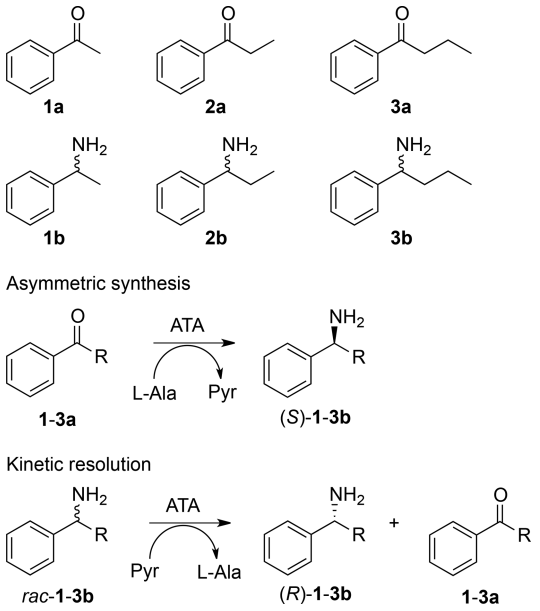
as useful biocatalyst due to its broad amino acceptor range was shown.[6−8] Despite the broad capabilities of the enzyme and detailed crystal structures, the *Cv*-ωTA still lacks the ability to accommodate amines with two bulky substituents.[9,10] The methyl group of (*S*)-phenylethylamine (**1b**) is well accepted in the SBP of the active site, but the only slightly bulkier ethyl (**2b**) or propyl (**3b**) substituents are already barely converted (Scheme 2) and hence only a minimal alteration of the aliphatic chain causes total loss of activity.

Several examples have shown how extensive protein engineering was capable to overcome the limitations of the SBP by using direct evolution and rational design approaches in different amine transaminases.[2,11] The first and most prominent example was published by Merck & Co. and Codexis for the chiral synthesis of (*R*)-sitagliptin employing a protein-engineered variant of the (*R*)-selective ATA117.[11] Inspired by this result, several examples of expanding the SBP also for (*S*)-selective ATAs from *Vibrio fluvialis* (*Vf*-ωTA) and *Ochrobactrum anthropi* followed.[5,12−14] Recently, Pavlidis et al. reported an engineered variant of the (*S*)-selective ωTA from *Ruegeria* sp. (3FCR) for the asymmetric synthesis of bulky chiral amines.[2] By applying the identified beneficial mutations, it was even possible to generate a general sequence motif, which led to the generation of six new TAs active toward bulky substrates. The common feature of those examples is the application of directed evolution, which requires a high screening effort due to large library sizes. Even the creation of statistics-based smart libraries, like the incorporation of the 3DM system for reducing the screening effort by focusing the mutations on a few identified positions with a specific set of amino acid exchanges, required an in vitro screening of larger subsets of variants.[15−17]

One alternative could be the application of quantum mechanics (QM) methods that are applied in the field of quantum chemistry for studying chemical reactions including a small set of atoms or molecules in silico. Quantum mechanical calculations are therefore based on electron density distributions of the system and rely rather on physical parameters than on experimental data. By focusing on the electron density of the system, QM-based methods can provide insights into

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

Scheme 2. Substrates and Reaction Modes Studied Using the Amine Transaminase (ATA) from *C. violaceum* and its Mutants

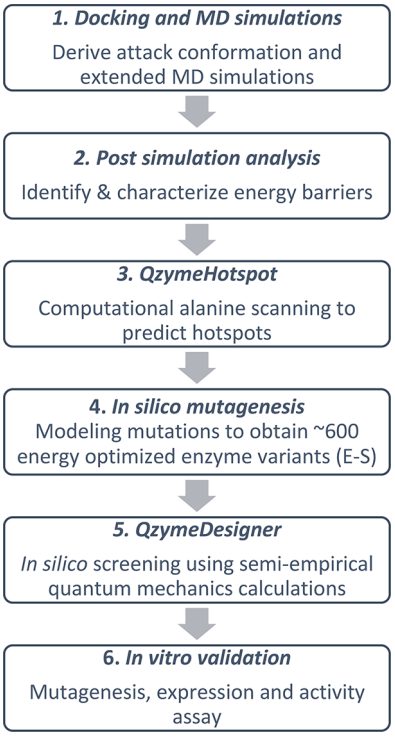


chemical reactions at the level of bond formation and cleavage, which are key aspects in the field of chemistry.[18] Examples have shown the successful application of QM methods for studying metal catalysts and the prediction of their enantioselectivity.[19] But still, such computational-power demanding calculations remain exclusive for small reaction systems including less than 250 atoms, due to the exponentially growing complexity.[18,20] Therefore, the adaption to enzyme-catalyzed reactions is only possible by considering a limited amount of residues in the QM calculation, usually the active site residues of the enzyme. The QM approach is therefore capable to study enzyme−substrate interactions, as recently exemplified by Cassimjee et al. by investigating the detailed reaction mechanism of the *Cv*-ωTA toward the well-accepted benchmark substrate (*S*)-**1b** with pyruvate (Pyr) as acceptor.[21,22]

The replacement of the enzymes active site and the fitting of the mutation in the overall enzyme structure can be realized by Molecular Mechanics (MM), based on classical mechanics to simulate atom−atom interactions.[18] In the MM simulations, the energy of the system is computed using force fields based on empiric data. This approximation allows to study molecular systems up to many thousands of atoms and therefore to simulate whole enzymes and the effect of mutations. In comparison to mutagenesis approaches focusing on residues in the active site to improve the formation of the Michaelis complex, the residues lining the substrate tunnel of the enzyme often also play a key role in the acceptance and conversion of artificial substrates. This is especially important for enzymes that have their active site deeply buried in the interior of the protein, as substrate diffusion into the protein and/or product release might become rate-determining steps.[23] Therefore, reshaping of the binding pocket and the optimization of the entrance tunnel might be equally important for engineering enzymes to accept unnatural (e.g., space-demanding) substrates.[24]

In this study, we outline the potential of semiempirical quantum mechanics (sQM)-based protein-engineering toward the *Cv*-ωTA. Our in silico screening aimed to predict variants capable of accepting the bulky (*S*)-phenylbutylamine **3b** (Scheme 2).

Scheme 3. General Overview of Our Rational Protein Design Concept



As outlined in Scheme 3, the first step of our approach is to determine plausible attack conformations of the substrate by docking experiments and conduct extended MD simulations of the *E*−*S* complexes obtained (step 1). Post simulation analyses were carried out to explore the distribution of the substrate conformation in the enzyme's entrance tunnel and the active site (step 2). These conformations of the substrates form clusters, which (in our case) are separated by barriers hindering the substrate to move toward the active site. Important E-S conformations from MD simulations were explored using sQM calculations to derive the energy profiles of the conformational transitions at the barriers. Computational alanine scanning using the snapshots derived from these calculations identifies relevant hotspot residues (step 3) that are then subjected to in silico mutagenesis (step 4). These variants are screened and ranked using the *QzymeDesigner* developed for this study (step 5). The variants are ranked according to the minimal interaction energies and substrate distances to the catalytic site, and the best variants were finally validated in the laboratory by applying them in the kinetic resolution mode as well as in asymmetric synthesis of the target molecule (step 6).

RESULTS AND DISCUSSION

Computational Analysis and in Silico Mutant Screening. To identify, which mutations would enable the acceptance of the sterically demanding ketones (**2a** and **3a**) by the ATA from *C. violaceum*, the ketones were in the first step docked into the enzyme's active site with the pyridoxamine-5′-phosphate (PMP) cofactor present (Scheme 1). The well-

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

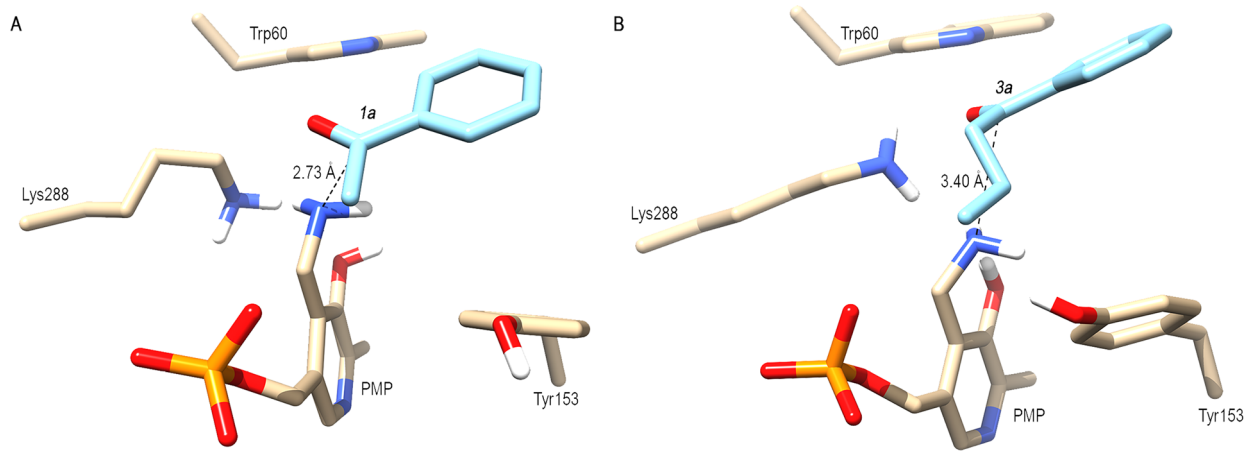


Figure 1. Plausible attack conformations (X_0) for the investigated ketones in the active site of the *Cv*-ωTA. A: Modeled complex for **1a** (E:PMP+**1a**). B: Complex for **3a** (E:PMP+**3a**) after docking and energy minimization. The carbons of the residue side chains and the PMP cofactor are colored in beige and the carbons of substrate are colored in blue. For clarity, nonpolar hydrogen atoms are hidden. The distances d_{1a} and d_{3a} between the carbonyl carbon atoms of ketones **1a** and **3a** and the cofactor's amino group is marked as a dotted line as well as the distance between the catalytic lysine (K288) and the proton of PLP's hydroxyl function.

accepted benchmark substrate **1a** served as control. On the basis of the orientation of the phenyl group and methyl or propyl side chain in the LBP and SBP respectively, the docked conformation of **1a** and **3a** in the active site is in-line with previous studies,[22] where the keto substrate's carbonyl group faces the amino group of PMP allowing a nucleophilic attack to form to external aldimine.

The resulting complexes (E:PMP+**1a** and E:PMP+**3a**) were energy minimized and equilibrated to serve as starting conformations for MD simulations. These conformations we will refer to as the X_0 conformation throughout this manuscript (Figure 1). In the MD simulations (step 2, Scheme 3), we observed that the bulky ketone **3a** moves out of the active site within the first 5 ns of the 50 ns simulation. Therefore, the simulated enzyme does not accommodate the substrate in the active site and fails to form a productive Michaelis complex, which is the starting point of the reaction. The distance d_{3a} between the cofactor PMP and the substrate ketone **3a** was measured between the PMP's exocyclic nitrogen atom (PMP-NH_2) and the carbonyl carbon atom of the ketone **3a**. During the simulations, d_{3a} showed two major increases of 4 Å each, the first increase at ∼0.3 ns and the second over a longer period between 3 and 4 ns. Beyond 5 ns of simulation, **3a** moved to a final distance of 12 Å to PMP (Figure 2B). The movement of **3a** out of the active site within 5 ns seems peculiar in the MD simulation, unless a steered force is applied. To confirm this, multiple MD analyses were performed with 50 and 20 ns of simulation time with similar results (Figures 2B and S2 of the Supporting Information, SI). The control experiment with the well-accepted ketone **1a** demonstrated the correct performance of the experiment, since **1a** stayed in the active site throughout the 20 ns of simulation at distance of ∼2.5 Å to PMP (Figure S1).

In an analysis of the MD simulations with the bulky ketone E:PMP+**3a** complex (step 3), we identified 5 clusters of conformations during the movement out of the active site (Figure 2A). As representatives of the first four clusters, we defined the conformations X_0 (starting conformation with 3.4 Å to PMP), X_1 (0.615 ns, 8 Å), X_2 (2.5 ns, 10 Å), and X_3 (4.4 ns, 12 Å) and extracted the coordinates of each conformation. X_0 represents a plausible attack conformation that could initiate the external aldimine formation.

The four representative conformations (X_0, X_1, X_2, and X_3) of the clusters are separated by barriers: areas where the ketone **3a** is observed in only few snapshots during the simulation. Assumingly, these barriers hinder the substrate to enter a catalytic active position in the enzymes' active site and mark high-energy conformations, which need to be overcome to reach the productive attack conformation (X_0). Therefore, mutating residues lining these barriers could alter the enzyme in a way that makes it easier for the substrate to reach the cofactor and to adopt the catalytically active conformation ready for external aldimine formation. This would be the first step facilitating the conversion of the challenging bulky ketone **3a**. To obtain realistic conformations of the E:PMP+**3a** complex and to derive detailed energies during the transition, sQM calculations were performed. The objective of this study was to obtain the energy profile (heat of formation, ΔH_f) of the three transitions ($X_3 \rightarrow X_2$, $X_2 \rightarrow X_1$ and $X_1 \rightarrow X_0$) and to identify the high energy conformational transitions, which are expected to be passed at the barriers identified in the MD simulations.

The enzyme seems to undergo conformational changes, especially in the active site to accommodate **3a**. The detailed sQM-based analysis of the $X_1 \rightarrow X_0$ transition showed the expected restraint of **3a** entering the catalytic active conformation, since it was observed that the initially decreasing ΔH_f energy started to fluctuate, marking the energy barrier (see Figure S3). To reach the active site, the propyl side chain rotates ∼180° to decrease d_{3a} from 3.9 to 3.4 Å and finally reaches the X_0 conformation. Assumingly, these unfavorable conformations, caused by the rotation of the propyl side chain, mark a significant hurdle for the acceptance of **3a** as substrate.

To identify hotspot positions positively affecting barrier transitions, sQM-based computational alanine scanning (QM-CAS) was carried out (step 4) by incorporating amino acid substitutions in the extracted coordinates of the $X_3 \rightarrow X_2$, $X_2 \rightarrow X_1$, and $X_1 \rightarrow X_0$ sQM calculations. For choosing target residues, we performed an atomic contact analysis using the data from the simulations (Methods in the SI), in addition to visual inspection. Selected residues were mutated in silico to

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

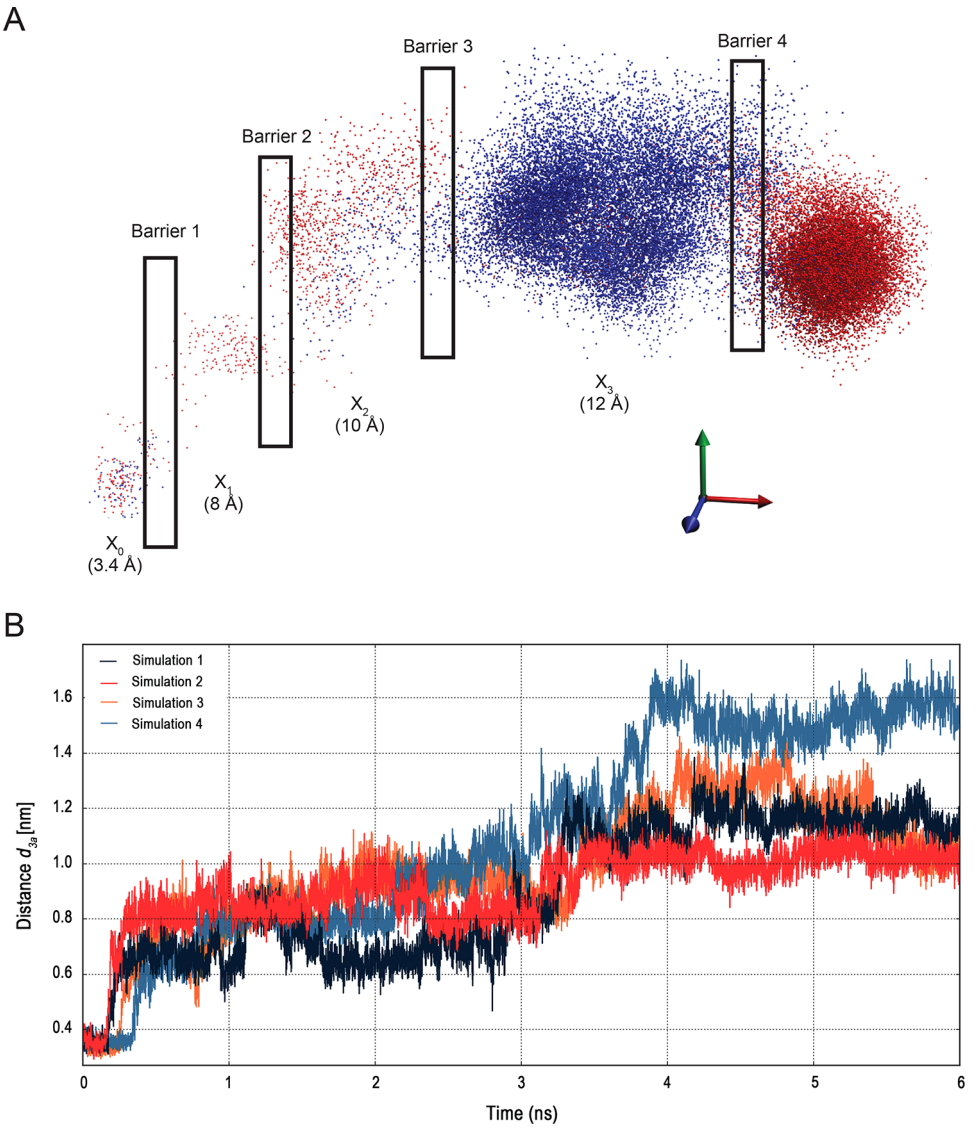


Figure 2. MD simulations with the bulky ketone **3a** in the *Cv*-ωTA wild-type structure. A: 2D projection of the observed positions of **3a** during the simulation. The dots represent the carbonyl carbon atoms of **3a**; blue dots: 50 ns trajectory I; red dots: 50 ns trajectory II. Five clusters of stable conformations were observed during the simulation. The X_0 to X_3 conformations were extracted as single representatives of the respective cluster. B: The distance of **3a** increases during the first 6 ns of the MD simulations. A clear increase of the distance values d_{3a} (defined in Figure 1) was observed at the beginning of the simulation (0.3 ns). The distance value d_{3a} increases further at around 3 ns. Four independent simulations were plotted: simulation 1:20 ns (trajectory I); simulation 2:20 ns (trajectory II); simulation 3:50 ns (trajectory I); and simulation 4:50 ns (trajectory II).

alanine one at a time. Alanine or proline residues were substituted to glycine to further decrease the size of their side chains and glycine residues were mutated to alanine (later referred to as GAP mutants). Next, the structure of the enzyme variants was optimized locally for each of the extracted QM coordinates. Then, the binding affinity of the substrate across the extracted coordinates were calculated. The variants were analyzed toward their interaction energy (see eq S1 in SI) and selected for variants with lower interaction energy than the wild-type enzyme (-56 ± 122 kcal mol^{-1}) for the $X_1 \rightarrow X_0$ transition. The higher standard deviation is due to the common partial charges of the substrate that was used for calculating the conformational transitions.

The QM-CAS analysis of the $X_1 \rightarrow X_0$ transition identified the single substitution variants W60A, F88A, S121A, and C418A as potentially beneficial since these variants showed lower interaction energy than the wild-type enzyme. On the basis of the QM-CAS analysis, the residue F88 seems to be the crucial hotspot for mutagenesis, because the F88A substitution is the only one which significantly lowered the interaction energy by ∼10 times compared to the wild-type enzyme (Table S1). Visualization studies revealed that F88A plays an important role in facilitating the movement of **3a** toward PMP. Additionally, double substitutions were carried out for all selected residues of the $X_1 \rightarrow X_0$ transition to predict whether combined mutations could further decrease the interaction energy. The six double mutants (W60A-F88A, F88A-C418A, F88A-S121A, W60A-S121A, W60A-C418A, and S121A-C418A) showed lower ΔG values compared to the wild-type enzyme and were selected for further studies (Table S4).

To guide the decision which amino acids should be used for in silico variant generation (step 5), sequence and structure

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

alignments were prepared using sequences similar to the ATA from *C. violaceum*. The three most frequently found amino acid residues occurring in the alignments in each of the four selected positions W60, F88, S121, and C418 were chosen for variant generation. All single and double mutations were created in silico.

QzymeBenchwork, which includes *QzymeHotspot* and *QzymeDesigner*, is an *in-house* computational tool developed by Quantumzyme LLP that uses a novel sQM based energy evaluation protocol and was used for screening enzyme variants in silico (see Methods in the SI). Variants having energy values lower than the wild-type were sorted based on the minimal d_{3a} obtained during the simulation. The top two double variants W60G/F88V and W60G/S121A were most promising based on the ranking and chosen for a subsequent evaluation of the enzymatic activity (Tables 1 and S7).

Table 1. Activities of the Purified *Cv*-ωTA Variants in the Adopted Acetophenone Assay Using (*S*)-3b or Asymmetric Synthesis Using 3a and the Calculated Interaction Energy

variant	activity[a] (U mg^{-1})	conversion[b] (%)	interaction energy[c] (kcal mol^{-1})
Cv-ωTA wt	0.018 ± 0.004	n.d.[d]	−56
W60G/S121A	0.003 ± 0.001	n.m.	−333
W60G/F88V	0.15 ± 0.01	n.m.	−722
F88A	0.20 ± 0.005	n.m.	−511
F88V	0.97 ± 0.03	62.0 ± 1.3	−747
F88L	2.27 ± 0.13	82.8 ± 0.6	−852
F88L/C418G	3.72 ± 0.19	89.8 ± 0.6	−847
F88L/C418L	4.12 ± 0.19	98.6 ± 0.1	−983

[a]Specific activity in the deamination of (*S*)-1-phenylbutylamine **3b** with pyruvate (spectrophotometric assay). [b]Conversion in an asymmetric synthesis of **3b** using the corresponding ketone **3a** and L-alanine as substrates and GDH/LDH for equilibrium shift after 16 h. [c]interaction energy (ΔG) of the $X_1 \rightarrow X_0$ transition, calculated by the *QzymeBenchwork* protocols. [d]n.d. = no detectable activity; n.m.= not measured.

The same *QzymeBenchwork* protocols were applied for the transitions $X_3 \rightarrow X_2$ and $X_2 \rightarrow X_1$ to identify additional positions and to predict beneficial single and double variants. None of the variants showed significant energy decreases compared to the wild-type enzyme, only one double variant (R405A-F22A in the $X_3 \rightarrow X_2$ transition) was identified as potentially interesting in the CAS protocol (Tables S2, S3, S5, and S6).

Although some of the modeled double variants only showed slight benefits, we constructed 216 variants in silico combining positions from the $X_3 \rightarrow X_2$ and $X_1 \rightarrow X_0$ transitions and the three GAP mutants to test for potential synergic effects. All variants contained amino acid substitutions at ten residues and were ranked by using the *QzymeBenchwork* protocol (Table S8). Although it is very difficult to accurately predict the effect of substituting ten residues at the same time, we wished to explore the potential of multiple substitutions using sQM methods. This also allows to recognize positive synergistic effects arising from multiple interactions of substituted amino acids. Finally, we chose the best 11 variants (Table S8) for experimental characterization.

Experimental Validation of sQM-Based Candidates. The in silico sQM-based evaluation identified ten positions and one to three amino acid alternatives per position as beneficial for converting the bulky substrate (*S*)-phenylbutylamine (**3a**). From a protein-engineering perspective, the simultaneous mutagenesis of ten residues within the active site region of an enzyme bears not only the risk of inactivity, but also the loss of the enzyme's folding and cofactor binding capacity. In addition to the variants containing mutations at all ten selected residues and therefore addressing several barriers simultaneously, double mutants were predicted covering only the $X_1 \rightarrow X_0$ transition. For validation of the theoretical results, all mutants were created, expressed in *E. coli* and purified to determine their activity (step 6) in the acetophenone assay.[25] Surprisingly, all predicted mutants could be expressed with good protein yield after purification. On the basis of their yellow color, the mutants were judged to have the PLP-cofactor correctly bound. Unfortunately, it turned out that all variants bearing ten mutations simultaneously (Table S8) exhibited no transaminase activity (Scheme 2, bottom reaction), which was not unexpected. However, the predicted double mutant W60G/F88V showed a 8-fold activity increase in the conversion of (*S*)-**3b** in the kinetic resolution mode (Table 1). Next, the corresponding single mutants were created and this revealed the F88V mutation as the beneficial one resulting in an even 50-fold activity improvement over the wild-type activity toward (*S*)-**3b** (Table 1). Besides the substitution to valine, alanine, and leucine were also suggested for the F88 residue. By incorporating these mutations in the *Cv*-ωTA, the F88L mutation roughly doubled the activity compared to F88V and led to an overall 120-fold increase of activity compared to the *Cv*-ωTA wt. Therefore, the F88 residue clearly is a hotspot for engineering the active site for the accommodation of bulkier substituents within the SBP. This finding is in-line with the sQM-results obtained from the QM-CAS analysis, where the F88A single-substitution showed a significant effect on the interaction energy.

The F88 position in *Cv*-ωTA was also identified by Deszcz et al. as beneficial to create serine:pyruvate ω-TA activity.[26] Humble et al. reported the F88A mutation as beneficial for the reversion of the enantioselectivity toward 1-aminotretralin and a decrease of the selectivity toward (*S*)-**1b**.[27] Humble et al. also pointed out, that the F88A mutation increases the space of the small binding pocket and therefore enables the coordination of the phenyl group of **1b** in both pockets. Still, the F88 residue was not considered so far to enable the conversion of bulky substrates by *Cv*-ωTA, although the corresponding residues in other (*S*)-selective ωTAs were reported as advantageous toward bulky substrates. Nobili et al. investigated the corresponding mutation F85L in the ωTA of *Vibrio fluvialis* (*Vf*-ωTA) and observed an activity of 0.32 U mg^{-1} toward the conversion of (*S*)-**3b**.[5] Nevertheless, the leucine substitution of the F85 residue led to a significantly lower impact on the activity toward (*S*)-**3b** compared to the corresponding mutation in *Cv*-ωTA. Pavlidis et al. identified the mutation Y87F in the scaffold of the ωTA from *Ruegeria* sp. (3FCR) as beneficial for the conversion of bulky amines in the context of other mutations.[2] Interestingly, the 3FCR scaffold shows increased activity by incorporating the *Cv*-ωTA wild-type residue phenylalanine.

Refining the In Silico Ranking. The results of the wet lab experiments were used to refine the ranking of the in silico experiments. Three insights were gained: (1) The major difference between W60G/S121A and W60G/F88V was the interaction energy (ΔG), which was significantly lower for the beneficial double mutant. Therefore, the minimal ΔG should have a higher impact in the rating compared to the minimal d_{3a}

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

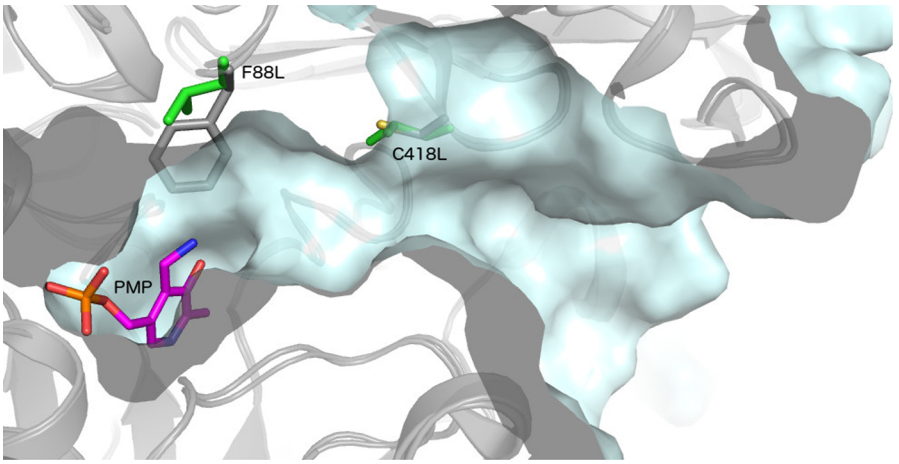


Figure 3. Location of the amino acid substitutions in variant F88L/C418L (shown in green). The amino acids of the wild-type are overlaid in gray. The backbone is shown as gray cartoon, a cut through the surface of the substrate entrance tunnel of the mutant structure is highlighted in light cyan.

during the sQM calculation. (2) F88L was identified as the most beneficial single mutant. Only double mutants with F88L were considered in the new ranking. (3) The S121A substitution was considered detrimental for the activity because the double variant containing S121A was not active toward **3b**. This result can be explained by the observation that S121 is part of the phosphate binding cup, as its hydroxyl side chain forms a hydrogen bond with PLP's phosphate. Variants containing S121 substitution were therefore excluded from the ranking.

On the basis of the refined ranking parameters, two variants F88L/C418G and F88L/C418L (out of the 54 double mutants; Table S7), were chosen with lower QM-ΔG and with minimal d_{3a}. C418 is far from the active site and F88 and C418 form the lining resides of the tunnel of **3a** (Figures 3 and S4).

Characterization of the Final Double Mutants. The proposed double mutants yielded high protein contents and remained in their active state after purification. The double mutants significantly increased the activity for the conversion of (*S*)-**3b** compared to the most active single mutant F88L. Both double mutants exhibit a >200-fold activity improvement compared to the wild-type activity (Table 1). Despite the high activity toward the bulky (*S*)-**3b** substrate, the double mutant F88L/C418G remained highly active toward the benchmark substrate (*S*)-**1a**, indicating the successfully extended substrate spectrum (Table S9).

Importantly, these variants also showed perfect enantioselectivity (>99%ee) concerning the asymmetric synthesis of **3b**. When enzyme activities of the characterized single and double mutants are compared with their calculated energy values (QM-ΔG), a clear trend was found (Table 1), indicating that energy values are an important predictor for activity toward the desired substrate. F88L/V is beneficial because it enlarges the active site (Figure 3) and allows a better positioning of the substrate to attack the cofactor. This is reflected in a lower K_M of 3 mM for **3b** for the single mutants in contrast to the wild-type ATA, where saturation with **3b** could only be estimated around 14 mM (Table S10). The C418L/G mutation further decreases the K_M for **3b**, with a drastic increase in the catalytic efficiency (k_{cat}/K_M). The residue C418 lies in the substrate entry tunnel (Figure 3), and mutating this residue might affect local dynamics of substrate and enzyme side chain interactions. MD simulations of the double variant, F88L/C418L complexed with **1a** and **3a** revealed that both substrates stayed in the active site throughout the simulation (Figure S5). This is in contrast to the simulation of the wild-type enzyme, where **3a** moved out 12 Å of the active site within the first 6 ns of the simulation as described above. This observation and the lowered ΔG for the F88L/C418L variant obtained by our sQM calculation is consistent with our hypothesis to remove an energy barrier of the substrate entrance ($X_1 \rightarrow X_0$ transition).

The residue C418, identified by our screening approach, drastically influenced the substrate specificity of the *Cv*-ωTA. The importance of this residue toward substrate specificity and activity was not reported before. Recently, Genz et al. identified the corresponding L417 residue in the scaffold of the *Vf*-ωTA by statistical analysis.[14,28] Surprisingly, L417 substitutions were shown to have no effect on the widening of the substrate binding pocket or increasing the activity toward bulky substrates in the *Vf*-ωTA scaffold.[14]

There are different methods to screen enzyme variants in silico, like MD simulations of transition state analogs[29] and simulations of near attack configurations to evaluate the enantioselectivity of enzymes.[30] Alternatively, a rapid and robust approach was developed to predict enzyme activity with a large number of substrates using geometric criteria based on the mechanism in combination with molecular docking.[31] Short simulations are integrated into the Quantitative Structure Activity Relationship Protocol (QSAR) to predict kinetic properties of enzymes.[32] In our study, we observed that the substrate is moving away from the active site; the most crucial point was that the wild type enzyme was unable to form the catalytic attack conformation. In this scenario, docking and short molecular dynamics studies—commonly used so far—would only help to choose residues in the active site region for engineering. The customized protocol used in this study is a systematic procedure with few manual interventions, and thus facilitates in silico screening of a considerable number of variants.

CONCLUSIONS

The aim of our research was the extension of the substrate spectrum of the (*S*)-selective *Cv*-ωTA, enabling the conversion of the sterically demanding benchmark substrate butyrophe-

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

none (**3a**). In contrast to the laboratory-intensive approach of directed evolution, our strategy represents a rational design approach, where MD and sQM-based protein engineering is the basis to predict beneficial enzyme variants. According to our analysis, residues near the active site or lining the substrate path influence the interaction energy for the substrate to enter the active site.

Our in silico screening identified variants with lowered interaction energies leading to an improved attack conformation according to our model, explaining the >200-fold increased activities in the conversion of **3a** by the F88L/C418(G/L) double mutants. Due to the sampling of the interaction energies of the *E*−*S* complex across the entry of the substrate, the identification of C418 was possible, which is ∼12 Å away from the catalytic site. To further explain the effect of the mutations in depth, we plan to perform QM/MM calculations to evaluate the impact of the identified mutations on the energetics of the transition states of the transamination reaction. This is beyond the scope of this in silico screening study, where we aimed to have a good balance between computational cost and accuracy and a possibly high throughout. The computational approach used in this work is tuned to screen ∼1000 variants in a week's time.

Compared to previous semirational approaches, we identified variants with similar activities toward the benchmark substrates but with lower laboratory screening effort. One experience of this study is that it is important to have an early feedback of lab experiments to validate assumptions of the calculations. Although predicted variants carrying ten substitutions could not be verified to improve activity, the prediction of single and double variants correlated well with the laboratory experiments and helped to refine and improve the ranking and interpretation of computational results.

METHODS

Docking, MD Simulations, and Semiempirical QM Calculations for the In Silico Screening of Enzyme Variants. The substrates **1a** and **3a** were docked in the active site of *Cv*-ωTA in separate docking experiments. The catalytic lysine (K288) was deprotonated and the PLP of the enzyme's crystal structure (PDB ID: 4A6T) was converted to PMP before docking. Postdocking, the conformations obtained from **1a** and **3a** were filtered to derive a plausible attack conformation using the E:PMP+**1a** complex obtained from Cassimjee et al.[22] as a reference template.

The docked complexes were used to conduct a series of molecular dynamics (MD) simulations. The charges for the substrates were obtained by quantum mechanics (QM) optimization protocol at HF/6-31G* level and the MD simulations were performed for either 20 or 50 ns for different *E*−*S* complexes, using 0.002 ps time steps. After the analysis, the enzyme complexed with **3a** was analyzed to determine the atomic contacts between the enzyme and the substrate to obtain a consolidated list of residues for the computational alanine scanning (CAS). The *QzymeHotspot* protocol identified several hotspots in the CAS analysis, which were mutated and combined (double mutations) for the in silico variant generation. Therefore, the beneficial hotspots were substituted with the most conserved residues identified in a multiple protein sequence alignment. This in silico mutagenesis approach resulted in ∼600 enzyme variants for computational screening. The modeled variants were screened using the *QzymeDesigner* protocol, which analyzes every enzyme variant using the sQM method and the variants were ranked using eq S1 (see SI).

Activity Determination in the Kinetic Resolution Mode. The activities of the purified *Cv*-ωTA variants were determined photometrically by applying the acetophenone assay on the TECAN Infinite 200 PRO reader.[25] The assay was performed with 2.5 mM pyruvate and 2.5 mM of the corresponding amino donor (*S*)-phenylethylamine (**1b**), (*S*)-phenylpropylamine (**2b**) or (*S*)-phenylbutylamine (**3b**) in 1.25% DMSO (5% DMSO for (*S*)-**3b**) in 50 mM HEPES buffer pH 8.2. The enzymes were diluted according to their activities. The formation of the corresponding ketones was detected by following the increase of absorption at 245 nm (**1a**) or 242 nm (**2a** and **3a**) over time.

Asymmetric Synthesis Experiments. The asymmetric synthesis reactions were performed in duplicates in a 1 mL scale with 1 mg of purified transaminase variant in 5 mM **3a**, 0.3 mg GDH, 90 U LDH, 2 mM NADH, 200 mM L-alanine, 220 mM D-glucose, 1 mM PLP, 10% DMSO in 50 mM HEPES buffer pH 8.0. The reaction was carried out at 30 °C with 900 rpm (ThermoMixer, Eppendorf). 60 μL samples were taken at defined time points and the reaction was stopped and basified with 10 μL 10 N NaOH. NaCl was added until saturation was observed and 220 μL of ethyl acetate (EtOAc) supplemented with 1 mM of *p*-bromoacetophenone (standard for quantification) was added for extraction. After vigorous shaking (3 min), the samples were centrifuged (17 000*g*, 5 min) and 160 μL of the organic phase was dried with $MgSO_4$. After centrifugation (17 000*g*, 5 min), 100 μL sample were obtained for subsequent GC analysis. The chiral analysis of the product and substrate was performed as described previously[28] with the following retention times: **3a**: 12.35 min; (*S*)-**3b**: 10.37 min; *p*-bromoacetophenone: 23.67 min. The conversion of **3a** was determined after normalization with *p*-bromoacetophenone in relation to the control experiment with known **3a**.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.8b03900.

Details on the methods for the docking experiments, MD and sQM calculations as well as the cloning, mutagenesis, expression, and purification of enzyme variants (PDF)

AUTHOR INFORMATION

Corresponding Authors

*Phone: (+49) 3834 420 4367. Fax: (+49) 3834 420 744367. Email: uwe.bornscheuer@uni-greifswald.de.

*Phone: (+49) 3834 420 4417. E-mail: matthias.hoehne@uni-greifswald.de.

*E-mail: pravinpaul2@gmail.com.

ORCID

Uwe T. Bornscheuer: 0000-0003-0685-2696

Author Contributions

U.T.B. and M.H. initiated and supervised the project. The in silico enzyme engineering framework was designed by P.K., developed and implemented by P.K., D.D., and A.K. N.K. was involved in the scientific discussions. M.G. performed initial experiments, M.V. and J.K. prepared and M.V. performed biochemical characterization of all enzyme variants. U.T.B., M.H., M.V., and P.K. analyzed the results. M.H. and M.V.

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

drafted the manuscript and M.V., P.K., M.H., and U.T.B. wrote the manuscript. The manuscript was revised and approved by all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the German Research Foundation (DFG) for financial support (U.T.B., grant number: Bo1862/16-1; M.H., grant number: Ho4754/4-1). We are also grateful to Hubert Kasprowski for additional computational analysis of the enzyme's scaffold and Stephan Schneiders for the help in enzyme variant generation and purification.

■ REFERENCES

(1) 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.; Bornscheuer, U. T. Bioinformatic Analysis of a PLP-dependent Enzyme Superfamily Suitable for Biocatalytic Applications. *Biotechnol. Adv.* **2015**, *33*, 566−604.

(2) Pavlidis, I. V.; Weiß, M. S.; Genz, M.; Spurr, P.; Hanlon, S. P.; Wirz, B.; Iding, H.; Bornscheuer, U. T. Identification of (*S*)-selective Transaminases for the Asymmetric Synthesis of Bulky Chiral Amines. *Nat. Chem.* **2016**, *8*, 1076−1082.

(3) Calvelage, S.; Dörr, M.; Höhne, M.; Bornscheuer, U. T. A Systematic Analysis of the Substrate Scope of (*S*)- and (*R*)-selective Amine Transaminases. *Adv. Synth. Catal.* **2017**, *359*, 4235−4243.

(4) Han, S.-W.; Park, E.-S.; Dong, J.-Y.; Shin, J.-S. Active Site Engineering of ω-Transaminase for Production of Unnatural Amino Acids Carrying a Side Chain Bulkier than an Ethyl Substituent. *Appl. Environ. Microbiol.* **2015**, *81*, 6994−7002.

(5) Nobili, A.; Steffen-Munsberg, F.; Kohls, H.; Trentin, I.; Schulzke, C.; Höhne, M.; Bornscheuer, U. T. Engineering the Active Site of the Amine Transaminase from *Vibrio fluvialis* for the Asymmetric Synthesis of Aryl−Alkyl Amines and Amino Alcohols. *ChemCatChem* **2015**, *7*, 757−760.

(6) Kaulmann, U.; Smithies, K.; Smith, M. E. B.; Hailes, H. C.; Ward, J. M. Substrate Spectrum of Omega-Transaminase from *Chromobacterium violaceum* DSM30191 and its Potential for Biocatalysis. *Enzyme Microb. Technol.* **2007**, *41*, 628−637.

(7) Ladkau, N.; Assmann, M.; Schrewe, M.; Julsing, M. K.; Schmid, A.; Bühler, B. Efficient Production of the Nylon 12 Monomer Omega-Aminododecanoic Acid Methyl Ester from Renewable Dodecanoic Acid Methyl Ester with Engineered *Escherichia coli*. *Metab. Eng.* **2016**, *36*, 1−9.

(8) Pressnitz, D.; Fuchs, C. S.; Sattler, J. H.; Knaus, T.; Macheroux, P.; Mutti, F. G.; Kroutil, W. Asymmetric Amination of Tetralone and Chromanone Derivatives Employing Omega-Transaminases. *ACS Catal.* **2013**, *3*, 555−559.

(9) Humble, M. S.; Cassimjee, K. E.; Håkansson, M.; Kimbung, Y. R.; Walse, B.; Abedi, V.; Federsel, H.-J.; Berglund, P.; Logan, D. T. Crystal Structures of the *Chromobacterium violaceum* ω-Transaminase Reveal Major Structural Rearrangements upon Binding of Coenzyme PLP. *FEBS J.* **2012**, *279*, 779−792.

(10) Sayer, C.; Isupov, M. N.; Westlake, A.; Littlechild, J. A. Structural Studies of *Pseudomonas* and *Chromobacterium*-Amino-transferases Provide Insights into their Differing Substrate Specificity. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 564−576.

(11) 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.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305−309.

(12) Midelfort, K. S.; Kumar, R.; Han, S.; Karmilowicz, M. J.; McConnell, K.; Gehlhaar, D. K.; Mistry, A.; Chang, J. S.; Anderson, M.; Villalobos, A.; Minshull, J.; Govindarajan, S.; Wong, J. W. Redesigning and Characterizing the Substrate Specificity and Activity of *Vibrio fluvialis* Aminotransferase for the Synthesis of Imagabalin. *Protein Eng., Des. Sel.* **2013**, *26*, 25−33.

(13) Han, S. W.; Park, E. S.; Dong, J. Y.; Shin, J. S. Mechanism-Guided Engineering of Omega-Transaminase to Accelerate Reductive Amination of Ketones. *Adv. Synth. Catal.* **2015**, *357*, 1732−1740.

(14) Genz, M.; Vickers, C.; van den Bergh, T.; Joosten, H.-J.; Dörr, M.; Höhne, M.; Bornscheuer, U. T. Alteration of the Donor/Acceptor Spectrum of the (*S*)-amine Transaminase from *Vibrio fluvialis*. *Int. J. Mol. Sci.* **2015**, *16*, 26953−26963.

(15) Nobili, A.; Gall, M. G.; Pavlidis, I. V.; Thompson, M. L.; Schmidt, M.; Bornscheuer, U. T. Use of 'Small but Smart' Libraries to Enhance the Enantioselectivity of an Esterase from *Bacillus stearothermophilus* Towards Tetrahydrofuran-3-yl Acetate. *FEBS J.* **2013**, *280*, 3084−3093.

(16) Kuipers, R. K.; Joosten, H.-J.; van Berkel, W. J. H.; Leferink, N. G. H.; Rooijen, E.; Ittmann, E.; van Zimmeren, F.; Jochens, H.; Bornscheuer, U.; Vriend, G.; Martins dos Santos, V. A. P.; Schaap, P. J. 3DM: Systematic Analysis of Heterogeneous Superfamily Data to Discover Protein Functionalities. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2101−2113.

(17) van den Bergh, T.; Tamo, G.; Nobili, A.; Tao, Y.; Tan, T.; Bornscheuer, U. T.; Kuipers, R. K. P.; Vroling, B.; de Jong, R. M.; Subramanian, K.; Schaap, P. J.; Desmet, T.; Nidetzky, B.; Vriend, G.; Joosten, H. J. CorNet: Assigning Function to Networks of Co-Evolving Residues by Automated Literature Mining. *PLoS One* **2017**, *12*, e0176427.

(18) Romero-Rivera, A.; Garcia-Borras, M.; Osuna, S. Computational Tools for the Evaluation of Laboratory-Engineered Biocatalysts. *Chem. Commun.* **2017**, *53*, 284−297.

(19) Ianni, J. C.; Annamalai, V.; Phuan, P. W.; Panda, M.; Kozlowski, M. C. A Priori Theoretical Prediction of Selectivity in Asymmetric Catalysis: Design of Chiral Catalysts by using Quantum Molecular Interaction Fields. *Angew. Chem., Int. Ed.* **2006**, *45*, 5502−5505.

(20) Lind, M. E. S.; Himo, F. Quantum Chemistry as a Tool in Asymmetric Biocatalysis: Limonene Epoxide Hydrolase Test Case. *Angew. Chem., Int. Ed.* **2013**, *52*, 4563−4567.

(21) Manta, B.; Cassimjee, K. E.; Himo, F. Quantum Chemical Study of Dual-Substrate Recognition in Omega-Transaminase. *ACS Omega* **2017**, *2*, 890−898.

(22) Cassimjee, K. E.; Manta, B.; Himo, F. A Quantum Chemical Study of the Omega-Transaminase Reaction Mechanism. *Org. Biomol. Chem.* **2015**, *13*, 8453−8464.

(23) Kaushik, S.; Marques, S. M.; Khirsariya, P.; Paruch, K.; Libichova, L.; Brezovsky, J.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Impact of the Access Tunnel Engineering on Catalysis is Strictly Ligand-Specific. *FEBS J.* **2018**, *285*, 1456−1476.

(24) Li, G.; Yao, P.; Gong, R.; Li, J.; Liu, P.; Lonsdale, R.; Wu, Q.; Lin, J.; Zhu, D.; Reetz, M. T. Simultaneous Engineering of an Enzyme's Entrance Tunnel and Active Site: the Case of Monoamine Oxidase MAO-N. *Chem. Sci.* **2017**, *8*, 4093−4099.

(25) Schätzle, S.; Höhne, M.; Redestad, E.; Robins, K.; Bornscheuer, U. T. Rapid and Sensitive Kinetic Assay for Characterization of ω-Transaminases. *Anal. Chem.* **2009**, *81*, 8244−8248.

(26) Deszcz, D.; Affaticati, P.; Ladkau, N.; Gegel, A.; Ward, J. M.; Hailes, H. C.; Dalby, P. A. Single Active-Site Mutants are Sufficient to Enhance Serine:Pyruvate α-Transaminase Activity in an ω-Trans-aminase. *FEBS J.* **2015**, *282*, 2512−2526.

(27) Humble, M. S.; Cassimjee, K. E.; Abedi, V.; Federsel, H. J.; Berglund, P. Key Amino Acid Residues for Reversed or Improved Enantiospecificity of an Omega-Transaminase. *ChemCatChem* **2012**, *4*, 1167−1172.

(28) Genz, M.; Melse, O.; Schmidt, S.; Vickers, C.; Dörr, M.; van den Bergh, T.; Joosten, H. J.; Bornscheuer, U. T. Engineering the Amine Transaminase from *Vibrio fluvialis* Towards Branched-Chain Substrates. *ChemCatChem* **2016**, *8*, 3199−3202.

(29) Zheng, F.; Yang, W. C.; Ko, M. C.; Liu, J. J.; Cho, H.; Gao, D. Q.; Tong, M.; Tai, H. H.; Woods, J. H.; Zhan, C. G. Most Efficient

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533

Cocaine Hydrolase Designed by Virtual Screening of Transition States. *J. Am. Chem. Soc.* **2008**, *130*, 12148−12155.
(30) Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B. Enantioselective Enzymes by Computational Design and *in silico* Screening. *Angew. Chem., Int. Ed.* **2015**, *54*, 3726−3730.
(31) Daniel, L.; Buryska, T.; Prokop, Z.; Damborsky, J.; Brezovsky, J. Mechanism-Based Discovery of Novel Substrates of Haloalkane Dehalogenases Using *in silico* Screening. *J. Chem. Inf. Model.* **2015**, *55*, 54−62.
(32) Kmuníček, J.; Luengo, S.; Gago, F.; Ortiz, A. R.; Wade, R. C.; Damborsky, J. Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10. *Biochemistry* **2001**, *40*, 8905−8917.

DOI: 10.1021/acscatal.8b03900
ACS Catal. 2018, 8, 11524−11533