

# Computational Prediction of $\omega$ -Transaminase Specificity by a Combination of Docking and Molecular Dynamics Simulations

Carlos Ramírez-Palacios, Hein J. Wijma, Sebastian Thallmair, Siewert J. Marrink, and Dick B. Janssen\*



Cite This: *J. Chem. Inf. Model.* 2021, 61, 5569–5580



Read Online

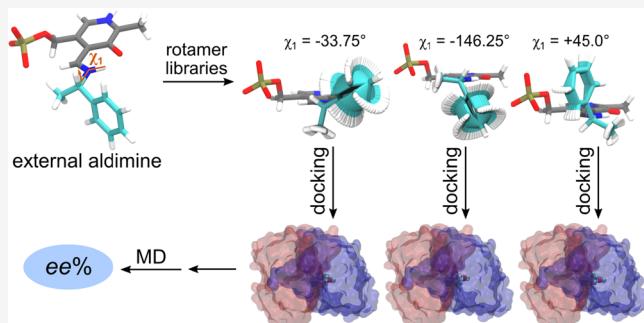
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:**  $\omega$ -Transaminases ( $\omega$ -TAs) catalyze the conversion of ketones to chiral amines, often with high enantioselectivity and specificity, which makes them attractive for industrial production of chiral amines. Tailoring  $\omega$ -TAs to accept non-natural substrates is necessary because of their limited substrate range. We present a computational protocol for predicting the enantioselectivity and catalytic selectivity of an  $\omega$ -TA from *Vibrio fluvialis* with different substrates and benchmark it against 62 compounds gathered from the literature. Rosetta-generated complexes containing an external aldimine intermediate of the transamination reaction are used as starting conformations for multiple short independent molecular dynamics (MD) simulations. The combination of molecular docking and MD simulations ensures sufficient and accurate sampling of the relevant conformational space. Based on the frequency of near-attack conformations observed during the MD trajectories, enantioselectivities can be quantitatively predicted. The predicted enantioselectivities are in agreement with a benchmark dataset of experimentally determined ee% values. The substrate-range predictions can be based on the docking score of the external aldimine intermediate. The low computational cost required to run the presented framework makes it feasible for use in enzyme design to screen thousands of enzyme variants.



## INTRODUCTION

$\omega$ -Transaminases ( $\omega$ -TAs) are pyridoxal-5'-phosphate-(PLP)-dependent enzymes that catalyze the reversible transfer of an amino group from an amino donor to a keto acceptor, yielding a keto product and a chiral amine in the process.<sup>1</sup> The high enantioselectivity of most  $\omega$ -TAs makes them attractive catalysts in the industrial production of chiral amines.<sup>2,3</sup> Classical chemical synthesis methods usually involve the use of expensive catalysts and hazardous conditions. Using  $\omega$ -TAs for the production of chiral amines can be advantageous because  $\omega$ -TAs offer high enantioselectivity, do not require external cofactors, allow high reactant concentrations, and give good product yields under mild reaction conditions.<sup>4</sup> However, the selectivity of  $\omega$ -TAs needs to be tailored to the substrate of interest for industrial application.

The range of substrates that can be accepted by  $\omega$ -TAs of the class III transaminases (fold-type I PLP enzymes) is mainly limited by the size of the small binding pocket, which rarely accommodates side chains larger than a methyl group. These  $\omega$ -TAs also have a large binding pocket that can easily accept more bulky substituents such as an aryl or alkyl group.<sup>5</sup> The size duality of the small and large binding pockets is what confers  $\omega$ -TAs their high enantioselectivity. Both binding pockets are hydrophobic, and both can be targeted to expand the substrate range of the enzymes.<sup>6,7</sup> Tailoring the reactivity of  $\omega$ -TAs to accept non-natural substrates is usually achieved via repetitive rounds of mutagenesis and testing. In general,

rational design and in silico approaches can be very effective for enzyme selectivity engineering but require case-specific fine-tuning to produce reliable results.<sup>8</sup> Therefore, we aimed to develop a computational framework that can be used in enzyme design for assessing the substrate scope and enantioselectivity of  $\omega$ -TA variants. The accuracy of the presented framework for the prediction of enantioselectivity and substrate scope was evaluated on a benchmark dataset of 62 compounds gathered from the literature.

The (*S*)-selective  $\omega$ -TA from *Vibrio fluvialis* (Vf-TA) was chosen as the model system for the benchmarking of the presented framework because it is a well-characterized enzyme. Vf-TA was first described in 1999 by Shin and Kim,<sup>9</sup> and a high-resolution crystal structure was reported in 2013.<sup>10</sup> Vf-TA has enzymatic activity toward both aliphatic and aromatic substrates,<sup>5</sup> generally producing highly enantiopure amines from ketones. Vf-TA tolerates mutations in both the small and large binding pockets. Tailoring Vf-TA to accept non-natural substrates via computational methods has previously been

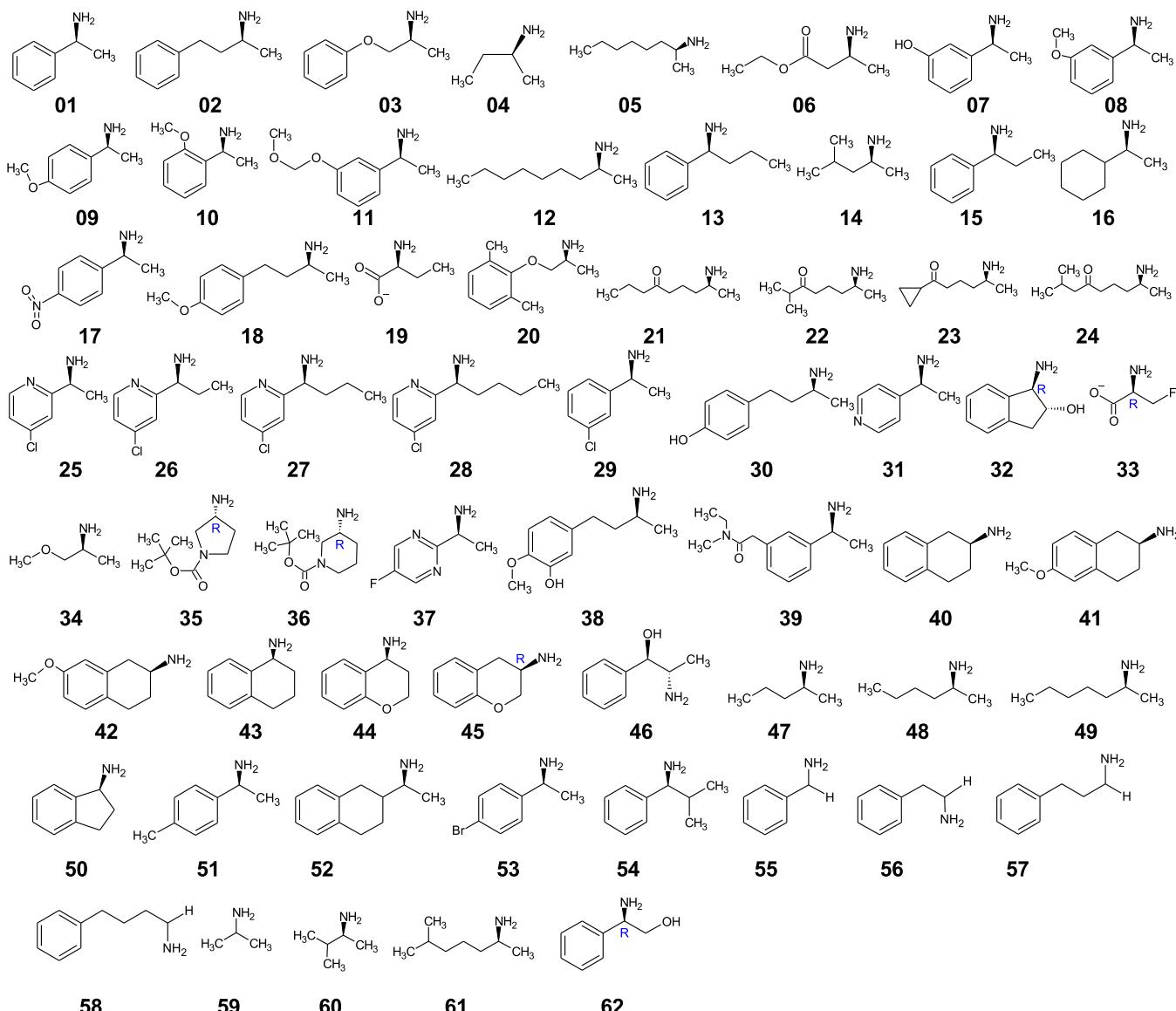
Received: May 28, 2021

Published: October 15, 2021



ACS Publications

© 2021 The Authors. Published by  
American Chemical Society



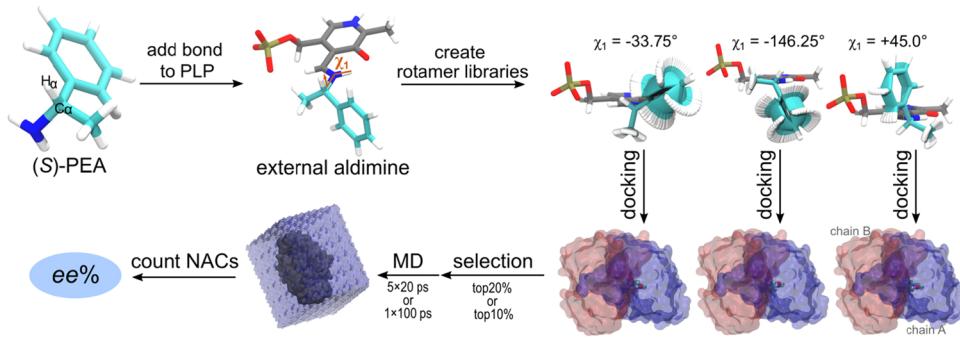
**Figure 1.** Compound dataset used in this study as benchmark to test the reliability of the developed protocol. The dataset was gathered from the existing literature, and sources are cited in Table 1. The enantiomeric excess for the asymmetric synthesis reactions used to obtain amines 01–49 from their respective ketones has been reported. Compounds 50–62 were only used for substrate-scope predictions because they are nonchiral or the experimental *ee*% for asymmetric synthesis was not found in the literature. All structures represent the preferred enantiomer of Vf-TA and are drawn with the bulkiest substituent to the left, i.e., in the position corresponding to the phenyl group of the (*S*)-enantiomer of 01. Following CIP rules, all compounds are (*S*)-amines, except 32, 33, 35, 36, 45, and 62 which formally are (*R*)-enantiomers.

attempted with varying degrees of success.<sup>11,12</sup> A common feature of these studies is their qualitative nature and the small range (less than 5) of screened substrates.<sup>11,13</sup>

Previous computational approaches aimed at understanding or engineering the reactivity of  $\omega$ -TAs have examined the keto substrate,<sup>11–14</sup> amino substrate,<sup>15</sup> quinonoid intermediate,<sup>12,16</sup> ketimine intermediate,<sup>17</sup> or external aldimine intermediate<sup>17</sup> as the ligand. The external aldimine intermediate was chosen for the current study because it is the first chiral intermediate in the second half reaction and is involved in the rate-limiting transamination step. Protonation of the quinonoid intermediate by the catalytic lysine leads to the formation of the external aldimine, which is subsequently cleaved to form the resting enzyme and the amine product. Cassimjee et al.<sup>1</sup> estimated the energy profiles of a half-transamination reaction catalyzed by an  $\omega$ -TA from *Chromobacterium violaceum* (Cv-TA) going

from (*S*)-1-phenylethylamine to acetophenone using density functional theory calculations of all the involved intermediates and transition states. The energy profile showed that the external aldimine intermediate is fenced by two large energy barriers corresponding to transition states in the forward or backward reaction direction, to form the quinonoid or the geminal diamine intermediates, respectively. Hence, we hypothesized that the enantiomer better accommodated by the enzyme in the external aldimine form would display a higher overall turnover rate.

To test the hypothesis that modeling the external aldimine intermediate is enough to predict the enantioselectivity (and substrate scope) of Vf-TA, we compared the results of the presented molecular modeling protocol against experimentally measured values for the asymmetric synthesis reactions (ketone  $\rightarrow$  amine). The approach makes three assumptions:



**Figure 2.** General overview of the presented framework for the enantioselectivity prediction of  $\omega$ -TAs. The  $\chi_1$  dihedral was set to fixed values,  $\chi_1 = \{-180^\circ, -168.75^\circ, -157.50^\circ, \dots, +180^\circ\}$ , adding up to 32 independent rotamer libraries per examined ligand.

(1) the energy profile for the half-transamination reaction is the same in Vf-TA as in Cv-TA; (2) the formation of all other reaction intermediates is not rate-limiting for the tested compounds; and (3) the forward and backward half-transamination reactions behave similarly. The first assumption is based on the fact that both Vf-TA and Cv-TA are (S)-selective  $\omega$ -TAs belonging to the PLP fold-type I superfamily class III in which active site groups and the catalytic mechanism are conserved.<sup>18</sup> The second assumption requires that the relative energy profile for the half-transamination reaction described for (S)-1-phenylethylamine to acetophenone<sup>1</sup> be maintained regardless of the substituents attached to the reacting C <sub>$\alpha$</sub> . The third assumption is based on the fact that the transamination reaction is reversible at every step of the catalytic cycle, and therefore, the principle of microscopic reversibility or detailed balance applies.<sup>19</sup> An additional advantage of modeling the external aldimine intermediate is the reduced computational cost. Because the C <sub>$\alpha$</sub>  is covalently bound via a Schiff base to the PLP cofactor, the search space in the docking stage is substantially reduced, thereby eliminating the need for initial placement algorithms.<sup>20</sup>

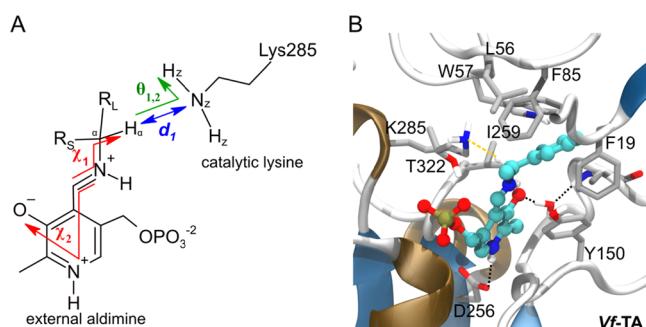
The results demonstrate the potential of using a combination of docking and molecular dynamics (MD) simulations for computational screening of different transaminase substrates. The developed protocol can give quantitative predictions of  $\omega$ -TA selectivity and enantioselectivity at a low computational cost, which can be employed in enzyme engineering efforts aimed at tailoring catalytic activity.

## METHODS

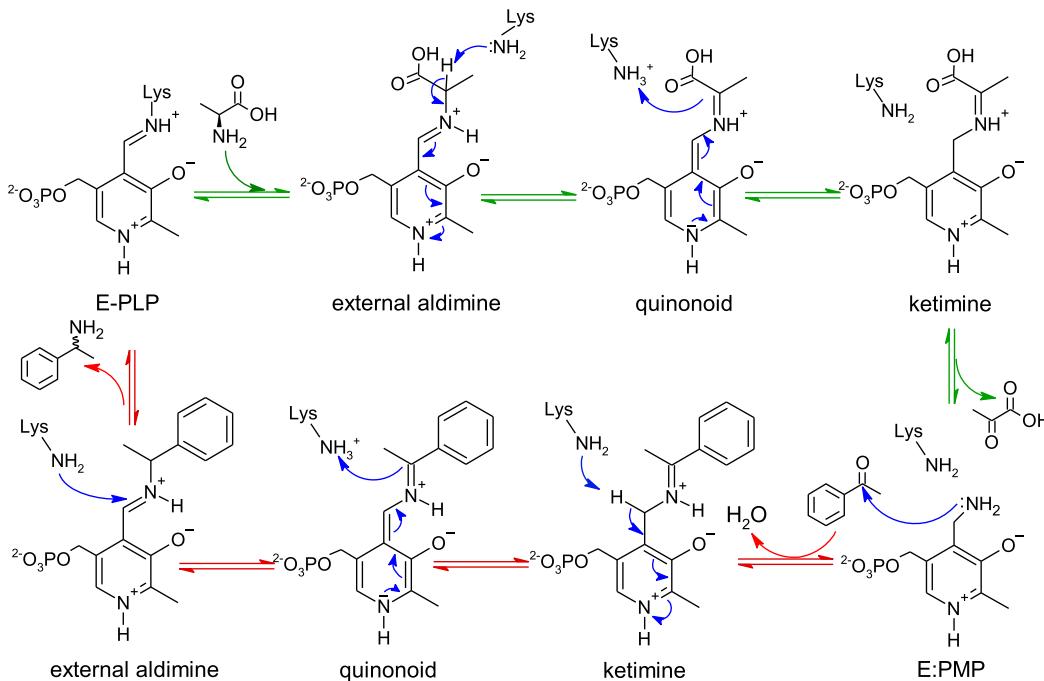
**Collection of the Benchmark Dataset.** A list of 62 compounds was gathered from existing literature to test the accuracy of the presented computational protocol (Figure 1). The gathered dataset contains experimentally determined values for enantiomeric excess (ee%) of amines (compounds 01–49) obtained in asymmetric synthesis catalyzed by Vf-TA. The ee% values obtained from kinetic resolution of racemic amines were not included in the dataset because of the dependence of observed ee% values on reaction progress. Additionally, data of compounds 50–62 were gathered for the substrate scope prediction task. The range of molecules explored is large, with the majority of the diversity coming from the substituent that binds to the large binding pocket. By contrast, the substituent that binds to the small binding pocket is a methyl or a short-chain alkyl group, with a few exceptions (e.g., compounds 28, 32, 33, 35, and 36).

**Preparation of Ligand Structures.** Initial atom coordinates for the (S)- and (R)-amine dataset (Figure 1) were

created using Avogadro software<sup>21</sup> and optimized with 500 steps of steepest descent minimization in the MMFF94 force-field.<sup>22</sup> The subsequent steps were made in YASARA ([www.yasara.org](http://www.yasara.org)).<sup>23</sup> Each amino compound was covalently bonded to the pyridoxal cofactor via a Schiff base to form the corresponding external aldimine complex, hereafter referred to simply as the ligand. The geometry of the ligand was then optimized in AM1 with implicit solvent.<sup>24</sup> Atomic charges were assigned for the ligand using the restrained electrostatic potential approach included in the Gaussian09 software package.<sup>25</sup> A library of relevant low-energy rotamers was generated for each compound by random perturbation of all torsion angles that contained at least one heavy atom (SampleDih YASARA routine), with the exception of ligand atoms originally belonging to the PMP cofactor. The latter were kept frozen during the rotamer generation stage. A total of 32 independent libraries were created for each ligand, with each library containing rotamers where the  $\chi_1$  dihedral was set to a fixed value ( $\chi_1 = \{-180^\circ, -168.75^\circ, -157.50^\circ, \dots, +180^\circ\}$ ) (Figure 2). The  $\chi_1$  dihedral indicates the orientation of the C <sub>$\alpha$</sub> -H <sub>$\alpha$</sub>  bond with respect to the PLP ring, and the precise definition is shown in Figure 3A. The initial size of each of the 32 libraries was 1000 rotamers, but after pruning for uniqueness (pair-wise RMSD between any two rotamers >0.005 Å) the size of each library ended in between 100 and 800 unique rotamers. In other words, the entire dataset



**Figure 3.** (A) An enzyme–ligand complex was considered to be in a NAC when the following criteria were simultaneously met:  $d_1 < 2.65$  Å,  $-75^\circ < \chi_1 < -105^\circ$ ,  $-15^\circ < \chi_2 < +15^\circ$ ,  $90^\circ < \theta_1 < 130^\circ$ , and  $90^\circ < \theta_2 < 130^\circ$ . Distances are colored in blue, angles in green, and dihedrals in red.  $\theta_1$  and  $\theta_2$  are the angles between the abstracting hydrogen H <sub>$\alpha$</sub>  the N <sub>$Z$</sub>  atom of Lys285, and either of the two H <sub>$Z$</sub>  atoms. R<sub>L</sub> = large substituent and R<sub>S</sub> = small substituent. (B) Structure of a NAC obtained from a simulation frame of the external aldimine of compound 01S.

Scheme 1. Transamination Reaction Mechanism<sup>a</sup>

<sup>a</sup>The fully reversible reaction consists of two half-transamination reactions.<sup>1</sup> In the first half reaction (green arrows), L-alanine is converted into pyruvate, resulting in the generation of the E/PMP intermediate (PMP is pyridoxamine-5'-phosphate). The second half (red arrows) consists of substrate entry and formation of the Michaelis complex with E:PMP. A nucleophilic attack on the carbonyl carbon of the substrate (acetophenone shown as an example) by the amino group of PMP leads to the formation of the ketimine intermediate. A further rearrangement involves the catalytic lysine (Lys285), acting first as a base and later as an acid, to form the quinonoid and external aldimine intermediates, respectively. Finally, Lys285 performs a nucleophilic attack on the iminium carbon of the external aldimine, resulting in the formation of amine (1-phenylethylamine) and an internal aldimine (E-PLP).

contained 49 unique chemical compounds, each compound had two external aldimine forms (one for the (S)- and one for the (R)-enantiomer), each of them had 32 unique rotamer libraries, and each of those libraries contained 100–800 unique rotamers.

**Preparation of the Enzyme Scaffold Structure.** The crystal structure of an (S)-selective  $\omega$ -TA from *Vibrio fluvialis* (PDB: 4E3Q) was used as scaffold.<sup>10</sup> Molecular modeling was performed using the dimer derived from subunits A and B. The dimer contains two PMP cofactors located in the subunit interface approximately 15 Å apart, forming two binding sites. Se-Met residues were converted into Met by the YASARA CleanObj routine. The protonation states of Asp, Glu, His, and Lys were predicted by a YASARA automated routine (*OptHyd*).<sup>26</sup> All Asp, Glu, and Lys were in their standard protonation states at neutral pH, and histidine residues were in the HIP (charge +1, both  $\delta$ - and  $\epsilon$ -nitrogens protonated; His at 26 and 178), HID (neutral,  $\delta$ -nitrogen protonated; His at 35, 151, and 362), and HIE (neutral,  $\epsilon$ -nitrogen protonated, His at 65, 83, 132, 183, 319, and 326) forms. All water molecules were removed from the crystal structure for the docking stage but were later added back in for the MD simulations.

**Rosetta Docking of the External Aldimines.** Docking of the external aldimine intermediates (ligand) into the active site of Vf-TA was performed with the Rosetta Enzyme Design application (build number 57927).<sup>27</sup> The following steps were performed on the binding site that is mainly composed of subunit A residues. Ligands were initially placed in the binding site by superimposing the pyridine ring heavy atoms to the equivalent PMP atoms found in the crystal structure. The

following command-line arguments were used when running Rosetta: `-enzdes`, `-cst_pseudock`, `-cst_design`, `-cut1 0.0`, `-cut2 0.0`, `-cut3 8.0`, `-cut4 10.0`, `-cst_min`, `-chi_min`, `-bb_min`, `-packing::use_input_sc`, `-packing::soft_rep_design`, `-design_min_cycles 3`, `-ex1:level 4`, `-ex2:level 4`, `-ex1aro:level 4`, `-ex2aro:level 4`.<sup>27</sup> For the enantioselectivity predictions, 10 docking structures were produced for each  $\chi_1$  dihedral, and from the resulting 320 structures ( $32 \times 10$ ), the top-scoring solutions, that is, the low energy conformations, were selected to serve as starting conformations for MD simulations. The selection was carried out on the basis of the Rosetta interface energy, adding up to a total of 64 (top20%) or 32 (top10%) starting structures for MD simulations. For substrate-range predictions, 200 Rosetta structures were generated for each enzyme–ligand complex (all with  $\chi_1 = -90^\circ$ ).

**MD Simulations of the Docked Structures.** Short MD simulations were performed using the top-ranking docked complexes to estimate the enantiomeric excess. All simulations were carried out in YASARA, using the knowledge-based YAMBER3 force field<sup>28</sup> for protein residues and the GAFF force field<sup>29</sup> for ligands. As for the identity of the ligands, the binding site of subunit A contained the docked external aldimine form of the query compound, while the binding site of subunit B contained the crystallographic PMP cofactor. The protein–ligand complexes were placed in a cubic simulation cell extending at least 10 Å beyond the protein. Crystallographic water molecules, formerly removed in the docking stage, were added back to the enzyme. Any water molecule with its center of mass closer than 1.1 Å from any non-water

**Table 1.** Comparison between the  $ee\%^{calc}$  and  $ee\%^{exp}$  Enantioselectivities for the Benchmark<sup>a</sup> Dataset<sup>15,47–65</sup>

no. <sup>b</sup>	%NACs		$ee\%^{calc}$	$ee\%^{exp}$	dev. <sup>c</sup> (kcal/mol)	no.	%NACs		$ee\%^{calc}$	$ee\%^{exp}$	dev. (kcal/mol)
	(R)	(S)					(R)	(S)			
01 <sup>15,47–50</sup>	0.156	1.518	82	99 <sup>d</sup>	1.8	26 <sup>50</sup>	0.171	1.058	72	99	2.1
02 <sup>47,51,52</sup>	0.155	1.782	84	84–99	0.0	27 <sup>50</sup>	0.230	2.511	83	99	1.7
03 <sup>47</sup>	0.035	2.958	98	97–99	0.0	28 <sup>50</sup>	0.671	1.919	49	81	0.7
04 <sup>47</sup>	1.361	4.859	56	93–99	1.2	29 <sup>50</sup>	0.003	1.836	100	99	0.0
05 <sup>47,52</sup>	0.085	1.920	92	92–96	0.0	30 <sup>53</sup>	0.073	1.462	91	95	0.4
06 <sup>47,54</sup>	0.272	0.746	47	96–99	1.7	31 <sup>50,55</sup>	0.333	2.608	78	99	1.9
07 <sup>54,56</sup>	0.008	0.999	99	98–99	0.0	32 <sup>57e</sup>	1.329	2.800	34	−98	−3.2
08 <sup>54,56</sup>	0.003	1.300	100	66–99	0.0	33 <sup>58e</sup>	7.301	0.163	−96	−98	−0.4
09 <sup>56</sup>	0.038	1.018	94	99	1.1	34 <sup>54</sup>	4.662	1.915	−42	40	1.0
10 <sup>54,56</sup>	0.000	1.386	100	96–99	0.0	35 <sup>51</sup>	0.189	0.468	42	−99	−3.7
11 <sup>56</sup>	0.000	2.031	100	95–99	0.0	36 <sup>51</sup>	1.059	1.255	9	−98	−2.8
12 <sup>53</sup>	0.044	3.338	98	90	−1.0	37 <sup>59</sup>	0.123	4.826	95	99	1.0
13 <sup>6</sup>	2.840	3.076	4	99	3.1	38 <sup>53</sup>	0.000	3.080	83	99	1.7
14 <sup>53</sup>	0.008	2.491	99	80	−1.8	39 <sup>60</sup>	0.000	2.103	100	99	0.0
15 <sup>15</sup>	2.089	0.955	−37	99	3.6	40 <sup>61</sup>	8.249	4.373	−31	18–54	0.6
16 <sup>53</sup>	0.013	1.106	98	99	0.4	41 <sup>61</sup>	1.687	2.673	23	76	0.9
17 <sup>53</sup>	0.000	0.931	100	100	0.0	42 <sup>61</sup>	1.800	1.855	1	99	3.1
18 <sup>53</sup>	0.138	1.600	84	90	0.3	43 <sup>61</sup>	0.967	0.769	−11	97–98	2.6
19 <sup>62</sup>	0.048	1.873	95	99	1.0	44 <sup>61</sup>	0.254	0.864	55	93–99	1.2
20 <sup>63</sup>	0.023	1.058	96	96	0.0	45 <sup>61e</sup>	9.819	2.502	−59	−40	0.8
21 <sup>64</sup>	0.069	1.726	92	99	1.3	46 <sup>65</sup>	0.021	0.137	68	99	2.2
22 <sup>64</sup>	0.003	3.591	100	99	0.0	47 <sup>53</sup>	1.168	4.177	56	15	−0.6
23 <sup>64</sup>	0.005	0.721	99	99	0.0	48 <sup>53</sup>	0.277	2.496	80	70	−0.3
24 <sup>64</sup>	0.036	2.685	97	98	0.2	49 <sup>53</sup>	0.264	3.806	87	90	0.2
25 <sup>50</sup>	0.000	4.403	100	99	0.0						

<sup>a</sup>The  $ee\%$  was calculated from the  $5 \times 20$  ps setup using eq 1. The  $ee\%^{calc}$  of most compounds closely matches the  $ee\%^{exp}$  compounds with deviations indicating an error in energy calculations larger than 2 kcal/mol shaded in gray. The formulas to calculate deviations are provided in the Supporting Information. <sup>b</sup>No: query compound number (Figure 1) with literature reference in superscript. <sup>c</sup>dev:  $\Delta\Delta G^\ddagger$  deviation (kcal/mol) between calculated and experimental values. <sup>d</sup>Positive  $ee\%$  values mean preference for the (S)-enantiomer. <sup>e</sup>Preferred enantiomer with a similar stereoconfiguration to (S)-01 but a different R/S notation because of the shift in CIP.

heavy atom was deleted from the simulation box. Enough TIP3P water molecules were added to fill the simulation cell, along with  $\text{Na}^+$  and  $\text{Cl}^-$  ions to a final concentration of 0.150 M. The  $pK_a$  values for residues Asp, Glu, His, and Lys were predicted using the YASARA *neutralization experiment* routine<sup>30</sup> at pH 8, resulting in standard protonation states for Asp, Glu, and Lys, and the following protonation states for histidine: HID (35, 151, and 362) and HIE (26, 65, 83, 132, 178, 183, 319, and 326). The protonation states of histidine residues 26 and 178 differed from the protonation states obtained using the *OptHyd* YASARA routine. The cell charge was neutralized with  $\text{Na}^+$  ions. Next, 25 steps of steepest descent minimization were carried out to remove conformational stress followed by simulated annealing to find an energy-minimum conformation. Simulated annealing was performed by slowly cooling down the system for 100 steps from 298 to 0 K by rescaling atom velocities by a factor of 0.9 each step, after which 100 steps of classical MD are performed at 298 K. After obtaining an energy-minimized conformation, the system was gradually warmed up from 0 to 298 K in 3.0 ps. A short equilibration of 2.0 ps (timestep = 2.0 fs) preceded the production run of 20.0 or 100.0 ps. The simulations were performed using the leap-frog integration step method of YASARA,<sup>31</sup> with a timestep of 2.5 fs. To gain performance, the nonbonded interactions were evaluated every second step. To enable such a large timestep for the intermolecular interactions, high-frequency bonds and angles (involving

hydrogens) were replaced by constraints. LINCS was used to constrain all covalent bonds involving hydrogen atoms with a relative geometric tolerance of  $10^{-4}$ .<sup>32</sup> Pressure (1.0 bar) and temperature (298 K) were maintained with a modified Berendsen barostat and thermostat, respectively.<sup>28,33</sup> Long-range electrostatic interactions were treated with the particle mesh Ewald algorithm and short-range interactions with direct-space Coulombic interactions with a cutoff of 8.0 Å.<sup>34</sup> The production run was 20 ps in length with five seeds ( $5 \times 20$  ps setup) or 100 ps with one seed ( $1 \times 100$  ps setup), bringing the total simulation time to 6400 ps per examined ligand in either setup.

**Calculation of  $ee\%$  from MD Trajectories.** The MD trajectories of each ligand were scored by counting the number of frames in which a near-attack-conformation (NAC) was observed (Figure 3). NACs were measured on-the-fly every 20 fs, adding up to 1000 datapoints for a 20 ps trajectory. For each compound (01–49 in Figure 1), the relative frequency of NACs produced by the docked complexes of the (S)- and (R)-enantiomer was used to calculate the expected  $ee\%$  and compared against the experimentally determined value.

**NAC Criteria.** NACs were defined by a combination of geometric parameters important for the transamination reaction to take place, in particular the nucleophilic proton abstraction of the external aldimine intermediate to form the quinonoid intermediate (Scheme 1). In order for Lys285 to perform a nucleophilic attack on the external aldimine, the

reacting atoms must be positioned at a short distance, where their van der Waals radii start overlapping ( $d_1$ ). Furthermore, the cleavable bond  $H_\alpha-C_\alpha$  needs to be positioned perpendicular to the Schiff base ( $\chi_1 \approx -90^\circ$ ) to allow the nascent p orbital to align with the electrons of the PLP-conjugated  $\pi$  system ( $\chi_2 \approx 0^\circ$ ). Additionally, the free electrons of the attacking lysine nitrogen ( $:N_Z$ ) need to be pointing directly toward  $H_\alpha$  ( $\theta_1, \theta_2 > 90^\circ$ ). The complete set of geometric criteria for a frame to be considered a NAC is listed in Figure 3. The precise criteria have been found to have a large influence on the absolute NAC values but not on the relative ratios observed between different states that are compared.<sup>35–43</sup>

## RESULTS

**General Approach.** As outlined in Figure 2, the first step of the approach for predicting the activity of Vf-TA toward a query compound was to generate enzyme–ligand conformations by Rosetta docking. The ligand is the external aldimine intermediate of the query compound. The generated docked structures were ranked based on the Rosetta interface energy, and the top-scoring structures (top10% or top20%) were used as starting conformations for MD simulations. The MD trajectories (1 × 100 ps or 5 × 20 ps) were scored based on a set of geometric parameters. In particular, we looked at the occurrence of NACs during the trajectories. NACs are ground-state conformers that lie close to the transition path of a chemical reaction.<sup>44</sup>

For prediction of stereoselectivity, an ee% value is obtained from MD simulations by combining the proportion of NACs found in the simulation for the (R)- and the (S)-enantiomer, using the following formula:

$$ee\%_{\text{calc}} = \left( \frac{NAC_S - NAC_R}{NAC_S + NAC_R} \right) \times 100\% \quad (1)$$

We thus use docking algorithms to reach the ground-state conformations of the external aldimine and MD simulations to sample the Boltzmann distribution of ground-state conformers and their propensity to adopt reactive poses. The proportion of ground-state conformers that are NACs gives an estimate of the free energy of NAC formation.<sup>45</sup> Comparing the free energy of NAC formation for the (R)- and (S)-enantiomer gives a direct estimate of the ee%. The use of NAC frequencies as indication of enzymatic selectivity has been adopted successfully in previous studies on a variety of enzymes.<sup>36–42,46</sup>

For the substrate-range prediction, we found the Rosetta interface energy between the enzyme and the external aldimine form of the query compound to be in agreement with the experimental activities of the whole transamination reaction. Conversely, there was no apparent correlation between the percentage of NACs and the enzymatic activity (results not shown).

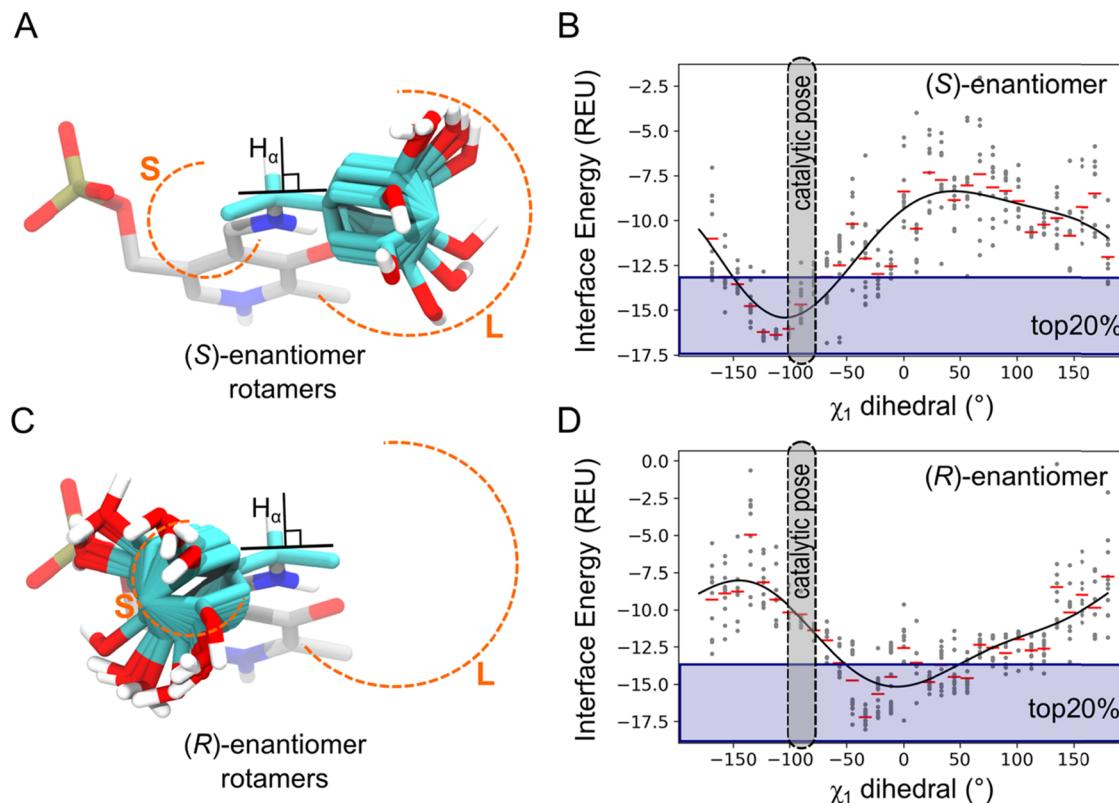
**Enantioselectivity Predictions with the Benchmark Compounds.** To estimate the ee% of compounds 01–49, the best Rosetta structures were selected to serve as starting conformations for short MD simulations, and the trajectories were scored by computing the NAC occurrence. The enantiomeric excess was calculated with eq 1, where the number of NACs produced by each enantiomer is thought of as being an indication of the ability of the enantiomer to adopt a conformation optimal for catalysis. The calculated ee% ( $ee\%_{\text{calc}}$ ) was then compared against the experimental ee% values

( $ee\%_{\text{exp}}$ ). In most cases, the  $ee\%_{\text{calc}}$  matched the expected values (Table 1).

The high enantioselectivity of  $\omega$ -TAs is an attractive feature in the synthesis of chiral amines but also hinders validation of the presented protocol because the dataset gathered from the literature is strongly unbalanced. Most of the compounds in the presented dataset have an  $ee\%_{\text{exp}}$  of around 99%, and the  $ee\%_{\text{calc}}$  values are correctly predicted to be in this range. The enantioselectivity calculations situated the numeric  $ee\%_{\text{calc}}$  values of most compounds as large and positive (a positive sign means (S)-enantiopreference, and a negative sign means (R)-enantiopreference), which is in agreement with the highly (S)-enantioselective nature of Vf-TA (Table 1). Compounds 32, 33, 35, 36, and 45 are the exception, but 32, 33, and 45 are only designated as (R)-amines because of CIP rules and retain a similar stereoconfiguration to (S)-01 (the bulky substituent binds in the large pocket). By contrast, compounds 35 and 36 do not maintain the stereoconfiguration of (S)-01. Here, the  $ee\%_{\text{calc}}$  for 35 (42%) and 36 (9%) was not in agreement with the  $ee\%_{\text{exp}}$  (−99 and −98%, respectively). While the predictions incorrectly favor the (S)-enantiomer, the experiments suggested that the (R)-enantiomer was preferred in both 35 and 36. The  $ee\%_{\text{exp}}$  reported by Höhne et al.<sup>51</sup> suggests that Vf-TA would strongly prefer to place the N-Boc substituent of compounds 35 and 36 in the small binding pocket to yield the corresponding amines. Nevertheless, structural analysis of the docked complexes of ligands 35(R/S) and 36(R/S) did not provide an explanation as to why Vf-TA should prefer one enantiomer over the other (Figure S1).

Additionally, it can be noted that there are a few compounds in the dataset with low absolute  $ee\%_{\text{exp}}$  values (i.e., 34, 40, 41, 45, 47, and 48) that were correctly predicted to have a poor enantioselectivity. For example, the  $ee\%_{\text{exp}}$  of 47 is 15%, which suggests that Vf-TA does not produce an enantiopure mixture of amines (ratio of (S)- to (R)-amine is 1.35:1), and the  $ee\%_{\text{calc}}$  has a low absolute value of 56% (3.5:1). On the other hand, there are a few compounds in the dataset with high absolute  $ee\%_{\text{exp}}$  values (i.e., 13, 15, 32, 35, 36, 42, and 43) that were incorrectly predicted to have much lower  $ee\%_{\text{calc}}$  values. For example, the  $ee\%_{\text{exp}}$  of 13 (99%) suggests a strong preference for the (S)-enantiomer, but the  $ee\%_{\text{calc}}$  (4%) would wrongly predict that Vf-TA produces both enantiomers in almost equal amounts. Compounds 35 and 36 were discussed earlier, and an explanation for the failure of the protocol in predicting the enantiopreference of compounds 13 and 15 could not be found.

The enantioselectivity predictions for compounds containing the indan or tetralin moieties (i.e., 32, 40–45) were generally not in agreement with the reported experimental values. For example, for compounds 42 and 43, Vf-TA was expected to have a strong preference toward the (S)-enantiomer, but the  $ee\%_{\text{calc}}$  values (1 and −11%, respectively) suggest no preference. Inspection of the docked structures provided no explanation for these inaccurate  $ee\%_{\text{calc}}$  values. Within the tetralin compounds, the  $ee\%_{\text{calc}}$  (−31%) for 40 deviates in 49 percentual points from the  $ee\%_{\text{exp}}$  (18%), which might lead to the conclusion that the prediction for 40 is incorrect. However, because the enantiomeric excess does not follow a linear scale, these deviations need to be interpreted accordingly. The  $ee\%_{\text{calc}}$  for compound 40 corresponds to a small enantiopreference for the (R)-enantiomer (enantiomeric ratio 0.52:1), whereas the  $ee\%_{\text{exp}}$  suggests a slight enantiopreference for the (S)-enantiomer (1.4:1). Accordingly, the

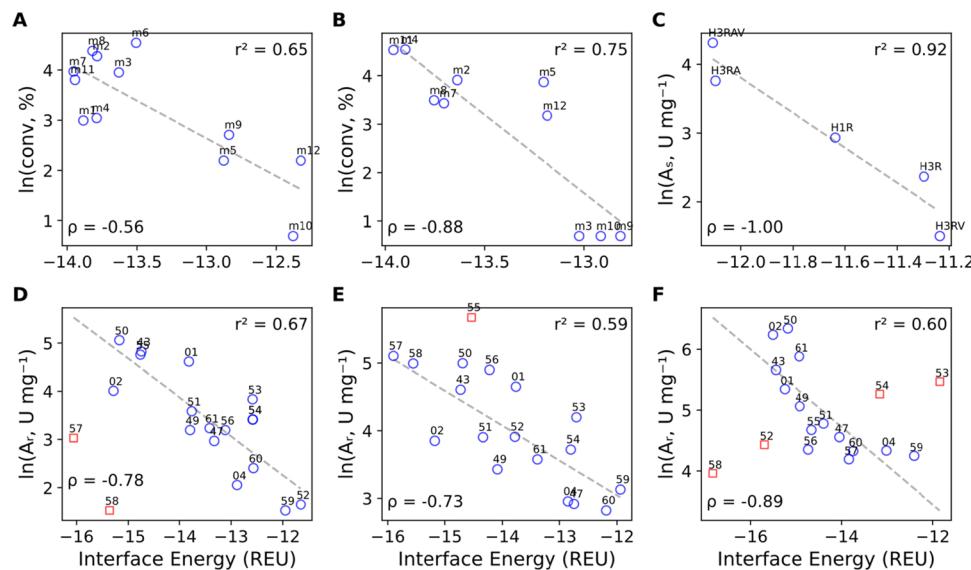


**Figure 4.** Comparison between the (R)- and (S)-enantiomers. (A) Preferred enantiomer (the (S)-enantiomer of compound 07) can position  $H_{\alpha}$  perpendicular to the plane of the PMP ring and ready for proton abstraction by Lys285. (B) Preference for the preferred enantiomer to adopt a catalytic orientation ( $\chi_1 = -90^\circ$ ) is evident when comparing the Rosetta score at different  $\chi_1$  values. (C) By contrast, the nonpreferred enantiomer is unable to adopt a catalytic orientation ( $\chi_1 = -90^\circ$ ), where the  $H_{\alpha}$  is positioned close to the catalytic lysine, because of steric hindrance with the small binding pocket (S: small, L: large binding pockets). (D) Scanning through the  $\chi_1$  dihedral shows that the energy minimum is not located near the optimal value for catalysis to take place. In panels (A) and (C), carbon atoms originally belonging to the cofactor are colored gray, while carbon atoms originally belonging to the amino compound are colored cyan. All panels correspond to compound 07, shown as an example.

energy deviations suggest that the enantioselectivity calculations are rather accurate (0.6 kcal/mol deviation), and the same holds for compounds 34, 45, and 47. Gawley<sup>66</sup> provides an excellent overview of the limitations of using enantiomeric excess as a metric to describe the enantiomeric composition. For that reason, we also calculated the  $\Delta\Delta G^{\ddagger}$  deviations from experimental data (see the Supporting Information for calculation details). In many cases (26 out of 49), the deviations are below  $\pm 1.0$  kcal/mol (Table 1), which corroborates the predictive power of the presented methodology for modeling the enantioselectivity of  $\omega$ -TAs.

The ability of Arg415 to form salt bridge interaction with carboxyl groups is corroborated by the enantio preference of Vf-TA toward compounds 19 (L-alanine) and 33 (3-fluoroalanine). Synthesis of (S)-19 and (R)-33 requires placing the carboxyl substituent in the large binding pocket, where Arg415 is located (Figure S2). The formation of a salt bridge between the guanidinium group of Arg415 and the  $\text{COO}^-$  group of the ligand to occur, the carboxyl substituent needs to be deprotonated. In fact, when the simulations (docking and MD) were performed with the carboxyl groups in the neutral form ( $\text{COOH}$ ), the  $ee\%_{\text{calc}}$  of compounds 19 and 33 was found to be 76 and -36%, respectively. The results presented in Table 1 are from simulations with the deprotonated form ( $\text{COO}^-$ ), where the preference is toward salt bridge formation ( $ee\%_{\text{calc}}$  was 95 and -96% for compounds 19 and 33, respectively).

**Additional Simulation Setups.** Additional simulation setups were used to test the robustness of the presented protocol (Table S1). We tested whether the number and length of the simulations ( $5 \times 20$  ps vs  $1 \times 100$  ps), or the fraction of docked structures selected for MD (top20% vs top10%), affected the outcome of the calculations. When comparing the  $5 \times 20$  ps against the  $1 \times 100$  ps setups, no large differences were found in the predicted  $ee\%$  values. In fact, NAC-producing frames were found to be evenly spaced across the MD trajectories (Figure S5), which means that the percentage of NACs produced by a trajectory was not dependent on the simulation length. Still, the length of the simulations should be long enough to generate conformational sampling of the ground-state region set at the docking stage, but short enough to maintain a conformation close to the initial structure. Our research group has previously shown that ps-scale MD simulations are long enough to allow meaningful counting of binding poses and that running multiple-independent MD simulations is better for sampling than performing a single long simulation.<sup>36,37,46,67</sup> Additionally, no large difference was found when comparing the  $ee\%$  obtained from using the top10% or top20% of Rosetta structures for MD simulations (Table S1). Again, only low-energy conformations should be used to avoid working with unfavorable ground states (unfavorable Rosetta interface energies), but the structures should be diverse enough to prevent single outliers from dominating the averaged frequency



**Figure 5.** Correlation between the Rosetta interface energy and substrate range. Panels A–C show the relation between the Rosetta score of the external aldimine of compounds **13**, **62**, and **54**, respectively, and the experimental conversion across different *Vf*-TA mutants. In panels D–F, we show the relation between the Rosetta score of wild-type *Vf*-TA, *Vf*-TA + W147G, and *Vf*-TA + W57G, respectively, and the experimental activity on a dataset of 17 compounds. All Rosetta calculations were performed using the external aldimine of the (*S*)-amine. The red squares are outliers, that is, datapoints that do not follow the overall trend and were not used to do the linear regression (dotted lines). Experimental data were obtained from the existing literature: panels A and B from Nobili et al.,<sup>6</sup> panel C from Genz et al.,<sup>15</sup> and panels D and F from Cho et al.<sup>5</sup> Abbreviations are as follows. For panels A and B: m<sub>1</sub>, wild-type; m<sub>2</sub>, mutant V153A; m<sub>3</sub>, mutant F85L; m<sub>4</sub>, mutant Y150F; m<sub>5</sub>, mutant Y150M; m<sub>6</sub>, mutant F85L\_V153A; m<sub>7</sub>, mutant F85L\_Y150F; m<sub>8</sub>, mutant F85L\_Y150F\_V153A; m<sub>9</sub>, mutant F85L\_Y150M; m<sub>10</sub>, mutant F85L\_Y150M\_V153A; m<sub>11</sub>, mutant Y150F\_V153A; m<sub>12</sub>, mutant Y150M\_V153A. For panel C: H<sub>1R</sub>, mutant L56V\_W57F\_F85V; H<sub>3R</sub>, mutant L56V\_W57C; H<sub>3RV</sub>, mutant L56V\_W57C\_F85V; H<sub>3RA</sub>, mutant L56V\_W57C\_V153A; H<sub>3RAV</sub>, mutant L56V\_W57C\_F85V\_V153A. For panels D, E, and F, the numbers correspond to the dataset presented in Figure 1. REU: Rosetta Energy Units., A<sub>s</sub>: specific activity, A<sub>r</sub>: relative specific activity, and ρ: Spearman's correlation coefficient.

of NAC conformations. Within the set of selected structures, the number of NACs that each trajectory produced was not correlated with the Rosetta interface energy of the starting structure (Figure S6). All in all, the choice of limiting the proportion of docking structures (top20%) to be used as starting conformations for multiple short MD simulations (5 × 20 ps) was adequate.

**Sampling of the  $\chi_1$  Dihedral Is Key to Modeling the Enantioselectivity of *Vf*-TA.** In this section, we provide further results that illustrate why sampling the  $\chi_1$  dihedral using the Rosetta energy function can be advantageous.

The most important difference between the preferred and nonpreferred enantiomers is that the former can position their cleavable H <sub>$\alpha$</sub> –C <sub>$\alpha$</sub>  bond perpendicular to the PMP plane, ready for proton abstraction by Lys285. By contrast, the nonpreferred enantiomers cannot position the H <sub>$\alpha$</sub>  close to Lys285, because this would require placing the large substituent in the small binding pocket, which is unfavorable (Figure 4). Therefore, the initial approach for modeling the enantioselectivity of *Vf*-TA was to dock the ligands of compounds **01**–**49** in a conformation that resembled as much as possible a catalytic conformation (setting  $\chi_1$  to  $-90^\circ$ ) and allow the nonpreferred enantiomer to reorient itself into a more favorable (and probably noncatalytic) conformation during the simulation (Figure 4A,C). In this way, the fraction of NACs produced by the nonpreferred enantiomer would be lower than the fraction produced by the preferred enantiomer (NAC criterium  $\chi_1$ ). However, we found that the nonpreferred enantiomer (i.e., the (*R*)-enantiomer) was not able to invert to a more favorable conformation in simulation time scales realistic for application in rapid computational screening (i.e.,

tens of ps). When the external aldimine form of both enantiomers was docked in a catalytic orientation ( $\chi_1 = -90^\circ$ ), followed by MD simulations, the ee%<sub>calc</sub> showed no correlation with the ee%<sub>exp</sub> values (Table S1). Hence, docking all external aldimine intermediates in a catalytic orientation was unsuitable for discerning between reactive and nonreactive enantiomers.

Longer MD simulations of the external aldimine complex of compounds **01R** and **01S** showed that the  $\chi_1$  dihedral evolves too slowly to obtain convergence in short time scales (Figure S7). We sorted out this limitation by using the Rosetta energy function to scan through the  $\chi_1$  dihedral and produce enzyme–ligand complexes with staggered  $\chi_1$  values ( $\chi_1 = \{-180^\circ, -168.75^\circ, -157.50^\circ, \dots, +180^\circ\}$ ) (Figure 2). The top-scoring Rosetta structures (top20%: 64 out of 320 total structures; top10%: 32 out of 320 total structures) were then used as starting conformations for MD simulations (Figure 4B,D). This way, the docking algorithm makes larger jumps in conformational space, and the short MD simulations subsequently sample the local region of said conformational space, yielding a fast and accurate protocol to model  $\omega$ -TA enantioselectivity.

**Influence of NAC Definitions on ee%<sub>calc</sub>.** Among the set of geometric criteria ( $d_1$ ,  $\theta_1$ ,  $\theta_2$ ,  $\chi_1$ , and  $\chi_2$ ), the dihedral  $\chi_1$  was the criterion with greater influence on the number of NACs produced in the trajectory because of its slow dynamics. The MD simulations (20–100 ps) were not of sufficient length as to allow  $\chi_1$  to move away from the starting value set at the docking stage (Figure S3). Hence, ligands docked in a catalytic conformation ( $\chi_1 = -90^\circ$ ) were more likely (but not guaranteed) to produce NACs than ligands docked in a

noncatalytic conformation ( $\chi_1 \neq -90^\circ$ ). Because a ligand docked in a catalytic conformation is not necessarily guaranteed to produce NACs (as NAC production also depends on other geometric parameters), performing MD simulations on the docked structures is still necessary to determine  $ee\%_{\text{calc}}$ . Similarly, a ligand docked in a noncatalytic conformation is still capable of producing NACs (Figure S4). Additionally, MD simulations allow the inclusion of explicit water molecules. All in all, the presented strategy of scanning the  $\chi_1$  dihedral landscape using the Rosetta energy function and then performing more accurate ps-scale MD simulations of the top-scoring docked structures enables a fast and accurate way of modeling the enantioselectivity of  $\omega$ -TAs.

**Substrate Scope Prediction Using the Rosetta Interface Energy.** After examining enantioselectivity, we investigated if the sole modeling of the external aldimine intermediate would suffice to predict the substrate range. Six independent sets of substrates were examined. The first three sets keep the query compound constant but explore different *Vf*-TA variants (Figure 5A–C), and the remaining three sets keep the enzyme variant constant but explore different compounds (Figure 5D–F). The Rosetta interface energy was found to be well correlated with the observed experimental activity for the six independent sets of compounds, once major outliers were omitted from the latter three datasets. The correlation is surprising because the interface energy only accounts for interactions between the ligand and the protein while ignoring other factors that may cause differences in catalytic activity, such as differences in the chemical potential of the reacting compounds or in desolvation free energy upon substrate binding across substrates or mutants.<sup>68</sup> Calculating only the interface energy of the external aldimine complex also ignores other steps in the catalytic cycle for a given substrate to undergo transamination, such as formation of the Michaelis complex or formation of reaction intermediates other than the external aldimine. However, because all these reaction intermediates are structurally similar (Scheme 1), a good interaction energy for the modeled intermediate might also reflect a good interaction energy for the nonmodeled intermediates. Whereas substrate binding (affinity) would mainly influence  $K_M$ , the affinity of an intermediate can also influence  $k_{\text{cat}}$ , which would be in agreement with inspecting the relationship between the Rosetta interface energy and experimental data, namely, yield and conversion. Another point to consider is the scale on which the observed correlation between Rosetta scores and enzymatic activity exhibits monotonicity. On short scales, two mutants with different activities may receive a similar Rosetta score (e.g., mutants m10 and m12 in Figure 5A, m3 vs m12 and m5 in Figure 5B, and H3RV vs H3R in Figure 5C). Despite these potential pitfalls, the good correlation found between interface energies and experimental substrate specificity suggests that the Rosetta score can be used as the objective function to guide the design of  $\omega$ -TA mutants.

## ■ DISCUSSION

Modeling the enzymatic enantioselectivity is a complicated task because of the small energy differences that need to be reproduced and the number of intermediates and transition states that need to be considered. Very accurate but computationally demanding methods, such as quantum mechanics (QM), molecular mechanics, empirical valence bond, and the quantum chemical cluster approach, have been

used successfully in the study of enzymatic enantioselectivity.<sup>69</sup> In this work, we explored a strategy that combines molecular docking and MD simulations to yield a computationally inexpensive but accurate framework for predicting the specificity of *Vf*-TA in the production of chiral amines. Modeling was performed using only the external aldimine intermediate of the query compounds. Docking allows a computationally efficient generation of ground-state conformations, while the MD simulations quantify the Boltzmann distribution of ground-state conformers, including the occurrence of reactive poses. Estimating the proportion of ground-state conformers that resemble a NAC provides an indication of the free energy of NAC formation<sup>45</sup> and a direct means for predicting the  $ee\%$  (Eq 1). The MD simulations are intended to account for the loss of accuracy arising from the discreteness of rotamers of docking approaches<sup>70,71</sup> and the lack of explicit water molecules.<sup>72</sup> The computational cost of the presented framework is around 37 CPU-h (using an HP workstation Z4 with Intel Xeon W-2135 Processor with six CPUs) per assayed compound (setup: 5 × 20 ps, top20%), and parallel computing is possible.

With six independent compound sets, it was shown that the Rosetta interface energy, measured between the enzyme variant and the external aldimine form of the query compound, correlates well with the experimental substrate scope of *Vf*-TA (Figure 5). At least for the presented datasets, this correlation shows that modeling the amino donor (typically L-alanine or isopropylamine) is not necessary for the reactivity prediction task, in contrast to the approach of Seo et al.,<sup>49</sup> where differential binding energies were considered for the substrates of both half-transamination reactions. We recently also reported the correlation between interface energies and experimental conversion in other *Vf*-TA and *Pj*-TA datasets<sup>73</sup> and used the Rosetta interface energy as the target function for the computational engineering of the *Pj*-TA substrate scope.<sup>68</sup> The protocol allowed the rapid design of *Pj*-TA variants accepting sterically hindered substrates. The computational cost of scanning for enzyme variants using Rosetta interface energy is circa 2 CPU-h per enzyme variant. Generating every enzyme–ligand structure takes approximately 60 s, but producing at least 100 structures per scanned variant is recommended. Other more sophisticated methodologies for predicting  $\omega$ -TA reactivity can be useful in later stages of the search for new mutants but tend to be more expensive. For example, the methodology of Voss et al.,<sup>13</sup> which used a combination of QM calculations and MD simulations, has a computational cost of around 60 CPU-h per enzyme variant.

Although validation was made solely on a *Vf*-TA dataset, the presented approach may be applicable to other (S)-selective  $\omega$ -TAs owing to their similarities. We have tested the substrate scope prediction task in *Pj*-TA<sup>68,73</sup> and *Cv*-TA.<sup>73</sup> In contrast, this approach may or may not be applicable to (R)-selective  $\omega$ -TAs, because the two enzymes evolved separately and belong to a completely different PLP fold type (fold type IV).<sup>18</sup>

## ■ CONCLUSIONS

We presented a framework for calculating the enantioselectivity of an  $\omega$ -TA from *V. fluvialis* toward a variety of compounds. The framework is computationally inexpensive and can assist in the enzyme design of  $\omega$ -TAs. In general, the retrospectively predicted  $ee\%$  values were in agreement with those found in the literature, but there were cases where the predictions did not correspond to the expected values. Furthermore, we

showed that the Rosetta interface energy of the external aldimine complex is a good predictor of the activity of Vf-TA across mutants or across different compounds. Using the Rosetta interface energy as the target scoring function can therefore aid in the search for enzyme variants with a broadened substrate range.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00617>.

Equations to calculate  $\Delta\Delta G^\ddagger$  deviations; structures of docked complexes; plots showing the variation of the  $\chi_1$  dihedral at the start and end of the MD simulations; 3D histogram of the dependency of NACs and docked dihedral; heatmaps showing the NAC count independency on the simulation time window and on the Rosetta score of the starting conformation; time evolution of the  $\chi_1$  dihedral in long MD simulations; and table comparing the  $ee\%$  obtained from different setups ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Author

Dick B. Janssen — *Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands;*  [orcid.org/0000-0002-0834-2043](https://orcid.org/0000-0002-0834-2043); Email: [d.b.janssen@rug.nl](mailto:d.b.janssen@rug.nl)

### Authors

Carlos Ramírez-Palacios — *Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands; Molecular Dynamics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands;*  [orcid.org/0000-0003-2724-0760](https://orcid.org/0000-0003-2724-0760)

Hein J. Wijma — *Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands;*  [orcid.org/0000-0003-0891-6972](https://orcid.org/0000-0003-0891-6972)

Sebastian Thallmair — *Molecular Dynamics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands; Frankfurt Institute for Advanced Studies, 60438 Frankfurt am Main, Germany;*  [orcid.org/0000-0002-3396-5840](https://orcid.org/0000-0002-3396-5840)

Siewert J. Marrink — *Molecular Dynamics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, 9747 AG Groningen, The Netherlands;*  [orcid.org/0000-0001-8423-5277](https://orcid.org/0000-0001-8423-5277)

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jcim.1c00617>

### Author Contributions

C.R.P. performed the computational work and wrote the manuscript. H.J.W. supervised docking and initial MD simulations. H.J.W. and D.B.J. conceived the project. S.T. supervised the MD simulation work. D.B.J. supervised the overall project. D.B.J. and S.J.M. revised the manuscript.

### Notes

The authors declare no competing financial interest.

Scripts necessary for running the presented protocol are publicly available at: <https://github.com/crp-mol/transaminases>. Usage instructions and example input and output files are provided.

## ■ ACKNOWLEDGMENTS

C.R.P. thanks CONACYT for the doctoral fellowship. We thank the Center for Information Technology (CIT) of the University of Groningen for providing access to the Peregrine high-performance computing cluster.

## ■ ABBREVIATIONS

$\omega$ -TA,  $\omega$ -transaminase; PLP, pyridoxal 5-phosphate; PMP, pyridoxamine 5-phosphate; E, enzyme; Vf-TA,  $\omega$ -TA from *Vibrio fluvialis*; Cv-TA,  $\omega$ -TA from *Chromobacterium violaceum*; Pj-TA,  $\omega$ -TA from *Pseudomonas jessenii*; NAC, near-attack conformation; ee%, percentual enantiomeric excess; REU, Rosetta energy units; MD, molecular dynamics; QM, quantum mechanics.

## ■ REFERENCES

- (1) Cassimjee, K. E.; Manta, B.; Himo, F. A Quantum Chemical Study of the  $\omega$ -Transaminase Reaction Mechanism. *Org. Biomol. Chem.* **2015**, *13*, 8453–8464.
- (2) Breuer, M.; Ditrich, K.; Habicher, T.; Hauer, B.; Kefeler, M.; Stürmer, R.; Zelinski, T. Industrial Methods for the Production of Optically Active Intermediates. *Angew. Chem., Int. Ed.* **2004**, *43*, 788–824.
- (3) Kelly, S. A.; Pohle, S.; Wharry, S.; Mix, S.; Allen, C. C. R.; Moody, T. S.; Gilmore, B. F. Application of  $\omega$ -Transaminases in the Pharmaceutical Industry. *Chem. Rev.* **2018**, *118*, 349–367.
- (4) Tufvesson, P.; Lima-Ramos, J.; Jensen, J. S.; Al-Haque, N.; Neto, W.; Woodley, J. M. Process Considerations for the Asymmetric Synthesis of Chiral Amines Using Transaminases. *Biotechnol. Bioeng.* **2011**, *108*, 1479–1493.
- (5) Cho, B.-K.; Park, H.-Y.; Seo, J.-H.; Kim, J.; Kang, T.-J.; Lee, B.-S.; Kim, B.-G. Redesigning the Substrate Specificity of  $\omega$ -Aminotransferase for the Kinetic Resolution of Aliphatic Chiral Amines. *Biotechnol. Bioeng.* **2008**, *99*, 275–284.
- (6) 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.
- (7) Genz, M.; Vickers, C.; van den Bergh, T.; Joosten, H.-J.; Dörr, M.; Höhne, M.; Bornscheuer, U. Alteration of the Donor/Acceptor Spectrum of the (S)-Amine Transaminase from *Vibrio fluvialis*. *Int. J. Mol. Sci.* **2015**, *16*, 26953–26963.
- (8) Chowdhury, R.; Maranas, C. D. From Directed Evolution to Computational Enzyme Engineering—A Review. *AIChE J.* **2020**, *66*, No. e16847.
- (9) Shin, J.-S.; Kim, B.-G. Asymmetric Synthesis of Chiral Amines with  $\omega$ -Transaminase. *Biotechnol. Bioeng.* **1999**, *65*, 206–211.
- (10) 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.
- (11) Sirin, S.; Kumar, R.; Martinez, C.; Karmilowicz, M. J.; Ghosh, P.; Abramov, Y. A.; Martin, V.; Sherman, W. A Computational Approach to Enzyme Design: Predicting  $\omega$ -Aminotransferase Catalytic Activity Using Docking and MM-GBSA Scoring. *J. Chem. Inf. Model.* **2014**, *54*, 2334–2346.
- (12) Dourado, D. F. A. R.; Pohle, S.; Carvalho, A. T. P.; Dheeman, D. S.; Caswell, J. M.; Skvortsov, T.; Miskelly, L.; Brown, R. T.; Quinn, D. J.; Allen, C. C. R.; Kulakov, L.; Huang, M.; Moody, T. S. Rational

- Design of a (*S*)-Selective-Transaminase for Asymmetric Synthesis of (1*S*)-1-(1',1'-biphenyl-2-yl)ethanamine. *ACS Catal.* **2016**, *6*, 7749–7759.
- (13) Voss, M.; Das, D.; Genz, M.; Kumar, A.; Kulkarni, N.; Kustosz, J.; Kumar, P.; Bornscheuer, U. T.; Höhne, M. In Silico Based Engineering Approach to Improve Transaminases for the Conversion of Bulky Substrates. *ACS Catal.* **2018**, *8*, 11524–11533.
- (14) Han, S.-W.; Kim, J.; Cho, H.-S.; Shin, J.-S. Active Site Engineering of  $\omega$ -Transaminase Guided by Docking Orientation Analysis and Virtual Activity Screening. *ACS Catal.* **2017**, *7*, 3752–3762.
- (15) 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.
- (16) 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.
- (17) López-Iglesias, M.; González-Martínez, D.; Rodríguez-Mata, M.; Gotor, V.; Bustó, E.; Kroutil, W.; Gotor-Fernández, V. Asymmetric Biocatalytic Synthesis of Fluorinated Pyridines through Transesterification or Transamination: Computational Insights into the Reactivity of Transaminases. *Adv. Synth. Catal.* **2017**, *359*, 279–291.
- (18) 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.
- (19) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; Series in Structural Biology; World Scientific, 2017; Vol. 9, 656.
- (20) DeLuca, S.; Khar, K.; Meiler, J. Fully Flexible Docking of Medium Sized Ligand Libraries with Rosetta Ligand. *PLoS One* **2015**, *10*, No. e0132508.
- (21) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *Aust. J. Chem.* **2012**, *4*, 17.
- (22) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (23) Krieger, E.; Vriend, G. YASARA View—Molecular Graphics for All Devices—from Smartphones to Workstations. *Bioinformatics* **2014**, *30*, 2.
- (24) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- (25) 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.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; 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, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision D.01*; 2013.
- (26) Krieger, E.; Dunbrack, R. L.; Hooft, R. W. W.; Krieger, B. Assignment of Protonation States in Proteins and Ligands: Combining pKa Prediction with Hydrogen Bonding Network Optimization. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer New York: New York, NY, 2012; Vol. 819, pp. 405–421.
- (27) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011**, *6*, No. e19230.
- (28) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. Making Optimal Use of Empirical Energy Functions: Force-Field Parameterization in Crystal Space. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 678–683.
- (29) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (30) Krieger, E.; Nielsen, J. E.; Spronk, C. A. E. M.; Vriend, G. Fast Empirical pKa Prediction by Ewald Summation. *J. Mol. Graphics Modell.* **2006**, *25*, 481–486.
- (31) Krieger, E.; Vriend, G. New Ways to Boost Molecular Dynamics Simulations. *J. Comput. Chem.* **2015**, *36*, 996–1007.
- (32) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (34) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (35) Hur, S.; Bruice, T. C. The near Attack Conformation Approach to the Study of the Chorismate to Prephenate Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12015–12020.
- (36) Wijma, H. J.; Marrink, S. J.; Janssen, D. B. Computationally Efficient and Accurate Enantioselectivity Modeling by Clusters of Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 2079–2092.
- (37) 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**, *127*, 3797–3801.
- (38) Noey, E. L.; Tibrewal, N.; Jiménez-Osés, G.; Osuna, S.; Park, J.; Bond, C. M.; Cascio, D.; Liang, J.; Zhang, X.; Huisman, G. W.; Tang, Y.; Houk, K. N. Origins of Stereoselectivity in Evolved Ketoreductases. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E7065–E7072.
- (39) Kiss, G.; Röthlisberger, D.; Baker, D.; Houk, K. N. Evaluation and Ranking of Enzyme Designs. *Protein Sci.* **2010**, *19*, 1760–1773.
- (40) Capoferro, L.; Leth, R.; ter Haar, E.; Mohanty, A. K.; Grootenhuis, P. D. J.; Vottero, E.; Commandeur, J. N. M.; Vermeulen, N. P. E.; Jørgensen, F. S.; Olsen, L.; Geerke, D. P. Insights into Regioselective Metabolism of Mefenamic Acid by Cytochrome P450 BM3 Mutants through Crystallography, Docking, Molecular Dynamics, and Free Energy Calculations. *Proteins* **2016**, *84*, 383–396.
- (41) Eichler, A.; Gricman, L.; Herter, S.; Kelly, P. P.; Turner, N. J.; Pleiss, J.; Flitsch, S. L. Enantioselective Benzylic Hydroxylation Catalysed by P450 Monooxygenases: Characterisation of a P450cam Mutant Library and Molecular Modelling. *ChemBioChem* **2016**, *17*, 426–432.
- (42) Bruice, T. C. A View at the Millennium: The Efficiency of Enzymatic Catalysis. *Acc. Chem. Res.* **2002**, *35*, 139–148.
- (43) Scheiner, S.; Lipscomb, W. N.; Kleier, D. A. Molecular Orbital Studies of Enzyme Activity. 2. Nucleophilic Attack on Carbonyl Systems with Comments on Orbital Steering. *J. Am. Chem. Soc.* **1976**, *98*, 4770–4777.
- (44) Sadiq, S. K.; Coveney, P. V. Computing the Role of Near Attack Conformations in an Enzyme-Catalyzed Nucleophilic Bimolecular Reaction. *J. Chem. Theory Comput.* **2015**, *11*, 316–324.
- (45) Bruice, T. C. Computational Approaches: Reaction Trajectories, Structures, and Atomic Motions. Enzyme Reactions and Proficiency. *Chem. Rev.* **2006**, *106*, 3119–3139.
- (46) Arabnejad, H.; Bombino, E.; Colpa, D. I.; Jekel, P. A.; Trajkovic, M.; Wijma, H. J.; Janssen, D. B. Computational Design of Enantiocomplementary Epoxide Hydrolases for Asymmetric Synthesis

- of Aliphatic and Aromatic Diols. *ChemBioChem* **2020**, *21*, 1893–1904.
- (47) Mutti, F. G.; Fuchs, C. S.; Pressnitz, D.; Turrini, N. G.; Sattler, J. H.; Lerchner, A.; Skerra, A.; Kroutil, W. Amination of Ketones by Employing Two New (S)-Selective  $\omega$ -Transaminases and the His-Tagged  $\omega$ -TA from *Vibrio fluvialis*. *Eur. J. Org. Chem.* **2012**, *2012*, 1003–1007.
- (48) Shin, J.-S.; Kim, B.-G. Comparison of the  $\omega$ -Transaminases from Different Microorganisms and Application to Production of Chiral Amines. *Biosci., Biotechnol., Biochem.* **2001**, *65*, 1782–1788.
- (49) Seo, J.-H.; Kyung, D.; Joo, K.; Lee, J.; Kim, B.-G. Necessary and Sufficient Conditions for the Asymmetric Synthesis of Chiral Amines Using  $\omega$ -Aminotransferases. *Biotechnol. Bioeng.* **2011**, *108*, 253–263.
- (50) López-Iglesias, M.; González-Martínez, D.; Gotor, V.; Bustos, E.; Kroutil, W.; Gotor-Fernández, V. Biocatalytic Transamination for the Asymmetric Synthesis of Pyridylalkylamines. Structural and Activity Features in the Reactivity of Transaminases. *ACS Catal.* **2016**, *6*, 4003–4009.
- (51) Höhne, M.; Kühl, S.; Robins, K.; Bornscheuer, U. T. Efficient Asymmetric Synthesis of Chiral Amines by Combining Transaminase and Pyruvate Decarboxylase. *ChemBioChem* **2008**, *9*, 363–365.
- (52) Tauber, K.; Fuchs, M.; Sattler, J. H.; Pitzer, J.; Pressnitz, D.; Koszelewski, D.; Faber, K.; Pfeffer, J.; Haas, T.; Kroutil, W. Artificial Multi-Enzyme Networks for the Asymmetric Amination of Sec-Alcohols. *Chem. – Eur. J.* **2013**, *19*, 4030–4035.
- (53) Schätzle, S. *Identification, Characterization and Application of Novel (R)-Selective Amine Transaminases*, Inaugural dissertation, University of Greifswald: Greifswald, 2011.
- (54) Mutti, F. G.; Kroutil, W. Asymmetric Bio-Amination of Ketones in Organic Solvents. *Adv. Synth. Catal.* **2012**, *354*, 3409–3413.
- (55) Holzer, A. K.; Hiebler, K.; Mutti, F. G.; Simon, R. C.; Lauterbach, L.; Lenz, O.; Kroutil, W. Asymmetric Biocatalytic Amination of Ketones at the Expense of NH<sub>3</sub> and Molecular Hydrogen. *Org. Lett.* **2015**, *17*, 2431–2433.
- (56) Fuchs, M.; Koszelewski, D.; Tauber, K.; Kroutil, W.; Faber, K. Chemoenzymatic Asymmetric Total Synthesis of (S)-Rivastigmine Using  $\omega$ -Transaminases. *Chem. Commun.* **2010**, *46*, 5500.
- (57) Yun, H.; Kim, J.; Kinnera, K.; Kim, B.-G. Synthesis of Enantiomerically Pure *trans*-(1*R*,2*R*)- and *cis*-(1*S*,2*R*)-1-amino-2-indanol by Lipase and  $\omega$ -Transaminase. *Biotechnol. Bioeng.* **2006**, *93*, 391–395.
- (58) Bea, H.-S.; Lee, S.-H.; Yun, H. Asymmetric Synthesis of (R)-3-Fluoroalanine from 3-Fluoropyruvate Using Omega-Transaminase. *Biotechnol. Bioprocess Eng.* **2011**, *16*, 291–296.
- (59) Meadows, R. E.; Mulholland, K. R.; Schürmann, M.; Golden, M.; Kierkels, H.; Meulenbroeks, E.; Mink, D.; May, O.; Squire, C.; Straatman, H.; Wells, A. S. Efficient Synthesis of (S)-1-(5-Fluoropyrimidin-2-yl)ethylamine Using an  $\omega$ -Transaminase Biocatalyst in a Two-Phase System. *Org. Process Res. Dev.* **2013**, *17*, 1117–1122.
- (60) Fuchs, M.; Koszelewski, D.; Tauber, K.; Sattler, J.; Banko, W.; Holzer, A. K.; Pickl, M.; Kroutil, W.; Faber, K. Improved Chemoenzymatic Asymmetric Synthesis of (S)-Rivastigmine. *Tetrahedron* **2012**, *68*, 7691–7694.
- (61) 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.
- (62) Shin, J.-S.; Kim, B.-G. Transaminase-Catalyzed Asymmetric Synthesis of L-2-Aminobutyric Acid from Achiral Reactants. *Biotechnol. Lett.* **2009**, *31*, 1595–1599.
- (63) Koszelewski, D.; Pressnitz, D.; Clay, D.; Kroutil, W. Deracemization of Mexiletine Biocatalyzed by  $\omega$ -Transaminases. *Org. Lett.* **2009**, *11*, 4810–4812.
- (64) Simon, R. C.; Grischek, B.; Zepeck, F.; Steinreiber, A.; Belaj, F.; Kroutil, W. Regio- and Stereoselective Monoamination of Diketones without Protecting Groups. *Angew. Chem., Int. Ed.* **2012**, *51*, 6713–6716.
- (65) Wu, X.; Fei, M.; Chen, Y.; Wang, Z.; Chen, Y. Enzymatic Synthesis of L-Norephedrine by Coupling Recombinant Pyruvate Decarboxylase and  $\omega$ -Transaminase. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7399–7408.
- (66) Gawley, R. E. Do the Terms "% ee" and "% de" Make Sense as Expressions of Stereoisomer Composition or Stereoselectivity? *J. Org. Chem.* **2006**, *71*, 2411–2416.
- (67) Marjanovic, A.; Ramírez-Palacios, C. J.; Masman, M. F.; Drenth, J.; Otzen, M.; Marrink, S. J.; Janssen, D. B. Thermostable D-Amino Acid Decarboxylases Derived from *Thermotoga maritima* Diaminopimelate Decarboxylase. *Protein Eng., Des. Sel.* **2021**, *34*, No. gzab016.
- (68) Meng, Q.; Ramírez-Palacios, C.; Capra, N.; Hoogwinkel, M. E.; Thallmair, S.; Rozeboom, H. J.; Thunnissen, A.-M. W. H.; Wijma, H. J.; Marrink, S. J.; Janssen, D. B. Computational Redesign of an  $\omega$ -Transaminase from *Pseudomonas jessenii* for Asymmetric Synthesis of Enantiopure Bulky Amines. *ACS Catal.* **2021**, *11*, 10733–10747.
- (69) Sheng, X.; Kazemi, M.; Planas, F.; Himo, F. Modeling Enzymatic Enantioselectivity Using Quantum Chemical Methodology. *ACS Catal.* **2020**, *10*, 6430–6449.
- (70) Gainza, P.; Roberts, K. E.; Donald, B. R. Protein Design Using Continuous Rotamers. *PLoS Comput. Biol.* **2012**, *8*, No. e1002335.
- (71) Kaufmann, K. W.; Lemmon, G. H.; DeLuca, S. L.; Sheehan, J. H.; Meiler, J. Practically Useful: What the ROSETTA Protein Modeling Suite Can Do for You. *Biochemistry* **2010**, *49*, 2987–2998.
- (72) Hu, X.; Maffucci, I.; Contini, A. Advances in the Treatment of Explicit Water Molecules in Docking and Binding Free Energy Calculations. *Curr. Med. Chem.* **2019**, *26*, 7598–7622.
- (73) Dong, L.; Meng, Q.; Ramírez-Palacios, C.; Wijma, H. J.; Marrink, S. J.; Janssen, D. B. Asymmetric Synthesis of Optically Pure Aliphatic Amines with an Engineered Robust  $\omega$ -Transaminase. *Catalysts* **2020**, *10*, 1310.