

A Computational Approach to Enzyme Design: Predicting ω -Aminotransferase Catalytic Activity Using Docking and MM-GBSA Scoring

Sarah Sirin,[†] Rajesh Kumar,[‡] Carlos Martinez,[‡] Michael J. Karmilowicz,[‡] Preeyantee Ghosh,[§] Yuriy A. Abramov,[‡] Van Martin,[‡] and Woody Sherman^{*,†}

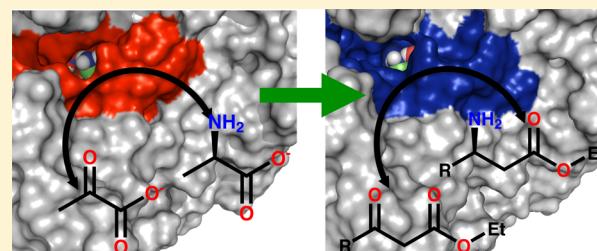
[†]Schrödinger, Inc., 120 West 45th Street, 29th Floor, New York, New York 10036, United States

[‡]Pfizer Worldwide Research and Development, Eastern Point Road, Groton, Connecticut 06340, United States

[§]Schrödinger, Sanali Infopark, 8-2-120/113, Banjara Hills, Hyderabad 500034, Andhra Pradesh, India

Supporting Information

ABSTRACT: Enzyme design is an important area of ongoing research with a broad range of applications in protein therapeutics, biocatalysis, bioengineering, and other biomedical areas; however, significant challenges exist in the design of enzymes to catalyze specific reactions of interest. Here, we develop a computational protocol using an approach that combines molecular dynamics, docking, and MM-GBSA scoring to predict the catalytic activity of enzyme variants. Our primary focuses are to understand the molecular basis of substrate recognition and binding in an *S*-stereoselective ω -aminotransferase (ω -AT), which naturally catalyzes the transamination of pyruvate into alanine, and to predict mutations that enhance the catalytic efficiency of the enzyme. The conversion of (R)-ethyl 5-methyl-3-oxo octanoate to (3*S*,5*R*)-ethyl 3-amino-5-methyloctanoate in the context of several ω -AT mutants was evaluated using the computational protocol developed in this work. We correctly identify the mutations that yield the greatest improvements in enzyme activity (20–60-fold improvement over wild type) and confirm that the computationally predicted structure of a highly active mutant reproduces key structural aspects of the variant, including side chain conformational changes, as determined by X-ray crystallography. Overall, the protocol developed here yields encouraging results and suggests that computational approaches can aid in the redesign of enzymes with improved catalytic efficiency.



1. INTRODUCTION

Enzymes are capable of catalyzing specific reactions with a remarkable degree of selectivity, fidelity, and efficiency—they have evolved into large complicated biomolecules that catalyze a variety of biochemical transformations that are essential for the proper functioning of an organism. However, enzymes can be modified and manipulated to alter their catalytic activity and target substrate(s) in order to streamline complicated synthetic processes or to act as therapeutic agents. One strategy to computationally design enzymes with desired properties is a *de novo* approach (designing entirely new enzyme functionality essentially from scratch), where a variety of computational methods have produced encouraging results for catalyzing single-step chemical transformations.^{1–3} For example, in recent work a computationally designed retro-aldolase displayed greater than 44-fold increase in activity; however, the crystal structure showed that the predicted active site was not consistent with the experimentally determined structure,⁶ indicating the difficulty and possible serendipity associated with *de novo* enzyme design. In addition, *de novo* designed enzymes have much lower catalytic efficiency than their naturally occurring counterparts.⁷

Perhaps a more tractable and practical approach to enzyme design is to modify naturally occurring enzymes with a catalytic function similar to the process of interest in order to tune the reaction toward the desired objectives (e.g., modified substrate, faster reaction, optimized performance at a different pH, etc.).^{8,9} For example, an existing alcohol dehydrogenase that operates on ethanol could be modified to catalyze the reaction of larger alcohols, like octanol. While this approach limits to some extent the scope of an enzyme design tool, it is more tractable with a higher likelihood of success as compared with the *de novo* enzyme design approach. This redesign approach could help explore untapped reactivities, and it could have many industrial applications in the biotechnology, biomedical, energy, and chemical industries.^{10,11}

Currently, the two most broadly used experimental approaches in enzyme design efforts can be classified as directed evolution^{12–14} and rational protein design.^{15,16} Directed mutagenesis relies on random genetic recombination to improve properties of interest in an enzyme—it capitalizes on the power of natural selection to screen for improved

Received: April 8, 2014



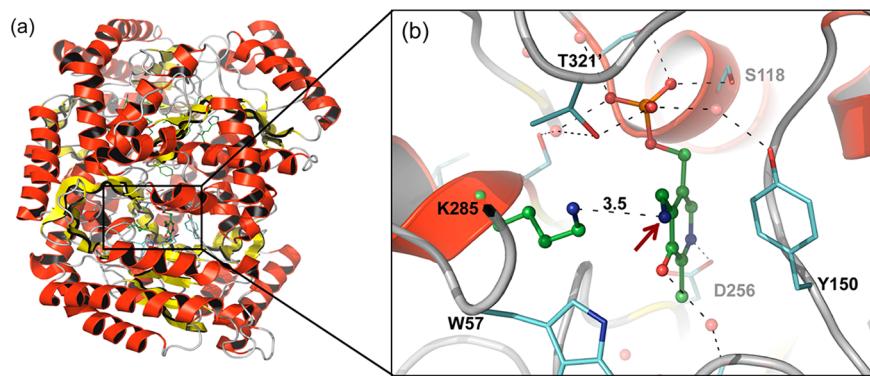


Figure 1. Illustration of X-ray crystal structure of wild-type ω -aminotransferase enzyme (PDB accession code 4E3Q⁵). Panel (a) illustrates the overall protein fold in the functional dimer, and panel (b) is a closeup view of the cofactor pyridoxamine phosphate (PMP), catalytic Lys285, and residues around the active site. PMP and reactive Lys285 side chain are colored using green carbon atoms with the reactive amine of PMP indicated using a red arrow. All other residues around the active site are shown using cyan carbons.

enzyme variants. Alternatively, rational protein design focuses on site-specific mutations, guided by computational or experimental data. This approach has the potential of being more efficient than directed evolution but requires a detailed understanding of the enzyme active site and the mechanism of the reaction of interest.

Two primary factors must be taken into consideration when trying to improve the efficiency of an enzyme: enzyme turnover rate (k_{cat}) and substrate binding (K_m). The enzyme efficiency can be defined as the ratio of first-order rate constant, or the catalytic constant, expressed in units of inverse time (k_{cat}) divided by the Michaelis–Menten constant (K_m), which is associated with the enzyme–substrate affinity. Either an increase in the enzyme turnover rate via transition state stabilization or an improvement in the Michaelis–Menten constant through increased enzyme–substrate affinity could be targeted to improve the overall efficiency of an enzyme. The approach of increasing enzyme–substrate affinity can be attractive in cases where the transition states are unknown and/or complicated with multiple intermediate steps, or when the reaction is primarily limited by binding of the substrate. For example, structure and genome information has been applied to predict binding of metabolites to an uncharacterized enzyme, which was then used to assign function to the enzyme as part of the Enzyme Function Initiative.^{17,18} Additionally, small-molecule docking, which was focused on substrate binding and not transition state stabilization, has been used to identify a catalytically active substrate for an iron oxygenase.¹⁹

In this work, we develop a general computational enzyme design protocol and apply it to predict the catalytic efficiency of an ω -aminotransferase (ω -AT) toward a target substrate of pharmaceutical interest (imagabalin precursor). ω -ATs are an attractive class of enzymes to study—they are stable proteins with a high turnover rate (k_{cat}) and broad substrate specificity, which also makes them an attractive target for biocatalysis and protein-based therapeutics.^{20–24} The functional units of ω -ATs are composed of homodimers, in which the active sites include residues from both complementary chains (see Figure 1a for an illustration of the overall protein fold). Each monomeric unit is made of one large and one small subdomain, where the large domain contains an $\alpha/\beta/\alpha$ sandwich with the catalytic lysine (Lys285) residue positioned within the central β -sheet and the small domain consists of residues from the N- and C-terminus and is disordered unless the monomeric units come together to

form a functional oligomer complex.²⁵ (This active site is shown in Figure 1b.) The cofactor pyridoxal 5-phosphate (PLP), more commonly referred to as pyridoxal-phosphate (PLP), is required for catalytic function, and binding of this cofactor to ω -AT induces a large structural rearrangement of the active site, as illustrated in Figure S1 of Supporting Information.

ω -ATs catalyze reversible S-stereoselective alanine metabolism/synthesis reactions via the Ping-Pong bi-bi reaction mechanism that is schematically outlined in Figure 2.²⁶ The

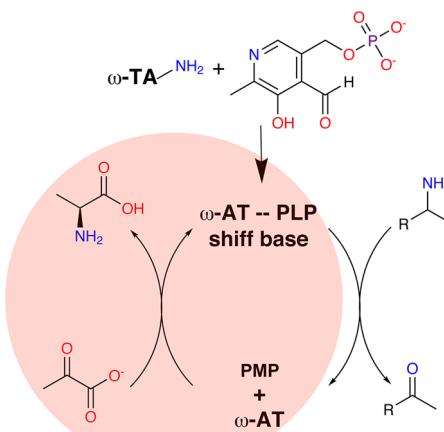


Figure 2. Schematic illustration of native reaction of pyruvate to alanine transformation as catalyzed by PLP-dependent ω -aminotransferase.

first step of this stereospecific amino acid metabolism/synthesis is initiated when PLP forms a Schiff base with the ω -AT Lys285 side chain. Subsequently, an amine donor can bind and be transformed into its ketone counterpart as PLP gets converted to pyridoxamine phosphate (PMP). ω -AT in complex with PMP can then react with a ketone donor, transforming it to an amine, while PMP gets converted back to an internal Schiff base.

In previous work, ω -AT variants with single point mutations, such as W57G and W147G, demonstrated increased catalytic activity with a concomitant change in substrate specificity toward aromatic amines without a loss in native activity or enantioselectivity.²¹ Additionally, aldolase activity was engi-

neered on a PLP-dependent alanine racemase by a single active site residue mutation, and cysteine sulfinate desulfinase activity was engineered on an aspartate aminotransferase scaffold, highlighting the versatility of PLP-dependent enzymes in catalyzing a range of biochemical transformations.^{8,27} More recently, ω -AT was designed to be used as a biocatalyst in the synthesis of chiral imagabalin, where the optimized variants catalyzed stereospecific transamination of (*R*)-ethyl 5-methyl 3-oxooctanoate to yield (3*S*,*S**R*)-ethyl 3-amino-5-methyloctanoate.⁹

Here, we develop a computational protocol that can be used to predict enzyme variants with increased activity and apply it toward ω -AT with the target substrate (*R*)-ethyl 5-methyl 3-oxooctanoate. We describe the important steps of the protocol, including substrate docking, conformational sampling of the enzyme–substrate complex, and scoring to assess the energetics of enzyme–substrate binding. Using this protocol, we have successfully identified the most active ω -AT variants that were reported in earlier experimental work⁹ and accurately classified other enzyme variants as active or inactive with a high degree of fidelity. This approach, which focuses on optimizing enzyme–substrate binding, could serve as a tool for other enzyme design efforts where catalytic efficiency can be improved through substrate binding. Finally, we describe the challenges associated with this approach and our ongoing efforts to apply the protocol to other enzyme design problems.

2. METHODS

2.1. Generation of Wild-Type Enzyme in Complex with Native and Target Substrates. A computational model of the reactant was prepared based on the wild-type X-ray crystal structure of ω -aminotransferase (ω -AT) in complex with the PMP cofactor (PDB accession code 4E3Q⁹). The residue numbering scheme used in this paper is the same as that in the work of Midelfort et al.⁹ The functional unit was identified as a dimer of identical subunits, which was used throughout this work. Missing side chains were built and hydrogen atoms were added using the Protein Preparation Wizard.^{28–30} Protonation states of ionizable amino acid side chains were determined using PROPKA,³¹ and hydrogen bond networks were subsequently optimized by flipping Asn/Gln/His residues and sampling hydroxyl/thiol hydrogens. The structure was then subjected to a restrained energy minimization to generate the structure to be used in the subsequent modeling calculations.

While the above protein preparation protocol is generally sufficient for determining a good starting structure, in the case of ω -AT there were a number of possible protonation state combinations for the enzyme and cofactor that necessitated further studies. To assess the most stable protonation states in the active site, short molecular dynamics (MD) simulations using Desmond^{32–34} were run with six plausible protonation state combinations, and the optimal state was chosen based on the simulation that resulted in the most stable MD trajectory over a period of 5 ns while reproducing key crystallographic active site interactions (see Table S1). The starting structures for the MD simulations were prepared with the System Builder in Desmond, which neutralized the system by adding Na⁺ ions and solvated the system with TIP3P³⁹ waters in a rectangular box with a 10 Å buffer distance, resulting in a system with approximately 80 000 atoms. This prepared system was subjected to the default equilibration protocol in Desmond (see below) with particle-mesh Ewald (PME) to treat long-range electrostatics and a 9-Å cutoff to separate the short-range

and long-range interactions. The OPLS2005^{35–38} force field was used to describe the protein, cofactor, and substrates. The default equilibration protocol in Desmond relaxes the system through a series of minimizations with harmonic restraints followed by a gradual heating of the system from 0 to 300 K over a cumulative 60 ps. The production run was simulated for 5 ns, for which the temperature was maintained at 300 K and the pressure at 1 atm in the NPT ensemble.

The binding pocket of the native enzyme was characterized using SiteMap^{40,41} and the native substrate (pyruvate) was docked into ω -AT active site using Glide^{42,43} with the enzyme treated rigidly. Docking of the target substrate, (*R*)-ethyl 5-methyl 3-oxooctanoate, was carried out using Induced Fit Docking (IFD)^{44,45} in order to account for protein flexibility while docking, which generates up to 20 initial poses using Glide with a softened potential that tolerates clashes of the substrate with the enzyme. Subsequently, each enzyme–substrate pose generated is refined and minimized using Prime⁴⁶ side-chain optimization. In the last step, the top scoring poses were redocked to the refined receptor. The IFD Score, which accounts for GlideScore and the Prime energy of the induced-fit receptor, was used to rank the final set of reported protein–ligand poses.

Here, we develop a modified IFD protocol that includes multiple iterations and geometric filtering based on the reactive PMP amine distance to the ketone group of the substrate to identify putative reactive enzyme–substrate poses, since substrate binding is likely to be less well determined than small molecule inhibitors—substrate binding is much weaker compared to inhibitor binding. Our IFD protocol began with a model of docked native substrate in complex with ω -AT, where the native substrate docking was performed using default Glide settings. Consistent reactive poses from this first IFD run were used in a second round of IFD in an effort to improve poses and identify conserved binding modes.

Metadynamics simulations were carried out as implemented in Desmond to explore energetically accessible binding modes for the native and target substrates within the context of wild-type ω -AT.^{32,47,48} Metadynamics describes an event (such as enzyme–substrate binding) by one or more collective variables (CVs) and enhances the sampling of the energy surface by biasing the simulation away from states that have already been sampled along the specified CVs. This biasing is achieved via addition of a series of repulsive Gaussian potentials inserted at regular intervals to the real energy landscape of the system. The added Gaussian potentials eventually fill up the current energy well and “push” the simulation to sample elsewhere. The free energy surface can then be extracted from the inserted Gaussians. While metadynamics can be sensitive to the choice of CVs, in this study metadynamics was used to enhance sampling for the generation of a representative conformational ensemble of the substrate within the binding site. The simulations were not used to directly assess the binding energetics or free energy surface (FES), and therefore the approach should be less sensitive to the choice of CVs.

Here, the distance between the center of mass of the PMP pyridine ring plus the atoms directly attached to this ring and the non-hydrogen atoms of the substrate was used as the first CV, while the angle between PMP amine nitrogen and the substrate ketone carbon and oxygen atoms was used as the second CV to describe substrate binding. The enzyme–substrate complex was prepared with the System Builder module in Desmond, as described above. The default relaxation

protocol was used prior to the production simulations, which, as mentioned above, involves a series of restrained minimizations followed by successive stages of heating the system from 0 to 300 K under successively weaker restraints. Production simulations were run at 300 K and 1 atm in the grand canonical ensemble (NPT) for 30 ns with no restraints. All simulations were run with the OPLS2005 force field. We extracted snapshots of the target substrate in complex with wild-type ω -AT every 10 ps and clustered them based on the RMSD of the substrate and PMP non-hydrogen atom coordinates. The cluster centers were then used to filter bound configurations based on the distance between PMP amine nitrogen and substrate ketone carbon atoms and the amine–ketone angle. The resultant ensemble of poses was further clustered based on substrate orientation (i.e., location of the ethyl tail and/or the direction of reactive ketone group) within the active site.

2.2. Generation of in Silico Enzyme Variants. Enzyme variants of interest were generated using the Residue Scanning module of BioLuminate,^{40,49} which uses a rotamer library in Prime⁴⁶ for side-chain sampling and the VSGB2.0^{50–52} implicit solvent model with the OPLS2005 force field for energy evaluations. The rotamer library used to generate mutated side-chain conformations has been parametrized based on a combination of a crystallographic database of 2239 side chains and structural prediction of 100 11–13 amino acid long loops and features physics-based corrections for hydrogen bonding as well as π – π , self-contact, and hydrophobic interactions.⁵⁰ An implicit solvent model is used based on the Surface Generalized Born (SGB)^{51,52} model with a variable dielectric (VD)⁵³ treatment of polarization from the protein side chains. Here, side-chain and backbone atoms of mutated residues were refined using Prime. The enzyme variants were generated using the following conditions: all residues within 5 Å of the mutated residues were refined using Prime side-chain prediction combined with backbone sampling. We also generated these enzyme variants using a rigid model, i.e. only the mutated residues were refined (cutoff distance = 0 Å). While for some systems using a rigid model is adequate, we find that for ω -AT allowing for some flexibility (cutoff distance = 5 Å) in the refinement phase gave results that were more consistent with experiments. Thus, the results presented here have been generated using the flexible model (cutoff = 5 Å) unless otherwise noted.

2.3. Scoring and Ranking. Here, we explored two general scoring schemes, namely MM-GBSA^{54,55} (Molecular Mechanics Generalized Born Surface Area) and Glide DockingScore, to rank enzyme variants with improved affinity toward target substrates compared with wild type. These two scoring functions were explored to determine if either has greater predictive capabilities for enzyme design applications, specifically related to ω -AT reactivity. These two tools are appreciably different: Glide DockingScore is an empirical scoring function that has been parametrized to separate binders from non-binders in virtual screening; Prime MM-GBSA is a physics-based method that computes the force field energies in implicit solvent of the bound and unbound molecules involved in the binding process. The MM-GBSA binding energy is calculated using eq 1, where E denotes energy and includes terms such as protein–ligand van der Waals contacts, electrostatic interactions, ligand desolvation, and internal strain (ligand and protein) energies, using VSGB2.0 implicit solvent model with the OPLS2005 force field. MM-GBSA has been shown to

perform better in scoring congeneric ligands than Glide.⁵⁵ It is important to note that the MM-GBSA method used here does not incorporate entropy terms related to the ligand or protein, which could play an important role in substrate binding and catalysis. However, solvent entropy is included in the VSGB2.0 energy model, as it is for other Generalized Born (GB) and Poisson–Boltzmann (PB) continuum solvent models. Since the application here involves neither virtual screening nor scoring congeneric compounds, we decided to test both methods of scoring in assessing binding of a substrate to multiple enzyme variants.

$$\Delta G_{\text{BINDING}} = E_{\text{COMPLEX}} - E_{\text{PROTEIN}} - E_{\text{LIGAND}} \quad (1)$$

Thus, we first used MM-GBSA and DockingScore to classify enzyme variants as “active” or “inactive” based on the observed experimental product yield. Variants with no experimental product yield were grouped as “inactive,” and all others were grouped as “active.” At this stage, the magnitude of the experimental activity was not considered, and the ability of scoring methods to correctly identify a variant as “active” or “inactive” was evaluated, analogous to virtual screening applications, where the objective is to distinguish active from inactive compounds,⁵⁶

The changes in MM-GBSA score and DockingScore for the mutants relative to wild-type enzyme were calculated and used to rank enzyme variants. An ensemble of wild-type ω -AT structures was used to generate the mutants using Residue Scanning (see above for details). The Boltzmann averaged Δ MM-GBSA binding affinity was determined for each ω -AT variant as outlined in eq 2, where i denotes the enzyme structure and E_i is the Prime energy of that structure. The Boltzmann weighted Δ MM-GBSA energy is simply referred to as “MM-GBSA” throughout the paper.

The ensembles of wild-type and mutant ω -AT structures were then used as input structures for Glide docking calculations to evaluate whether there is a predictive advantage in using Glide DockingScore to rank variants. The highest ranked Glide predictions were filtered based on geometric orientation of substrate and cofactor reactive groups to ensure reactive geometry, as judged by a geometric criterion such as the distance between the reactive amine group of cofactor and the carbonyl group of the reacting substrate. For each structure (i), the Boltzmann weighted docking score (D_i) was calculated using eq 3, where k denotes the pose for the docked substrate, D is the Glide score, and E_k is the energy of the docked substrate. Subsequently, the Boltzmann weighted docking score (D) for wild type and mutant ω -AT variants was determined using eq 4, where $\langle D_i \rangle$ is the docking score calculated for the enzyme structure in the ensemble and E_i is the Prime energy property for the enzyme structure. Finally, the change in the docking score (ΔD) for each mutant was calculated using in eq 5, where $\langle D_{\text{MUT}} \rangle$ and $\langle D_{\text{WT}} \rangle$ are the Boltzmann weighted docking scores for the mutant and wild type, respectively. (ΔD is commonly referred to as “DockingScore” throughout the paper.) In addition, we also evaluated the predictive capability of using consensus scoring, since consensus scoring approaches generally boost virtual screening campaigns.⁵⁷ Hence, we determined and rank-ordered Z-scores for MM-GBSA and DockingScore individually, and then we examined the predictive power of using their averaged Z-score rank in correctly identifying “active” enzyme variants. The averaged Z-

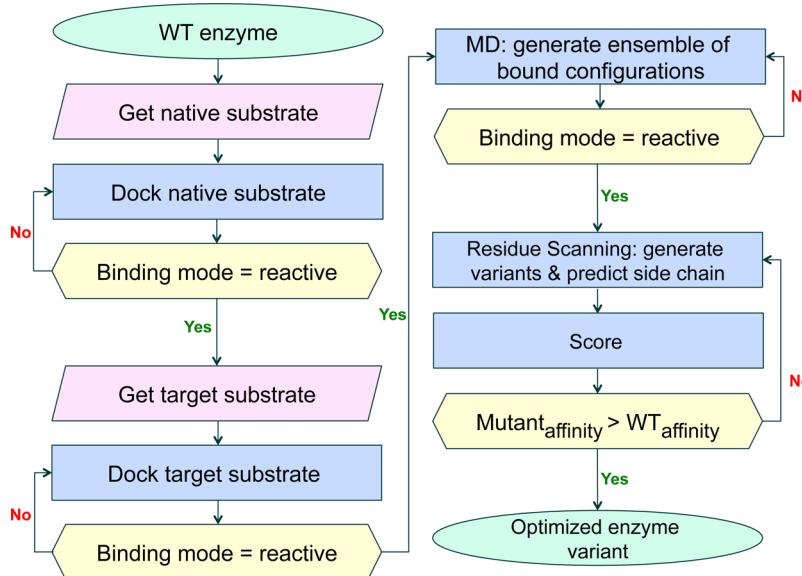


Figure 3. Flowchart for the computational workflow.

score rank used here is commonly referred to as “Z-score” throughout the paper.

$$\langle \Delta \text{MM-GBSA} \rangle = \frac{\sum_i \Delta \text{MM-GBSA} \cdot e^{-E_i/k_B T}}{\sum_i e^{-E_i/k_B T}} \quad (2)$$

$$\langle D_i \rangle = \frac{\sum_k D_i e^{-E_i/k_B T}}{\sum_k e^{-E_i/k_B T}} \quad (3)$$

$$\langle D \rangle = \frac{\sum_i \langle D_i \rangle e^{-E_i/k_B T}}{\sum_i e^{-E_i/k_B T}} \quad (4)$$

$$\Delta D = \langle D_{\text{MUT}} \rangle - \langle D_{\text{WT}} \rangle \quad (5)$$

The performances of MM-GBSA, DockingScore, and Z-score predictions were evaluated at different thresholds using the area under the curve (AUC) of the receiver-operating characteristic (ROC) as a metric. The 95% confidence intervals of the AUC predictions were determined by bootstrapping. Also, the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC)⁵⁸ metric and enrichment factor (EF) values at 5 and 10% of the total data set were calculated using the *enrichment.py* script available from Schrödinger Script Center (www.schrodinger.com/scriptcenter). Here, the EF values specify the percentage of “active” enzyme variants found in a defined fraction of the ranked list and are reported here as the percentage of actives in the top 5 and 10% of the results. EF and BEDROC provide an assessment for the percentage of the computational enzyme predictions that should be carried forward for experimental testing.

In addition to classifying enzymes as “active” and “inactive” as described above, DockingScore was also used to rank enzymes into three groups as “most active,” “less active,” and “inactive” in an effort to get a finer assessment of the degree of activity. The predictions were classified as true positive, false positive, true negative, and false negative. A true positive is a variant that is both computationally and experimentally determined to be “most active;” a false positive is a variant predicted to be “most active” but is experimentally determined

to be classified as “less active/inactive;” a true negative is a variant computationally and experimentally determined to be “less active/inactive;” a false negative is a variant computed to be “less active/inactive” but experimentally determined to be “most active.” The accuracy of the predictions was calculated as the proportion of correct predictions; the sensitivity (true positive rate) was calculated as the proportion of true positives predictions for “most active” enzymes, and the specificity (true negative rate) was calculated as proportion of true negative predictions for enzymes that are inactive. Here, a random scoring scheme would result in a computed accuracy of 0.33, whereas the two-category scheme described previously in this section a random prediction would result in an accuracy of 0.5.

2.4. Automated Workflow. The computational approach described above is outlined in Figure 3. To facilitate the rapid testing of many enzyme variations, an automated KNIME⁵⁹ workflow was implemented (see Figure S2). The workflow automates the generation of *in silico* enzyme variants based on a list of residue positions and possible amino acid substitutions. The computational steps, which included the MD simulations, were run outside of the KNIME workflow. Currently, the automated workflow takes an ensemble of wild-type enzyme–substrate structures and introduces the residue substitutions that are specified by the human designer. Subsequently, *in silico* enzyme variants are generated automatically, and the change in binding activity is calculated. While the workflow has been implemented to optimize binding, it can be easily utilized to optimize enzymes toward a putative transition state.

2.4. Experimental Screening. ω -AT from *Vibrio fluvialis* gene was synthesized by DNA 2.0 and cloned into the pet28b vector that was transformed into BL21(DE3) Gold cells. Cell pellets obtained from overnight incubation were resuspended in phosphate buffer (pH 7.0, 100 mM) and used for activity testing. Screening for activity was performed at 30 °C and pH 7.0 using a 96-well plate (2 mL capacity with conical bottom) containing 100 mM of ketone substrate, 100 mM of (S)-1-phenylethanamine as an amine donor (used as 1 M solution adjusted to pH 7.0 using hydrochloric acid), 2 mM pyridoxal phosphate, and a 80g/L *E. coli* cells suspension. Reaction samples were collected at 4 and 24 h time points and analyzed

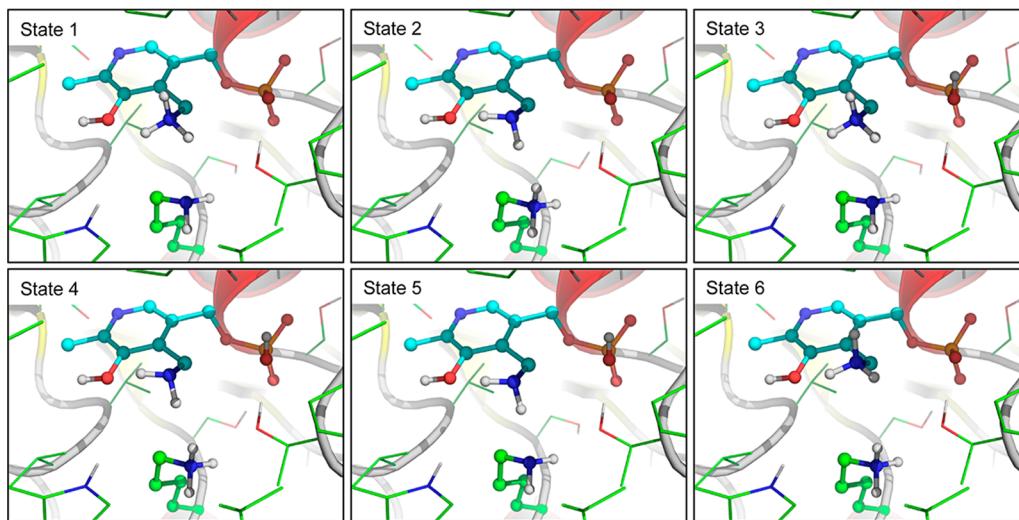


Figure 4. Illustration of the explored PMP and Lys285 protonation states within the context of wild-type apo ω -AT.

for product formation via Marfey's derivatization followed by UPLC analysis of the samples. Further details of the experimental methods have been published recently and can be found in Midelfort et al.⁹

3. RESULTS AND DISCUSSION

3.1. Structural Analysis of Wild-Type ω -Aminotransferase. A thorough understanding of the wild-type enzyme active site, including key interactions for native substrate recognition and catalysis, is an integral part of any rational enzyme design effort. Hence, we performed structural analysis on the wild-type ω -AT structure using a combination of structural visualization and molecular simulations. Figure 1a illustrates the overall ω -aminotransferase structure, and Figure 1b illustrates the active site with the PMP cofactor. The reactive Lys285 residue and pyridoxal phosphate (PLP/PMP) cofactor are required for the native ω -AT catalytic function (the reversible transformation of pyruvate to alanine) as illustrated in Figure 2 (a more detailed reaction scheme is illustrated in Figure S3). Additionally, cofactor binding induces the active site to undergo an open-to-close conformational conversion and adopt an active configuration (Figure S1). The reversible transamination reaction is initiated once a PLP cofactor binds and reacts with ω -AT to form an internal aldimine via an Schiff base linkage. This cofactor–enzyme complex can then react with an amine donor, in which the donor amine displaces the amine group of Lys285—forming an external aldimine. The ketone product is formed and released via hydrolytic cleavage with concurrent transformation of PLP to pyridoxamine phosphate (PMP). In the second step of the reaction, a ketone donor can bind and react to form a chiral amine with simultaneous conversion of PMP back to its internal aldimine PLP form. Here, we focus our modeling efforts on the PMP form of the cofactor in complex with ω -AT, since our main interest is in modifying the ketone selectivity of the enzyme, illustrated as the highlighted region in Figure 2.

To model the reactive state of the ω -AT enzyme properly, protonation states of the PMP cofactor and active site residues were carefully considered based on empirical pK_a predictions using PROPKA.³¹ The most plausible ionization states of binding site residues and the PMP cofactor were retained after visual analysis. However, we could not determine an

unambiguously correct protonation state of PMP and reactive residue Lys285 based on visualization alone, so the plausible states were probed further using MD simulations. Figure 4 shows the states considered in this work, and the corresponding ionization states are also listed in Table S1. Analysis of MD trajectories (see Methods for details) shows that state 5, where the PMP and Lys285 amine groups were neutral and the phosphate group carried a net -1 charge, is the most stable reactive conformation, as judged by the reproduction of the crystallographic orientation of the reactive PMP amine group and the Lys285 side chain. State 5 forms an average of 6.0 hydrogen bonds across the trajectory for the active site residues, including PMP, as compared with 5.3, 6.7, 4.4, 5.0, and 5.9 average hydrogen bonds for states 1, 2, 3, 4, and 6, respectively. While, state 2 forms a greater number of hydrogen bonds on average compared with state 5, the Lys285 side chain adopts an unreactive configuration by forming a salt bridge with the PMP phosphate tail (see Table S2 for average distances of Lys285 amine nitrogen to the PMP phosphorus atom). This might be a viable state for some part of the binding process but does not correspond to a reactive state. Also, in states 1 and 6, PMP forms an intramolecular salt bridge between its positively charged amine and negatively charged phosphate groups, producing unreactive configurations. In addition, as compared with the other states, state 5 had the lowest average $C\alpha$ active site RMSD from the crystal structure (Figures S4 and S5), suggesting that it is the preferred state that is most consistent with available X-ray crystal structures. Finally, state 5 also maintains key X-ray crystal structure distances throughout the simulation trajectory (Table S2). For these reasons, we chose to carry out subsequent docking and scoring calculations on state 5, where Asp256 was modeled to be in its neutral (protonated) state, reactive Lys285 and the PMP amine group were neutral (deprotonated), and the phosphate group was modeled to carry a net -1 charge.

3.2. Generation of Enzyme–Substrate Reactive Poses.

An essential part of predicting enzymatic activity is predicting the reactive binding mode of the substrate. To do so, we first modeled the substrate using a rigid receptor docking protocol in order to get an initial guess for the enzyme–substrate complex, in which structural waters hydrogen bonded with Arg415 and Trp57 were removed to enable substrate binding to

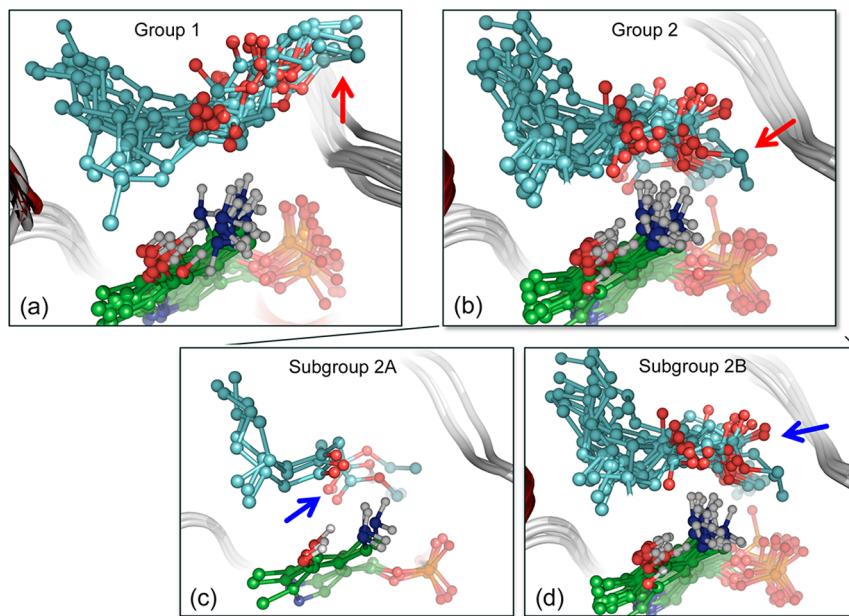


Figure 5. Snapshots along a 30 ns metadynamics simulation were clustered into 100 bins based on (*R*)-ethyl 5-methyl-3-oxooctanoate and pyridoxal 5'-phosphate RMSD. Configurations corresponding to the bound state were selected and further analyzed: 27 representative structures were further grouped into two subsets (panels a and b) based on the orientation of the ethyl tail indicated using red arrows. Group 2 was further divided into subgroup 2A (panel c) and subgroup 2B (panel d) based on the orientation of the carboxyl ketone group indicated using blue arrows. The number of structures that were characterized as belonging to group 1 is 11; group 2 is 16.

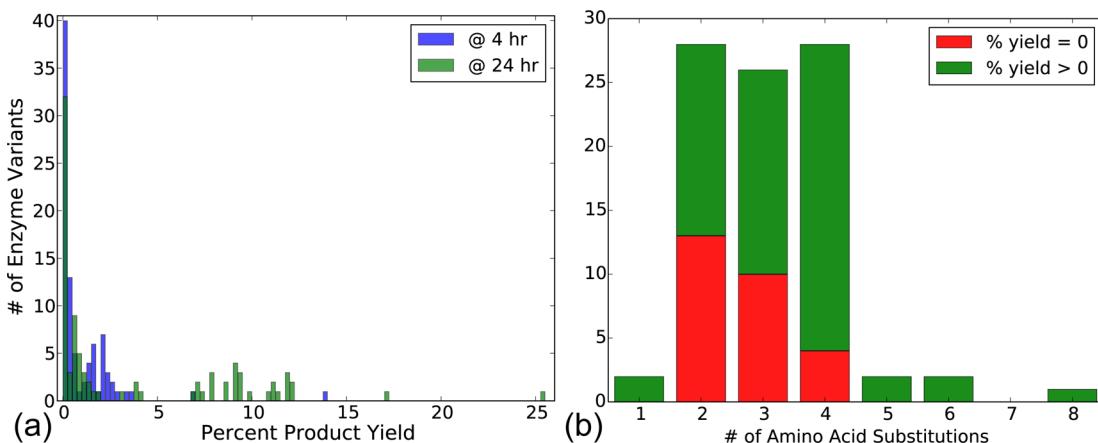


Figure 6. Analysis of the experimentally generated enzyme variants: (a) Histogram plot of the percent product yield measured at 4 and 24 h. In total, 89 enzyme variants were analyzed with experimental data collected at 4 and 24 h. (b) Histogram plot of the number of concerted amino acid substitutions introduced in each enzyme variant. Twenty-seven variants were inactive (red), while 62 variants had some activity toward the target substrate (green) at 4 h.

the active site, and the water bridging Tyr150 hydroxyl group and the cofactor phosphate moiety was kept during initial modeling (Figure S6). The carboxyl and ketone functional groups of the docked native substrate form hydrogen bonds with Arg415 and Trp57, and the methyl group occupies the volume characterized as hydrophobic by SiteMap (Figure S7), showing a good fit between the nature of the active site and the identified binding mode. We were also able to identify a second binding mode, in which the ketone group is rotated by approximately 180° and faces the PMP amine group. Figure S8 shows that most bound structures adopted this conformation, in which the ketone oxygen points toward the PMP amine group (Figure S8a), while the ketone group points in the opposite direction in a few structures visited during the simulation (Figure S8b).

Subsequently, we used Induced Fit Docking (IFD)^{44,45} to generate initial poses for the target substrate. Because substrate binding is likely to be less well-defined than that of a small molecule inhibitor, owing to its relatively weaker binding, we modified the default IFD protocol to generate substrate poses that were consistent with the viable reaction. We developed a modified IFD protocol that includes multiple iterations and geometric filtering to identify putative enzyme–substrate poses (see Methods for details). An example of optimized wild-type ω -AT interactions with native and target substrates is illustrated in Figure S9.

In an effort to validate the predicted enzyme–substrate complexes from IFD, we ran metadynamics simulations and analyzed the bound configurations. We had to rely on computational tools to validate the predicted binding modes,

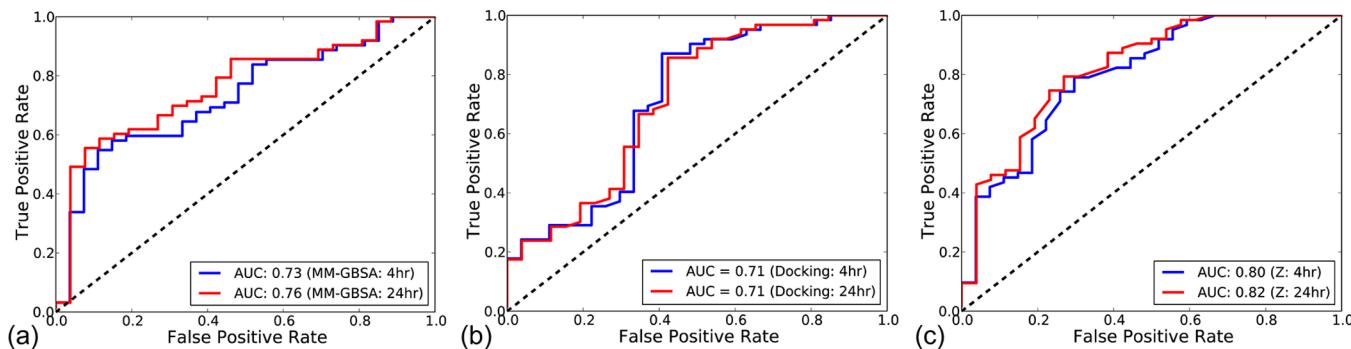


Figure 7. Receiver operating characteristic (ROC) curve for Boltzmann weighted (a) MM-GBSA, (b) DockingScore, and (c) average rank ordered Z-score predictions at 4 and 24 h, shown using blue and red, respectively. Shown results correspond to the enzyme variants that were generated using a 5-Å cutoff in the refinement stage.

since there are no available crystal structures of ω -AT in complex with any substrates (native substrate, target substrates, or substrate analogs). Here, metadynamics enabled us to explore a diverse ensemble of possible substrate binding modes for native and target substrates, where multiple binding/unbinding events were observed during the simulation time. Configurations along this metadynamics simulation of the wild-type enzyme in complex with target substrate ((R)-ethyl 5-methyl 3-oxooctanoate) were clustered into 100 bins, based on substrate and cofactor (PMP) RMSD, and from these clusters we extracted configurations corresponding to the bound states. The bound states were defined as configurations in which the PMP amine nitrogen to substrate carbonyl carbon distance was measured to be less than 4 Å, and these bound state structures were further grouped (see Methods for details) based on the substrate geometry at the active site (Figure 5). Using this approach we extracted a diverse set of configurations that showed some flexibility in the binding mode and agreed with IFD predictions.

3.3. Prediction of Activity for Enzyme Variants. Our training data set includes 89 unique ω -AT variants that have been studied experimentally, where the product yield measurements were taken at 4 and 24 h. The Pearson correlation coefficient (r) between measurements taken at 4 and 24 h is 0.85 (R^2 of 0.73). The product yield measurements taken at 4 h ranged from 0 to 13.78% with a mean of 0.98 ± 1.76 ; the median is 0.29 with a mode of 0%. Meanwhile, the data range for product yield measured at 24 h was 0 to 32.84% with a mean of 4.10 ± 5.87 ; the median is 0.79 with a mode of 0%. A histogram of the product yields is illustrated in Figure 6a. Over 90% of the enzyme variants in our data set had two to four mutations (Figure 6b). While about one-third of the ω -AT variants showed no reactivity toward the target substrate, this reactivity was more likely to be observed as the number of simultaneous amino acid substitutions was increased (Figure 6b). In total, there were 27 inactive and 62 active variants for the data collected at 4 h, while at 24 h there were 26 inactive and 63 active ω -AT variants.

We studied the effects of residue mutations on substrate activity using the bound snapshots from metadynamics simulations (see Methods section for details and Table S3 for key geometric measurements in the reactant state). Previous work has determined that for this substrate, enzyme mutations impact binding (i.e., K_m) and not the activation barrier (i.e., k_{cat}) during substrate reactivity.⁹ As such, we studied the effects of residue mutations on the predicted relative binding energy of substrates. We focused on analyzing our predictions using a

binning classification, rather than seeking numerical correlation, since the available experimental data set was not uniformly distributed and about 30% of the enzyme population was experimentally determined to be inactive toward the target substrate.

In the first round of scoring and ranking, we performed Prime MM-GBSA calculations to classify enzyme variants as “inactive” or “active” based on the computed binding energy. Prime MM-GBSA has shown good success in rank-ordering congeneric series of ligands and can identify “hot-spot” mutations in protein–protein complexes with high accuracy.^{49,55} In this binary classification, an enzyme variant was experimentally grouped as active if it showed any product yield and a variant that showed no product yield was grouped into the inactive category. Figure 7a, illustrates the receiver operating characteristic (ROC) curve for this binary classification at different thresholds. The calculated AUC, defined as the area under the ROC curve, was around 75% for correctly identifying active enzyme variants using MM-GBSA as shown in Figure 7a (see Figure S10 for the ROC curve generated using a 0-Å cutoff shell). We have also evaluated the predictive power of MM-GBSA components such as Coulomb energy, van der Waals, solvation, and intramolecular energies and found that the Coulomb energy is similarly predictive in binning active/inactive variants. Next, DockingScore was analyzed using the ROC metric, and the calculated AUC was above 70% (Figure 7b). Next, we investigated whether using a normalized consensus-scoring scheme would have any predictive advantage: MM-GBSA and DockingScore predictions were converted into their respective Z-scores and then rank ordered, after which the ranks for each mutant were averaged and analyzed using the ROC metric. The AUC for the Z-score was computed to be around 80% (Figure 7c), highlighting the general predictive advantage in using consensus score rather than a single scoring scheme. Histograms depicting the overlap of the MM/GBSA, DockingScore, and Z-score predictions for active and inactive ω -AT variants are shown in Figure S11.

While, both MM-GBSA and the consensus (Z-score) scoring approach give a higher AUC compared with the DockingScore, the DockingScore is better at ranking active variants higher in the list (see Table 1). The computed BEDROC ($\alpha = 20$) metric is 0.97 for DockingScore, while it is 0.85 and 0.95 for MM-GBSA and Z-score, respectively. Additionally, when the predictions are rank-ordered and the percentage of “active” ω -AT variants recovered in the top 10% of our data set is analyzed, we find that Glide DockingScore slightly outperforms MM-GBSA and Z-score metrics (Table 1). Figure S12

Table 1. Enrichment Values (BEDROC with $\alpha = 20$, Enrichment Factor (EF) in the Top 5 and 10%) Using Boltzmann Weighted MM-GBSA, DockingScore, and Z-score^a

	MM-GBSA		DockingScore		Z-score	
	4 h	24 h	4 h	24 h	4 h	24 h
BEDROCK ($\alpha = 20$)	0.87	0.87	0.97	0.97	0.95	0.95
EF (5%)	4.8	4.7	6.3	6.2	6.5	6.3
EF (10%)	12.7	12.5	14.3	14.1	11.3	11.1

^aThe EF reported is the percentage of “active” ω -AT variants in the top 5(10) percent of the predictions.

illustrates the average AUC values for the three scoring metrics (MM-GBSA, DockingScore, and Z-score) and their corresponding 95% confidence intervals. While all methods clearly distinguish active from inactive enzymes, Z-score performs slightly better compared with the ranking scheme, when a single scoring scheme is used.

Next, we ran an experiment to confirm the suggestion that target reactivity is attributable more to improved binding than to greater transition state stabilization. In this experiment, we explored whether using a high-energy transition state conformation in our modeling provided any predictive enhancements when using the MM-GBSA scoring approach. In this effort, we generated an ensemble of putative transition state conformers of wild-type ω -AT in complex with the target substrate, based on the currently accepted PLP-dependent transamination mechanism.⁶⁰ Prime implicit solvent minimization with constraints was used to mimic this high-energy state (see Table S3 for a list of key active site distances averaged over all input structures). The ROC plot for these MM-GBSA predictions indicates that there is no predictive power when these high-energy transition state conformations are used (Figure S13), thus reinforcing our assessment regarding the nature of ω -AT design problem.

While all three scoring schemes performed well in a binary prediction of active vs inactive enzyme variants using the reactant conformation (which would be valuable in significantly reducing the number of enzyme variants to carry forward for experimental testing), DockingScore performed slightly better at rank-ordering the enzyme variants as indicated by a slightly higher enrichment factor and the ability to clearly distinguish active vs inactive variants illustrated by the histogram overlay of the active vs inactive scores in Figure S11. Thus, we were interested in evaluating the predictive performance of DockingScore in a three-class grouping scheme (“inactive,” “active,” and “most active”), since the Glide DockingScore is slightly better at rank ordering active enzyme variants and is faster than Prime MM-GBSA. A variant was considered “most active” if its experimental product yield was greater than 2.5 at 4 h and 5% at 24 h and “inactive” if it showed no product yield. Any activity between 0 to 2.5% at 4 h and 0 to 5% at 24 h was binned into the “active” (average activity) category, suggesting the mutations have some beneficial effect and improve activity toward the target substrate. Using this scheme, two enzyme variants were considered to be “most active,” 30 considered as “active,” and 27 considered “inactive” as judged by percent product yield measurements taken at 4 h. Meanwhile, three enzyme variants were considered to be “most active,” 60 considered as “active,” and 26 considered “inactive,” when percent product yield measurements taken at 24 h were

considered. DockingScore is able to classify experimental enzyme activity at 4 h with an accuracy of 0.64, while experimental enzyme activity at 24 h was described with a similar accuracy of 0.62. The truth matrix tabulating the predictions for experimental activities is shown in Table 2 (see

Table 2. Ability of DockingScore to Correctly Classify “Most Active” and “Less Active” Variants for Measurements Taken at 4 and 24 h

predicting activity 4 h	2 true positive 10 false positive	0 false negative 77 true negative
predicting activity 24 h	2 true positive 10 false positive	1 false negative 76 true negative

Table S4 for the more detailed contingency matrix). In classifying enzyme activity at 4 h, there were two true positive, 10 false positive, 0 false negative, and 77 true negative predictions in identifying the most active enzyme variants (see Methods for details). And, for classifying enzyme activity at 24 h, there were two true positive, 10 false positive, one false negative, and 76 true negative predictions. The complete list of ω -AT variants and the corresponding experimental yield as well as computational predictions is listed in Table S5. Additionally, a scatter plot of experimental activity vs predicted DockingScore is illustrated in Figure 8, which shows that computa-

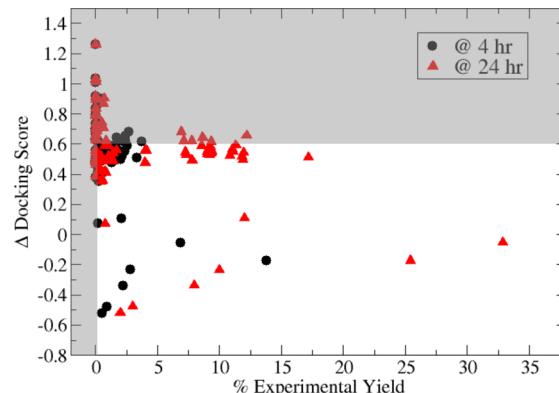


Figure 8. Scatter plot of experimental product yield and the corresponding Boltzmann weighted DockingScores.

ally highly ranked ω -AT variants are also the best mutations observed experimentally. Lastly, overlay of the computationally predicted structures of the most active ω -AT variant and the corresponding X-ray crystal structure shows a good overlap (Figure 9). The all-atom RMSD between the computational model and X-ray crystal structure for one of the most active enzyme variants (# 414) ranged from 1.0 to 1.4 Å.

Next, all residues within 8 Å of the bound target substrate (61 residues in total) were selected for mutagenesis, where all possible natural amino acid perturbations were allowed, including possible protonation states of histidine residues. Figure 10 shows a heat map of the predicted changes in DockingScore after these single-point mutations. Residues for which no mutation resulted in a strong signal have been omitted for clarity (see Figure S14 for the complete DockingScore map). While some single point substitutions at residues 57, 153, and 259 show some increase in activity over the wild type, we find, computationally, that most single-point mutations (46 residues) do not lead to significant changes in

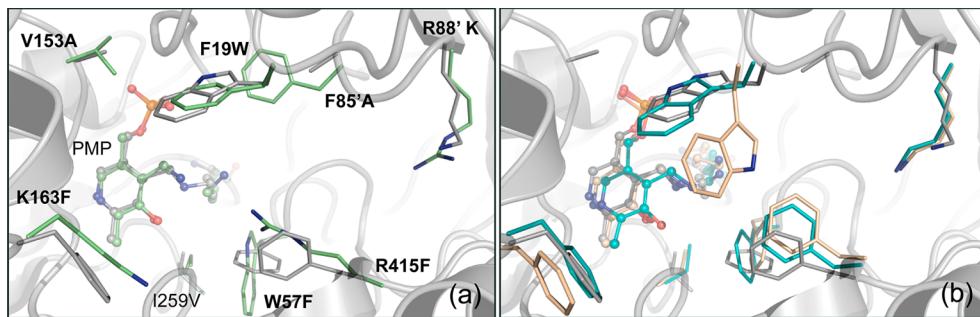


Figure 9. (a) Overlay of the wild type (green carbons) and the most active (#414) ω -AT variant (gray carbons), PDB accession codes 4E3Q and 4E3R, respectively. The wild-type structure is illustrated using green, and the ω -AT variant is illustrated using navy carbon atoms. (b) Superimposing of #414 variant X-ray crystal structures and the computational prediction with the lowest and highest RMSD, illustrated using cyan and tan colored carbon atoms, respectively. The all-atom RMSD between ω -AT variant #414 X-ray crystal structure and our computational prediction ranged from 1.0 to 1.4 Å. For simplicity and clarity, the mutated residues are labeled only in panel a and alanine predictions hidden in panel b.

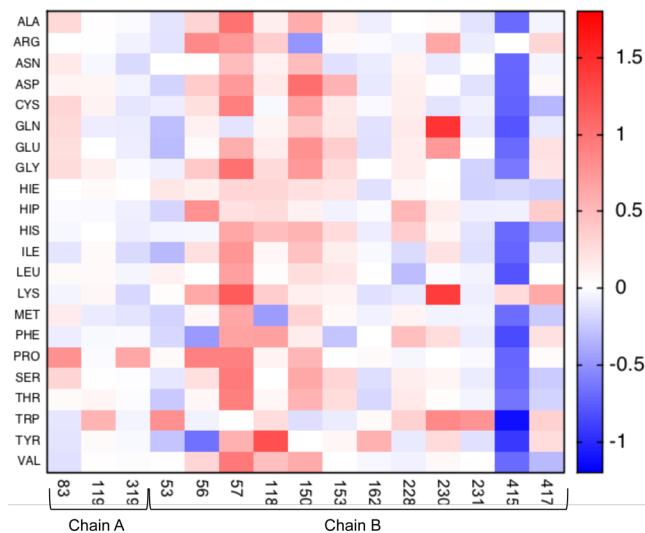


Figure 10. ω -AT active site substrate binding activity map: Boltzmann weighted change in docking scores (DockingScore) for single point mutations of residues within 8 Å of the bound target substrate. In total, 61 residues were perturbed and only 15 residues have one or more single point mutation, where the calculated DockingScore was greater or less than 0.5 kcal/mol. A total of 46 residues were omitted from the activity map for simplicity.

activity—consistent with experimental observations. An initial overview of the binding landscape indicates that residue 415 substitutions should improve binding most significantly. While only seven out of 89 enzyme variants tested experimentally included substitutions at residue 415, every enzyme variant perturbed at this residue was highly active compared with the wild type. Thus, using computational screens to triage and select new residues for mutagenesis could lead to examination of previously overlooked residues and result in design diversity. In addition, the heat map suggests that Tyr150 mutations would decrease binding activity, which is consistent with experimental observations—the addition of Tyr150 mutations to active variants leads to diminished activity. While most predictions are consistent with experiments, predictions for residue 57 perturbations suggest that mutating this site would lead to a decrease in binding affinity, whereas the first experimental improvement toward the target substrate was seen upon residue 57 perturbations. However, this map also shows that residue 57 perturbations are among the most divergent from wild-type activity. Hence this residue highlights

the importance of not only targeting computational mutations that improve binding toward the target substrate but also focusing on residues that give large deviations away from wild-type activity.

In summary, we have generated an activity map of ω -AT active site and identified residues as potential targets for experimental mutagenesis, which differ from already explored perturbations and could serve as new leads, adding diversity to design experiments.

4. CONCLUSIONS

The ability to rationally design enzymes with improved substrate binding (K_m) and/or catalytic turnover (k_{cat}) is a highly desirable objective in biotechnology and pharmaceutical research. In this work, we developed a structure-based approach to predict the activity of designed enzymes toward novel substrates that are related to the native substrate. We use ω -aminotransferase (ω -AT), which is a pyridoxal-phosphate (PLP)-dependent enzyme natively involved in stereospecific pyruvate-to-alanine transformation, as a model system to develop, implement, and validate a computational protocol that can be used to predict the structure and activity for enzyme variants with improved activity. We focus specifically on active site complementarity and substrate recognition to predict the activity of previously designed ω -AT variants; for instance, ω -AT has recently been engineered to catalyze the transformation of β -keto esters into a chiral β -amino esters that could be used as a precursor for imagabalin synthesis.⁹

The protocol that we developed in this work has been demonstrated predictive at separating highly active from less active enzyme variants. We tested the protocol on approximately 90 ω -AT variants whose activity for the imagabalin precursor (*R*)-ethyl 5-methyl 3-oxooctanoate has been measured experimentally. The predictions correctly classify the ω -AT variants as “active” or “inactive” with good accuracy, and most significantly, we identify the two ω -AT variants with the greatest activity (20- to 60-fold improvement over wild type) as the top ranked hits in the computational screen.

The workflow developed here relies on the hypothesis that the activity of ω -AT can be altered via perturbing active site residues to improve substrate binding. The successes we describe using the model derived from this hypothesis corroborate its validity for this system. The computational protocol uses a combination of protein sampling with molecular dynamics to obtain an ensemble of energetically accessible states, followed by docking and Boltzmann weighting

of the scores to get the final prediction of activity. The computational approach used here can complement experimental enzyme engineering efforts and can help interpret experimental activity at a molecular level from a substrate binding perspective. The approach described here could also be used for enzyme library design, identifying previously unexplored residues with putative beneficial effects on activity, or to help enhance sequence diversity in design experiments.

The protocol presented here relies on binding affinity optimization (K_m) to drive catalytic efficiency, which is appropriate for ω -AT since improvements in enzyme activity were shown to be derived from improved substrate binding (K_m), rather than an increased catalytic turnover.⁹ However, for many enzymes, the catalytic efficiency is limited by the turnover rate (k_{cat}), which has not been explored here. Optimizing enzymes for transition state stabilization could be used to predict k_{cat} , which may require optimizing for multiple transition states and intermediates, and a detailed understanding of the reaction energetics, while also assuming reactivity remains constant as active site amino acids and the substrates evolve. We are currently exploring other enzyme systems, where the catalytic turnover rate is a critical determinant of enzyme efficiency, to see if the protocol can be augmented to account for transition state stabilization. Finally, even in cases where substrate binding has been determined as a critical modulator of catalytic efficiency, optimizing binding too much can result in reduced catalytic efficiency, illustrating the challenges associated with enzyme design and the need for additional research in this direction.

The prospective validation of computational enzyme design strategies requires careful experimental design in order for the results to meaningfully indicate whether the method is providing value. For example, if the baseline rate of determining enzymes with enhanced activity is one in 100, then a protocol that improves that rate by a factor of 2 would still need to test approximately 100 compounds to get a statistically meaningful assessment of the predictive capabilities of the method. Other challenges include whether the computational approach adequately samples the combinatorial sequence space and correctly predicts structures as the native enzyme evolves and reflects multiple mutations. Additionally, improving enzyme activity is ultimately a multiobjective optimization problem, where affinity, inhibition, reaction energetics, stability, higher oligomer formation, cofactor and/or metal dependence, and solubility could all be contributing to enzyme activity. These challenges, among other inherent problems—such as whether an enzyme structure is evolvable^{61,62}—remain open-ended questions that must be evaluated carefully for each design problem. Finally, the applicability of an optimization workflow to systems that are already catalytically prolific—such as OMP decarboxylase, which has significant catalytic efficiency due to transition state stabilization^{63,64}—is beyond the scope of the current work.

ASSOCIATED CONTENT

Supporting Information

Tables listing the molecular dynamics simulations carried out with various protonation states for cofactor pyridoxamine phosphate (PMP) and reactive lysine ionizable groups; calculated key active site distances averaged over 5 ns molecular dynamics simulations; key geometric measurements for the reactant and transition state input structures for wild-type ω -AT; contingency matrix of DockingScore in classifying enzyme

variants as “most active,” “active,” and “inactive;” mutations to wild-type ω -AT and the corresponding experimental activity and the computational predictions as judged via docking score. Figures showing the overlay of the open-to-close conformation switch induced by cofactor binding; the automated KNIME workflow that can perform *in silico* mutagenesis and compute changes in binding activity relative to wild-type enzyme variant; detailed illustration of the generally accepted pyruvate to alanine transamination mechanism; RMSD from the crystal coordinates for the C- α atoms in ω -AT dimer during MD simulations; RMSD from the crystal coordinates for the active site non-hydrogen atoms in ω -AT dimer during MD simulations; X-ray crystallographic water molecules that bridge PMP phosphate and amine groups with Tyr150, Arg415, and Trp57 side chains; characterized ω -AT binding pocket and docked native substrates; sampled coordinates native substrate adopts in the bound configuration during 30 ns metadynamics simulations; ligand interaction diagrams for the native and target substrates in complex with wild-type ω -AT; receiver operating characteristic (ROC) curve for MM-GBSA when 0-Å-distance cutoff was used in the refinement step; overlay histogram of MM-GBSA, DockingScore, and Z-score predictions for active and inactive ω -AT variants; receiver operating characteristic (ROC) curve for MM-GBSA predictions when a transition state model was used; Boltzmann weighted change in DockingScore for 61 single point mutations of residues within 8 Å of the bound target substrate. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Phone: +1 212 295 5800. Fax: +1 212 295 5801. E-mail: woody.sherman@schrodinger.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Schrödinger IT for providing computational resources and support, Jean-Christophe Mozziconacci and Chris Higgs for KNIME support, and David Pearlman and Michelle Lynn Hall for useful discussions.

ABBREVIATIONS

ω -AT, ω -aminotransferase; PLP, pyridoxal-phosphate; PSP, pyridoxal-5-phosphate; PMP, pyridoxamine phosphate; MD, molecular dynamics; MM-GBSA, molecular mechanics–generalized Born surface area; AUC, area under the curve; ROC, receiver-operating characteristic; BEDROC, Boltzmann-enhanced discrimination of receiver operating characteristic; EF, enrichment factor

REFERENCES

- (1) Kiss, G.; Celebi-Olcum, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem., Int. Ed. Engl.* **2013**, *52*, 5700–5725.
- (2) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011**, *6*, e19230.
- (3) Khersonsky, O.; Rothlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. Evolutionary Optimization of Computationally Designed Enzymes: Kemp Eliminases of the Ke07 Series. *J. Mol. Biol.* **2010**, *396*, 1025–1042.

- (4) Althoff, E. A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J. K.; Wang, Z.; Smith, M.; Hari, S.; Kast, P.; Herschlag, D.; Hilvert, D.; Baker, D. Robust Design and Optimization of Retroaldol Enzymes. *Protein Sci.* **2012**, *21*, 717–726.
- (5) Kries, H.; Blomberg, R.; Hilvert, D. De Novo Enzymes by Computational Design. *Curr. Opin. Chem. Biol.* **2013**, *17*, 221–228.
- (6) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494–498.
- (7) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82*, 447–470.
- (8) Seebeck, F. P.; Hilvert, D. Conversion of a Plp-Dependent Racemase into an Aldolase by a Single Active Site Mutation. *J. Am. Chem. Soc.* **2003**, *125*, 10158–10159.
- (9) Midelfort, K. S.; Kumar, R.; Han, S.; Karmilowicz, M. J.; McConnell, K.; Gehlhaar, D. K.; Mistry, A.; Chang, J. S.; Anderson, M.; Villalobos, A. 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.
- (10) Schmid, A.; Dordick, J.; Hauer, B.; Kiener, A.; Wubbolt, M.; Witholt, B. Industrial Biocatalysis Today and Tomorrow. *Nature* **2001**, *409*, 258–268.
- (11) Wiseman, A. *Handbook of Enzyme Biotechnology*; Ellis Horwood Ltd.: Chichester, U.K., 1975.
- (12) Otten, L. G.; Quax, W. J. Directed Evolution: Selecting Today's Biocatalysts. *Biomol. Eng.* **2005**, *22*, 1–9.
- (13) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866–876.
- (14) Cramer, A.; Raillard, S.-A.; Bermudez, E.; Stemmer, W. P. DNA Shuffling of a Family of Genes from Diverse Species Accelerates Directed Evolution. *Nature* **1998**, *391*, 288–291.
- (15) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302*, 1364–1368.
- (16) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329*, 309–313.
- (17) Gerlt, J. A.; Allen, K. N.; Almo, S. C.; Armstrong, R. N.; Babbitt, P. C.; Cronan, J. E.; Dunaway-Mariano, D.; Imker, H. J.; Jacobson, M. P.; Minor, W. The Enzyme Function Initiative. *Biochemistry* **2011**, *50*, 9950–9962.
- (18) Zhao, S.; Kumar, R.; Sakai, A.; Vetting, M. W.; Wood, B. M.; Brown, S.; Bonanno, J. B.; Hillerich, B. S.; Seidel, R. D.; Babbitt, P. C. Discovery of New Enzymes and Metabolic Pathways by Using Structure and Genome Context. *Nature* **2013**, *502*, 698–702.
- (19) Esmieu, C.; Cherrier, M. V.; Amara, P.; Gireggi, E.; Marchi-Delapierre, C.; Oddon, F.; Iannello, M.; Jorge-Robin, A.; Cavazza, C.; Ménage, S. An Artificial Oxygenase Built from Scratch: Substrate Binding Site Identified Using a Docking Approach. *Angew. Chem.* **2013**, *125*, 4014–4017.
- (20) Shin, J.-S.; Kim, B.-G. Exploring the Active Site of Amine: Pyruvate Aminotransferase on the Basis of the Substrate Structure-Reactivity Relationship: How the Enzyme Controls Substrate Specificity and Stereoselectivity. *J. Org. Chem.* **2002**, *67*, 2848–2853.
- (21) 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 Ω -Aminotransferase for the Kinetic Resolution of Aliphatic Chiral Amines. *Biotechnol. Bioeng.* **2008**, *99*, 275–284.
- (22) 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. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305–309.
- (23) Martin, A. R.; DiSanto, R.; Plotnikov, I.; Kamat, S.; Shonnard, D.; Pannuri, S. Improved Activity and Thermostability of (S)-Aminotransferase by Error-Prone Polymerase Chain Reaction for the Production of a Chiral Amine. *Biochem. Eng. J.* **2007**, *37*, 246–255.
- (24) Mathew, S.; Yun, H. Ω -Transaminases for the Production of Optically Pure Amines and Unnatural Amino Acids. *ACS Catal.* **2012**, *2*, 993–1001.
- (25) Humble, M. S.; Cassimjee, K. E.; Håkansson, M.; Kimburg, 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.
- (26) Shin, J.-S.; Yun, H.; Jang, J.-W.; Park, I.; Kim, B.-G. Purification, Characterization, and Molecular Cloning of a Novel Amine: Pyruvate Transaminase from *Vibrio Fluvialis* Js17. *Appl. Microbiol. Biotechnol.* **2003**, *61*, 463–471.
- (27) Fernandez, F. J.; de Vries, D.; Peña-Soler, E.; Coll, M.; Christen, P.; Gehring, H.; Vega, M. C. Structure and Mechanism of a Cysteine Sulfinate Desulfinate Engineered on the Aspartate Aminotransferase Scaffold. *Biochim. Biophys. Acta, Proteins Proteomics* **2012**, *1824*, 339–349.
- (28) Prime, version 3.2; Schrödinger, LLC: New York, 2013.
- (29) *Protein Preparation Wizard* 2013-3, Epik version 2.4; Schrödinger, LLC: New York, 2013.
- (30) Sastry, G. M.; Inakollu, V. S.; Sherman, W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from 2d Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* **2013**, *53*, 1531–1542.
- (31) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. Propka3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (32) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *SC 2006 Conference, Proceedings of the ACM/IEEE*; IEEE: Piscataway, NJ, 2006; pp 43–43.
- (33) *Desmond Molecular Dynamics System*, version 3.6, D. E. Shaw Research: New York, 2013.
- (34) *Maestro-Desmond Interoperability Tools*, version 3.6, Schrödinger, LLC: New York, 2013.
- (35) Guvench, O.; MacKerell, A. D., Jr. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In *Molecular Modeling of Proteins*; Springer: New York, 2008; pp 63–88.
- (36) Jorgensen, W. L.; Tirado-Rives, J. The Opls [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (37) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the Opls-Aa Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (38) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the Opls Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509–1519.
- (39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
- (40) *SiteMap*, version 2.9; Schrödinger, LLC: New York, 2013.
- (41) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- (42) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (43) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

- (44) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534–553.
- (45) Sherman, W.; Beard, H. S.; Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–84.
- (46) Prime, version 3.4; Schrödinger, LLC: New York, 2013.
- (47) Liu, M.; Bender, S. A.; Cuny, G. D.; Sherman, W.; Glicksman, M.; Ray, S. S. Type Ii Kinase Inhibitors Show an Unexpected Inhibition Mode against Parkinson's Disease-Linked Lrrk2 Mutant G2019s. *Biochemistry* **2013**, *52*, 1725–1736.
- (48) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (49) Beard, H.; Cholleti, A.; Pearlman, D.; Sherman, W.; Loving, K. A. Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. *PLoS One* **2013**, *8*, e82849.
- (50) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The Vsgb 2.0 Model: A Next Generation Energy Model for High Resolution Protein Structure Modeling. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 2794–2812.
- (51) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.
- (52) Yu, Z.; Jacobson, M. P.; Friesner, R. A. What Role Do Surfaces Play in Gb Models? A New Generation of Surface-Generalized Born Model Based on a Novel Gaussian Surface for Biomolecules. *J. Comput. Chem.* **2006**, *27*, 72–89.
- (53) Zhu, K.; Shirts, M. R.; Friesner, R. A. Improved Methods for Side Chain and Loop Predictions Via the Protein Local Optimization Program: Variable Dielectric Model for Implicitly Improving the Treatment of Polarization Effects. *J. Chem. Theory Comput.* **2007**, *3*, 2108–2119.
- (54) Greenidge, P. A.; Kramer, C.; Mozziconacci, J.-C.; Wolf, R. M. Mm/Gbsa Binding Energy Prediction on the Pdbbind Data Set: Successes, Failures, and Directions for Further Improvement. *J. Chem. Inf. Model.* **2012**, *53*, 201–209.
- (55) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and Mm-Gbsa Scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.
- (56) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening—an Overview. *Drug Discovery Today* **1998**, *3*, 160–178.
- (57) Sastry, G. M.; Inakollu, V. S.; Sherman, W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* **2013**, *53*, 1531–1542.
- (58) Truchon, J.-F.; Bayly, C. I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
- (59) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. *KNIME: The Konstanz Information Miner*; Springer: New York, 2008.
- (60) Oliveira, E. F.; Cerqueira, N. M.; Fernandes, P. A.; Ramos, M. J. Mechanism of Formation of the Internal Aldimine in Pyridoxal 5'-Phosphate-Dependent Enzymes. *J. Am. Chem. Soc.* **2011**, *133*, 15496–15505.
- (61) Brustad, E. M.; Arnold, F. H. Optimizing Non-Natural Protein Function with Directed Evolution. *Curr. Opin. Chem. Biol.* **2011**, *15*, 201–210.
- (62) Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. Evolutionary Optimization of Computationally Designed Enzymes: Kemp Eliminases of the Ke07 Series. *J. Mol. Biol.* **2010**, *396*, 1025–1042.
- (63) Miller, B. G.; Wolfenden, R. Catalytic Proficiency: The Unusual Case of Omp Decarboxylase. *Annu. Rev. Biochem.* **2002**, *71*, 847–885.
- (64) Lee, J. K.; Houk, K. A. Proficient Enzyme Revisited: The Predicted Mechanism for Orotidine Monophosphate Decarboxylase. *Science* **1997**, *276*, 942–945.