SubTuner: a Physics-Guided Computational Tool for Modifying Enzymatic Substrate Preference and Its Application to Anion Methyltransferases

Qianzhen Shao[1#], Asher C. Hollenbeak[3#], Yaoyukun Jiang[1], Brian O. Bachmann[1,3]*, and Zhongyue J. Yang[1-5,]*

[1]*Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States*

[2]*Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States*

[3]*Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37235, United States*

[4]*Data Science Institute, Vanderbilt University, Nashville, Tennessee 37235, United States*

[5]*Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, United States*

[#]*These two authors contribute equally to this work.*

**Corresponding author's email: brian.bachmann@vanderbilt.edu;zhongyue.yang@vanderbilt.edu*

Abstract

Engineering enzymes to catalyze non-native substrates is critical for chemical synthesis and drug development. Although general-purpose computational tools exist, a significant challenge is to create a tool specialized in shifting an enzyme's activity toward a specified non-native substrate. We developed SubTuner, a physics-guided computational tool that automates enzyme engineering for catalyzing desired non-native substrates. To test the performance of SubTuner, we designed three tasks – all aiming to identify beneficial anion methyltransferase mutants for synthesis of non-native *S*-adenosyl-L-methionine analogs: first in the conversion of ethyl iodide from a pool of 190 *At*HOL1 single-point mutants for an initial test of accuracy and speed; second of ethyl iodide and *n*-propyl iodide from a pool of 600 *acl*-MT multi-point mutants for a test of generalizability; and eventually of bulkier substrates (*n*-propyl iodide, isopropyl iodide, and allyl iodide) combined with experimental characterization for a test of *a priori* predictivity. All the tests demonstrate SubTuner's ability to accelerating the discovery of function-enhancing mutants for non-native substrates. Moreover, utilizing molecular simulation data derived from SubTuner, we elucidated

how beneficial mutations promote catalysis. SubTuner, with its solid physical hypothesis, quantitative accuracy, and mechanism-informing ability, holds a significant potential to aid enzyme engineering for substrate scope expansion.

1. Introduction

Enzymes catalyze their native reactions with evolutionarily optimized proficiency but show diminished activities using non-native substrates. The ability to modify an enzyme from favoring its native substrate to favoring a desirable non-native substrate has great utility in a wide range of applications. Directed evolution techniques have been previously applied to engineer enzyme mutants, such as cytochrome P450,[1-4] ketoreductases,[5] imine reductases,[6] and terpene synthases,[7] to modify their substrate preference.[8-15] Despite these successes in producing mutant enzymes optimized for non-native reactions, the mutagenesis-based screening underlying directed evolution approaches have some significant drawbacks: high resource and labor costs deriving from low rates of identifying beneficial mutations, as well as uncertainty in designing a functional readout both complicate study design logistically. Even successful identification of beneficial mutations does not prohibit falling into an evolutionary dead end (e.g., screening $>10^9$ mutants does not yield any functional improvements[16]), nor are we informed of the sequence-structure-function relationships underlying the mutational screening. However, a physics-based *in silico* enzyme engineering tool that screens mutants virtually would ameliorate the aforementioned complications: labor cost, readout, interpretability, etc. Therefore, one pressing challenge is how to develop a computational tool that specializes in predicting enzyme mutants that accommodate a desired non-native substrate.

Several *in silico* enzyme engineering tools, such as HotSpotWizard[17], FuncLib[18], AsiteDesign[19], and CASCO[20] have been developed for enzyme engineering. However, none of

them specializes in improving enzymatic activity targeting a specific non-native substrate. FuncLib has been arguably the most applied computational enzyme engineering tool in the community, assisting to engineer 20 enzymes in 19 research works by August 2024.[18, 21-38] FuncLib was shown to successfully discover mutants that shift a wild-type enzyme's native activity toward unspecified substrates.[18, 21-24] When the users intend to engineer enzymes for catalyzing a specified substrate, however, FuncLib demonstrated moderate to feeble performance. There are 4 research works finding FuncLib designs to be deleterious or neutral,[25-28] 10 to be moderate in functional enhancement even with multiple mutations (~2 fold relative to the wild-type).[29-38] This is likely caused by the inability of FuncLib to predict the impact of a substrate on an enzyme's mutational landscape, albeit its accurate inference of mutational hotspots from evolutionary information and protein thermostability. In addition to FuncLib, HotSpotWizard specializes in identifying mutational hotspots and is widely applied to guide the selection of initial sites for site-saturation library construction. AsiteDesign specializes in introducing catalytic residues into a binding pocket for hydrolase design. CASCO specializes in altering the enantio- or regioselectivity of the enzyme. These specializations fall out of the need of identifying beneficial mutants targeting a specific non-native substrate.

To address this urgent need, we developed SubTuner, a physics-guided computational tool for modifying enzymatic substrate preferences towards a desired substrate. SubTuner is built on the hypothesis that when designing an enzyme for non-native reactions, beneficial mutants need to be: 1) thermostable, 2) effective at binding to a rate-limiting transition state (TS), and 3) good at stabilizing the developing dipole of the TS with interior enzyme electric field. Both the TS binding and electric field hypotheses about beneficial mutants have previously been utilized to design improved Kemp eliminases.[39-41] We translated each of the hypotheses into a computational

screening step in SubTuner and automated the computational workflow using EnzyHTP,[42-46] a Python library previously developed by our group.

Nonnatural *S*-adenosyl-L-methionine (SAM) analogs have been used for late-stage methyltransferase-catalyzed modification of small molecules and tagging of macromolecules such as proteins and DNA.[47] However, the difficulty of synthesizing SAM analogs and their reduced stability in solution makes biocatalytic transformations using non-natural SAM analogs difficult.[48] Anion methyltransferases (or halide methyltransferase, HMT) can overcome these issues by catalyzing the transfer of alkyl groups from iodoalkanes to *S*-adenosyl-L-homocysteine – forming alkyl-SAM analogs – concurrent with turnover of a promiscuous methyltransferase. We tested SubTuner using two HMTs, the *Arabidopsis thaliana* Harmless to Ozone Layer 1 (*At*HOL1) enzyme[49] and *Aspergillus clavatus* methyltransferase (*acl*-MT)[50]. We first evaluated SubTuner's ability to identify beneficial mutants from 190 *At*HOL1 mutants that were experimentally screened by saturation mutagenesis for modifying the wild-type *At*HOL1 to catalyze ethyl iodide (EtI)[51]. To test SubTuner's generalizability with multi-mutations, substrates, and enzymes, we then tasked it to identify beneficial *acl*-MT mutants from a Spread-Out Low Diversity (SOLD) library of 600 mutants for ethyl iodide (EtI) and *n*-propyl iodide (*n*PrI) catalysis.[52] Finally, to test SubTuner's capability of guiding experimental engineering, we applied SubTuner to identify beneficial *At*HOL1 mutants for catalyzing bulkier substrates including *n*-propyl iodide (*n*PrI), *i*-propyl iodide (*i*PrI), and allyl iodide (allyl-I), and experimentally characterized the top five mutants predicted for each *n*PrI, *i*PrI, and allyl-I, and discovered previously unidentified mutants with improved turnover. These results demonstrate SubTuner's capability of accelerating enzyme engineering for targeted substrate scope expansion.

2. Results

2a. An overview of the SubTuner workflow

SubTuner takes in the enzyme-substrate complex structure (i.e., considered as the "wild-type" structure) and predicts beneficial mutations that accelerate the enzymatic reaction of a specified non-native substrate. Empowered by the EnzyHTP[42] library, SubTuner serves as an automatically operated command line tool. The workflow involves four steps: *Mutation*, *Thermostability Screening*, *TS-Analog Binding Screening*, and *Reactivity Ranking* (Figure 1). *Mutation* determines the "mutant library" to be screened, accommodating multiple library construction strategies: site-saturation mutagenesis, random mutagenesis, rational design, or any structure-based selection. *Thermostability Screening* filters out thermally unstable mutants from the "mutant library" by the change of folding free energy ($\Delta\Delta G_{fold}$), where the folding free energy is computed by cartesian_ddG[53, 54] in Rosetta.[55] *TS-Analog Binding Screening* eliminates the mutants that involve unfavorable transition state (TS) binding, where binding enthalpy and binding entropy are approximated using the MMPBSA energy and active site RMSD, respectively – both are derived from MD trajectories. From the remaining variants, *Reactivity Ranking* selects the top ten beneficial mutant candidates for experimental characterization, in which electrostatic stabilization energy of the transition state, $\Delta G_{ele}^{\ddagger}$,[56, 57] is used as the metric to assess the reactivity of each mutant (see details of the workflow in Section 4: Implementation).

To assess the performance of SubTuner, we designed two metrics: hit rate and function-enhancing speed (FES). The hit rate is calculated by dividing the number of mutant hits (characterized experimentally) by the total number of predicted mutants (i.e., ten mutants), representing the probability of observing function-enhancing mutants from the predicted candidates. The function-enhancing speed is calculated by first identifying the mutant with the greatest increase in turnover, and then dividing its improvement in activity fold by the time

consumption, representing how fast the function enhancement can be achieved through experimental screening or computation-guided experimental discovery. Based on these metrics, we tested SubTuner for its ability to accelerate the discovery of anion methyltransferase mutants that form non-native SAM analogs in the following sections.

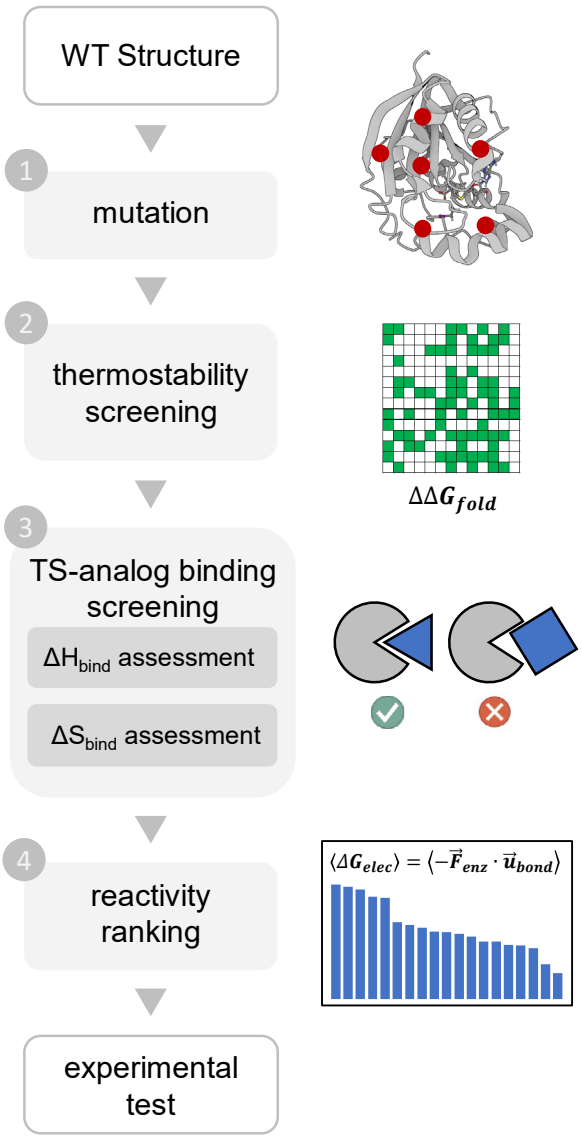


Figure 1. An overview of the SubTuner workflow. The workflow consists of four steps: 1. *Mutation*, which determines the initial mutant library; 2. *Thermostability Screening*, which screens the mutant library for stable folding; 3. *TS-Analog Binding Screening*, which screens the

thermostable mutants for favorable binding of a TS-analog; and 4. *Reactivity Ranking*, which ranks remaining mutants based on predicted reactivity. The reactivity is assessed by the electrostatic stabilization energy of each mutant relative to the wild type ($\Delta\Delta G_{ele}^{\ddagger}$), which informs the enzymatic activity change upon mutations.

2b. Test of SubTuner in identifying beneficial *At*HOL1 mutants to catalyze Ethyl-SAM formation

In this test, SubTuner was tasked to alter the substrate preference of *At*HOL1 toward a non-native substrate ethyl iodide (EtI). *Arabidopsis thaliana* Harmless to Ozone Layer 1 (*At*HOL1), first discovered by Nagatoshi *et al.*[49], catalyzes the conversion of *S*-adenosyl-L-homocysteine (SAH) and methyl iodide (MeI) to SAM. Previously reported, *At*HOL1 mutants generating *S*-adenosyl-L-ethionine (ethyl-SAM) from SAH and EtI (Figure 2A) were identified via a NNK single site-saturation mutagenesis performed by Tang *et al.* on ten residues in the active site (i.e., P20, V23, L27, W47, W36, V140, C143, Y139, Y172, R214, Figure 2B), leading to a 190-mutant library where 4 potentially beneficial mutants were identified (i.e., V140T, V140C, V23T, W36F).[51] We tested to what extent SubTuner can identify beneficial mutants in its final suggestion using the experimental results as the ground truth. The prediction outcomes allowed us to estimate to what magnitude the function enhancement could speed up if SubTuner were to assist the experimental screening.

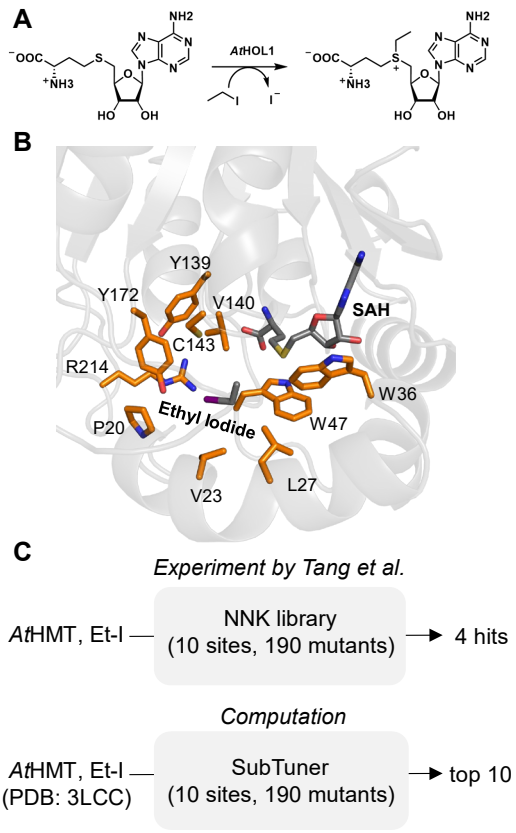


Figure 2. Reaction schematic, active site structure, and screening conditions for *At*HOL1, the model enzyme used in this study. (A) The target enzymatic reaction in *At*HOL1. (B) The structure of the active site (PDB ID: 3LCC). The residue sites involved in mutations in the test are highlighted as orange sticks. EtI and SAM are shown in gray sticks. (C) The details of the experimental directed evolution and the setup of the SubTuner test.

SubTuner takes the *At*HOL1-EtI complex as input and predicts ten beneficial mutant candidates as output (Figure 2C). The 190 mutants in the initial mutant library were first screened by thermostability. Using the cutoff defined in Implementation (pass rate: 40%), SubTuner selected 76 mutants (labeled as green squares, *Thermostability Screening*, Figure 3 and see Table S3 for more details). These stable mutants then underwent the TS-analog binding test through classical MD simulations to filter for mutants with favorable TS-binding energy (i.e., low ΔH_{bind}

value) and activation entropy (i.e., low active site RMSD), where an S_N2-type TS-analog is modeled by restraining the reaction coordinates between EtI and SAH.[58] The top 40 mutants were ranked separately for ΔH_{bind} and active site RMSD values. The intersection of the two rankings results in 27 mutants (*TS-analog Binding Screening*, Figure 3). SubTuner ranked the remaining 27 mutants by calculating $\Delta\Delta G_{ele}^{\ddagger}$, the change of electrostatic stabilization energy caused by the mutation and selecting the top ten mutant candidates (green box, *Reactivity Ranking*, Figure 3) as the predicted beneficial mutants. Among the ten scored mutants, we observed three experimental hits, which are ranked in the 1st (V23T), 2nd (V140C), and 6th (V140T) place, improving the wild-type *At*HOL1 ethyl-SAM activity by 1.6, 3.6, and 5.3-fold, respectively. W36F, an experimental hit whose activity improvement is merely 1.1-fold, falls off from the top ten and ranks in the 19th place.

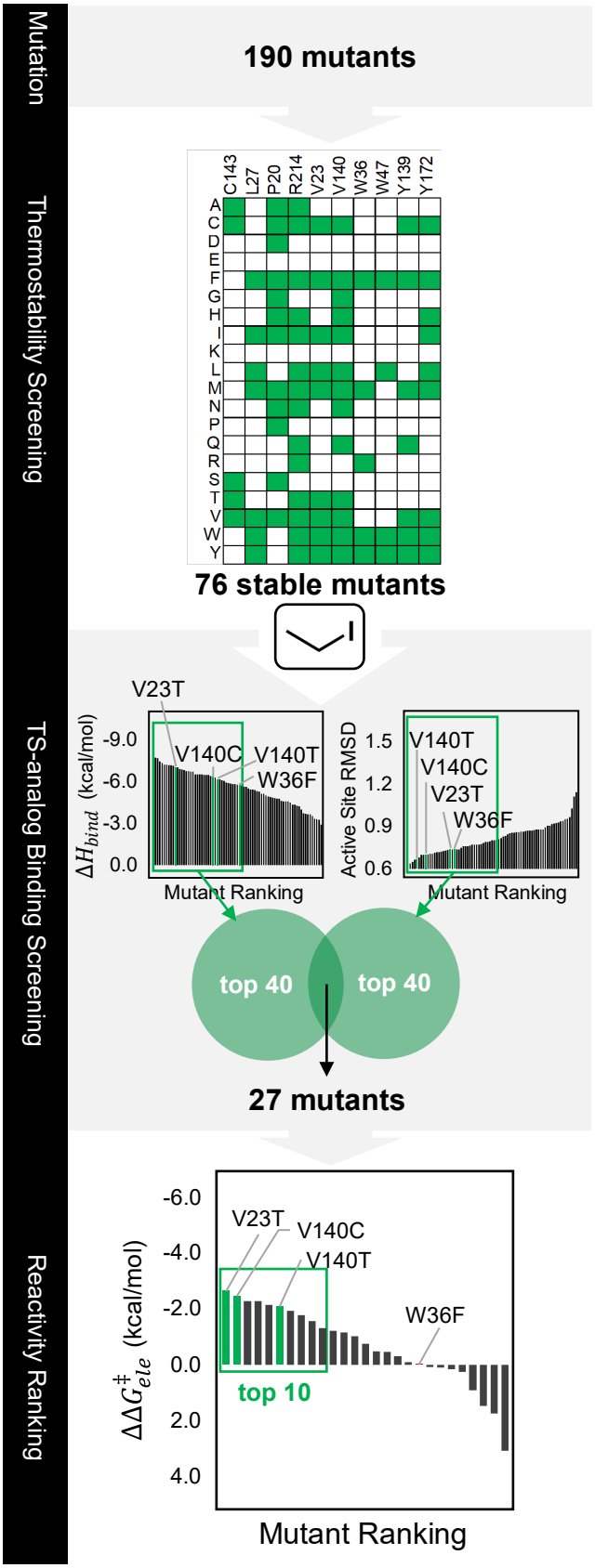


Figure 3. SubTuner screening results. Ten beneficial mutant candidates were predicted for *At*HOL1 toward EtI. *Thermostability Screening*: the thermostability matrix shows the result of the screening for each mutant. The column of the matrix stands for the residue sites on different variants of mutation. The row of the matrix stands for different target residues of the mutation. The color of each matrix element indicates thermostable (green) or unstable (white). (See detailed of the screening methods in Implementation) *TS-analog Binding Screening*: 2 bar graphs are used to present the ranking of the MMPBSA binding enthalpy (ΔH_{bind}) and the Active Site RMSD. (See detailed of the screening methods in Implementation) The x-axis of both graphs indicates the ranking of the mutants from high to low. For ΔH_{bind}, the y-axis is inverted to show a taller bar for a more negative number, which indicates a tighter binding. The name of the experimental hits is noted next to the corresponding value. The green box shows the top-40 mutants from each ranking. The Venn diagram represents mutants that are ranked as top-40 in both rankings. *Reactivity Ranking*: the bar graph represents the $\Delta\Delta G_{ele}^{\ddagger}$ ranking of the mutants from the previous screening rounds. The y-axis is inverted to show a taller bar for a more negative value, which suggests a higher activity relative to the wild type. The name of the experimental hits is noted next to the corresponding value. The green box shows the top ten mutants that compose the final mutant suggestion.

To evaluate SubTuner's performance, we calculated the hit rate and function-enhancing speed (detailed in Section 2a) on the *At*HOL1–EtI reaction, comparing each of these metrics with experimental estimates. First, SubTuner identified three experimentally validated beneficial mutants from ten predicted mutants, exhibiting a hit rate of 30%. In contrast, the experimental hit rate is 2.1%. (four hits from 190 candidates) The hit rate of SubTuner thus outperforms random experimental screening by 14-fold. Second, it takes ~2 days for SubTuner to generate the ten

mutants (MD is most time-consuming[44]; see hardware specifications in Text S8), followed by ~3 days for a protein engineer to complete site-directed mutagenesis and screen through ten mutants (i.e., 1 day for transferring 30 vectors, 1 day for expression, and 1 day for performing the crude enzyme activity assay). With a total effort of five days, SubTuner-facilitated experimental screening achieves a function-enhancing speed of 1.06-fold per day (i.e., V140T with an improvement of 5.3-fold). In contrast, the total effort of site-saturation mutagenesis is estimated to be ~34 days, including 1 day for the site-saturation PCR with the NNK codon, 2 days for cloning, 3 weeks for expressing at least 224 variants from each plate to reach 99% coverage of the mutants,[59] 3 days for screening, 7 days for sequencing the hits. As such, site-saturation mutagenesis exhibits a function-enhancing speed of 0.2-fold per day. The SubTuner tool substantially speeds up the function-enhancing speed by about 5-fold.

2c. Test of SubTuner in identifying beneficial multi-point *acl*-MT mutants to catalyze Ethyl- and Propyl-SAM formation

To further test the generalizability of SubTuner across various substrates, enzyme sequences, and types of mutations (single and multi-mutations), we tasked SubTuner to expand the substrate scope of anion methyltransferase, *acl*-MT, towards ethyl iodide (EtI) and *n*-propyl iodide (*n*PrI) (Figure 4A). Experimentally screening a Spread-Out Low Diversity (SOLD) library of 600 multi-point *acl*-MT mutations has yielded ~10% active mutations (57 hits for ethyl iodide; 65 hits for n-propyl iodide), in which the SOLD library was constructed by combining mutations from five residue sites through MD-guided rational engineering (i.e., V11, L30, W31, W41, L39, W41, Figure 2B).[52] Similar to the *At*HOL1 test described in Section 2b, we evaluated the performance of SubTuner based on its hit rate among the final list of ten predictions for each substrate and its functional enhancement speed.

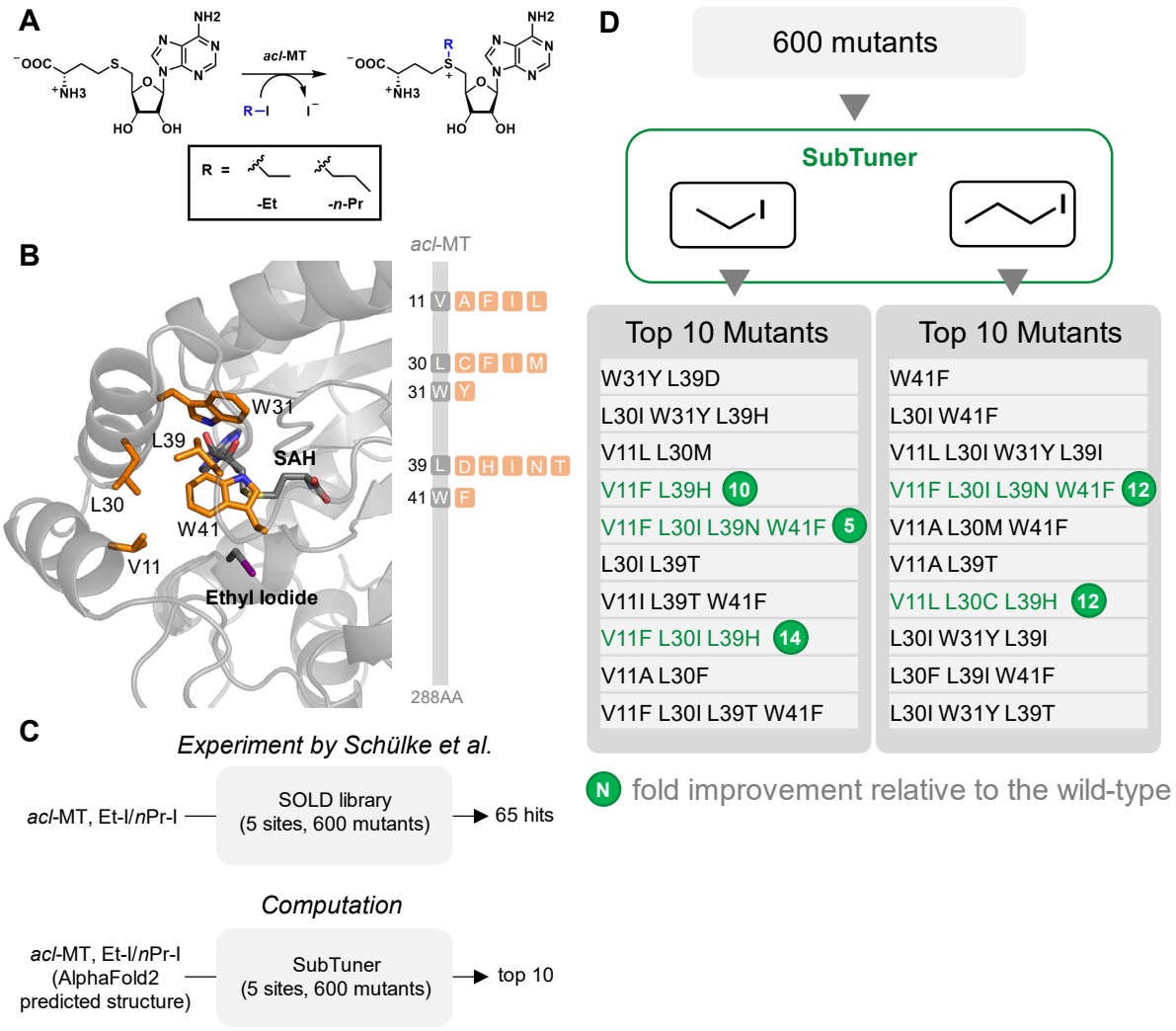


Figure 4. Reaction schematic, active site structure, screening conditions, and SubTuner results for *acl*-MT, the model enzyme used in this study. (A) The target enzymatic reaction in *acl*-MT. (B) The structure of the active site (AlphaFold DB, access ID: A1CIS5). The residue sites involved in mutations in the test are highlighted as orange sticks. EtI is used here as an example of the substrate. EtI and SAM are shown in gray sticks. (C) The details of the experimental directed evolution and the setup of the SubTuner test. (D) SubTuner screening results. The mutant's activity improvement relative to the wild-type is shown as the number in the green circle.

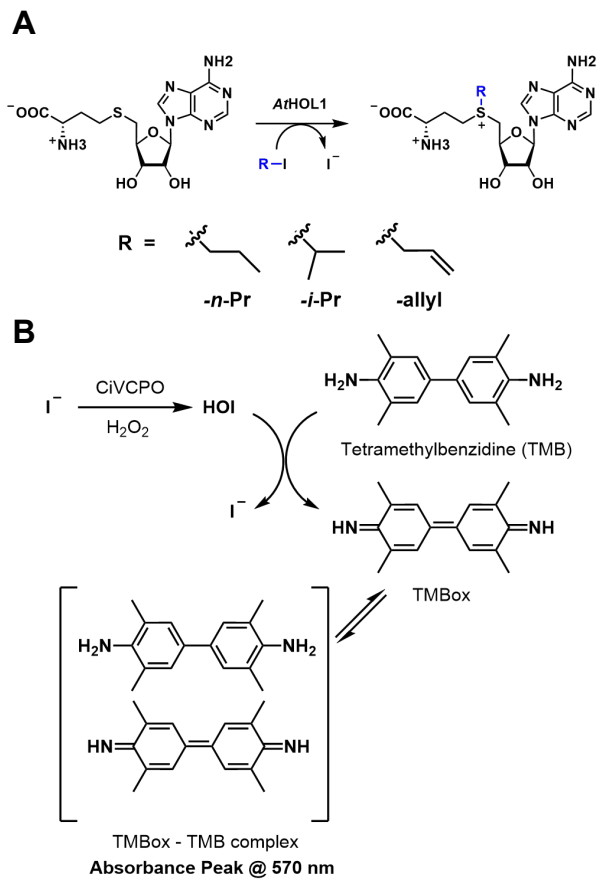
SubTuner takes the two enzyme-substrate complexes (*acl*-MT-EtI and *acl*-MT-*n*PrI) as input separately and creates two independent tasks, each predicting ten beneficial mutant candidates as output (Figure 2C). The 600 mutants in the initial mutant library were first screened by thermostability to yield 240 stable mutants. These stable mutants then underwent the TS-analog binding test, which gave 22 mutants for ethyl iodide and 14 mutants for n-propyl iodide (Table S3). SubTuner ranked these mutants by their activity ($\Delta\Delta G_{ele}^{\ddagger}$) for each substrate and yielded the top ten mutants as the final suggestion. Among them, we found three beneficial mutants for ethyl iodide, which are: V11F-L39H, V11F-L30I-L39N-W41F, and V11F-L30I-L39H, improving the *acl*-MT activity by 10-, 5-, and 14-fold, respectively, toward catalyzing ethyl iodide. For *n*-propyl iodide, we found two beneficial mutants, V11F-L30I-L39N-W41F and V11L-L30C-L39H, in the top ten suggestions, both improving the *acl*-MT activity by 12 folds toward propyl iodide.

To evaluate SubTuner's performance, we calculated the hit rate and function-enhancing speed for both substrates. For ethyl iodide, SubTuner identified three hits with the greatest activity improvement of 14 fold (i.e., V11F-L30I-L39H). With the additional site-directed mutagenesis experiment, the total discovery time would be 8 days (5 days for running SubTuner + 3 days for site-directed mutagenesis and screening of 10 mutants). This gives the overall hit rate of 30% and a function-enhancing speed of 1.8-fold/day. For *n*-propyl iodide, SubTuner identified two hits with the greatest activity improvement of 12 folds (i.e., V11L L30C L39H) from ten mutant suggestions in 8 days, leading to an overall hit rate of 20% and function-enhancing speed of 1.5 fold/day. As a reference, experimentally screening the SOLD library of 600 mutants with UHPLC-ESI-MS found 57 hits for ethyl iodide with the greatest activity improvement of 16 folds in an estimated 3 months, corresponding to an overall hit rate of 9.5% and a function-enhancing speed of 0.2-fold/day. In contrast, SubTuner can enrich the concentration of beneficial mutants in the smart library and

accelerate the speed of identifying function-enhancing mutants. These results demonstrate SubTuner's generalizability in accelerating the enzyme engineering processes for substrate scope modification.

2d. Apply SubTuner in identifying beneficial *At*HOL1 mutants to catalyze bulkier substrates

Although the two tests described in section 2c and 2d validated the efficiency and generalizability of SubTuner, they are a posterior in nature. As such, we designed a prior assessment, which employed experimental assays to characterize the predicted mutants identified by SubTuner *in priori*. This test serves as a proof-of-principle to demonstrate SubTuner's capability of guiding experimental engineering. Specially, we applied SubTuner to engineer *At*HOL1 toward three bulkier substrates (Scheme 1A), *n*-propyl iodide (*n*PrI), isopropyl iodide (*i*PrI), and allyl iodide (allyl-I). V140T was experimentally shown to be active toward these substrates even though it was discovered in the site-saturation mutagenesis screening for EtI.[51] Through separately testing each substrate against the library of 76 stable mutants, we also investigated whether SubTuner can consistently predict V140T among the ten recommended mutant candidates.



Aslan-Uzel *et al.*, 2020 & Tang *et al.*, 2021

Scheme 1. Schematics of *At*HOL1-catalyzed syntheses of non-native SAM analogs and the biochemical assays for mutagenesis screening. (A) *At*HOL1-catalyzed reactions of SAH with three bulkier substrates: *n*PrI, *i*PrI, and allyl-I. The substituent differences of these substrates are highlighted in blue in the reaction scheme. (B) The tetramethylbenzidine (TMB) assay for characterization enzyme turnover. The iodide anion released from substrate haloalkane is oxidized by *Curvularia inaequalis* Vanadium-Dependent Chloroperoxidase (*Ci*VCPO) to hypoiodous acid, spontaneously oxidizing TMB to TMBox. TMBox forms a complex with TMB with a characteristic absorbance peak at 570 nm. (See Methods for more details)

Three separate SubTuner tasks, corresponding to each of the bulkier iodoalkane substrates, were conducted to recommend the top ten mutant candidates for each substrate (Figure 5). Since

folding free energy on apo-*At*HOL1 is substrate-independent, SubTuner directly recycled the 76 mutants involved in the thermostability screening stage of EtI (Figure 3). These mutants underwent TS-analog binding screening, resulting in 27 mutants for *i*PrI, 26 mutants for *n*PrI, and 23 mutants for allyl-I (TS-analog Binding Screening, Figure 5, Table S3). SubTuner ranked these mutants by their reactivity ($\Delta\Delta G_{ele}^{\ddagger}$) for each substrate and selected the top ten mutant candidates (green box, Reactivity Ranking, Figure 5). SubTuner succeeded in identifying V140T among the recommended mutants for two substrates, ranking it 1st place for *n*PrI and 8th place for allyl-I. For *i*PrI, however, V140T was ranked at the 15th place. Though falling off from the top ten list, the result does not necessarily suggest a failure of SubTuner in *i*PrI, because most mutants recommended by SubTuner had not been characterized experimentally in its original publication.[51] Therefore, we performed experimental tests on the top five recommended mutants for each substrate to validate whether beneficial mutants were correctly identified and discover new *At*HOL1 mutants that catalyze these bulkier substrates.

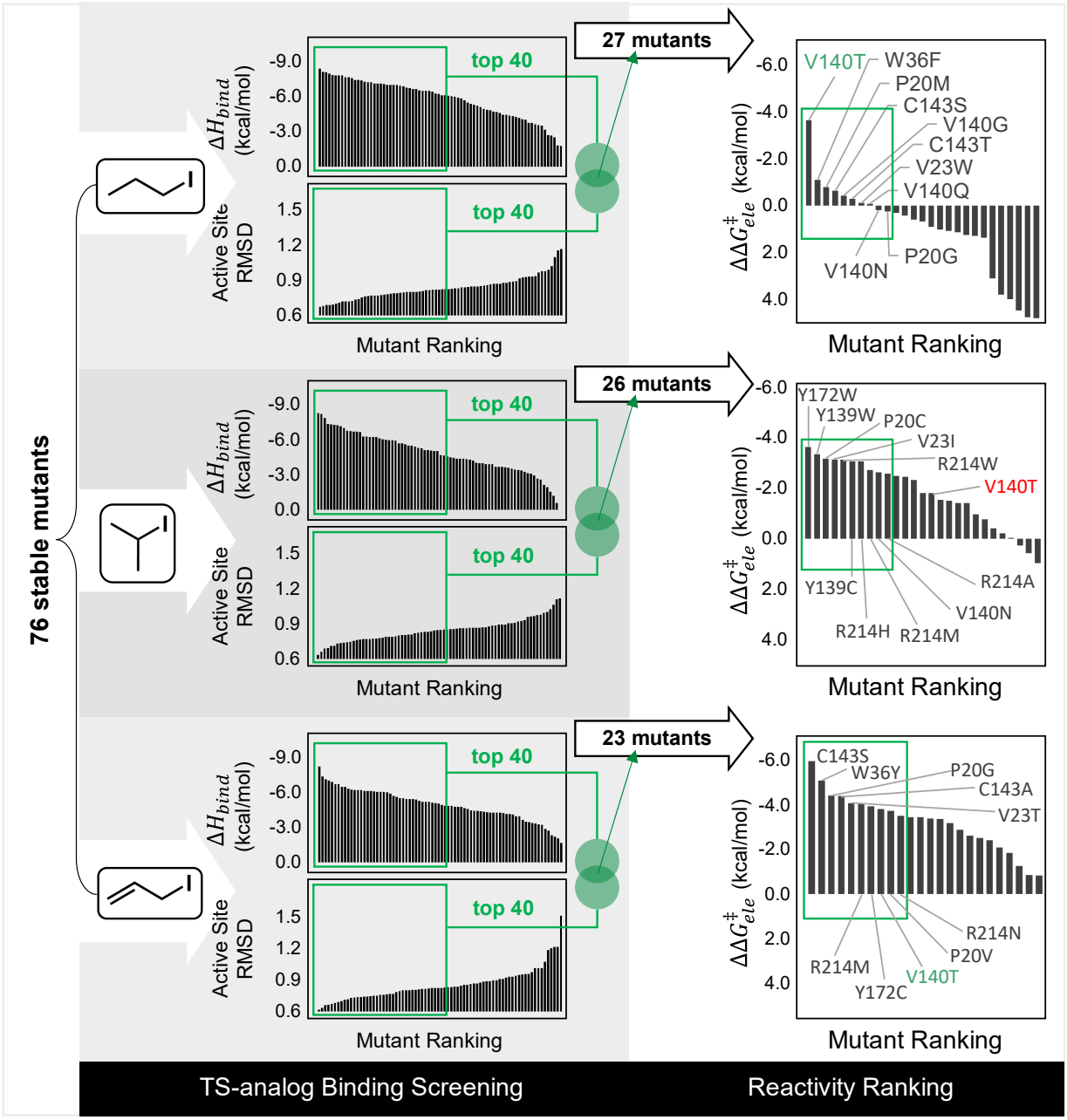


Figure 5. Screening results of the SubTuner to predict ten beneficial *At*HOL1 mutant candidates toward EtI, *n*PrI, *i*PrI, and allyl-I. The top ten candidate mutations are labeled in the Reactivity Ranking section. V140T, a beneficial mutant reported by Tang *et al.*, is recommended for *n*PrI and allyl-I (colored in green), but not for *i*PrI (colored in red).

We synthesized, expressed, and purified the top five predicted *At*HOL1 mutants for each substrate. The activity was determined using a chloride-insensitive, iodide-sensitive *Ci*VCPO-TMB assay based on the absorbance at 570 nm (Scheme 1B and the Method section). For *n*PrI, two out of the five SubTuner-suggested mutations show improved activity (V140T and P20M, Figure 6), with V140T demonstrating the greatest turnover improvement of 4.9-fold higher than the wild-type. For *i*PrI, even though V140T is not included in the final suggestions, three out of the five SubTuner suggestions did show improved activity (P20C, R214W, Y139W, Figure 6), with P20C exhibiting the greatest turnover improvement of 3.1-fold higher than the wild-type (slightly greater than V140T, Table S2). Notably, for allyl-I, none of the five mutants demonstrated improved activity compared to the wild-type, which already exhibits high efficiency.

SubTuner identified 5 hits with the greatest activity improvement of 4.9-fold (i.e., V140T for *n*PrI) from 15 mutant suggestions in 10 days (6 days for running SubTuner + 4 days for site-directed mutagenesis and screening of 15 mutants). This gives the overall hit rate of 33.3 and a function-enhancing speed of 0.5-fold/day, outperforming experimental screening estimated from EtI screening (function-enhancing speed 0.2-folds/day). Notably, this proof-of-principle test only involves one round of SubTuner engineering with single-point mutations. In practice, multiple rounds of SubTuner engineering with multi-mutations are expected, which likely yields higher activity improvements.

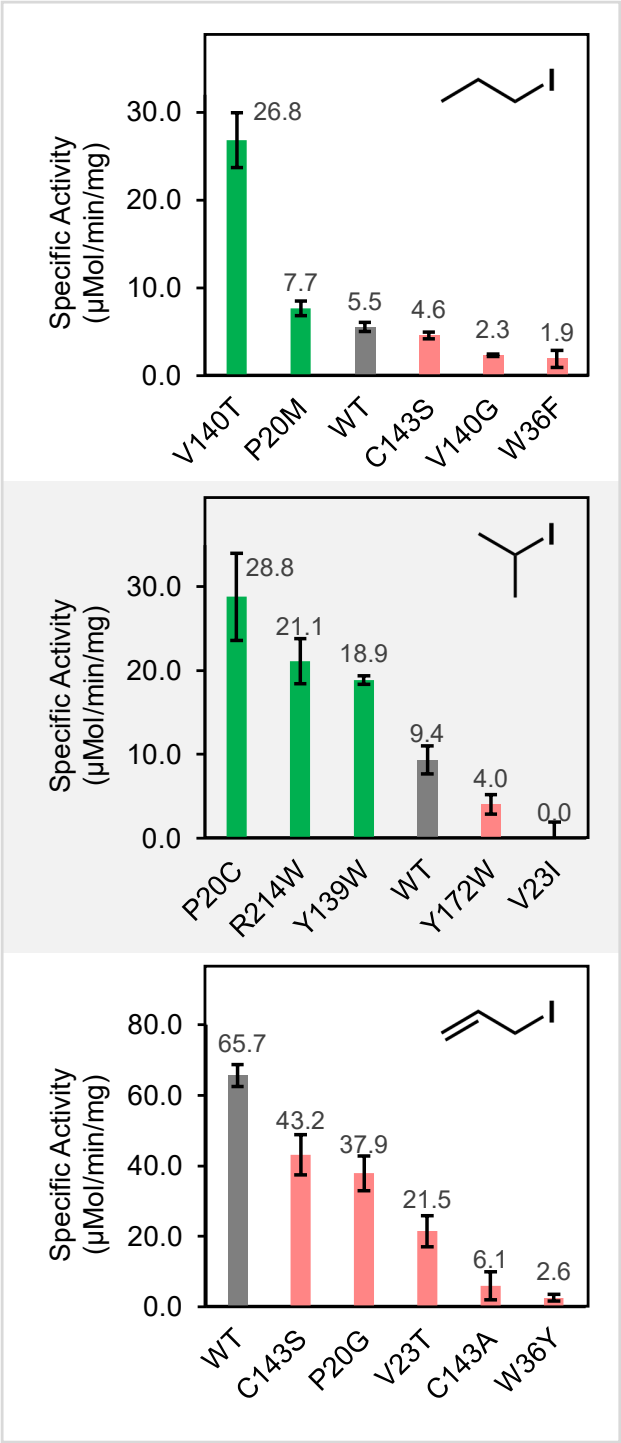


Figure 6. Turnover of top five SubTuner predicted *At*HOL1 mutants and wild-type against various iodoalkane substrates at 3.2 mMol. The corresponding substrate of the test is noted in the top right

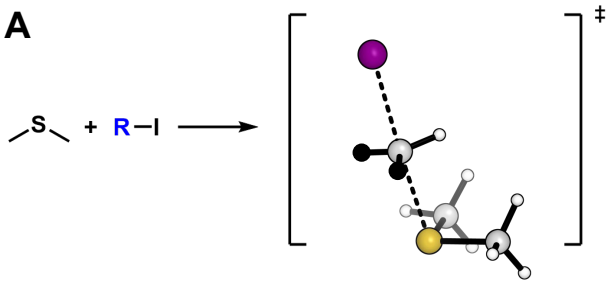
corner. The mutants that have improved or deteriorated activity relative to the wild type are shown in green or red, respectively. The wild type is shown in gray.

2e. Mechanistic insights behind the beneficial *At*HOL1 mutants

The ability to derive mechanistic, atomic-level insight into the impact of beneficial mutations on biocatalysis is valuable beyond the ability to enable enzyme optimization. To this end, we utilized molecular simulation data derived from SubTuner to elucidate the impact of beneficial mutations on catalysis. Three mechanistic questions stand out. First, why does wild-type (WT) fail to maintain high catalytic activity on bulkier substrates (i.e., EtI, *n*PrI, *i*PrI, and allyl-I)? Second, why does V140T exhibit promiscuous activities toward the tested bulkier substrates? Third, why does P20C, an uncommon mutation target, enhance *At*HOL1's activity toward *i*PrI?

To investigate the failure of wild-type *At*HOL1 to catalyze bulkier substrates, we started by performing density functional theory (DFT) calculations on the intrinsic reaction mechanisms between dimethyl sulfide (used as an analog of SAH) and various substrates (i.e., MeI, EtI, *n*PrI, *i*PrI, and allyl-I), elucidating the energetic reasons behind WT's low activity toward non-native substrates (Text S12). The results show that the intrinsic reaction barrier in water increases from 17.5 kcal/mol for MeI and allyl-I, to 20.4 kcal/mol for EtI and *n*PrI, and further to 21.2 kcal/mol for *i*PrI (Figure 7A). From MeI, EtI, *n*PrI, to *i*PrI, the trend of barrier height increasing, driven by the decreasing electrophilicity of the reacting carbon, is correlated with the experimentally observed rate reduction (i.e., MeI > EtI > *n*PrI ~ *i*PrI, Table S2). However, the intrinsic preference of the electronic structure does not explain the low activity of WT for allyl-I, which has a similar barrier to MeI but reacts 10^{-4} fold slower (specificity activity, Table S2). Through inspecting the

active site binding pocket in the crystal structure of *At*HOL1 with docked MeI, we identified the small cavity size as another factor behind WT's failure to accommodate a bulkier substrate. Specifically, enlarging the carbon chain on the substrate will cause steric hindrance with W36 and V140 (Figure 7B), hampering binding (worse MMPBSA binding energies, Table S2) and misaligning the substrate with the interior enzyme electric field (worse electrostatic stabilization energy, Table S2). The steric clash causes EtI, *n*PrI, and *i*PrI to react even slower than what is expected from their intrinsic barrier (lower specific activity, Table S2). Guided by these findings, we hypothesize that V140T and P20C rescue activities of HMT on bulkier substrates by reshaping the active site cavity to favor binding and electrostatic stabilization.



R	$\Delta G^{\ddagger}$ (kcal/mol)
Me	17.5
Et	20.4
*n*Pr	20.4
*i*Pr	21.2
allyl	17.4

PBE0-D3/def2-TZVP/SMD(H_2O)//PBE0-D3/def2-SVP/SMD(H_2O)

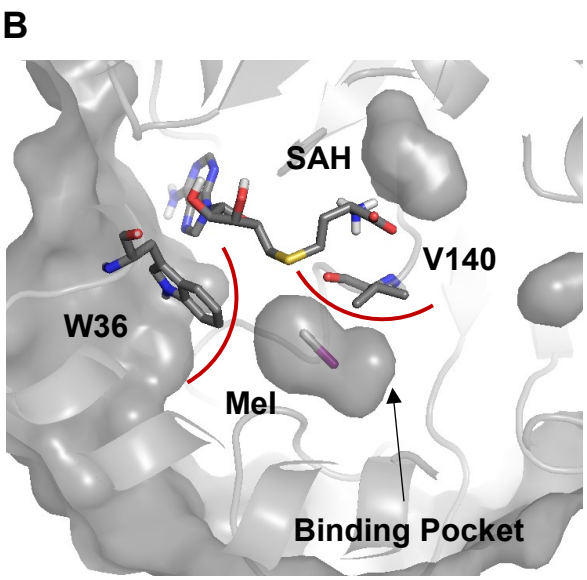


Figure 7. Intrinsic reaction barriers between dimethyl sulfide and iodoalkyl compounds (i.e., MeI, EtI, *n*PrI, *i*PrI, and allyl-I) and the active-site structure of wild-type *At*HOL1. (A) The free energy barrier of the intrinsic reaction is calculated by the DFT method (optimization: PBE0-D3/def2-SVP; single-point energy: PBE0-D3/def2-TZVP; both treated with SMD implicit water model).

The letter "R" in blue indicates different substituents on I. The cartoon shows a representative TS, with the black balls showing sites of substituents of the "R" groups in the table. Details about the QM calculation are described in Text S3. (B) The active site binding pocket in the crystal structure of *At*HOL1 (PDB ID: 3LCC) with MeI docked and solvent-exclusive surface (gray) is shown. V140 and W36 are highlighted as sticks with non-polar hydrogens hidden.

To understand the promiscuity of the V140T variant, we looked into the MD trajectory and the EF calculation generated during the SubTuner screening against EtI, *n*PrI, *i*PrI, and allyl-I. Different from the WT, where residue Val140 involves a flexible conformation characterized by a wider distribution of dihedral angles (Figure S3B, dihedral distribution), mutation to Thr140 results in H-bonding to the carboxylic acid group on SAH, forcing it to consistently adopt a conformation that reduces steric hindrance with each binding substrate (Figure S3). In addition, the Thr140-SAH interaction positions each substrate's C–I bond along a favorable interior enzyme electric field for stabilization (Table S1 and Figure S4). Taking EtI as an example, due to the formation of Thr140-SAH hydrogen bond, V140T better aligns the substrate's C–I bond with the direction of interior enzyme electric field (average angle with EF, V140T: 16.6° vs. WT: 41.2°), leading to a stronger projected EF with a narrower distribution (V140T: -43.4 ± 5.7 MV/cm vs WT: -35.2 ± 9.6 MV/cm, Figure 8A). These changes correspond to an enhancement of $\Delta\Delta G_{ele}^{\ddagger}$ by 2.2 kcal/mol for EtI (V140T: -7.8 kcal/mol vs WT: -5.6 kcal/mol), 3.6 kcal/mol for *n*PrI, 1.8 kcal/mol for *i*PrI and 3.8 kcal/mol for allyl-I (Table S2). These results suggest that the promiscuity of the V140T variant originates from its creation of a larger active site binding pocket and the subsequent positioning of the substrate's C–I bond along the enzyme's electric field for stabilization, facilitated by the Thr140-SAH hydrogen bonding.

The final question we aim to answer is what causes P20C, an unusual mutation located in the second shell of the active site, to be stronger in activity toward *i*PrI than WT and V140T. Due to hindered peptide bond isomerization caused by the conformation of the cyclic side chain, proline residues usually adopt a *cis* conformation induced β–turn and are therefore not a common mutation target as mutation to a less rigid amino acid can disturb the fold significantly. Based on the MD data collected from SubTuner, we observed that P20C, though not directly interacting with the substrate, causes a dynamic shift of the α-helix (residue 20-40, PDB: 3LCC) and makes a nearby loop more flexible (Figure S2). This dynamic shift repositions the α-helix containing V140 away from the substrate, thereby reducing steric repulsion in the active site (MMPBSA energy, Table S2; see the six components of the dynamic shift described in Text S10 and Figure S2). These dynamic shifts better align *i*PrI's C–I bond with a favorable EF in the enzyme (P20C: -42.7 ± 11.0 MV/cm vs WT: -28.8 ± 9.0 MV/cm, Figure 8B), allowing P20C to enhance $\Delta\Delta G_{ele}^{\ddagger}$ by 2.2 kcal/mol (P20C: -6.9 kcal/mol and WT: -3.7 kcal/mol) These results are consistent with P20C promoting catalysis of *i*PrI through a dynamic shift of loops and α-helixes, informing a new hypothesis to further improv the activity of *At*HOL1 through optimizing substrate positioning.

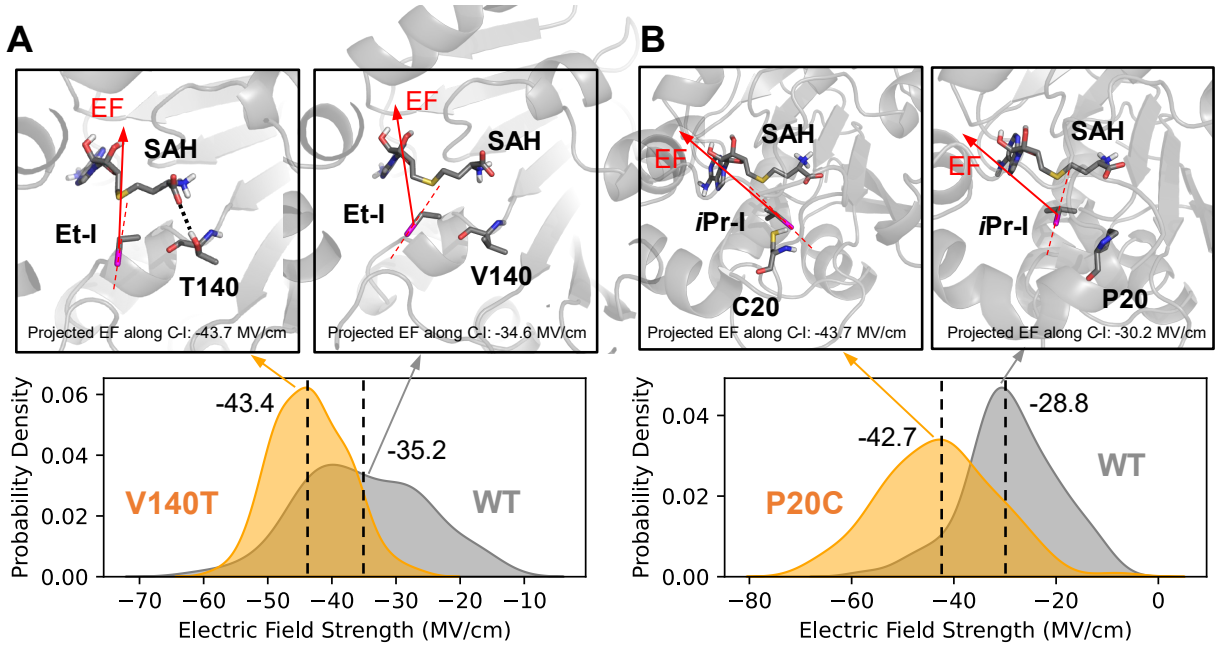


Figure 8. (A) The projected electric field strength (EF) distribution of V140T and WT in complex with EtI. (B) The projected electric field strength (EF) distribution of P20C and WT in complex with *i*PrI. For both subfigures, the kernel density distribution is fitted using the *gaussian_kde* function from the *scipy* package. The fitting data is obtained from EF calculations of 100 snapshots sample from the MD trajectory. Representative *At*HOL1 structures (top) are shown with projected EF values (labeled) with the direction of EF (red arrow) and direction of C-I bond (red dashed) indicated. The probability density function of the electric field strength (bottom) for wild-type *At*HOL1 (gray) and mutant *At*HOL1 (orange) with indicated average value of the distribution (dotted line, labeled).

3. Discussion

In this work, we developed a physics-guided computational tool, SubTuner, to predict beneficial mutations for catalyzing non-native substrates. SubTuner begins with the enzyme-

substrate complex as an input, screens a mutant library through three selection steps (i.e., assessing thermostability, transition state (TS) analog binding, and electrostatic stabilization), and provides ten beneficial mutant candidates as an output. We evaluated SubTuner for its ability to identify mutants that improve *At*HOL1's activity towards EtI and other bulkier substrates (*n*PrI, iPrI, and allyl-I), demonstrating its accuracy, generalizability, and mechanistic interpretability. Below, we discuss three questions regarding SubTuner's technical strengths and limitations, as well as its scientific implications.

How does SubTuner guide enzyme engineering? SubTuner predicts ten beneficial mutant candidates for each enzyme-substrate complex, forming a "smart library" to accelerate enzyme engineering. Besides applying the predicted mutations for experimental testing, users can identify the mutation hotspots by ranking the frequency of each mutation site involved in the final list (e.g., V140 with a frequency of 22.2%, Table S3), and design combinatorial site-saturation mutagenesis accordingly. In addition, SubTuner, enabled by its generated molecular modeling data, informs atomic-level insight into the catalytic actions of beneficial mutants that enhance enzyme activity toward non-native substrates (see Section 4d). These insights serve as an engineering principle to guide iterative rounds of activity optimization.

Though tested in *At*HOL1, SubTuner is generalizable to other enzyme targets. The overall strategy is independent of specific applications. This complements experimental protocols, which can vary significantly across different enzymes, involving gene expression, protein purification, and activity readout designs. However, one potential limitation of SubTuner lies in the choice of the target transition state (TS). While this may not be a problem for enzymatic reactions with a well-defined rate-limiting transition state like methyltransferases, multi-step enzymatic reactions

may have altered rate-limiting steps upon mutation.[60, 61] This serves as a key challenge for the next step development of SubTuner.

How does SubTuner compare with FuncLib? FuncLib is among the most widely applied *in silico* enzyme engineering tools in the community. With the input of the enzyme structure, FuncLib recommends mutations that are most evolutionarily frequent in nature and involve a high-ranking score of Rosetta energy function. These mutations, though predicted without dependence on the information of any target substrate, have been extensively employed to screen against certain substrates of interest for improved activities.[25-38] Due to FuncLib's high accessibility and popularity in the field, we compared the hit rate of FuncLib-predicted mutations to that of SubTuner. Here, we tested the performance of FuncLib to identify beneficial mutants for 1) *At*HOL1 to catalyze the conversion of ethyl iodide, and 2) *acl*-MT of ethyl iodide and *n*-propyl iodide, and compared the outcome with that of SubTuner (Table 1). We ran FuncLib on the same mutant library from *At*HOL1 and *acl*-MT and obtained the top ten mutant suggestions for each enzyme. For *At*HOL1, FuncLib found one beneficial mutant, V140C, while the other mutants are shown to be inactive or less active than the wild-type. For *acl*-MT, all FuncLib mutant suggestions are deleterious for both ethyl iodide and *n*-propyl iodide. In contrast, SubTuner found three hits for *At*HOL1 against ethyl iodide (best activity improvement: 5.3-fold), three hits for *acl*-MT against ethyl iodide (best activity improvement: 14 fold), and two hits for *acl*-MT against *n*-propyl iodide (best activity improvement: 12 fold). As such, SubTuner complemented FuncLib in the field and filled in the gap of *in silico* enzyme engineering targeting the activity of a specific non-native substrate.

	AtHOL1 – EtI		acl-MT – EtI/*n*PrI	
	Funclib Suggestion	Exp. Activity (relative to WT)	Funclib Suggestion	Exp. Activity (relative to WT)
1	V140L	<1	V11I L30I L39M	<1, <1
2	V140N	<1	V11M L39M	<1, <1
3	V140D	<1	L30V W41M	<1, <1
4	V140Q	<1	V11D L30V L39M	<1, <1
5	V140C	3.6	V11M	<1, <1
6	V140F	<1	L39Q	<1, <1
7	V140A	<1	W41M	<1, <1
8	V140G	<1	L30I W41M	<1, <1
9	V140E	<1	L39Q W41Y	<1, <1
10	V140S	<1	V11N L30I L39M	<1, <1

Table 1. FuncLib's top ten mutant suggestions for *At*HOL1 and *acl*-MT and their experimental activity against ethyl iodide and *n*-propyl iodide. The FuncLib result is obtained from the FuncLib webserver (funclib.weizmann.ac.il, Text S11). The experimental activity is obtained from the experimental screening by Tang *et al.* for AtHOL1[51] and Schülke *et al.* for *acl*-MT[52]. For *acl*-MT, the experimental activity is listed for both ethyl iodide and *n*-propyl iodide separated by comma.

Do we need three stages of screening? To answer this question, we tested the performance of SubTuner in the absence of each screening stage using *At*HOL1-EtI system as an example. First, without thermostability screening, all 190 mutants go to the TS-analog binding screening, directly slowing down the workflow by 2.5 folds (increasing the number of MD simulations from 76 to 190). Furthermore, some unstable mutants show TS-analog binding ability due to stability-activity compensation,[62] excluding real hits from the final output. As a result, the top ten beneficial mutant candidates do not contain any real hits (Table S4). These results demonstrate that thermostability screening is vital to SubTuner's success.

Second, without the TS-analog binding screening and its associated MD simulations, the 76 mutants from thermostability screening will directly undergo TS stabilization ranking ($\Delta\Delta G_{ele}^{\ddagger}$) using their static structures, rather than the conformational ensemble. As a result, mutants that hinder binding but facilitate stabilizing the reacting dipole of the TS in an arbitrary conformation are included in the top-ranking list, resulting in no hits found in the final list (Table S5). These results show the importance of the TS-analog binding step in beneficial mutant screening.

Finally, without the TS stabilization ranking, the final list of mutant candidates will comprise 27 unranked mutants (in the *At*HOL1-EtI case) resulting from the TS-analog binding screening. The user has to randomly select if they still want ten final candidates. In this case, with the random selection, the expected value of the hit rate will be 14.8% (Text S9), dropping the original hit rate of SubTuner by half (i.e., 30%). Notably, compared to the MD simulations, the single-point QM cluster calculation is resource efficient and costs relatively omittable time.[44] Therefore the computational time will be similar. In the same time, the expected value of the greatest activity improvement from ten random selections in these 27 mutants is 3.1-fold (Text S9). This gives 59.3% of the original function-enhancing speed of SubTuner (5.3-fold). As such, reactivity ranking significantly improves the accuracy of SubTuner with minimal computational overhead.

How do we balance computational accuracy and efficiency? Balancing computational accuracy and efficiency in SubTuner involves considering various computational setups and parameters, such as MD sampling methods, time length, force fields, QM theories, and optimization approaches. Unlike traditional computational chemistry missions (e.g., thermochemistry predictions) with a well-established benchmark set, computational enzyme engineering deals with a moving target that hinges on experimental screening effort. The decision

to adopt more sophisticated computational methods depends on whether they would improve the hit rate and function-enhancing speed compared to pure experimental screening. For example, in the *Ruegeria sp.* TM1040 amine transaminase-phenylethylamine system, 12 beneficial mutants were identified from experimentally screening 57 mutants in 2 rounds of site-saturation mutagenesis, leading to a screening hit rate of 21%.[63] With such a high hit-rate from pure experimental screening, incorporating a higher level of QM theory, polarizable force field, and free energy perturbation for binding calculation into SubTuner will likely result in marginal to no improvement in hit rate, but undoubtedly cost more computational time. For intrinsically difficult enzyme engineering missions (i.e., have a low hit-rate even though all mutations are screened, or involve a complex mechanism), however, a combination of expensive molecular modeling methods can be critical to improve the number of hits from 0 to 1. The development of the critical assessment of computational enzyme engineering tools will be necessary, considering the complexity of balancing computational accuracy and efficiency, and will be foundational for comparing computational enzyme engineering approaches and achieving computational optimality.

4. Implementation

SubTuner takes the enzyme-substrate complex structure (i.e., considered as the "wild-type" structure) as input and predicts beneficial mutations as output. The SubTuner workflow consists of four steps, including Mutation, Thermostability Screening, TS-Analog Binding Screening, and Reactivity Ranking (Figure 1). These steps operate in an automated manner using our high-throughput enzyme modeling platform EnzyHTP.[42] Below, we detail each step in the workflow.

Step-1: Mutation. This step serves to determine the "mutant library" that will be screened by the following workflow. The workflow starts with preparing the input structure (i.e., the enzyme-substrate complex) for molecular modeling (detailed in Text S1). Using the prepared

structure as the "wild-type", candidate mutations are generated by site-saturation mutagenesis, random mutagenesis, rational design, or any empirical selection desired by the users (by the *assign_mutant* API in EnzyHTP), thus forming a "mutant library" for the remaining steps. In this work, we employed site-saturation mutagenesis.

Step-2: Thermostability Screening. This step serves to filter thermally unstable mutants from the "mutant library" constructed in *Step-1*. The change of the folding free energy ($\Delta\Delta G_{fold}$) relative to the wild type is used as the metric to assess the impact of mutations on thermostability. The $\Delta\Delta G_{fold}$ values are calculated using the *cartesian_ddg* protocol[53, 54] from Rosetta[55] and is automated by the *get_rosetta_ddg* API in EnzyHTP. (Text S4) The mutants are ranked based on their $\Delta\Delta G_{fold}$ values, and a fraction of unstable mutants are discarded. In this study, the workflow filters out 60% of the mutants – they are predicted to be thermally unstable relative to the wild type (i.e., has a positive $\Delta\Delta G_{fold}$).

Step-3: TS-analog binding screening. This step further eliminates the mutants that involve weak transition state (TS) binding. From a short list of mutants out of *Step-2*, the workflow first creates the enzyme mutant-substrate complex structure of the wild-type by replacing the side chain of target residues in EnzyHTP (powered by PyMol[64] and Amber[65]). These structures are then used as input for MD-based conformational sampling with constraints added on the reaction coordinates (i.e.: bond distance, angle, etc. of the forming bonds, Figure S1) to approximate the binding of a TS as the pre-reaction complex[66, 67] (or a TS-analog, see details in Text S2 and Text S7). For each mutant, the binding enthalpy of this TS-analog is calculated by the MM-PBSA method using the sampled conformational ensemble. In addition, the workflow also evaluates the flexibility of the active site by calculating the Root Mean Square Deviation (RMSD) of the active site (using

residues within 5 Å from any atom in the substrate, Text S5). A more rigid active site benefits catalysis by reducing the activation entropy.[68-70] The change of the RMSD of the active site is used as a descriptor of the change of the activation entropy upon mutations (i.e., TS binding entropy). The mutants are ranked separately based on the MMPBSA energy and the active site RMSD. The MMPBSA energies are ranked from negative to positive, i.e., strong to weak binding. The active site RMSD is ranked from low to high, i.e., rigid to flexible. The workflow selects mutants that are top in both ranks. In this study, mutants are selected that are among the 52% MMPBSA energy and active site RMSD values. In this work, the number is selected based on the acceptable amount of experimental effort.

Step-4: Reactivity ranking. This step ranks the remaining mutants from the previous screening steps by their chemical reactivities. The workflow uses the electrostatic stabilization energy of the transition state, $\Delta G_{ele}^{\ddagger}$,[56, 57] as the metric to assess the reactivity of each mutant. The difference in electrostatic stabilization energy between a mutant and the wild type ($\Delta\Delta G_{ele}^{\ddagger}$) reflects the stabilization of enzyme interior electric field on a reacting dipole, which has been shown to correlate with the change of activation energy or enthalpy of the corresponding chemical step.[39, 56, 57, 71-74] Distinct from former steps, this step considers the impact of mutation on the electronic structure of ethe nzyme active site. To calculate $\Delta\Delta G_{ele}^{\ddagger}$ from each MD-sampled snapshot, the workflow first constructs a QM cluster consisting of the reacting species and residues, and then performs single point energy calculations using a Gaussian16[75] interface in EnzyHTP. (Text S3) This yields the electronic structure of the active site, from which the dipole moment of the breaking bond is estimated using the Multiwfn[76] interface in EnzyHTP. The $\Delta G_{ele}^{\ddagger}$ is computed as the negative dot product between the bond dipole and the electric field strength (see Text S6 for more details).

The mutants' $\Delta\Delta G_{ele}^{\ddagger}$ values are ranked from negative to positive, with negative meaning a more stabilized TS compared to the wild type. The top ten mutant candidates from this ranking are recommended for the experimental test. Notably, the number of mutants selected from the final ranking is flexible and is based on how much time the user wants to devote to the experimental test.

5. Experimental Methods

Chemicals, strains, and plasmids. Chemicals were purchased from the Sigma Chemical corporation (St. Louis, MO, USA) unless otherwise specified. The gene encoding CiVCPO from *Curvularia inaequalis* was generously provided to us in the pBADVCPO vector (acknowledgment to Uwe Bornscheuer) and transformed into *E. coli* TOP10 for protein expression. The gene encoding *At*HOL1 and mutants were synthesized and subcloned into pET-28a (+) vectors (pET28a_*At*HOL1_mut) via Genscript (Piscataway, NJ, USA), and transformed into *E. coli* TOP10 for storage and *E. coli* BL-21 (DE3) for protein expression. Antibiotic markers kanamycin (pET28a_*At*HOL1, pET28a_*At*HOL1_mut) and ampicillin (pBADVCPO) were used at a final concentration of 50 μg/mL (50mg/mL stocks) for transformation and culturing.

AtHOL1 Protein Expression and Purification. Plasmid pET28a_*At*HOL1_mut was transformed into chemically competent *E. coli* BL-21 (DE3) and plated (LB + kanamycin) for overnight incubation @ 37 °C. single colonies were picked and inoculated into 10 mL LB media with kanamycin in 50 mL Falcon Tubes to incubate (37 °C, rotating at 200 rpm for 16 h). An aliquot of 500 μL of culture was inoculated into 500 mL LB + kanamycin in a 2 L baffled flask and incubated (37 °C for 3 h) until OD_{600} reached 0.6. IPTG (500 mL of a 100 mM stock solution) was added to a final concentration of 0.1 mM, the temperature was lowered to 25 °C, and the flasks

were further incubated for 20 h. The cells were then harvested via centrifugation (2500 x g for 30 min), the supernatant was discarded, and the pellet was stored at -20 °C prior to lysis. The pellets were resuspended in 25 mL lysis buffer (20 mM imidazole, 500 mM NaCl, 20 mM Tris Base, pH 7.5) and one vial of 25 mL Culture Vial of CelLytic™ Express, and incubated (37 °C for 30 min). The cells were further lysed via French Pressure cell (2 times at >19,000 psi) and the lysate was ultracentrifuged (10,000 x g for 20 min). FPLC Procedure: the supernatant solution was loaded onto a Cytiva 5ml HisTrap™ FF column previously equilibrated with 5 column volumes (CV) lysis buffer. Elution was conducted via a linear gradient of imidazole (20 mM to 250 mM). The eluted fractions were monitored via A_{280}, fractions combined, concentrated to ~1 mL using a 15mL Amicon Ultracentrifugal Filter 10 kDa MWCO, and desalted into 50 mM $NaHPO_4$ (pH 7.5) The final protein concentration was determined via Nanodrop™ One (Thermo Fisher, Hennigsdorf, Germany). The purified protein was immediately used for activity assays.

CiVCPO Protein Expression and Purification. Plasmid pBADVCPO was transformed into chemically competent *E. coli* BL-21 (DE3) and plated on LB + ampicillin plates for incubation (37 °C for 16 h). Single colonies were picked and inoculated into 10 mL LB media + ampicillin in 50 mL Falcon Tubes at 37 °C, with shaking at 200 rpm for 16 h. A 500 µL of culture was inoculated into 500 mL LB media containing ampicillin in a 2 L baffled flask and incubated with shaking at 37 °C for 3 h until OD_{600} reached ~0.6. IPTG was added to a final concentration of 0.1 mM, the temperature was lowered to 20 °C, and the flasks were further incubated for 20 h. The cells were then harvested via centrifugation (2500 x g for 30 min), the supernatant discarded, and the pellet stored at -20 °C. The pellets were resuspended in 25 mL 50 mM Tris $H2_SO_4$ (pH 8), one vial of 25 mL Culture Vial of CellLytic™ Express, and one tablet of cOmplete™ Protease Inhibitor (Roche, Basel, Switzerland). The cells were further lysed via French Pressure cell (2 times at > 16,000psi)

and the lysate was ultracentrifuged (10,000 x g for 20 min). The supernatant was added to 25 mL isopropanol, incubated on ice for 30 m, and ultracentrifuged (10,000 x g for 20 m). FPLC Procedure**:** the supernatant solution was loaded onto a DEAE Column previously equilibrated with 5 column volumes (CV) 50 mM Tris H_2SO_4 (pH 8) buffer. The column was washed with 5 CVs of 50 mM Tris H_2SO_4 100 mM NaCl (pH 8). The CiVCPO was eluted with 50 mM Tris H_2SO_4 1M NaCl. Desalting via dialysis was conducted in 1L 50 mM $NaHPO_4$ 100 μM sodium orthovanadate (pH 8) for 4 h, the dialysis buffer was refreshed and further dialyzed for 16 h. The dialyzed CiVCPO was stored at 4 °C in a 10 mL Falcon Tube.

CiVCPO Activity Determination. CiVCPO stock activity was determined via the monochlorodimedone assay (Hager, "Chloroperoxidase II. Utilization of Halogen Anions", 1966). The bromination of monochlorodimedone leads to near total loss of absorbance at A_{290} ($\Delta\varepsilon = 20000$ M^{-1}, cm^{-1}). The assay mixture (1 mL, pH 6) consisted of monochlorodimedone (42 μM), $NaHPO_4$ (20 mM), sodium orthovanadate (1 mM final), NaBr (100 μM final), 0.9 μL of 30% hydrogen peroxide, and was initiated by adding 10 μL purified CiVCPO. The loss of A_{290} was monitored in a quartz cuvette, and the change in absorbance converted to μM monochlorodimedone brominated/uL CiVCPO solution via Beer's Law.

Activity Assay. The assay used for following alkyl iodide addition was adapted from a previously reported iodide detection assay.[77] Iodide is converted to hypoiodous acid via CiVCPO, consuming hydrogen peroxide. Hypoiodous acid chemically oxidizes the dye 3,3'5,5'-tetramethylbenzidine to a two-electron oxidized diimine, via the one-electron oxidized radical cation TMB, leading to an absorbance peak at 570 nm not observed in TMB. Hydrogen peroxide and TMB were included in the TMB Liquid Substrate System for Membranes. *At*HOL1 mutant protein preparations were diluted to 4.0 mg/mL. *S*-adenosylhomocysteine (stock: 40 mM in DMSO)

was added to the *At*HOL1 protein preparations (2 mM final). *At*HOL1 protein variant preparations (50 µL) were added to the wells of a 96-well polystyrene, flat-bottomed, clear plate. Stock solutions (500 µL each, 400 mM) of alkyl iodides (1-iodopropane, 2-iodopropane, allyl iodide) were prepared in DMSO. The alkyl iodide solutions were serially diluted in DMSO ($1{:}10^{-5}$) to produce a 126.5 mM stock solution and a 40 mM stock solution. An aqueous 6.32 mM alkyl iodide stock solution and an aqueous 2 mM alkyl iodide stock solution were produced by a 1:20 dilution of the 126.5 mM alkyl iodide stock (in DMSO) and 40 mM alkyl iodide stock in phosphate buffer (50 mM $NaHPO_4$, pH 8). To each well containing the *At*HOL1 mutant protein preparations, 50 µL of either the 6.32 mM aqueous alkyl iodide stock solution (3.16 mM final) or the 2 mM aqueous alkyl iodide stock solution (1 mM final) was added to initiate the reaction. To wells containing only phosphate buffer, 50 µL of either the 6.32 mM aqueous alkyl iodide stock solution (3.16 mM final) or the 2 mM aqueous alkyl iodide stock solution (1 mM final) was added to monitor iodoalkyl autohydrolysis. Each reaction condition was performed in triplicate. The plate was incubated at 25 °C with shaking, for 16 h. A standard curve from 3.13 µM to 400 µM of KI (eight dilutions) in 50 mM $NaHPO_4$ (pH 8) was prepared for each plate. Following incubation, 2 µL of each reaction mixture and standard curve wells were added to 98 µL of TMB Liquid Substrate System + CiVCPO (final activity 1.37 U/ mL). The A_{570} was monitored every 30 s for 30 m. The initial velocity of 570 nm absorbance increase was measured. Autohydrolysis of the iodoalkyl compound varies by iodoalkyl compound and by concentration. Therefore, to eliminate the effect of autohydrolysis, the average initial velocity of 570 nm absorbance associated with the autohydrolysis at each concentration for each iodoalkyl compound was subtracted from the initial velocity of 570 nm absorbance of the corresponding *At*HOL1-containing wells. The resulting adjusted initial velocity of 570 nm absorbance associated with the enzymatic activity was then

compared to the KI standard curve to convert absorbance units to μM iodide. Specific activity was calculated as μMols iodide produced/min/mg *At*HOL1.

ASSOCIATED CONTENT

Supporting Information. Protein preparation and mutation; Molecular dynamics; The constraints applied in the MD simulation; Quantum mechanics; Thermostability calculation; MMPBSA and RMSD calculation; Electrostatic stabilization energy calculation; The use of the pre-reaction complex as the TS-analog; Hardware specifications; Statistical analysis for the random selections; The 6 components of the dynamic shift in P20C; The 6 components of the dynamic shift in P20C; The V140/T140 sidechain conformation of WT and V140T; The average C-CA-CB-HB dihedral angle of T140 in the MD simulation with each substrate; The dihedral distribution of residue 140 in the MD simulation with allyl-I; Computational metrics and specific activity for the key variants of AtHOL1; Details of the FuncLib test input; List of mutants and the computational metrics in different stages of SubTuner; List of mutants in the modified SubTuner workflow without the thermostability screening; List of mutants in the modified SubTuner workflow without the TS-analog binding screening; List of mutants in the modified SubTuner workflow without the reactivity; Coordinates of optimized reactants and transition states of the intrinsic reaction between the SAH-analog and different substrates (PDF)

Input files for SubTuner: input structures of *At*HOL1 and *acl*-MT; the main python script of SubTuner; job submission script of the workflow script (ZIP from https://doi.org/10.5281/zenodo.13381714).

Data and Software Availability. The main script of SubTuner and input structures are provided as part of the SI files. All original data is available from Zenodo

(https://doi.org/10.5281/zenodo.13381714). EnzyHTP is available from https://github.com/ChemBioHTP/EnzyHTP/tree/develop. AMBER 18 is available from http://ambermd.org/. Gaussian 16 is available from https://gaussian.com/.

AUTHOR INFORMATION

Corresponding Author

*Email: zhongyue.yang@vanderbilt.edu; Phone: 615-343-9849

*Email: brian.bachmann@vanderbilt.edu; Phone: 615-322-8865

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This research was supported by the startup grant from Vanderbilt University and the fellowship of Vanderbilt Institute of Chemical Biology. The PI thanks the sponsorship from Rosetta Commons Seed Grant Award. This work was also supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM146982. This work used SDSC Dell Cluster with AMD Rome HDR IB at Expanse from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants BIO200057.[78]

Reference

1. Cheng, L.; Li, D.; Mai, B. K.; Bo, Z.; Cheng, L.; Liu, P.; Yang, Y., Stereoselective amino acid synthesis by synergistic photoredox-pyridoxal radical biocatalysis. *Science* **2023**, *381*, 444-451.
2. Li, A.; Acevedo-Rocha, C. G.; D'Amore, L.; Chen, J.; Peng, Y.; Garcia-Borràs, M.; Gao, C.; Zhu, J.; Rickerby, H.; Osuna, S.; Zhou, J.; Reetz, M. T., Regio- and Stereoselective Steroid Hydroxylation at C7 by Cytochrome P450 Monooxygenase Mutants. *Angew. Chem. Int. Ed.* **2020**, *59*, 12499-12505.
3. Pinto, G. P.; Corbella, M.; Demkiv, A. O.; Kamerlin, S. C. L., Exploiting enzyme evolution for computational protein design. *Trends in Biochemical Sciences* **2022**, *47*, 375-389.
4. Renata, H., Engineering Catalytically Self-Sufficient P450s. *Biochemistry* **2023**, *62*, 253-261.
5. Shanbhag, A. P., Stairway to Stereoisomers: Engineering Short‐ and Medium‐Chain Ketoreductases To Produce Chiral Alcohols. *ChemBioChem* **2023**, *24*.
6. Gilio, A. K.; Thorpe, T. W.; Turner, N.; Grogan, G., Reductive aminations by imine reductases: from milligrams to tons. *Chemical Science* **2022**, *13*, 4697-4713.
7. Whitehead, J. N.; Leferink, N. G. H.; Johannissen, L. O.; Hay, S.; Scrutton, N. S., Decoding Catalysis by Terpene Synthases. *ACS Catalysis* **2023**, *13*, 12774-12802.
8. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185-194.
9. Burton, S. G.; Cowan, D. A.; Woodley, J. M., The search for the ideal biocatalyst. *Nature Biotechnology* **2002**, *20*, 37-45.
10. Madhavan, A.; Arun, K. B.; Binod, P.; Sirohi, R.; Tarafdar, A.; Reshmy, R.; Kumar Awasthi, M.; Sindhu, R., Design of novel enzyme biocatalysts for industrial bioprocess: Harnessing the power of protein engineering, high throughput screening and synthetic biology. *Bioresource Technology* **2021**, *325*, 124617.
11. Ju, S.; Kuzelka, K. P.; Guo, R.; Krohn-Hansen, B.; Wu, J.; Nair, S. K.; Yang, Y., A biocatalytic platform for asymmetric alkylation of α-keto acids by mining and engineering of methyltransferases. *Nat. Commun.* **2023**, *14*.
12. Prakinee, K.; Phintha, A.; Visitsatthawong, S.; Lawan, N.; Sucharitakul, J.; Kantiwiriyawanitch, C.; Damborsky, J.; Chitnumsub, P.; Van Pée, K.-H.; Chaiyen, P., Mechanism-guided tunnel engineering to increase the efficiency of a flavin-dependent halogenase. *Nature Catalysis* **2022**, *5*, 534-544.
13. Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T., Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem. Int. Ed.* **2021**, *60*, 88-119.
14. Chen, K.; Arnold, F. H., Engineering new catalytic activities in enzymes. *Nature Catalysis* **2020**, *3*, 203-213.
15. Turner, N. J., Enantioselective Oxidation of C–O and C–N Bonds Using Oxidases. *Chem. Rev.* **2011**, *111*, 4073-4087.
16. Blazeck, J.; Karamitros, C. S.; Ford, K.; Somody, C.; Qerqez, A.; Murray, K.; Burkholder, N. T.; Marshall, N.; Sivakumar, A.; Lu, W.-C.; Tan, B.; Lamb, C.; Tanno, Y.; Siddiqui, M. Y.; Ashoura, N.; Coma, S.; Zhang, X. M.; McGovern, K.; Kumada, Y.; Zhang, Y. J.; Manfredi, M.; Johnson, K. A.; D'Arcy, S.; Stone, E.; Georgiou, G., Bypassing evolutionary dead ends and switching the rate-limiting step of a human immunotherapeutic enzyme. *Nature Catalysis* **2022**, *5*, 952-967.

17. Sumbalova, L.; Stourac, J.; Martinek, T.; Bednar, D.; Damborsky, J., HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res.* **2018**, *46*, W356-W362.
18. Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D. L.; Prilusky, J.; Amengual-Rigo, P.; Guallar, V.; Tawfik, D. S.; Fleishman, S. J., Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Molecular Cell* **2018**, *72*, 178-186.e5.
19. Roda, S.; Terholsen, H.; Meyer, J. R. H.; Cañellas-Solé, A.; Guallar, V.; Bornscheuer, U.; Kazemi, M., AsiteDesign: a Semirational Algorithm for an Automated Enzyme Design. *The Journal of Physical Chemistry B* **2023**, *127*, 2661-2670.
20. Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B., Enantioselective Enzymes by Computational Design and In Silico Screening. *Angew. Chem. Int. Ed.* **2015**, *54*, 3726-3730.
21. Gomez De Santos, P.; Mateljak, I.; Hoang, M. D.; Fleishman, S. J.; Hollmann, F.; Alcalde, M., Repertoire of Computationally Designed Peroxygenases for Enantiodivergent C–H Oxyfunctionalization Reactions. *J. Am. Chem. Soc.* **2023**, *145*, 3443-3453.
22. Barber-Zucker, S.; Mateljak, I.; Goldsmith, M.; Kupervaser, M.; Alcalde, M.; Fleishman, S. J., Designed High-Redox Potential Laccases Exhibit High Functional Diversity. *ACS Catalysis* **2022**, *12*, 13164-13173.
23. Bengel, L. L.; Aberle, B.; Egler‐Kemmerer, A. N.; Kienzle, S.; Hauer, B.; Hammer, S. C., Engineered Enzymes Enable Selective N‐Alkylation of Pyrazoles With Simple Haloalkanes. *Angew. Chem. Int. Ed.* **2021**, *60*, 5554-5560.
24. Ospina, F.; Schülke, K. H.; Soler, J.; Klein, A.; Prosenc, B.; Garcia‐Borràs, M.; Hammer, S. C., Selective Biocatalytic N‐Methylation of Unsaturated Heterocycles. *Angew. Chem. Int. Ed.* **2022**, *61*.
25. Liu, Y.; Khusnutdinova, A.; Chen, J.; Crisante, D.; Batyrova, K.; Raj, K.; Feigis, M.; Shirzadi, E.; Wang, X.; Dorakhan, R.; Wang, X.; Stogios, P. J.; Yakunin, A. F.; Sargent, E. H.; Mahadevan, R., Systems engineering of Escherichia coli for n-butane production. *Metab. Eng.* **2022**, *74*, 98-107.
26. Sun, R.; Zheng, P.; Chen, P.; Wu, D.; Zheng, J.; Liu, X.; Hu, Y., Enhancing the Catalytic Efficiency of D‐lactonohydrolase through the Synergy of Tunnel Engineering, Evolutionary Analysis, and Force‐Field Calculations. *Chemistry – A European Journal* **2024**, *30*.
27. Deshmukh, F. K.; Ben-Nissan, G.; Olshina, M. A.; Füzesi-Levi, M. G.; Polkinghorn, C.; Arkind, G.; Leushkin, Y.; Fainer, I.; Fleishman, S. J.; Tawfik, D.; Sharon, M., Allosteric regulation of the 20S proteasome by the Catalytic Core Regulators (CCRs) family. *Nat. Commun.* **2023**, *14*.
28. Liu, Z.; Li, K.; Liu, X.; Zhao, J.; Yu, Y.; Wang, L.; Kong, Y.; Chen, M., Production of microhomogeneous glycopeptide by a mutated NGT according FuncLib with unique sugar as substrate. *Enzyme Microb. Technol.* **2022**, *154*, 109949.
29. Saez-Jimenez, V.; Scrima, S.; Lambrughi, M.; Papaleo, E.; Mapelli, V.; Engqvist, M. K. M.; Olsson, L., Directed Evolution of (R)-2-Hydroxyglutarate Dehydrogenase Improves 2-Oxoadipate Reduction by 2 Orders of Magnitude. *ACS Synthetic Biology* **2022**, *11*, 2779-2790.
30. Li, Z.; Zhu, Z.; Xu, G.; Wei, L.; Liu, J.; Wang, H.; Lu, C.; Li, Y.; Zhu, D.; Shen, Y., Functional Application of the Single-Module NRPS-like <scp>d</scp>-Alanyltransferase in Maytansinol Biosynthesis. *ACS Catalysis* **2024**, *14*, 8062-8072.

31. Ma, L.; Gao, J.; Qiu, Y.; Liu, Y.; Wu, X.; Zhang, X.; Tong, Y.; Gao, W.; Zhang, Y.; Liu, C., Rational Design and Site‐Directed Mutagenesis Promote the Heterologous Production of the Antitumor Natural Product Dehydroabietic Acid. *ChemistrySelect* **2024**, *9*.
32. Trudeau, D. L.; Edlich-Muth, C.; Zarzycki, J.; Scheffen, M.; Goldsmith, M.; Khersonsky, O.; Avizemer, Z.; Fleishman, S. J.; Cotton, C. A. R.; Erb, T. J.; Tawfik, D. S.; Bar-Even, A., Design and in vitro realization of carbon-conserving photorespiration. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E11455-E11464.
33. Vos, P. D.; Rossetti, G.; Mantegna, J. L.; Siira, S. J.; Gandadireja, A. P.; Bruce, M.; Raven, S. A.; Khersonsky, O.; Fleishman, S. J.; Filipovska, A.; Rackham, O., Computationally designed hyperactive Cas9 enzymes. *Nat. Commun.* **2022**, *13*.
34. Klaus, M.; Buyachuihan, L.; Grininger, M., Ketosynthase Domain Constrains the Design of Polyketide Synthases. *ACS Chemical Biology* **2020**, *15*, 2422-2432.
35. Risso, V. A.; Romero-Rivera, A.; Gutierrez-Rus, L. I.; Ortega-Muñoz, M.; Santoyo-Gonzalez, F.; Gavira, J. A.; Sanchez-Ruiz, J. M.; Kamerlin, S. C. L., Enhancing ade novoenzyme activity by computationally-focused ultra-low-throughput screening. *Chemical Science* **2020**, *11*, 6134-6148.
36. Huang, J.; Xie, X.; Zheng, W.; Xu, L.; Yan, J.; Wu, Y.; Yang, M.; Yan, Y., In silico design of multipoint mutants for enhanced performance of Thermomyces lanuginosus lipase for efficient biodiesel production. *Biotechnology for Biofuels and Bioproducts* **2024**, *17*.
37. Huang, A.; Zhang, X.; Yang, Y.; Shi, C.; Zhang, B.; Tuo, X.; Shen, P.; Jiao, X.; Zhang, N., Biocatalytic Synthesis of Ruxolitinib Intermediate via Engineered Imine Reductase. *J. Org. Chem.* **2024**, *89*, 11446-11454.
38. Zhao, M.; Zhou, B.; Jia, X.; Wang, M.; Liu, Z.; Zheng, Y., Increasing catalytic efficiency of SceCPR by semi-rational engineering towards the asymmetric reduction of D-pantolactone. *J. Biotechnol.* **2023**, *373*, 34-41.
39. Welborn, V. V.; Ruiz Pestana, L.; Head-Gordon, T., Computational optimization of electric fields for better catalysis design. *Nature Catalysis* **2018**, *1*, 649-655.
40. Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; Dechancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190-195.
41. Rakotoharisoa, R. V.; Seifinoferest, B.; Zarifi, N.; Miller, J. D. M.; Rodriguez, J. M.; Thompson, M. C.; Chica, R. A., Design of Efficient Artificial Enzymes Using Crystallographically Enhanced Conformational Sampling. *J. Am. Chem. Soc.* **2024**, *146*, 10001-10013.
42. Shao, Q.; Jiang, Y.; Yang, Z. J., EnzyHTP: A High-Throughput Computational Platform for Enzyme Modeling. *J. Chem. Inf. Model.* **2022**, *62*, 647-655.
43. Jiang, Y.; Stull, S. L.; Shao, Q.; Yang, Z. J., Convergence in determining enzyme functional descriptors across Kemp eliminase variants. *Electronic Structure* **2022**, *4*, 044007.
44. Shao, Q.; Jiang, Y.; Yang, Z. J., EnzyHTP Computational Directed Evolution with Adaptive Resource Allocation. *J. Chem. Inf. Model.* **2023**, *63*, 5650-5659.
45. Juarez, R. J.; Jiang, Y.; Tremblay, M.; Shao, Q.; Link, A. J.; Yang, Z. J., LassoHTP: A High-Throughput Computational Tool for Lasso Peptide Structure Construction and Modeling. *J. Chem. Inf. Model.* **2023**, *63*, 522-530.

46. Yang, Z. J.; Shao, Q.; Jiang, Y.; Jurich, C.; Ran, X.; Juarez, R. J.; Yan, B.; Stull, S. L.; Gollu, A.; Ding, N., Mutexa: A Computational Ecosystem for Intelligent Protein Engineering. *J. Chem. Theory Comput.* **2023**, *19*, 7459-7477.
47. Tang, Q.; Pavlidis, I. V.; Badenhorst, C. P. S.; Bornscheuer, U. T., From Natural Methylation to Versatile Alkylations Using Halide Methyltransferases. *ChemBioChem* **2021**, *22*, 2584-2590.
48. Dalhoff, C.; Lukinavičius, G.; KlimašAuskas, S.; Weinhold, E., Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol.* **2006**, *2*, 31-32.
49. Nagatoshi, Y.; Nakamura, T., Characterization of three halide methyltransferases in Arabidopsis thaliana. *Plant Biotechnol.* **2007**, *24*, 503-506.
50. Schülke, K. H.; Ospina, F.; Hörnschemeyer, K.; Gergel, S.; Hammer, S. C., Substrate Profiling of Anion Methyltransferases for Promiscuous Synthesis of S - Adenosylmethionine Analogs from Haloalkanes. *ChemBioChem* **2022**, *23*.
51. Tang, Q.; Grathwol, C. W.; Aslan‐Üzel, A. S.; Wu, S.; Link, A.; Pavlidis, I. V.; Badenhorst, C. P. S.; Bornscheuer, U. T., Directed Evolution of a Halide Methyltransferase Enables Biocatalytic Synthesis of Diverse SAM Analogs. *Angewandte Chemie International Edition* **2021**, *60*, 1524-1527.
52. Schülke, K. H.; Fröse, J. S.; Klein, A.; Garcia‐Borràs, M.; Hammer, S. C., Efficient Transferase Engineering for SAM Analog Synthesis from Iodoalkanes. *ChemBioChem* **2024**, *25*.
53. Frenz, B.; Lewis, S. M.; King, I.; DiMaio, F.; Park, H.; Song, Y., Prediction of Protein Mutational Free Energy: Benchmark and Sampling Improvements Increase Classification Accuracy. *Frontiers in Bioengineering and Biotechnology* **2020**, *8*.
54. Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; Baker, D.; Dimaio, F., Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016**, *12*, 6201-6212.
55. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. Rosetta3. In; Elsevier: 2011, pp 545-574.
56. Fried, S. D.; Boxer, S. G., Electric Fields and Enzyme Catalysis. *Annual Review of Biochemistry* **2017**, *86*, 387-415.
57. Fried, S. D.; Boxer, S. G., Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Accounts of Chemical Research* **2015**, *48*, 998-1006.
58. Lankau, T.; Ken, H. C.; Chang, H. M.; Yu, C. H., A Computational Study of the Promiscuity of the SAM-Dependent Methyltransferase AtHTMT1. *ACS Omega* **2022**, *7*, 12753-12764.
59. Nov, Y., When Second Best Is Good Enough: Another Probabilistic Look at Saturation Mutagenesis. *Appl. Environ. Microbiol.* **2012**, *78*, 258-262.
60. Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J. T.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao, R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.; Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione, G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D., Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *Journal of the American Chemical Society* **2012**, *134*, 16197-16206.

61. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329*, 309-313.
62. Xie, W. J.; Warshel, A., Natural Evolution Provides Strong Hints about Laboratory Evolution of Designer Enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2207904119.
63. Ao, Y. F.; Pei, S.; Xiang, C.; Menke, M. J.; Shen, L.; Sun, C.; Dörr, M.; Born, S.; Höhne, M.; Bornscheuer, U. T., Structure‐ and Data‐Driven Protein Engineering of Transaminases for Improving Activity and Stereoselectivity. *Angewandte Chemie International Edition* **2023**, *62*.
64. Schrodinger, LLC, In; 2015.
65. D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman Amber 2021.
66. 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.
67. Luo, S.; Liu, L.; Lyu, C.-J.; Sim, B.; Liu, Y.; Gong, H.; Nie, Y.; Zhao, Y.-L., Understanding the effectiveness of enzyme pre-reaction state by a quantum-based machine learning model. *Cell Reports Physical Science* **2022**, *3*, 101128.
68. Yabukarski, F.; Biel, J. T.; Pinney, M. M.; Doukov, T.; Powers, A. S.; Fraser, J. S.; Herschlag, D., Assessment of enzyme active site positioning and tests of catalytic mechanisms through X-ray–derived conformational ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 33204-33215.
69. Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A., Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat. Commun.* **2020**, *11*.
70. Riziotis, I. G.; Ribeiro, A. J. M.; Borkakoti, N.; Thornton, J. M., Conformational Variation in Enzyme Catalysis: A Structural Study on Catalytic Residues. *J. Mol. Biol.* **2022**, *434*, 167517.
71. Vaissier, V.; Sharma, S. C.; Schaettle, K.; Zhang, T.; Head-Gordon, T., Computational Optimization of Electric Fields for Improving Catalysis of a Designed Kemp Eliminase. *ACS Catalysis* **2018**, *8*, 219-227.
72. Welborn, V. V.; Head-Gordon, T., Fluctuations of Electric Fields in the Active Site of the Enzyme Ketosteroid Isomerase. *Journal of the American Chemical Society* **2019**, *141*, 12487-12492.
73. Bhowmick, A.; Sharma, S. C.; Head-Gordon, T., The Importance of the Scaffold for de Novo Enzymes: A Case Study with Kemp Eliminase. *J. Am. Chem. Soc.* **2017**, *139*, 5793-5800.
74. Zheng, C.; Ji, Z.; Mathews, I. I.; Boxer, S. G., Enhanced active-site electric field accelerates enzyme catalysis. *Nature Chemistry* **2023**.

75. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
76. Lu, T.; Chen, F., Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* **2012**, *33*, 580-592.
77. Aslan‐Üzel, A. S.; Beier, A.; Kovář, D.; Cziegler, C.; Padhi, S. K.; Schuiten, E. D.; Dörr, M.; Böttcher, D.; Hollmann, F.; Rudroff, F.; Mihovilovic, M. D.; Buryška, T.; Damborský, J.; Prokop, Z.; Badenhorst, C. P. S.; Bornscheuer, U. T., An Ultrasensitive Fluorescence Assay for the Detection of Halides and Enzymatic Dehalogenation. *ChemCatChem* **2020**, *12*, 2032-2039.
78. Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N., XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16*, 62-74.