Rate-enhancing Single Amino Acid Mutation for Hydrolases: A Statistical Profiling

Bailu Yan,[1] Xinchun Ran,[1] Yaoyukun Jiang,[1] Sarah K. Torrence,[4] Li Yuan,[4] Qianzhen Shao,[1] and Zhongyue J. Yang[1-4,]*

[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*

ABSTRACT: We reported the statistical profiling for rate-enhancing mutant hydrolases with single amino acid substitution. We constructed an integrated structure-kinetics database, IntEnzyDB, which contains 3,907 experimentally characterized hydrolase kinetics and 2,715 hydrolase Protein Data Bank IDs. The hydrolase kinetics data involve 9% rate-enhancing mutations. Mutation to nonpolar residues with a hydrocarbon chain shows a stronger preference for rate acceleration than to polar or charged residues. To elucidate the structure-kinetics relationship for rate-enhancing mutations, we categorized each mutation into one of the three spatial shells of hydrolases. We defined the spatial shells by reference to either the active site or the center-of-mass of the enzyme. In either case, mutations in the first shell (i.e., closest to the reference point) appear on average more rate-deleterious than those in the other two shells (i.e., ~1.0 kcal/mol in $\Delta\Delta G^{\ddagger}$). Under the active-site reference, mutations in the third shell (i.e., most distal to the active site) exhibit the highest likelihood of rate enhancement. This propensity is significant for larger-sized hydrolases. In contrast, under the center-of-mass reference, mutations in the second shell (i.e., 33.3th to 66.7th percentile rank of spatial proximity to the center-of-mass of the enzyme) show the highest likelihood of rate enhancement. This trend is significant for smaller-sized hydrolases. The studies reveal the statistical features for identifying rate-enhancing mutations in hydrolases, which will potentially guide hydrolase discovery in biocatalysis.

Keywords: Biocatalysis; Enzymology database; Mutations; Rate enhancement; Statistical analysis; Hydrolases

1. Introduction

Hydrolases, such as esterases, glycosidases, peptidases, and nucleosidases, are the building blocks for modern pharmaceutical, food, and laundry industries.[1-3] They serve as promising candidates for biodegradation of environmental wastes, such as PFAS[4-5] and poly(ethylene terephthalate) (PET).[6] The design and discovery of new hydrolases that enable efficient conversion of natural and non-natural chemical transformations have been largely advanced by the development of directed evolution strategies[7-10] and *de novo* enzyme design algorithms.[8, 11-12]

Identifying beneficial mutant hydrolases with enhanced rate, selectivity, stability, solubility, and expressibility is critical for the prediction of new enzyme variants for challenging chemical transformations. Among these, the rate-enhancing mutation is arguably the most difficult to attain because enzyme kinetics are globally encoded across the entire protein sequence and are highly substrate-dependent (e.g., Tn5 transposon-derived kinase[13] and amidase[14]).[15] Genome sequencing techniques and high-throughput assay (e.g., deep mutational scanning[16-19] and high-throughput microfluidic enzyme kinetics[20]) have largely boosted the discovery of rate-enhancing mutants via simultaneously screening ten thousands of mutations for specific reactions. However, given the gigantic combination number of possible enzyme variants and the low yield of beneficial mutants from screening (i.e., typically less than 5% according to deep mutational scanning[21] and directed evolution experiments[22]), it remains a critical challenge in the community to develop new strategies for designing rate-enhancing mutations.

Statistical modeling,[23] among other molecular simulation-based computational strategies,[8, 11-12, 24-27] have been extensively augmented with experiments to build metabolic models[28] and to guide the discovery of rate-enhancing mutations by reducing the experimental testing candidates *a priori*.[7] The statistical models have been constructed to inform the population and spatial distribution of rate-enhancing mutations for specific enzymes with a few substrates (e.g., amiE[14] and PafA[16]).[16, 22, 29-31] Remote mutations (>10Å from active site) have been widely reported to be critical for rate-enhancement,[32] while a statistical study of 55 rate-enhancing enzyme variants by Morley et al.[29] show that close mutations can also be rate-enhancing. For hydrolases, Lim, Fernandes, and coworkers have reported the statistical studies for activation free energy and enzyme efficiency in 339 wild-type hydrolases.[3, 33] However, across a diverse range of hydrolase sequences, functions, and substrate types, the properties and spatial distributions for rate-enhancing mutations remain unexplored. This is primarily caused by the challenges for integrating enzyme structural data and kinetics data – they are stored in different databases (e.g., PDB,[34] UniProt,[35-36] BRENDA,[37] and SABIO-RK[38]) with a diverse range of data formats and standards, which is very difficult to collect and clean.

Here we built a database, IntEnzyDB, that stores clean and tabulated structural and catalytic data for hydrolases. Using IntEnzyDB, we curated 3,907 kinetics parameters for investigating what types of amino acids are more likely to induce rate-enhancement, and 505 kinetics-structure pairs for studying the spatial distribution of rate-enhancing mutations. The study shows that mutation to bulky nonpolar residues is more likely to induce rate enhancement than to the polar or charged residues. The study reveals the spatially-resolved and protein-size-dependent likelihood for identifying rate-enhancing mutations.

2. Computational Details

Database Construction We built a relational database for hydrolases, IntEnzyDB, with the flattened data structure to facilitate statistical analysis and data-driven modeling. The database is publically accessible through the MongoDB Compass connection string: mongodb+srv://access_1:Aa123@cluster0.5ey45.mongodb.net/test. Unlike Protein Data Bank (PDB) and other object-oriented database that stores information of a protein using one individual data file, IntEnzyDB uses one table for hydrolase kinetics and three tables for structure data of different scale, including chain table, amino acid table, and atom table. Each table contains entries of all hydrolases stored. Different tables are connected by the keywords (i.e., foreign keys): UniProtKB, PDB ID and enzyme commission (EC) number. To ensure precise mapping of enzyme kinetics to structure, we manually aligned the mutation residue sequence reported in kinetics database (or labeled in Uniprot) with the PDB structure.

The kinetics data of hydrolases were collected from BRENDA[37] and SABIO-RK.[38] Enzyme entries lacking the wild-type or mutant k_{cat} values, substrate information, or unknown experimental temperature were excluded. The kinetics table of the database contains 3,907 entries for 411 hydrolases. The kinetics table stores turnover number k_{cat}, reaction type, substrate name, mutation type, experimental conditions (e.g., temperature, pH, and pressure), and so on. The protein structure data were collected from RCSB Protein Databank.[34] The structure table of IntEnzyDB contains 2,715 structures. The enzyme chain table contains structure name, sequence, resolution, missing residue, global stoichiometry, organism, and FASTA sequence. The enzyme amino acid table contains amino acid type, sequence number, coordinates of the C_α, and center-of-mass coordinate of the entire amino acid. The atom table contains atom type, sequence number, amino acid to which the atom belongs, coordinates, and atomic mass.

Data Curation Using UniProtKB as the key,[35-36, 39] the kinetics data and the structural data were paired. The data entries were excluded that involve either missing kinetics or PDB structure data. The kinetics data (i.e., turnover number, k_{cat}, s^{-1}) for hydrolases were curated based on the following filtration criteria: 1) the data entry has been assigned a UniProtKB and there is at least one known PDB structure under the UniProtKB, 2) the data entry stores mutation (or wild-type) and substrate information, 3) the data entry includes temperature, 4) the data entry corresponds to either a wild-type enzyme of a single-amino-acid-substitution mutant, and 5) there is at least one wild-type enzyme/mutant pair with shared substrate and temperature condition. This yields 1,500 kinetics data entries for hydrolase-substrate complexes that consist of 221 unique hydrolases (i.e., UniProtKBs), 910 mutant hydrolases with single-amino-acid substitution, and 362 substrates. Among the curated data, 95% of the enzymatic kinetics were experimentally measured in the temperature range from 295.15 to 343.15 K (Supporting Information, Figure S1). IntEnzyDB contains nine types of hydrolases that act on ester bonds (EC 3.1), N-glycosidic bond (EC 3.2), ether bonds (EC 3.3), peptide bonds (EC 3.4), carbon-nitrogen bonds other than peptide bonds (EC 3.5), acid anhydrides (EC 3.6), carbon-carbon bonds (3.7), halide bonds (3.8), phosphorus-nitrogen bonds (EC 3.9, Supporting Information, Figure S2).

Under a specific UniProtKB, the PDB structure of hydrolases was selected that allows a successful matching of the mutation and active-site spots labeled in the kinetic database to the PDB sequence. The distribution of the number of the missing residues and the resolution for each curated structure is shown in the Supporting Information, Figure S3a and S3b, respectively. Only one chain was selected to represent the whole structure. This yields 80 different PDB structures, corresponding to 80 UniProtKBs.

The cleaned kinetics data table and the paired kinetic-structure data table, along with the code we used for data cleaning, can be found in the zip file of the Supporting Information. All statistical analysis, including the histogram, boxplot, bar graph, percentage bar graph, were generated using R package.

3. Results and Discussion

3a. IntEnzyDB Integrates Structure and Kinetics Data for Hydrolases

We have built a new hydrolase database, IntEnzyDB, which integrates clean and tabulated structure and kinetics data in one place (Figure 1). Unlike the Protein Data Bank (PDB) that stores protein data files individually, IntEnzyDB adopts a relational architecture with the flattened data structure, in which each data table stores all hydrolase entries. We have created one data table for hydrolase kinetics, and three separate tables for different scales of hydrolase structure information, including enzyme chain table, amino acid table, and atom table. These tables share the keywords: PDB ID, UniProtKB, and EC Number, which can be used for mapping hydrolase kinetics-structure pairs (Figure 1). Besides easy pairing of enzyme kinetics and structure data, IntEnzyDB is also advantaged by the efficiency of data processing – IntEnzyDB outperforms PDB by hours of efficiency when processing a large number of enzyme structures (e.g., 1.2 hours faster for 1,000 hydrolase structures, Supporting information, Table S1). This is because IntEnzyDB avoids repetitive file I/O operations of individual structure files by loading all data entries simultaneously to the CPU memory.

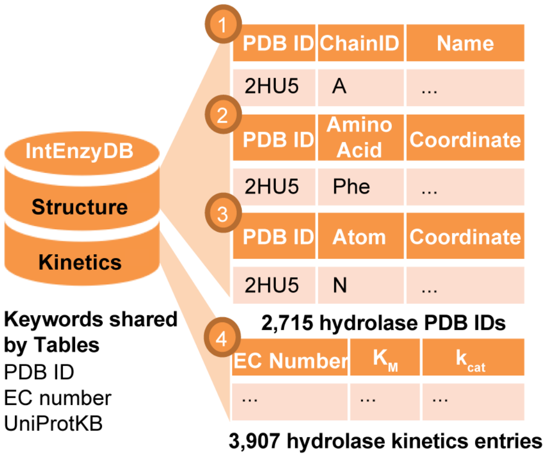


Figure 1. Design architecture for the integrated structure-kinetics database, IntEnzyDB. The database consists of four tables: three tables are used for storing structural data (i.e., protein chain table, protein amino acid table, and protein atom table), and one table for kinetics data. The tables share keywords, including PDB ID, EC Number, and UniProtKB.

Enabled by IntEnzyDB, we have curated two datasets of mutant hydrolases for analysis. The first dataset involves kinetics parameters for 1,500 mutant hydrolases-catalyzed reactions, consisting of 211 unique hydrolases, 910 mutant hydrolases with single amino acid substitution, and 362 substrates. This dataset will be applied to investigate the percentage of rate-enhancing mutations among all mutations and the rate-enhancing propensity of mutation to a certain type of the 20 canonical amino acids. The second dataset involves kinetics-structure pairs for 505 mutant hydrolases-catalyzed reactions where the active-site residues information is labeled either in UniProt or PDB. The dataset consists of 80 unique hydrolases, 350 mutant hydrolases with single amino acid substitution, and 136 substrates. This dataset enables us to statistically profile the spatial distributions for the rate-enhancing mutants.

3b. The Rate-enhancing Mutant Hydrolases Occupies 9% in IntEnzyDB

We first investigate the percentage of the rate-enhancing mutations. Figure 2 shows the histogram of free energy barrier changes upon mutation (i.e., $\Delta\Delta G^{\ddagger}$) for all hydrolase variants, in which the $\Delta\Delta G^{\ddagger}$ is converted from $k_{cat}^{mutation}/k_{cat}^{wild\text{-}type}$ using the Eyring's equation. R is the gas constant, T is the temperature, and k_{cat} is the apparent turnover number for the enzymatic reaction:

$$\Delta\Delta G^{\ddagger} =- RT \ln \frac{k_{cat}^{mutant}}{k_{cat}^{wild-type}}$$

The distribution conforms to a Gaussian shape with a heavier right tail, ranging from –4.1 to 8.6 kcal/mol. The average of the $\Delta\Delta G^{\ddagger}$ is 0.9 kcal/mol – this is consistent with the common experimental observation that mutations likely increase the activation free energy barrier and reduce the turnover number. To characterize the rate-perturbing effects, the mutations are categorized to be rate-enhancing (i.e., $\Delta\Delta G^{\ddagger} \leq -0.5$ kcal/mol), rate-neutral (i.e., $\Delta\Delta G^{\ddagger} > -0.5$ and ≤ 0.5 kcal/mol), and rate-deleterious (i.e., $\Delta\Delta G^{\ddagger} > -0.5$ kcal/mol). The proportion of rate-enhancing mutations is 9% (Figure 2), which is a smaller composition compared to the rate-neutral and -deleterious mutations. Nonetheless, this percentage biases towards overestimating the natural abundance of the rate-enhancing mutations due to the exclusion of mutations that are non-expressible or abolish hydrolase activity. In contrast, the percentage of beneficial single-mutation has been reported to be around 5% for an aliphatic amide hydrolase amiE with three different substrates (i.e., acetamide, propionamide, and isobutyramide),[14] and to be only 0.01–1% observed in the directed evolution experiments,[9] albeit both are derived from a fitness metric rather than from $\Delta\Delta G^{\ddagger}$.

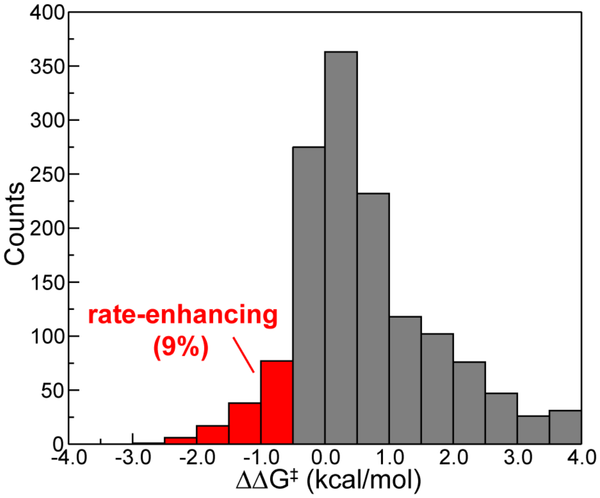


Figure 2. The histogram of the activation free energy change upon mutation (i.e., $\Delta\Delta G^{\ddagger}$) for 1,500 hydrolase variants-catalyzed reactions. The bin size is 0.5 kcal/mol. Colored in red are for rate-enhancing mutations with their $\Delta\Delta G^{\ddagger}$ less than or equal to –0.5 kcal/mol. Colored in grey are for rate-neutral (i.e., $\Delta\Delta G^{\ddagger} > -0.5$ and ≤ 0.5 kcal/mol) and rate-deleterious (i.e., $\Delta\Delta G^{\ddagger} > -0.5$ kcal/mol) mutations.

3c. Mutation to Nonpolar Residue Has the Highest Likelihood to Accelerate the Hydrolase-Catalyzed Reactions

We characterized the rate-enhancing propensity of mutation to a certain type of the 20 canonical amino acids (Figure 3). For each type of amino acid, we calculated and ranked the percentage of rate-enhancing mutations (Figure 3a) and the median $\Delta\Delta G^{\ddagger}$ for all mutations of the same amino acid type (Figure 3b). The percentage of rate-enhancing mutations ranges broadly from ~26% for Val to ~2% for Glu (Figure 3a), while the median $\Delta\Delta G^{\ddagger}$ ranges from –0.1 kcal/mol for Tyr to 1.2 kcal/mol for Asn (Figure 3b). Among the five amino acids with the highest percentage of rate-enhancing mutations, three are bulky nonpolar residues with a hydrocarbon side chain (i.e., Val, Ile, and Phe). Val and Ile are also among the top three amino acid types that involve the lowest median $\Delta\Delta G^{\ddagger}$. For both ranks shown in Figure 3, the common residues among the top ten involve four nonpolar (i.e., Val, Ile, Phe, and Pro), two polar (i.e., Tyr and Thr), and one charged (i.e., Arg) residue(s), while those among the bottom ten involve one nonpolar (i.e., Ala), two polar (i.e., His and Gln), and three charged (i.e., Glu, Asp, and Lys) residue(s). Notably, the only nonpolar residue in the bottom ten, Ala, is more humanly biased towards rate-deleterious than other residues because of its extensive use in alanine scanning to replace catalytically-competent residues for testing biochemical hypothesis. Quantitatively, mutation to bulky nonpolar residues with a hydrocarbon side chain has a 7% higher likelihood to enhance turnover number than other types of residues for hydrolases (Supporting information, Figure S4). Mutation from polar or charged residues to hydrocarbon chain-containing nonpolar residues is found to be the most rate-enhancing (~16%, Supporting information, Figure S5). These statistical results emphasize the important roles of mutation to nonpolar residues in accelerating the hydrolase-catalyzed reactions.

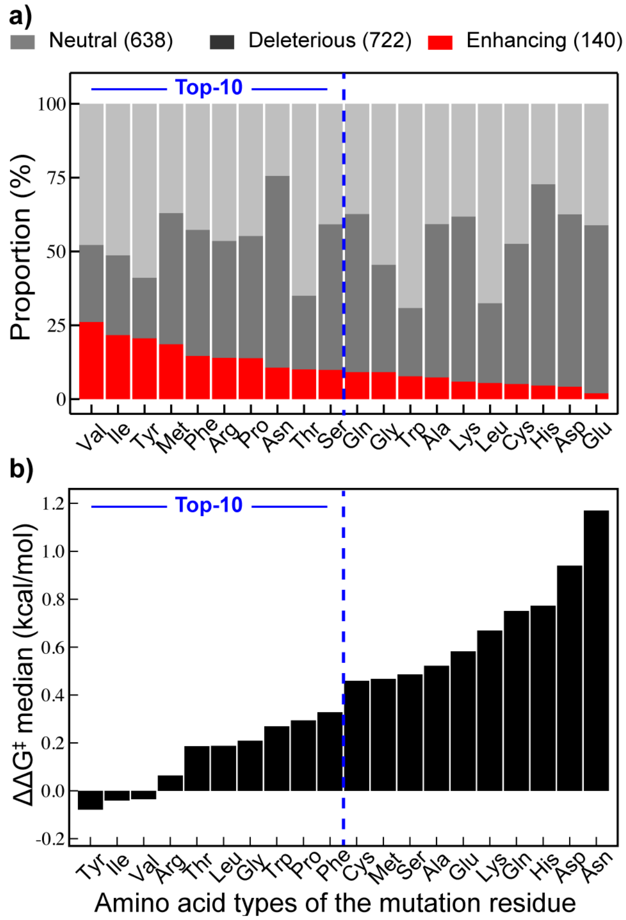
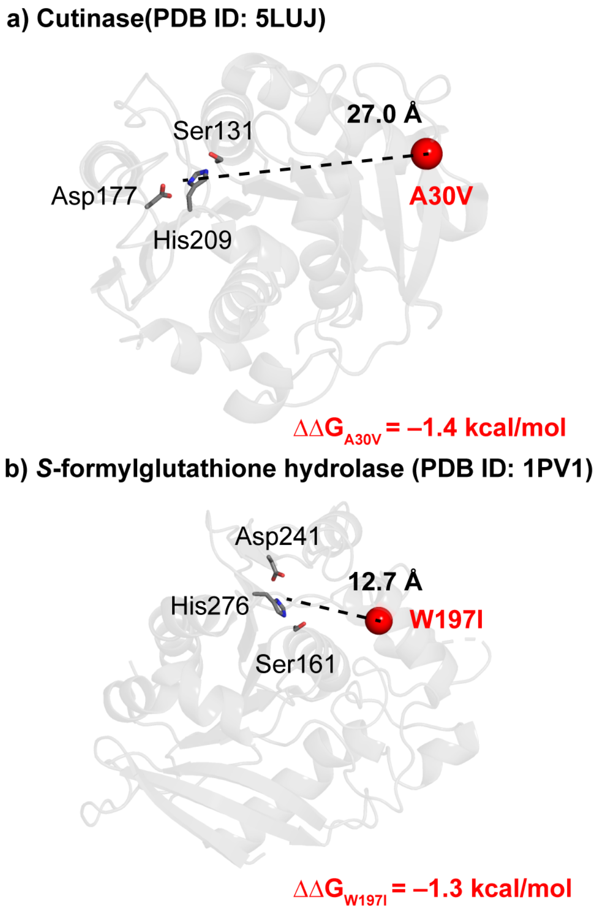


Figure 3. The rate-perturbing effects of mutation to a certain type of the 20 canonical amino acids. a) The normalized proportion of rate-enhancing (red), -deleterious (dark grey), and -neutral (grey) mutation for each amino acid type of the mutation residue, ranked by the proportion of the rate-enhancing mutations. b) The ranked median $\Delta\Delta G^{\ddagger}$ for each amino acid type of the mutation residue.

The nonpolar residues with a hydrocarbon chain are chemically inert. They are not able to form strong and directional hydrogen bonding or electrostatic interactions with local residues but have the capability of tuning protein dynamics through hydrophobic interactions and steric frictions. Mutation to nonpolar residues has been reported to change enzyme conformational population and dynamics, substrate positioning, and the shape of the active-site, in ways to stabilize the transition state.[40-43] As an example, we show the structures for two typical hydrolases, cutinase (Thc_Cut2)[44-45] and *S*-formylglutathione hydrolase (SFGH),[46] which involve significant barrier reduction (i.e., $\Delta\Delta G^{\ddagger} \leq -1.3$ kcal/mol) upon single amino acid substitution to Val and Ile, respectively (Figure 4). Both enzyme mutants involve the same type of catalytic triad (Ser-His-Asp)[2] and show a similar magnitude of rate enhancement for the substrate 4-nitrophenylbutyrate (i.e., $\Delta\Delta G^{\ddagger} = -1.4$ kcal/mol for Thc_Cut2 *vs* -1.3 kcal/mol for SFGH). The mutation residues are located with distinct spatial proximity and orientation to the catalytic triad (i.e., 27.0 Å for Thc_Cut2 and 12.7 Å for SFGH). Inspired by the observation, we investigated the rate-enhancing likelihood for mutations of various spatial proximity to the active site (section 3d) and the center-of-mass of the enzyme (section 3e).


a) Cutinase(PDB ID: 5LUJ)

27.0 Å
Ser131
Asp177
His209
A30V
$\Delta\Delta G_{A30V} = -1.4$ kcal/mol

b) *S*-formylglutathione hydrolase (PDB ID: 1PV1)

Asp241
12.7 Å
His276
W197I
Ser161
$\Delta\Delta G_{W197I} = -1.3$ kcal/mol

Figure 4. Typical mutant hydrolases that involve significant rate-enhancement upon single amino acid substitution to bulky nonpolar residues. a) Thc_Cut2 cutinase with A30V mutation, in which the active-site catalytic triad is Ser131-His209-Asp177. b) *S*-formylglutathione hydrolase (SFGH) with W197I mutation, in which the active-site catalytic triad is Ser161-His276-Asp241. For both enzymes, the substrate is 4-nitrophenylbutyrate. The distances between the C_α of the variant residue and the center of mass of the catalytic triad are labeled.

3d. Structure-Kinetics Relationship with Reference to the Active-Site

We studied the structure-kinetics relationship for mutant hydrolases with reference to the active-site (Figure 5). We employed two types of criteria to characterize the spatial proximity of the mutation residues to the active site residues (labeled in UniProt or PDB). The distance-criterion categorizes a mutation into the inner-shell (i.e., ≤ 10 Å), mid-shell (i.e., >10 and ≤ 20 Å), or outer-shell (i.e., >20 Å) based on the distance of its C_α coordinate to the geometric center of the active-site residues C_α coordinates (Figure 5a). This results in 153, 260, and 92 mutations in the inner-, mid-, and outer shell. The percentile-criterion categories a mutation residue into the first- (i.e., > 66.7th percentile), second- (i.e., >33.3th and ≤ 66.7th percentile), or third-shell (i.e., ≤ 33.3th percentile) based on its percentile rank of spatial proximity to the active-site among all residues of a hydrolase (Figure 5d). This results in 358, 95, and 52 mutations in the first-, second-, and third-shell. Notably, the shells defined here are intended to reflect a spatial cutoff – they are relevant but not identical to the well-known coordination shells defined based on the layers of contact residues surrounding the substrate.

Using the distance-criterion, we first investigated the distribution of $\Delta\Delta G^{\ddagger}$ in each shell (Figure 5b). The mutations in the inner-shell involve higher median $\Delta\Delta G^{\ddagger}$ (by 0.8 kcal/mol) than those in the mid- or outer-shell. Similarly, when shifting to the percentile-criterion, the mutations in the first-shell also involve significantly higher $\Delta\Delta G^{\ddagger}$ (by 0.6 kcal/mol) than those of the second- and third-shell (Figure 5e). These results show that mutations are statistically more deleterious when they are spatially proximal to the active site. This is intuitive because the active-site residues constitute the hydrolase catalytic functions, which include but are not limited to participate in bond arrangement (i.e., catalytic triad), stabilize the oxyanion hole, transfer proton, or bind substrate. Many of these residues are evolutionarily conserved, the mutation of which commonly causes a large increase of the $\Delta\Delta G^{\ddagger}$, if not abolish the catalytic activity entirely.

We further characterized the spatially-resolved likelihood of rate enhancement. As pointed out by Morley et al. (Paper14), the globular shape of the enzyme makes the mid- and outer-shells involve a larger population of residues than the inner-shell, which results in the observation of more mutations and accordingly more beneficial mutations remote to the active site. To normalize this effect, we computed the proportion of rate-enhancing mutations over all the mutations populated in a certain shell. Under the distance-criterion, the highest proportion of rate-enhancing mutations is found in the mid-shell, which contains residues between 10 and 20 Å away from the active site (12%, Figure 5c). The favorability of mid-shell mutations is consistent with the statistical survey by Morley et al. that indicates particularly a high population for activity-enhancing mutations located in the 9–20 Å from the active site.[29]

Under the percentile-criterion, however, a significantly higher propensity for rate-enhancing mutations is found in the third-shell which homes to the 33.3% of protein residues most distal to the active-site in each hydrolase (15%, Figure 5f). This indicates that the distal mutations for each hydrolase, however distant they actually are from the active site, hold the best likelihood for rate enhancement. This is consistent with the widely-reported observation of the beneficial roles played by remote mutations in directed evolution experiments.[32]

Noticeably, despite being more deleterious in terms of median $\Delta\Delta G^{\ddagger}$, the mutations in the inner-shell or in the first-shell show a decent proportion of rate-enhancing mutations (~10%, Figure 5c and 5f, respectively), which is comparable to the average percentage of rate-enhancing mutations across shells (9%, Figure 2). Moreover, when separately analyzing the rate-enhancing mutations, the distribution of $\Delta\Delta G^{\ddagger}$ for the inner-shell mutation has no significant difference from that for the other shells (Supporting Information, Figure S6). These results reveal that the inner-shell contains mutations with a diverse range of rate-perturbing effects, where the opportunities for rate enhancement and the tendency for activity abolishment co-exist. We also observed a particularly low population of the rate-neutral mutations in the inner-shell, which indicates their low likelihood of inducing neutral drift[21, 47-49] (Supporting Information, Figure S7).

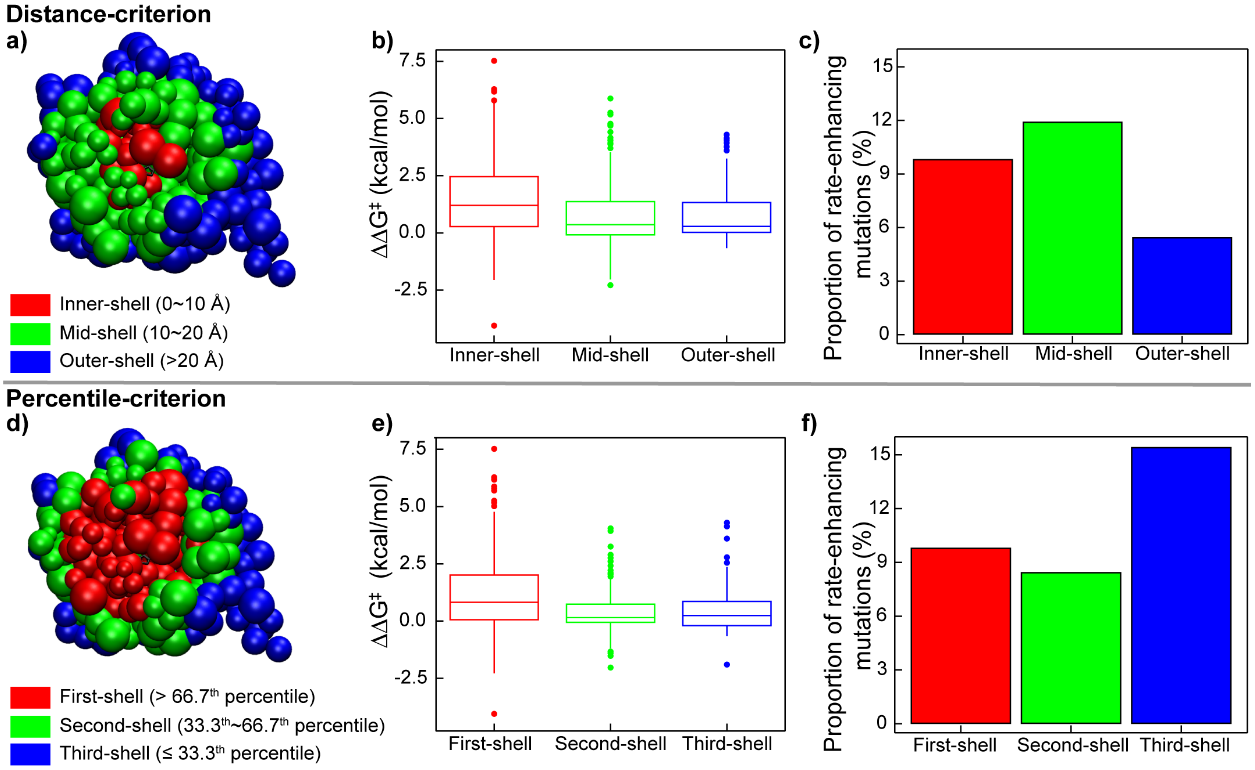


Figure 5. The spatially-resolved characterization for rate-enhancing hydrolase mutations with reference to the active-site. a) Three spatial shells, inner- (red), mid- (green), and outer-shell (blue) defined based on the distance of the mutation residue C_α coordinate to the geometric center of the active-site residues C_α coordinates. b) The distribution of $\Delta\Delta G^\ddagger$ for the mutation residues in the inner-, mid-, and outer-shell. c) The proportion of the rate-enhancing mutations located in the inner-, mid-, and outer-shell. d) Three spatial shells, first- (red), second- (green), and third-shell (blue) defined based on the mutation residue's percentile rank of spatial proximity to the active-site. e) The distribution of $\Delta\Delta G^\ddagger$ for the mutation residues located in the first-, second-, and third-shell. f) The proportion of the rate-enhancing mutations located in the first-, second-, and third-shell. The boxplots include the median, the 25th quantile, and the 75th quantile as the middle bar, the lower bound, and the upper bound of the box. The graphic illustration for the three-shells with serum paraoxonase as the model hydrolase (PDB ID: 1V04).

Although the distance-criterion has a straightforward chemical meaning and has been frequently applied to discuss the spatial distribution of protein mutations, the distance cutoff is not capable of accounting for the diverse range of protein size and shape. For instance, the number of residues in the outer-shell is much less for smaller-sized proteins than that for larger-sized proteins. Consequently, in our following analysis, we adopted the percentile-criterion, which normalizes the difference in protein size. For each hydrolase, the percentile-criterion allows an approximately equal number of enzyme residues populated in each protein shell.

Using the percentile-criterion, we studied how the spatial distribution of rate-enhancing mutations depends on the protein size. We evenly divided the hydrolases into larger-sized (i.e., sequence length > 324) and smaller-sized groups (i.e., sequence length ≤ 324) ranked by their sequence lengths (Supporting Information, Figure S8). In contrast, larger-sized and smaller-sized hydrolases show the distinct spatial distribution for rate-enhancing mutations. For larger-sized hydrolases, the mutations in the third-shell involve a three-times higher proportion of rate-enhancing mutations than those in the other two shells (21% in the third-shell versus 6-7% in the inner- or the mid-shell Figure 6a). For smaller-sized hydrolases, however, mutations in the three shells involve a relatively similar proportion of rate enhancement with the third-shell slightly more advantaged than the other two shells (12%, 11%, and 10% in the third-, second- and first-shell Figure 6b). These results show that larger-sized hydrolases have a stronger preference towards rate-enhancing distal mutations.

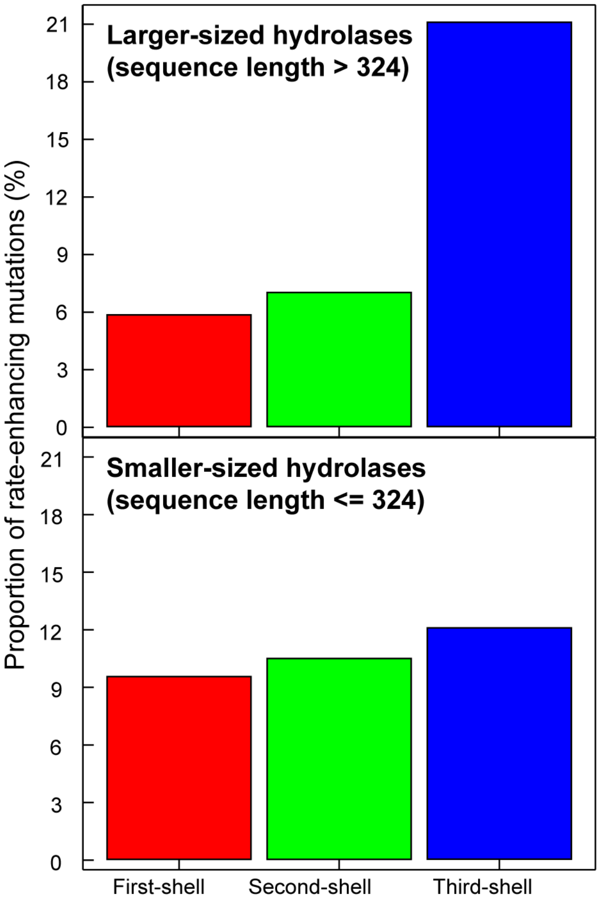


Figure 6. The proportion of the rate-enhancing mutations located in the first- (red), second- (green), and third-shell (blue) for larger-sized and smaller-sized hydrolases. The three shells are

defined based on the mutation residue's percentile rank of spatial proximity to the active site of the hydrolase.

3e. Structure-Kinetics Relationship with Reference to the Protein Center-of-Mass

We investigated the structure-kinetics relationship for mutant hydrolases with reference to the center-of-mass of the hydrolase. Unlike the active-site whose position varies in different hydrolases (e.g., buried inside or close to the surface), the center-of-mass characterizes an interior geometric center for hydrolases with a folded globular shape. These two references are complementary: the active site reference relates to the enzyme function, while the center-of-mass reference concerns the protein geometry.

Using the percentile-criterion, we categorized the mutation residues into the three shells based on the percentile rank of its spatial proximity to the center-of-mass of the hydrolase among all residues (Figure 7). This results in 287, 153, and 65 mutations in the first-, second-, and third-shell. The mutations in the first-shell involve higher $\Delta\Delta G^{\ddagger}$ by about 1.0 kcal/mol than those in the other two shells (Figure 7a). This observation is very similar to those defined using the active-site references (Figure 5e) because there is a significant overlap in the first-shell mutations defined using these two references (i.e., 258 mutations in common, Supporting Information, Figure S9). This indicates that mutations of the residues located in the protein geometric core is likely to be deleterious.

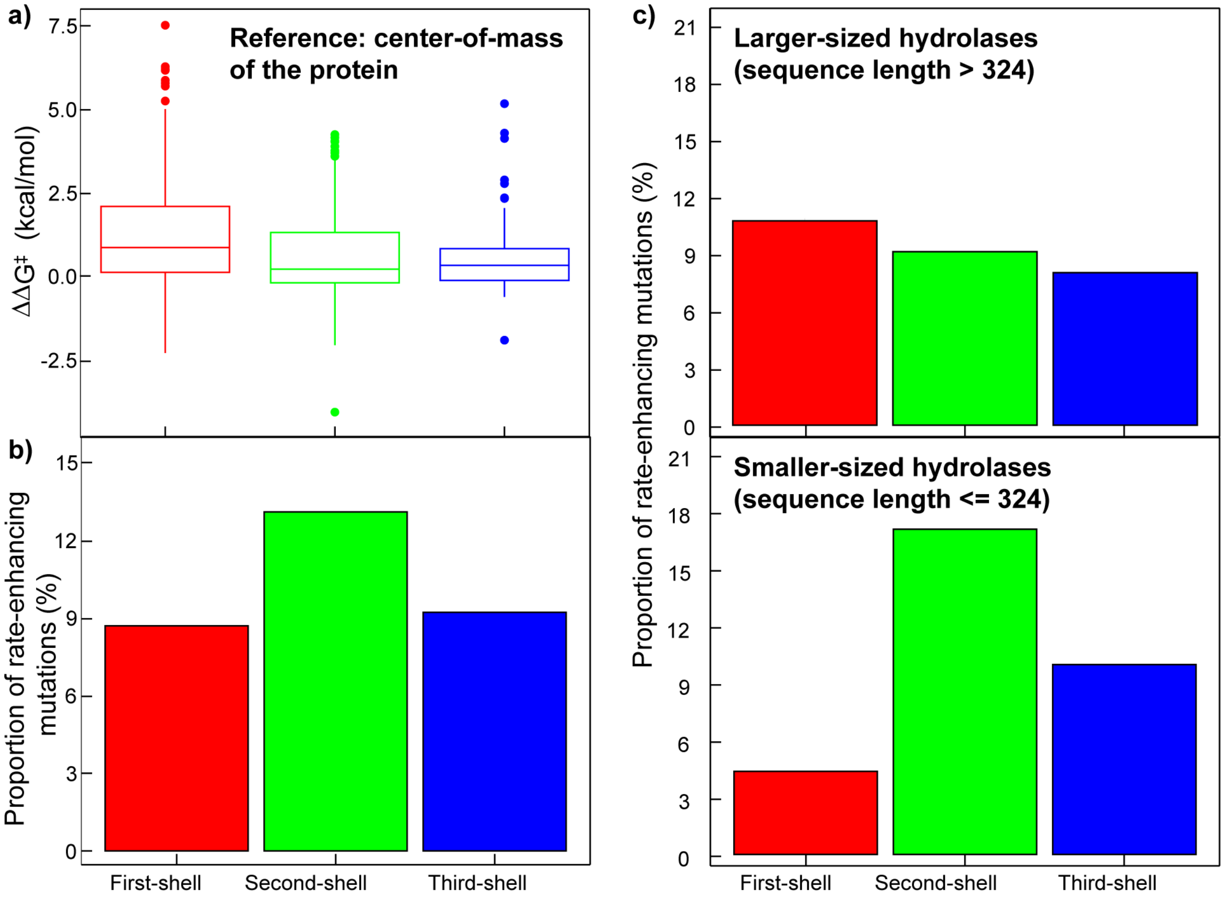


Figure 7. The spatially-resolved characterization for rate-enhancing hydrolase mutations with reference to the center-of-mass of the hydrolase, where the three spatial shells, first- (red), second- (green), and third-shell (blue) defined based on the mutation residue's percentile rank of spatial proximity to the center-of-mass of the hydrolase. a) The distribution of $\Delta\Delta G^{\ddagger}$ for the mutation residues and b) the proportion of the rate-enhancing mutations in the three shells. c) The proportion of the rate-enhancing mutations for larger-sized and smaller-sized hydrolases in the three shells.

To characterize the spatially resolved likelihood of rate enhancement, we computed the proportion of rate-enhancing mutations for each shell. Distinct from the observation of the third-shell residues being most rate-enhancing under the active-site reference (Figure 5f), the mutations in the second-sell have the highest proportion of rate-enhancing mutations under the center-of-mass reference (13%, Figure 7b). When categorizing the hydrolases into two groups based on the sequence lengths as described in the section 2d, the mutations in the second-shell for the smaller-sized hydrolases were found to involve a two- to three-fold higher proportion of rate-enhancing mutations than those in the other two shells (i.e., 17% in second-shell versus 10% and 4% in the third- and first-shell, Figure 7d), while mutations in the three shells for the larger-sized hydrolases involve a relatively similar proportion of rate-enhancement (8%, 9%, and 11% in the third-, second- and first-shell Figure 7c). These results show that smaller-sized hydrolases exhibit a stronger preference towards rate-enhancing second-shell mutations under the center-of-mass reference. The protein-center-oriented description can thus complement the active-site-oriented description to provide new metrics for evaluating the spatial dependence of the rate-enhancing mutations.

4. Conclusions

We constructed a hydrolase database, IntEnzyDB, which stores clean and tabulated structure and kinetics data by adopting a relational architecture with the flattened data structure. The database allows the easy pairing of hydrolase structure and kinetics data and exhibits superior efficiency when processing a large amount of data. With IntEnzyDB, we curated two datasets of mutant hydrolases to statistically characterize the rate-enhancing single amino acid mutations. One dataset consists of 1,500 distinct kinetics entries and the other of 505 kinetics-structure pairs.

Using the kinetics dataset, we first converted $k_{cat}^{mutation}/k_{cat}^{wild\text{-}type}$ to $\Delta\Delta G^{\ddagger}$, and then categorized each mutation to be rate-enhancing, -neutral, and -deleterious. We found 9% of the mutations are rate-enhancing, which overestimates the natural occurrence but provides abundant data for our analysis. Among the rate-enhancing mutations, we observed a particularly strong rate-enhancing propensity of mutation to bulky nonpolar residues with a hydrocarbon side chain. Since these nonpolar residues do not directly participate in forming strong polar or charged interactions with local residues, we suspect they play significant roles in tuning protein dynamics for promoting reaction rate.

Using the kinetics-structure dataset, we studied the spatially-resolved likelihood of rate enhancement for mutation. With the active site as the reference point, we compared two criteria for defining the spatial shells. The distance-criterion categorizes a mutation into one of the spatial shells based on its distance to the active site. Despite being chemically intuitive, this criterion is incapable of normalizing the difference in protein size and shape. Rather, we adopted a percentile-criterion that categorizes a mutation into a spatial shell based on its percentile rank of spatial proximity to the active site. We observed the highest likelihood for locating rate-enhancing mutations in the third shell, which is most distal to the active-site. This trend enhances for larger-sized hydrolases. When shifting the reference to the center-of-mass of the enzyme, we observed the highest likelihood for locating rate-enhancing mutations in the second shell, which contains residues that are 33.3th to 66.7th percentile rank of proximity to the center-of-mass of the enzyme. This trend enhances for smaller-sized hydrolases. Under either reference, mutations in the first shell (i.e., closest to the reference point) appear significantly more rate-deleterious than those in the other two shells (i.e., ~1.0 kcal/mol in $\Delta\Delta G^{\ddagger}$).

In a summary, this study provides a meta-analysis of the amino-acid-type and spatial distributions for mutations that are prone to induce rate enhancement for hydrolases. The study has the potential of guiding the identification of single amino acid mutations that accelerate the hydrolase-catalyzed reactions. We hope the study will inspire further investigations on how specific enzyme functions and substrate properties influence the discovery of rate-enhancing mutant hydrolases. The IntEnzyDB we constructed will provide clean and tabulated structural and kinetics data, enabling easy construction of predictive models for enzyme kinetics based on statistical modeling or machine learning methods.

ASSOCIATED CONTENT

Supporting Information. Distribution of operating temperature for enzymatic reactions; distribution of hydrolase enzyme commission number; distribution of the number of the missing residue and resolution in hydrolases; benchmark of data processing time for IntEnzyDB against Protein Data Bank; analysis of rate-enhancing mutations for bulky nonpolar residues with a hydrocarbon chain and other types of residues; analysis of rate-enhancing mutations for nonpolar-nonpolar, nonpolar-charged/polar, charged/polar-nonpolar, charged/polar-charged/polar residues; distribution of $\Delta\Delta G^{\ddagger}$ for rate-enhancing mutations in different spatial shells; analysis of proportion of rate-neutral mutations in different spatial shells; distribution of hydrolase sequence length; estimation of active site and center of mass layers overlap. (PDF)

Brenda kinetic data extraction code; Sabio-RK kinetic data extraction code; kinetics data cleaning for Brenda and Sabio-RK; PDB structural data extraction code; cleaned kinetics data table; kinetics-structural data mapping table; general structural data table; structural atom coordinate data table. (ZIP)

AUTHOR INFORMATION

Corresponding Author

*email: zhongyue.yang@vanderbilt.edu phone: 615-343-9849

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This research was supported by the startup grant from Vanderbilt University and the scholarship from the Vanderbilt Institue of Chemical Biology. This work was carried out in part using computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number TG-BIO200057.[50]

References

1. Pavlidis, I. V. In *Identification and Evolution of Biocatalysts of Interest*, Advanced Nanotechnologies for Detection and Defence against CBRN Agents, Dordrecht, 2018//; Petkov, P.; Tsiulyanu, D.; Popov, C.; Kulisch, W., Eds. Springer Netherlands: Dordrecht, 2018; 477-485.
2. Rauwerdink, A.; Kazlauskas, R. J. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS Catal.* **2015,** *5*, 6153–6176.
3. Sousa, S. F.; Ramos, M. J.; Lim, C.; Fernandes, P. A. Relationship between Enzyme/Substrate Properties and Enzyme Efficiency in Hydrolases. *ACS Catal.* **2015,** *5*, 5877–5887.
4. Liu, J.-Q.; Kurihara, T.; Ichiyama, S.; Miyagi, M.; Tsunasawa, S.; Kawasaki, H.; Soda, K.; Esaki, N. Reaction Mechanism of Fluoroacetate Dehalogenase from Moraxella sp. B*. *J. Biol. Chem.* **1998,** *273*, 30897-30902.
5. Schulz, E. C.; Mehrabi, P.; Müller-Werkmeister, H. M.; Tellkamp, F.; Jha, A.; Stuart, W.; Persch, E.; De Gasparo, R.; Diederich, F.; Pai, E. F.; Miller, R. J. D. The Hit-and-Return System Enables Efficient Time-resolved Serial Synchrotron Crystallography. *Nat. Methods.* **2018,** *15*, 901–904.
6. Yoshida, S.; Hiraga, K.; Takehana, T.; Taniguchi, I.; Yamaji, H.; Maeda, Y.; Toyohara, K.; Miyamoto, K.; Kimura, Y.; Oda, K. A Bacterium that Degrades and Assimilates Poly(ethylene terephthalate). *Science* **2016,** *351*, 1196.
7. Zeymer, C.; Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018,** *87*, 131–157.
8. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008,** *319*, 1387.
9. Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **2009,** *10*, 866–876.

10. Gerlt, J. A.; Babbitt, P. C. Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annu. Rev. Biochem.* **2001,** *70*, 209–246.

11. 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. *J. Am. Chem. Soc.* **2012,** *134*, 16197–16206.

12. 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.

13. Melnikov, A.; Rogov, P.; Wang, L.; Gnirke, A.; Mikkelsen, T. S. Comprehensive Mutational Scanning of a Kinase in vivo Reveals Substrate-dependent Fitness Landscapes. *bioRxiv* **2014**, 004317.

14. Wrenbeck, E. E.; Azouz, L. R.; Whitehead, T. A. Single-mutation Fitness Landscapes for an Enzyme on Multiple Substrates Reveal Specificity is Globally Encoded. *Nat. Commun.* **2017,** *8*, 15695.

15. Nannemann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O. Assessing Directed Evolution Methods for the Generation of Biosynthetic Enzymes with Potential in Drug Biosynthesis. *Future Med. Chem.* **2011,** *3*, 809-819.

16. Markin, C. J.; Mokhtari, D. A.; Sunden, F.; Appel, M. J.; Akiva, E.; Longwell, S. A.; Sabatti, C.; Herschlag, D.; Fordyce, P. M. Revealing Enzyme Functional Architecture via High-throughput Microfluidic Enzyme Kinetics. *bioRxiv* **2020**, 2020.11.24.383182.

17. Fowler, D. M.; Fields, S. Deep Mutational Scanning: a New Style of Protein Science. *Nat. Methods.* **2014,** *11*, 801–807.

18. Araya, C. L.; Fowler, D. M. Deep Mutational Scanning: Assessing Protein Function on a Massive Scale. *Trends Biotechnol.* **2011,** *29*, 435–442.

19. Chen, J. Z.; Fowler, D. M.; Tokuriki, N. Comprehensive Exploration of the Translocation, Stability and Substrate Recognition Requirements in VIM-2 Lactamase. *bioRxiv* **2020**, 2020.02.19.956706.

20. Zheng, J.; Guo, N.; Wagner, A. Selection Enhances Protein Evolvability by Increasing Mutational Robustness and Foldability. *Science* **2020,** *370*, eabb5962.

21. Soskine, M.; Tawfik, D. S. Mutational Effects and the Evolution of New Protein Functions. *Nat. Rev. Genet.* **2010,** *11*, 572–582.

22. Yang, G.; Miton, C. M.; Tokuriki, N. A Mechanistic View of Enzyme Evolution. *Protein Sci.* **2020,** *29*, 1724–1747.

23. Barak, Y.; Nov, Y.; Ackerley, D. F.; Matin, A. Enzyme Improvement in the Absence of Structural Knowledge: a Novel Statistical Approach. *ISME J.* **2008,** *2*, 171–179.

24. Amrein, B. A.; Steffen-Munsberg, F.; Szeler, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L. CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCrJ* **2017,** *4*, 50–64.

25. Amrein, B. A.; Runthala, A.; Kamerlin, S. C. L. In Silico-Directed Evolution Using CADEE. In *Computational Methods in Protein Evolution*, Sikosek, T., Ed. Springer New York: New York, NY, 2019; 381-415.

26. Davey, J. A.; Damry, A. M.; Goto, N. K.; Chica, R. A. Rational Design of Proteins that Exchange on Functional Timescales. *Nat. Chem. Biol.* **2017,** *13*, 1280–1285.

27. Kipnis, Y.; Baker, D. Comparison of Designed and Randomly Generated Catalysts for Simple Chemical Reactions. *Protein Sci.* **2012,** *21*, 1388–1395.
28. Heckmann, D.; Lloyd, C. J.; Mih, N.; Ha, Y.; Zielinski, D. C.; Haiman, Z. B.; Desouki, A. A.; Lercher, M. J.; Palsson, B. O. Machine Learning Applied to Enzyme Turnover Numbers Reveals Protein Structural Correlates and Improves Metabolic Models. *Nat. Commun.* **2018,** *9*, 5252.
29. Morley, K.; Kazlauskas, R. Improving enzyme properties: when are closer mutations better? Trends Biotechnol. 23, 231-237. *Trends Biotechnol.* **2005,** *23*, 231–7.
30. Miton, C.; Tokuriki, N. How Mutational Epistasis Impairs Predictability in Protein Evolution and Design. *Protein Sci.* **2016,** *25,* 1260–1272.

31. Kaltenbach, M.; Jackson, C. J.; Campbell, E. C.; Hollfelder, F.; Tokuriki, N. Reverse Evolution Leads to Genotypic Incompatibility Despite Functional and Active Site Convergence. *Elife* **2015,** *4*, e06492.
32. Wilding, M.; Hong, N.; Spence, M.; Buckle, A. M.; Jackson, C. J. Protein Engineering: the Potential of Remote Mutations. *Biochem. Soc. Trans.* **2019,** *47*, 701–711.
33. Sousa, S. F.; Calixto, A. R.; Ferreira, P.; Ramos, M. J.; Lim, C.; Fernandes, P. A. Activation Free Energy, Substrate Binding Free Energy, and Enzyme Efficiency Fall in a Very Narrow Range of Values for Most Enzymes. *ACS Catal.* **2020,** *10*, 8444–8453.
34. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28*, 235–242.
35. The UniProt, C., UniProt: a Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019,** *47*, D506–D515.
36. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. L. UniProt: the Universal Protein Knowledgebase. *Nucleic Acids Res.* **2004,** *32*, D115–D119.
37. Jeske, L.; Placzek, S.; Schomburg, I.; Chang, A.; Schomburg, D. BRENDA in 2019: a European ELIXIR Core Data Resource. *Nucleic Acids Res.* **2019,** *47*, D542–D549.
38. Wittig, U.; Kania, R.; Golebiewski, M.; Rey, M.; Shi, L.; Jong, L.; Algaa, E.; Weidemann, A.; Sauer-Danzwith, H.; Mir, S.; Krebs, O.; Bittkowski, M.; Wetsch, E.; Rojas, I.; Müller, W. SABIO-RK–Database for Biochemical Reaction Kinetics. *Nucleic Acids Res.* **2012,** *40*, D790–6.
39. Gray, V. E.; Kukurba, K. R.; Kumar, S. Performance of Computational Tools in Evaluating the Functional Impact of Laboratory-Induced Amino Acid Mutations. *Bioinformatics* **2012,** *28*, 2093–2096.
40. Dušan, P.; Shina Caroline Lynn, K. Molecular Modeling of Conformational Dynamics and its Role in Enzyme Evolution. *Curr. Opin. Struct. Biol.* **2018,** *52*, 50–57.
41. Otten, R.; Liu, L.; Kenner, L. R.; Clarkson, M. W.; Mavor, D.; Tawfik, D. S.; Kern, D.; Fraser, J. S. Rescue of Conformational Dynamics in Enzyme Catalysis by Directed Evolution. *Nat. Commun.* **2018,** *9*, 1314.
42. Campitelli, P.; Modi, T.; Kumar, S.; Ozkan, S. B. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* **2020,** *49*, 267–288.
43. Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of Conformational Dynamics in the Evolution of Novel Enzyme Function. *ChemComm* **2018,** *54*, 6622–6634.

44. Ribitsch, D.; Hromic, A.; Zitzenbacher, S.; Zartl, B.; Gamerith, C.; Pellis, A.; Jungbauer, A.; Łyskowski, A.; Steinkellner, G.; Gruber, K.; Tscheliessnig, R.; Herrero Acero, E.; Guebitz, G. M. Small Cause, Large Effect: Structural Characterization of Cutinases from *Thermobifida cellulosilytica*. *Biotechnol. Bioeng.* **2017,** *114*, 2481–2488.

45. Herrero Acero, E.; Ribitsch, D.; Dellacher, A.; Zitzenbacher, S.; Marold, A.; Steinkellner, G.; Gruber, K.; Schwab, H.; Guebitz, G. M. Surface Engineering of a Cutinase from *Thermobifida cellulosilytica* for Improved Polyester Hydrolysis. *Biotechnol. Bioeng.* **2013,** *110*, 2581–2590.

46. Legler, P. M.; Kumaran, D.; Swaminathan, S.; Studier, F. W.; Millard, C. B. Structural Characterization and Reversal of the Natural Organophosphate Resistance of a D-Type Esterase, Saccharomyces cerevisiae S-Formylglutathione Hydrolase. *Biochemistry* **2008,** *47*, 9592–9601.

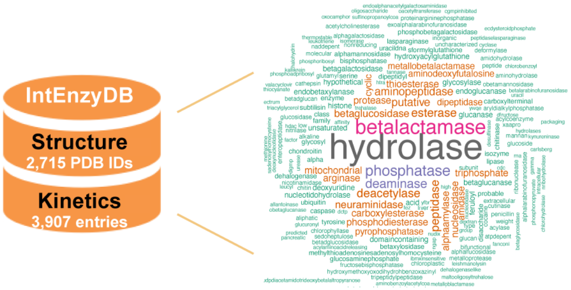
47. Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S. M.; Roodveldt, C.; Tawfik, D. S. The 'Evolvability' of Promiscuous Protein Functions. *Nat. Genet.* **2005,** *37*, 73–76.

48. Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLoS Comput. Biol.* **2008,** *4*, e1000002.

49. Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D. S. The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J. Mol. Biol.* **2007,** *369*, 1318–1332.

50. 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. *COMPUT SCI ENG.* **2014,** *16*, 62–74.

Table of Contents Graphic


IntEnzyDB
Structure
2,715 PDB IDs
Kinetics
3,907 entries
endoalphanacetylgalactosaminidase
oligosaccharide
oacetyltransferase
cgmpinhibited
oxocamphor
sulfinopropanoylcoa
proteinargininephosphatase
acetylcholinesterase
exoalphalarabinofuranosidase
ecdysteroidphosphate
phosphobetagalactosidase
thermostable
alphagalactosidase
inorganic
peptidaselasparaginase
leukotriene
isomerase
lasparaginase
naddepent
nonreducing
uncharacterized
cyclase
uracildna
sformylglutathione
deformylase
molecular
alphamannosidase
hydroxyacylglutathione
amidohydrolase
halohydrin
phosphoribosyl
bisphosphatase
kallikrein
betagalactosidase
metallobetalactamase
peptide
chlorobenzoyl
phosphoadpribosyl
glutamyl
serine
tannase
dipeptidyl
aminodeoxyfutalosine
aminohydrolase
valacyclovir
cathepsin
hypothetical
cyclic
htra
thioesterase
glycosylase
obetamannosidase
thiocyanate
endobetaxylanase
aminopeptidase
endoglucanase
betalarabinofuranosidase
uracil
ectrum
betadglucan
enzyme
protease
putative
dipeptidase
carboxylterminal
triacylglycerol
subtilisin
histone
trehalase
ywqe
aryldialkylphosphatase
glucosidase
class
betaglucosidase
esterase
glucanase
dfructose
family
affinity
acylcoenzyme
packaging
hdac
low
xaapro
nitrilase
unsaturated
betalactamase
desulfinase
chitinase
hydrolases
methionine
sadenosylhomocysteine
deoxynucleotidase
enteropeptidase
alkaline
mannan
kynureninase
factor
glycosyl
hydrolase
glucoside
chondroitin
carlsberg
isozyme
alpha
rna
dgmp
urease
lipase
mitochondrial
phosphatase
subunit
cdc
triphosphate
ribonuclease
alphalarabinofuranosidase
gamma
intracellular
phosphonopyruvate
nribohydrolase
dehalogenase
arginase
nicotinamidase
deaminase
betaglucanase
leucyl
chitin
deoxyuridine
high
nucleotidohydrolase
deacetylase
peptidase
nucleosidase
amidase
nucleotidase
feruloyl
probable
allantoinase
ubiquitin
acid
yfbr
extracellular
obetaglucanase
neuraminidase
fez
liver
cutinase
chlorohydrolase
caspase
dctp
carboxylesterase
disaccharide
cocaine
penicillin
aliphatic
dextran
weight
glucuronyl
tyrosine
phosphodiesterase
alphaamylase
type
acylase
betaglycosidase
chlorophyllase
group
predicted
sedoheptulose
pyrophosphatase
nudix
atpdepent
pancreatic
glucan
betadglucosidase
domaincontaining
ngap
fanconi
acylaminoacidreleasing
betaxylosidase
bifunctional
methylthioadenosinesadenosylhomocysteine
alphalfucosidase
glucosaminephosphate
ibmxinsensitive
metalloprotease
fructosebisphosphatase
chloroplastic
leishmanolysin
hydroxymethoxyoxodihydrohbenzoxazinyl
dehalogenaselike
tripeptidylpeptidase
maltooligosyltrehalose
udpdiacetamidotrideoxybetalaltropyranose
aminobenzoylacetylcoa
metalloblactamase