Supporting Information

IntEnzyDB: an Integrated Structure-Kinetics Enzymology Database

Bailu Yan,[1,6] Xinchun Ran,[1] Anvita Gollu,[1] Zihao Cheng,[1] Xiang Zhou,[1] Yiwen Chen,[4] 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, 37205, United States* [6]*Department of Biostatistics, Vanderbilt University, Nashville, Tennessee, 37205, United States*

Corresponding Author

*email: zhongyue.yang@vanderbilt.edu

Table S1. Enzymology Databases.

Name	Type	UniProt	EC Number	PDB ID	Website
BRENDA	Kinetics	Yes	Yes	Part	https://www.brenda-enzymes.org/
SABIO-RK	Kinetics	Yes	Yes	No	http://sabio.h-its.org/
STRENDA DB	Kinetics with uniform data standard	Yes	Yes	No	https://www.beilstein-strenda-db.org/strenda/
PDB	Structure	Yes	Yes	Yes	https://www.rcsb.org/
AlphaFold DB	Predicted Structure	Yes	No	Part	https://alphafold.ebi.ac.uk/
UniProt	Protein Sequence and function	Yes	Yes	Yes	https://www.uniprot.org/
KEGG	Enzyme nomenclature	No	No	No	https://www.genome.jp/kegg/
IntEnz	Enzyme nomenclature	No	No	No	https://www.ebi.ac.uk/intenz/
ProtaBank	Kinetics, structure, and sequence	Part	Part	Part	https://www.protabank.org/
Design2Data	Kinetics and structure	Yes	Yes	Yes	https://siegel.ucdavis.edu/design-2-data
SFLD	Enzyme superfamily mapped to sequence and structure	Yes	Yes	Part	http://sfld.rbvi.ucsf.edu/archive/django/index.html
BioCatNet	Sequence and binding parameters	No	No	No	https://www.biocatnet.de/
BioModels	kinetic models of biochemical and cellular systems	No	No	No	https://www.ebi.ac.uk/biomodels/
JWS Online	kinetic models	No	No	No	http://jjj.biochem.sun.ac.za/
DOQCS	Quantitative cellular signaling	No	No	No	https://doqcs.ncbs.res.in/
DDBJ/EMBL/GenBank	Nucleotide sequence	No	No	No	https://www.insdc.org/
ProBiS-Dock	Ligand binding sites	No	No	No	http://probis-dock-database.insilab.org/

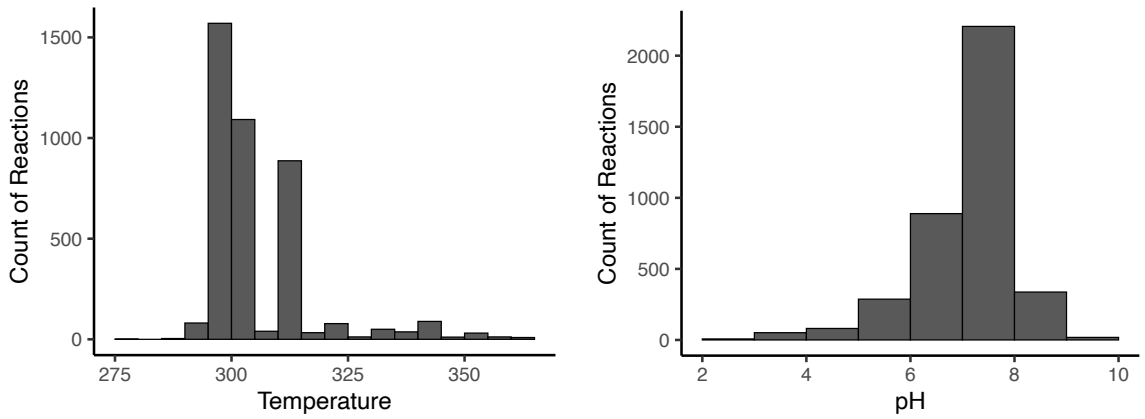

Figure S1. Distribution of experimental temperature (*left*) and pH (*right*) for the enzymatic reactions. The x axis shows the temperature or pH under which the enzymatic reaction kinetics was measured in Kelvin. The y axis shows the number of enzymatic reactions under a certain property range.

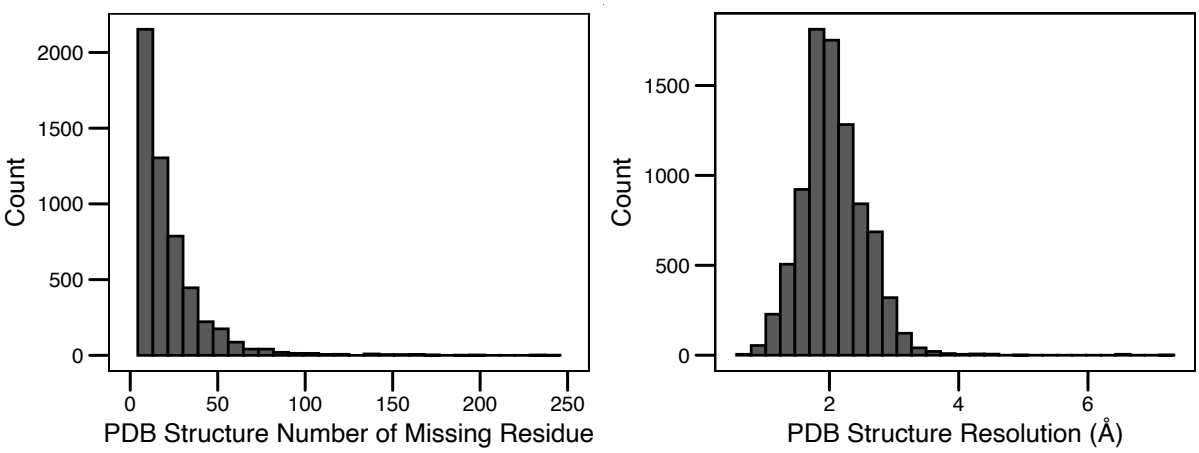

Figure S2. (Left) Distribution of the number of missing residues for the curated enzyme structures; x axis shows the sequence length of each PDB structure and y axis shows the number of structures for a corresponding sequence length range. The bin size is 5 residues. (Right) Distribution of enzyme structure resolution; x axis shows the resolution of each PDB structure (in Å) and y axis shows the count of structures for a resolution range. The bin size is 0.1 Å.

Table S2. Proportion of efficiency-enhancing, neutral, and deleterious enzyme mutant for close and distal mutations under various distance cutoff.

Rate-Perturbing Type	Distal or Close	Proportion	Distance (Å)
Deleterious	close	0.8358	10
Deleterious	distal	0.5245	10
Enhancing	close	0.0448	10
Enhancing	distal	0.0979	10

Neutral	close	0.1194	10
Neutral	distal	0.3776	10
Deleterious	close	0.8172	11
Deleterious	distal	0.5177	11
Enhancing	close	0.0457	11
Enhancing	distal	0.1003	11
Neutral	close	0.1371	11
Neutral	distal	0.3820	11
Deleterious	close	0.7773	12
Deleterious	distal	0.5092	12
Enhancing	close	0.0601	12
Enhancing	distal	0.0965	12
Neutral	close	0.1626	12
Neutral	distal	0.3943	12
Deleterious	close	0.7583	13
Deleterious	distal	0.4972	13
Enhancing	close	0.0688	13
Enhancing	distal	0.0924	13
Neutral	close	0.1729	13
Neutral	distal	0.4104	13
Deleterious	close	0.7500	14
Deleterious	distal	0.4728	14
Enhancing	close	0.0717	14
Enhancing	distal	0.0921	14
Neutral	close	0.1783	14
Neutral	distal	0.4351	14
Deleterious	close	0.7283	15
Deleterious	distal	0.4720	15
Enhancing	close	0.0804	15
Enhancing	distal	0.0818	15
Neutral	close	0.1913	15
Neutral	distal	0.4463	15
Deleterious	close	0.7227	16
Deleterious	distal	0.4564	16
Enhancing	close	0.0788	16
Enhancing	distal	0.0846	16
Neutral	close	0.1985	16
Neutral	distal	0.4590	16
Deleterious	close	0.6928	17

Deleterious	distal	0.4598	17
Enhancing	close	0.0866	17
Enhancing	distal	0.0675	17
Neutral	close	0.2206	17
Neutral	distal	0.4727	17
Deleterious	close	0.6987	18
Deleterious	distal	0.3962	18
Enhancing	close	0.0835	18
Enhancing	distal	0.0731	18
Neutral	close	0.2177	18
Neutral	distal	0.5308	18
Deleterious	close	0.6835	19
Deleterious	distal	0.3973	19
Enhancing	close	0.0854	19
Enhancing	distal	0.0639	19
Neutral	close	0.2310	19
Neutral	distal	0.5388	19
Deleterious	close	0.6723	20
Deleterious	distal	0.3636	20
Enhancing	close	0.0814	20
Enhancing	distal	0.0788	20
Neutral	close	0.2463	20
Neutral	distal	0.5576	20

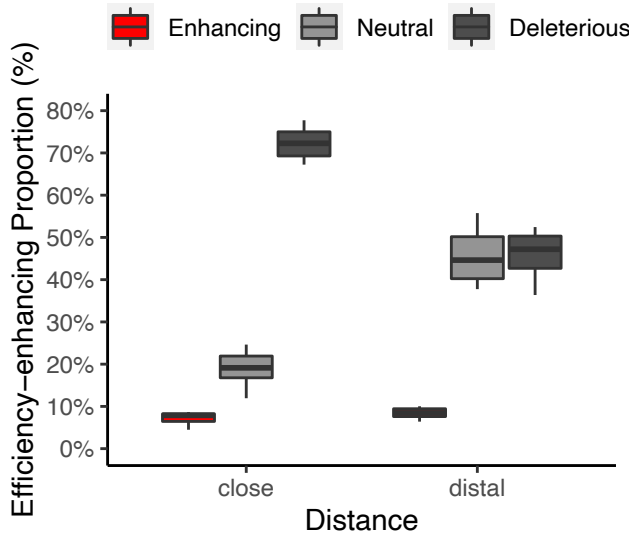

Figure S3. Distribution of proportions for efficiency-enhancing (red), -neutral (grey), and -deleterious (dark grey) mutations of six enzyme types (EC 1-6) using close and distal cutoff values

ranging from 10 to 20 Å with 1 Å interval. The lower bar of the box represents the 25% percentile of the distribution, the bold bar in the middle of the box represents the median of the distribution, the upper bar of the box represents the 75% percentile of the distribution of the proportions. The lower point of the vertical line for each box represents the minimum of the distribution, and the upper point of the vertical line for each box represents the maximum of the distribution.

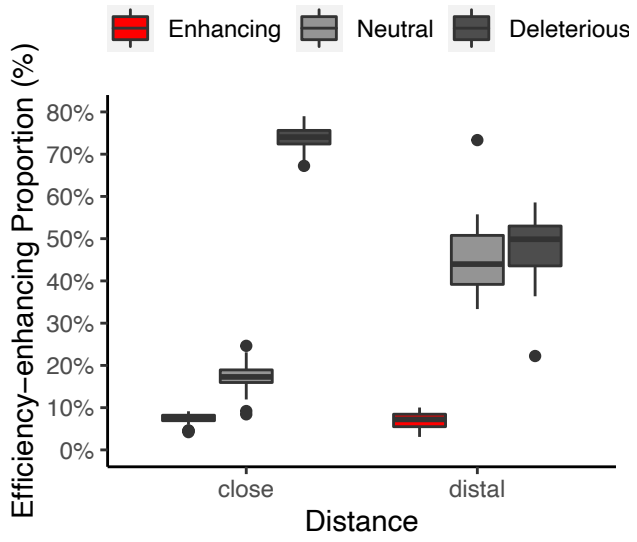

Figure S4. Distribution of proportions for efficiency-enhancing (red), -neutral (grey), and -deleterious (dark grey) oxidoreductases mutations using close and distal cutoff values ranging from 10 to 20 Å with 1 Å interval. The lower bar of the box represents the 25% percentile of the distribution, the bold bar in the middle of the box represents the median of the distribution, the upper bar of the box represents the 75% percentile of the distribution of the proportions. The lower point of the vertical line for each box represents the minimum of the distribution, and the upper point of the vertical line for each box represents the maximum of the distribution.

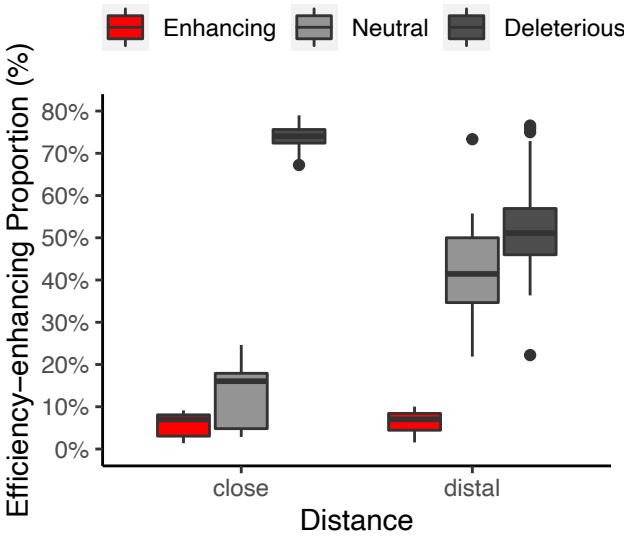

Figure S5. Distribution of proportions for efficiency-enhancing (red), -neutral (grey), and -deleterious (dark grey) transferase mutations using close and distal cutoff values ranging from 10 to 20 Å with 1 Å interval. The lower bar of the box represents the 25% percentile of the distribution, the bold bar in the middle of the box represents the median of the distribution, the upper bar of the box represents the 75% percentile of the distribution of the proportions. The lower point of the vertical line for each box represents the minimum of the distribution, and the upper point of the vertical line for each box represents the maximum of the distribution.

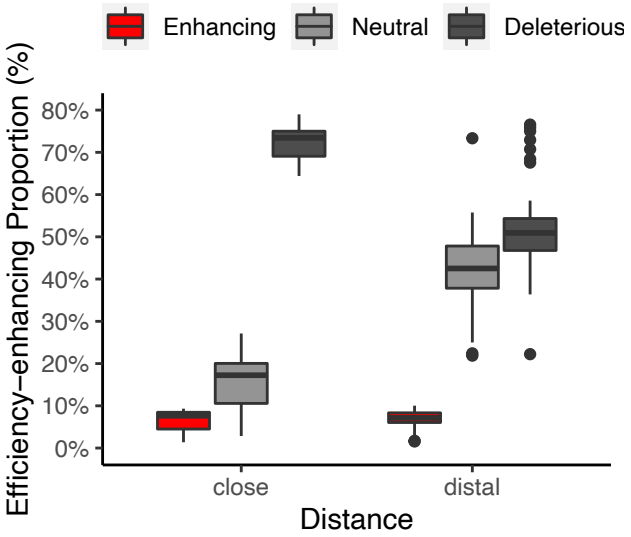

Figure S6. Distribution of proportions for efficiency-enhancing (red), -neutral (grey), and -deleterious (dark grey) hydrolase mutations using close and distal cutoff values ranging from 10 to 20 Å with 1 Å interval. The lower bar of the box represents the 25% percentile of the distribution, the bold bar in the middle of the box represents the median of the distribution, the upper bar of the box represents the 75% percentile of the distribution of the proportions. The lower point of the

vertical line for each box represents the minimum of the distribution, and the upper point of the vertical line for each box represents the maximum of the distribution.

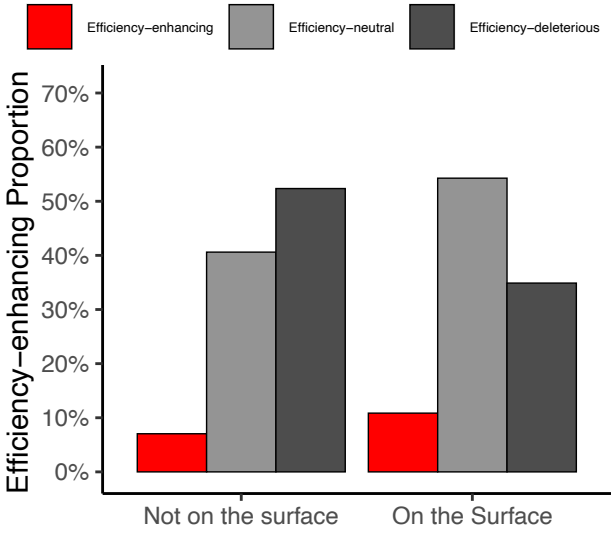

Figure S7. The proportion of efficiency-enhancing (red), -neutral (light grey), and -deleterious (dark grey) distal mutations (>15 Å from the active site) that are buried inside (*left*) or reside on the surface (*right*). The surface residues are defined to be those whose relative solvent-accessible surface area is >0.5. The relative solvent-accessible surface area was computed using Pymol. Otherwise, the residues are considered to be not on the surface. Efficiency-enhancing mutant is defined as $\Delta\Delta G^{\ddagger}$ less or equal to -0.5 kcal/mol (red), efficiency-neutral mutant is defined as $\Delta\Delta G^{\ddagger}$ greater than -0.5 kcal/mol and less or equal to 0.5 kcal/mol (light grey), efficiency-deleterious mutant is defined as $\Delta\Delta G^{\ddagger}$ greater than 0.5 kcal/mol (dark grey).

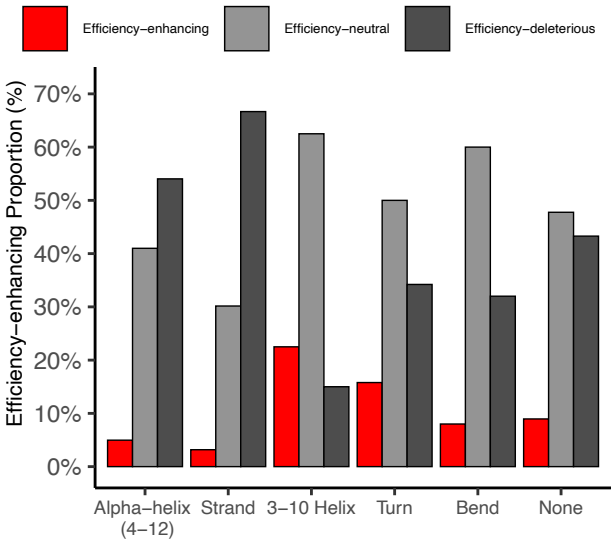

Figure S8. The proportion of efficiency-enhancing (red), -neutral (light grey), and -deleterious (dark grey) distal mutations (>15 Å from the active site) that reside on "alpha-helix", "beta-

strand", "3-10 helix", "beta-turn", "bend", or "none" (i.e., coil). The 8-class protein secondary structures are calculated using DSSP software. Efficiency-enhancing mutant is defined as $\Delta\Delta G^{\ddagger}$ less or equal to -0.5 kcal/mol (red), efficiency-neutral mutant is defined as $\Delta\Delta G^{\ddagger}$ greater than -0.5 kcal/mol and less or equal to 0.5 kcal/mol (light grey), efficiency-deleterious mutant is defined as $\Delta\Delta G^{\ddagger}$ greater than 0.5 kcal/mol (dark grey).