Design to Data for mutants of **β**-glucosidase B from *Paenibacillus polymyxa*: M319C, T431I, and K337D

Peishan Huang[1,2], Stephanie C. Contreras[1], Eliana Bloomfield[1], Kristine Schmitz[1], Augustine Arredondo[4], Justin B. Siegel[1,3][4]

[1]Genome Center, University of California, Davis, California, United States of America

[2]Biophysics Graduate Group, University of California, Davis, California, United States of America

[3]Department of Biochemistry & Molecular Medicine, University of California, Davis, California, United States of America

[4]Department of Chemistry, University of California, Davis, California, United States of America

ABSTRACT: The use of computational tools has become an increasingly popular tool for engineering protein function. While there are numerous examples of computational tools enabling the design of novel protein functions, there remains room for improvement in both prediction accuracy and success. To improve algorithms for functional and stability predictions, we have initiated the development of a data set designed to be used for training new computational algorithms for enzyme design. To date our dataset is composed of over 129 mutants with associated expression levels, kinetic data, and thermal stability for the enzyme β-glucosidase B (BglB) from *Paenibacillus polymyxa*. In this study, we introduced three new variants (M319C, T431I, and K337D) to our existing dataset with the goal of cultivating a larger dataset to train new design algorithms and more broadly explore structure-function relationships in BglB.

INTRODUCTION

Computational tools to engineer protein catalytic efficiency, stability, or substrate specificity are becoming crucial for the development of industrial and therapeutic proteins. Despite recent success in using rational design for thermal stability,[1–4] or creating *de novo* enzymes[5–8] not found in nature, current algorithms are inconsistent in their performance on different biological systems and often have a lower success than desirable.[9]
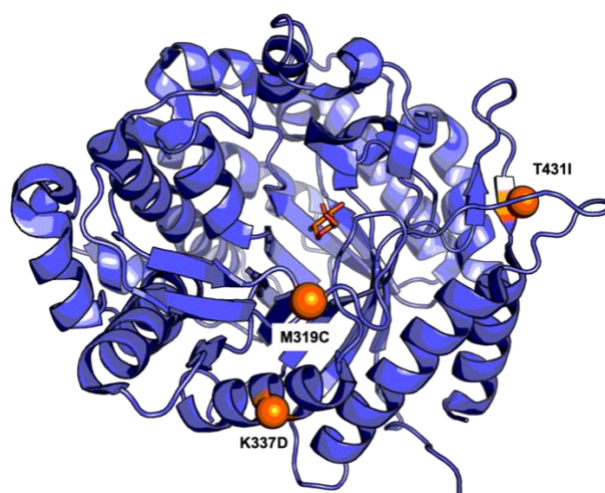
Figure 1. Rendering of BglB from *Paenibacillus polymyxa*. The three mutational sites that were made in study are highlighted in orange spheres on the cartoon illustration of BglB (PDB ID: 2JIE), rendered using PyMOL.[10]

One potential reason for inconsistent and low prediction accuracies is that the current algorithms are trained primarily on protein structure as opposed to function. This is particularly challenging for enzymes, many of which are believed to adopt multiple conformations to accommodate the reaction from of a substrate

product.[11] Furthermore, for enzymes, there are no large datasets that quantitatively measure structure-function relationships on potentially modellable physical properties (e.g. T_M, k_{cat}, K_M, and k_{cat}/K_M). Therefore, we have initiated an effort to systematically characterize a family 1 glycoside hydrolase (GH1) from *Paenibacillus polymyxa*, Beta-glucosidase B (BglB, Uniprot ID: P22505). Here we present three new variants (M319C, T431I, and K337D) to explore the mutational effects surface residues have on kinetic activity and thermal stability. To date, our dataset consists of over 129 mutants that have been characterized for T_{50}, k_{cat}, K_M, and k_{cat}/K_M[12] and we have added these three new variants to continue the expansion of our dataset for training an algorithm to have better predictive functional properties and thermal stability. Currently, the majority of the variants are found near the active-site of the BglB, to explore the mutational effects of surface residues on kinetic activity and thermal stability, M319C, T431I, and K337D (Fig 1) were chosen for testing.

METHODS

Design and building of new variants. Using a previously described method,[9] three single-point mutations, M319C, T431I, and K337D were selected and modeled using FoldIt.[13] The three surface mutational changes were scored by the Rosetta energy function and are given a total system energy score (TSE).[14] No mutants were selected if the changes between the wild type and variant score is predicted to be greater than +5 to enrich for proteins likelihood to fold and function.

Generating mutant plasmids. The mutant plasmids were generated by Kunkel mutagenesis[15] as previously described.[12] The plasmids were electroporated into Dh5α competent *Escherichia coli* (E.*coli*) cells and plated onto Luria-Bertani (LB) agar plate with 50 μg/mL kanamycin for overnight growth. Single colonies were picked and sequence to identify the desired mutants.

Protein production and purification. The sequence-verified BglB mutant DNA were transformed, grown, and expressed using the previously described method.[12] Briefly, following autoinduction the cells were lysed and proteins purified using immobilized metal ion affinity chromatography. The protein purity was analyzed using 4-12% SDS-PAGE (Life Technologies) and all samples used in the study were > 80 % pure. The protein concentrations were determined by measuring absorbance at 280 nm and calculating the expected molarity based on the predicted extinction coefficient from ProtParam.[16]

Michaelis-Menten kinetics and temperature melting assays (T_M). The kinetic characterization was done in the same manner as previous study.[9] Briefly, the activity was obtained by monitoring the production of 4-nitrophenyl at absorbance 420 nm from the substrate *p*-nitrophenyl-beta-D-glucoside (pNPG) at a range of substrate concentrations. The data was then fitted using the Michaelis-Menten equation to determine k_{cat} and K_M.

The thermal stability (T_m) for the mutants were determined using the Protein Thermal shift (PTS)™ kit made by Applied BioSystem ® from Thermo Fisher. Following the standard protocol by the manufacturer, purified proteins were diluted from 0.1 to 0.5 mg/mL and fluorescence reading was monitored using QuantaStudio™ 3 System from 20 °C to 90 °C. The T_M values were then determined using the two-state Boltzmann model from the Protein Thermal Shift™ Software 1.3 by Applied BioSystem ® from Thermo Fisher.

RESULTS

Molecular modeling of three mutants using FoldIt. All mutants were hypothesized to maintain similar kinetic activity based on the residues being on the surface of the enzyme and are 19-20 Å from the active-site (Fig 1). Furthermore, based on the change of total system energy (ΔTSE) between the WT and the mutant, M319C (ΔTSE of -2.55) and K337D (ΔTSE of -2.368), we predicted a slight increase in thermal stability (T_M) compared to WT (Fig 2). On the other hand, the change from hydroxyl phenol group with hydrogen bonding capability to an alkyl group for mutant T431I we predicted a small decrease in stability (ΔTSE of +0.86).
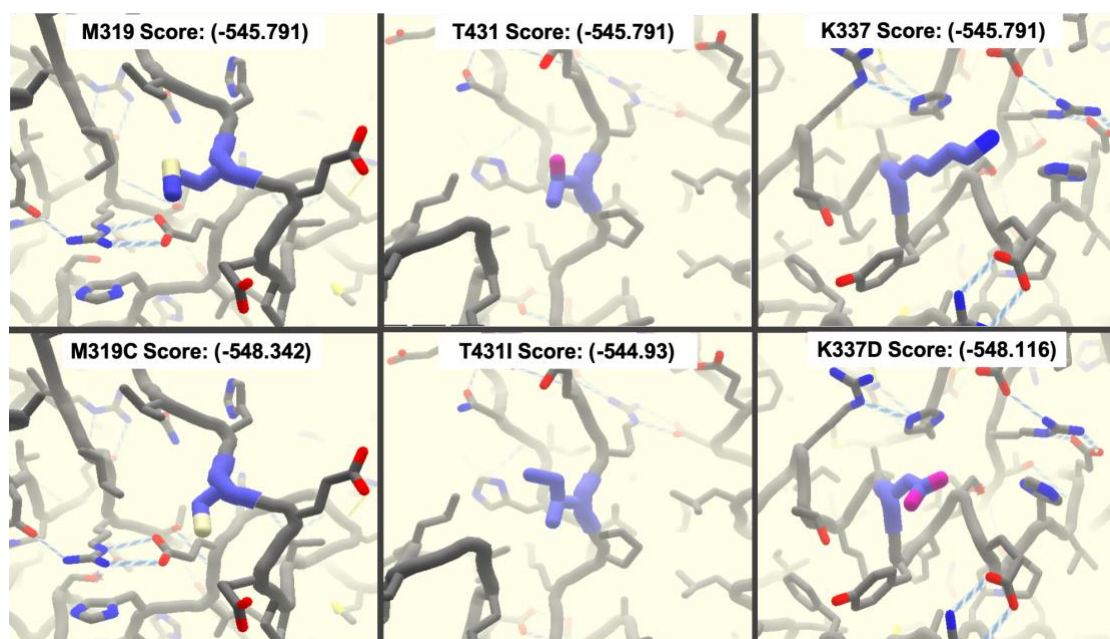
Figure 2. Molecular modeling of the three variants. Using the FoldIt modeling tool, the three mutants: M319C, T431I, and K337D are shown from left to right, respectively. Top row shows each mutational sites of the wild type with its initial total system energy score (TSE of -545.791). The bottom row shows the new molecular environment and TSE upon mutation.

Protein expression and purity. All protein samples were able to express as soluble proteins as shown in Fig 3. Using 4-12% SDS-PAGE gel, the purity of all samples was greater than 80% for kinetic characterization and thermal stability assay.
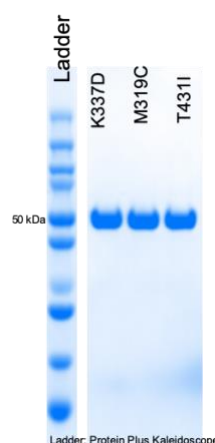

Figure 3. Purity and expression of three variants. The purity of all protein samples were evaluated SDS-PAGE. All samples are showing at 50 kDa from left to right: K337D, M319C, and T431I.

Michaelis-Menten kinetic parameters. The Michaelis-Menten parameters (k_{cat}, K_M, and k_{cat}/K_M) for the three mutants were measured using the reporter substrate p-nitrophenyl-beta-D-glucoside (pNPG). The colorimetric assay for each mutant was done in triplicates. In this study, the k_{cat}, K_M, and k_{cat}/K_M for wild type are 1070 ± 10 Min^{-1}, 6.5 ± 0.2 mM, and 165 ± 35 mM^{-1} Min^{-1}, respectively. All mutants exhibit no significant changes in catalytic efficiency compared to WT (SI Table 1). No significant change in K_M was observed for the mutants with 4.3mM, 5.1mM, and 6mM for K337D, T43ID, and M319C respectively (Fig 4). The kinetic constants between wild type and all three mutant proteins are consistent with modeling predictions of being distal from the active site and relatively small changes in the TSE.
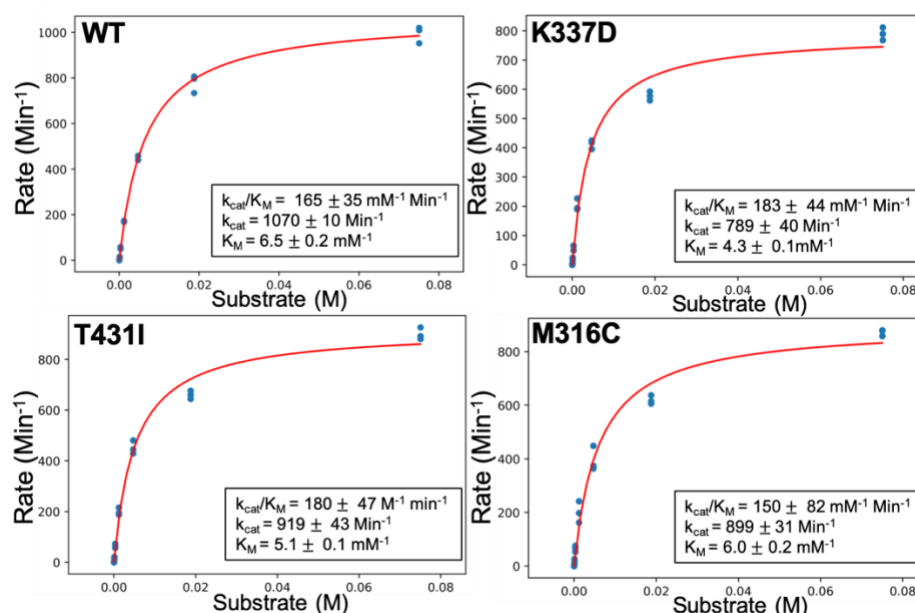
Figure 4. Michaelis-Menten constants (k_{cat}, K_M, and k_{cat}/K_M) of the WT and three mutants. All distal mutants showed no significant changes in kinetic activity compared to wild type which is consistent with modeling predictions.

Thermal stability of the designed mutants. The observed T_M for both M319C (45.8 ± 0.2 °C) and T431I (43.6 ± 0.1) °C remained fairly close to the WT (45.3 ± 0.4 °C), both having less than a 2 °C change (Fig 5). The mutant with the most drastic change in thermal stability is K337D (38.8 ± 0.8 °C) with roughly a 6 °C decrease from WT. Based on the Foldit simulations both M319C and K337D mutants were hypothesized to be slightly more stable than WT, while T431I was predicted to be less stable due loss of hydrogen bonding interactions. These predictions are consistent with the experimental observations that M319C will have a slight increase in stability, while T431I is slightly less stable than WT. However, for mutant K337D, the prediction did not match with experimental T_M result since the mutant was hypothesized to have a small increase in stability.
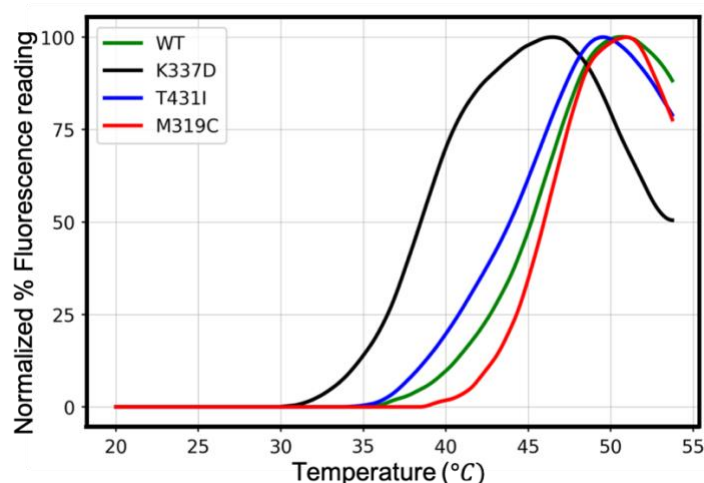
Figure 5. Temperature melting curves for WT and three mutants. The normalized percent fluorescence readings of the melting curve using the dye-based Protein Thermal Shift assay. The green, black, blue, and red depicts the average data of three replicates observed for WT, K337D, T431I, and M319C, respectively.

CONCLUSION

The three mutants in this study were all chosen on the surface of the protein and distal from the active-site. These variants did not exhibit unexpected kinetic activity changes compared to WT, which supports our

hypothesis due to the mutants being surface residues. Furthermore, the stability changes predicted for two mutants from FoldIt-based molecular modeling was also in agreement with measured changes in thermal stability, albeit relatively small changes. Unexpectedly, the FoldIt predictions did not match with the thermal stability results for K337D. A potential reason for the inaccurate predications could be due to the mutational changes causing an electrostatic repulsion with another nearby aspartic acid residue. The electrostatic term within the Rosetta energy function is a statically-derived approximation[14] and therefore for the specific environment of K337D effect of this interaction may have been underestimated.

Overall, this data supports efforts to focus on mutations in either the core or interface of the enzyme-substrate complex to obtain the most significant changes. However, for subtle improvements, optimization of surface residues should not be overlooked as the additive effects of M319C could potentially increase thermal stability by almost 1°C. More importantly, these mutants continue to add to the Design to Data efforts to quantitatively catalog the effects of a mutation on BglB in a consistent and systematic fashion. Datasets such as these have the potential to be the foundation for developing new algorithms that accurately predict a mutations' effect on protein thermal stability and kinetic activity.

REFERENCES

(1)   Borgo, B.; Havranek, J. J. Automated Selection of Stabilizing Mutations in Designed and Natural Proteins. *Proc. Natl. Acad. Sci.* **2012**, *109* (5), 1494–1499. https://doi.org/10.1073/pnas.1115172109.

(2)   Lehmann, M.; Loch, C.; Middendorf, A.; Studer, D.; Lassen, S. F.; Pasamontes, L.; van Loon, A. P. G. M.; Wyss, M. The Consensus Concept for Thermostability Engineering of Proteins: Further Proof of Concept. *Protein Eng.* **2002**, *15* (5), 403–411.

(3)   Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87* (1), 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102.

(4)   Soh, L. M. J.; Mak, W. S.; Lin, P. P.; Mi, L.; Chen, F. Y. H.; Damoiseaux, R.; Siegel, J. B.; Liao, J. C. Engineering a Thermostable Keto Acid Decarboxylase Using Directed Evolution and Computationally Directed Protein Design. *ACS Synth. Biol.* **2017**, *6* (4), 610–618. https://doi.org/10.1021/acssynbio.6b00240.

(5)   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.; et al. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science (80-.).* **2010**, *329* (5989), 309–313. https://doi.org/10.1126/science.1190239.

(6)   Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; et al. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453* (7192), 190–195. https://doi.org/10.1038/nature06879.

(7)   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.; et al. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **2012**, *134* (39), 16197–16206. https://doi.org/10.1021/ja3037367.

(8)   Mak, W. S.; Siegel, J. B. Computational Enzyme Design: Transitioning from Catalytic Proteins to Enzymes. *Curr. Opin. Struct. Biol.* **2014**, *27* (1), 87–94. https://doi.org/10.1016/j.sbi.2014.05.010.

(9)   Carlin, D. A.; Caster, R. W.; Wang, X.; Betzenderfer, S. A.; Chen, C. X.; Duong, V. M.; Ryklansky, C. V.; Alpekin, A.; Beaumont, N.; Kapoor, H.; et al. Kinetic Characterization of 100 Glycoside Hydrolase Mutants Enables the Discovery of Structural Features Correlated with Kinetic Constants. *PLoS One* **2016**, *11* (1), 1–14. https://doi.org/10.1371/journal.pone.0147596.

(10)   Schrödinger, L. The PyMOL Molecular Graphics System http://www.pymol.org.

(11)   Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D. L.; Prilusky, J.; et al. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **2018**, *72* (1), 178-186.e5. https://doi.org/10.1016/j.molcel.2018.08.033.

(12)   Carlin, D. A.; Hapig-Ward, S.; Chan, B. W.; Damrau, N.; Riley, M.; Caster, R. W.; Bethards, B.; Siegel, J. B. Thermal Stability & Kinetic Constants for 129 Variants of a Family 1 Glycoside Hydrolase Reveal That Enzyme Activity & Stability Can Be Separately Designed. *PLoS One* **2017**, *12* (5), 1–13. https://doi.org/10.1371/journal.pone.0176255.

(13)   Kleffner, R.; Flatten, J.; Leaver-Fay, A.; Baker, D.; Siegel, J. B.; Khatib, F.; Cooper, S. Foldit Standalone: A Video Game-Derived Protein Structure Manipulation Interface Using Rosetta. *Bioinformatics* **2017**, *33* (17), 2765–2767. https://doi.org/10.1093/bioinformatics/btx283.

(14)   Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031–3048. https://doi.org/10.1021/acs.jctc.7b00125.

(15)   Kunkel, T. A. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Natl. Acad. Sci.* **1985**, *82*, 488–492. https://doi.org/10.1177/019263654703114907.

(16)   Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A. *Protein Identification and Analysis Tools on the ExPASy Server*; Humana Press, 2005. https://doi.org/10.1134/s0006297906060150.

SI Table 1. Total system energy (TSE), kinetic parameters, and thermal stability (T$_M$) of WT and three distal mutants.

Protein	TSE	k_{cat} (Min^{-1})	K_M (mM)	k_{cat}/K_M (mM^{-1} Min^{-1})	T$_M$ (°C)
WT	-545.79	1070 ± 10	6.5 ± 0.2	165 ± 35	45.3 ± 0.4
M319C	-548.34	899 ± 31	6.0 ± 0.5	150 ± 82	45.8 ± 0.2
T431I	-544.93	919 ± 43	5.1 ± 0.1	180 ± 47	43.6 ± 0.1
K337D	-545.12	789 ± 40	4.3 ± 0.1	183 ± 44	38.8 ± 0.8

The estimated errors are the s.d of three independent measurements.