Design to Data for mutants of β-glucosidase B from *Paenibacillus polymyxa*: L171G, L171V and L171W

Thomas To

Dr. Justin Siegel, PhD & Ashley Vater, MS.

Genome Center

Department of Biochemistry and Molecular Medicine

Department of Chemistry

University of California, Davis

Keywords Biochemistry, Bioengineering, Protein Design, Molecular Modeling, Protein Engineering, Enzyme Engineering, Rational Design, Enzyme Functionality

Abstract Employing computational tools to predict enzyme stability and catalytic efficiency is a growing method in protein engineering. To improve the predictive accuracy of enzyme modeling software, high-quality libraries of Michaelis-Menten and thermal stability (T_M) data are needed for training the algorithms. As of 2019, 129 mutants have been successfully characterized with Michaelis-Menten, thermal stability (T_{50}) and expression constants of ß-glucosidase B from *Paenibacillus polymyxa*. To contribute to the expansion of this dataset, we designed, built and tested three-point mutations: L171G, L171V and L171W. The enzyme variants were modeled with FoldIT software, built using Kunkel Mutagenesis methods in *Escherichia coli*, and the purified proteins were tested for kinetic activity and T_M. This variant set was produced to investigate changes in amino acid size on overall structural stability by mutating non-polar, small amino acids to larger ones. The addition of these mutations to the Design2Data (D2D) Course-based Undergraduate Research Experience database contribute to improved understanding of the structure-function relationship of ß-glucosidase B and expand the potential for improving the accuracy of computational modeling tools for protein design.

Submitted: 30 October 2020

I. Introduction

Enzymes are a class of proteins that facilitate biochemical reactions.[1] By engineering enzymes, it is possible to design for novel or improved biocatalysts through the use of computational modeling methods. However, despite the success of computational enzyme designs,[2–9] the current computational modeling algorithms for designing mutants commonly have a low success rate in modeling the intended function.[10] In a previous study characterizing 100 mutant variants of β-glucosidase B (BglB) from *Paenibacillus polymyxa*, suggests improvements of the modeling algorithm may be accomplished by supplementing structural parameters with empirically determined functional data on catalytic efficiency (k_{cat}/K_M), apparent substrate affinity (estimated by K_M), and turnover rate (k_{cat}).[10] This dataset became the basis for the Desing2Data (D2D) database. To date, the D2D database consists of over 129 BglB mutants that have been characterized for T_{50}, k_{cat}, K_M, and k_{cat}/K_M. The continued expansion of this structural-functional database aims to aid in improved predictive computational model accuracy of BglB and might provide insight to enzymes across the family.[11]

Here, three single-point mutant variants of BglB: L171G, L171V and L171W, were characterized with a design-build-test methodology to acquire functional data and contribute to the D2D database[12]. The mutants are first designed and visualized using FoldIT[13] and overall stability was quantified through a measure of total system energy score (TSE)[14], mutant genes were integrated into pET plasmids using Kunkel Mutagenesis[15], and the purified proteins were tested for kinetic activity and T_M. In this study, the variant set was produced to investigate changes in amino acid size on overall structural stability by mutating non-polar, small amino acids to larger ones. Given the trends we've seen on the 171 site in the existing dataset, solely based on the space filling trend of increasing non-polar amino acids, we hypothesize: L171G and L171V will have an decrease in kinetic activity and increased thermostability, and conversely L171W will have increased kinetic activity and decreased thermostability relative to wild type.

This narrative reports only on results that capture progression in the workflow, but it fails to meet the project objective and does not include the functional outcomes of this set of mutations.[†]

II. Methods

The design-build-test methodology begins by using a previously described method[10] with FoldIT[13] to score the designed mutant's energy function and TSE using the Rosetta energy function[14]. After visually reviewing the structure, we obtained the nucleotide sequence encoding the single point mutations flanked native sequence to create a 33mer oligo L171V, L171W and L171G. Then, we performed Kunkel Mutagenesis to integrate the mutations into the BglB gene on a pET plamids.[15] The designed plasmids were transformed into Dh5α competent *Escherichia coli* (*E.coli*) cells and plated onto Luria-Bertani (LB) agar plates with 50 μg/mL kanamycin for overnight growth at 37 °C. Single colonies were picked and sequenced to identify mutant colonies. Sequence data was analyzed in benchling.[16] Mutant protein data bank files, produced in FoldIT[13], were further analyzed in PyMol.[17]

[†] Experimentation was halted due to COVID-19 social distancing restrictions, we expect to resume when it is safe for undergraduate students to return to lab. As such, the findings presented here are preliminary.

For sequence-verified[†], successfully built mutants, the plasmids would have been transformed, grown, expressed using a previously described method[10] then kinetic and thermal stability assays would have been done fit the Michaelis Menten model.[19,†] The kinetic characterization of the mutant strains of BglB are determined by detecting changes in the enzyme-substrate-product solution in buffer at pH 7.5 PNP over time using an A420 spectrophotometer assay.

Next, the thermal stability (T_M) for the variants were determined using a Protein Thermal Shift™ 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 by Applied BioSystem ® from Thermo Fisher.[†]

III. Results

The three mutations, L171G, L171W and L171V probe the effects of changes at location: L171 approximately 8.9 Å from the active site (**Fig 2**). Sequence verification of Kunkel Mutagenesis for these mutations (**Fig 1**) were successful for L171G and L171W. However, for L171V a mixed population of L171V (dominant) and the native L171 (minor) appear to be present indicating a contaminated sample and more isolated colonies should be sequenced. (**Fig 1**).

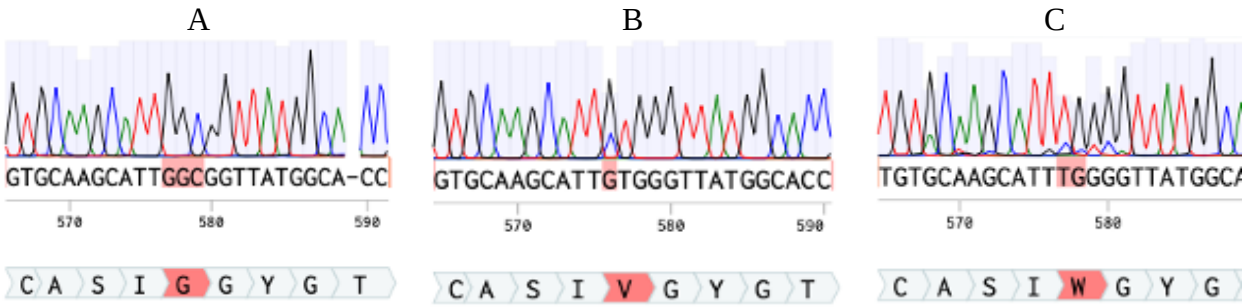

Fig 1. Chromatogram visualization of representative forward sequences of **(A)** L171G, **(B)** L171V, **(C)** L171W. Colored lines: black, blue, green and red represent Guanine (G), Cytosine (C), Adenine (A), and Thymine (T) respectively. **Fig 1A** shows clear peaks for the respected codon highlighted in red; whereas reads in **Fig 1B** indicate contamination due to overlapping cytosine peaks at the mutation site. **Fig 1C** appears to have noise (<10% intensity) from a medium quality read in the underlying minor peaks at and around the mutation site.

[†] Experimentation was halted due to COVID-19 social distancing restrictions, we expect to resume when it is safe for undergraduate students to return to lab. As such, the findings presented here are preliminary

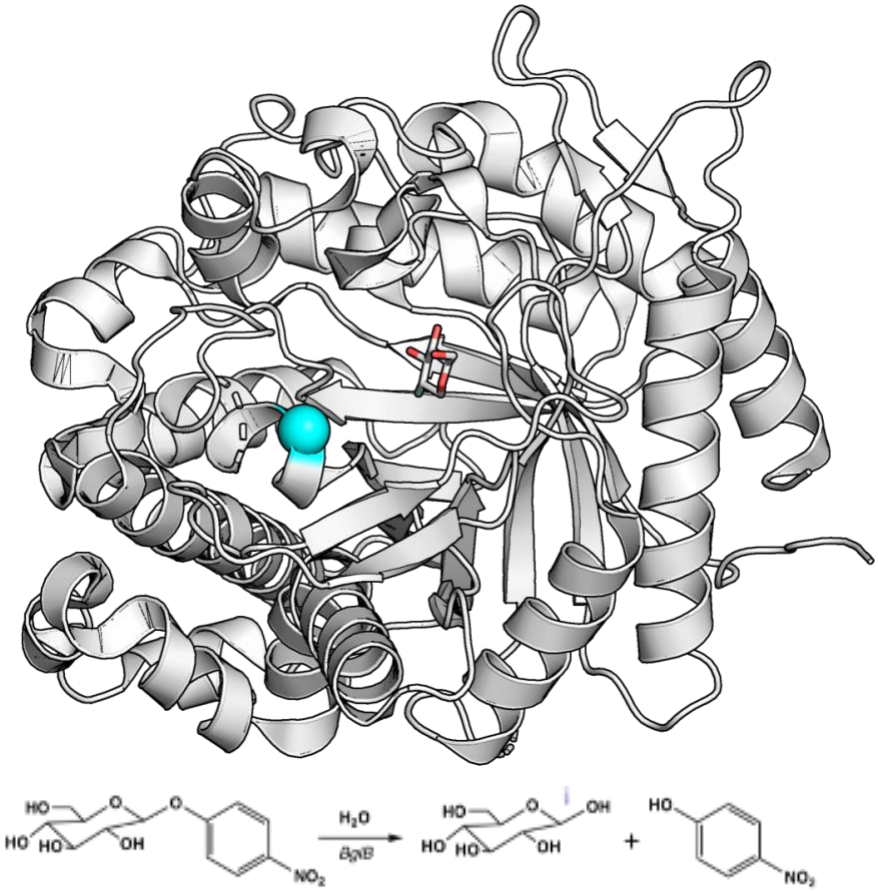

Fig 2. Overview of the modeled BglB-pNG complex (PDB ID: 2JIE) showing the L171 position in cyan using PyMOL[17]. Below, the reaction scheme of the hydrolysis of pNPG by BglB was used to determine functional T_M and kinetic parameters, k_{cat}, K_M, and k_{cat}/K_M.

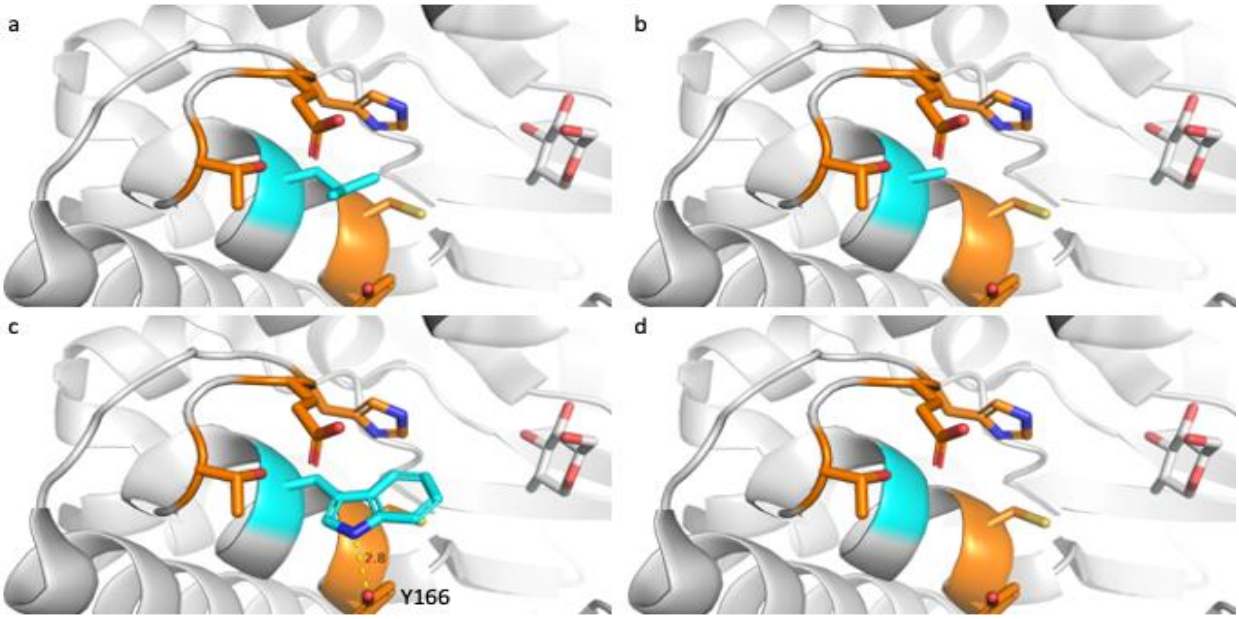

Fig 3. Overview of **a)** wild type (L171) **b)** L171A **c)** L171W **d)** L171G single point mutations of BglB-pNG complex (PDB ID: 2JIE) showing the L171 position in cyan and nearby residues in orange using PyMOL[17]. A novel hydrogen bond was formed with Y166 approximately 2.8 Angstroms from the Tryptophan (W) shown in **c)**.

IV. Discussion

In examining the sequencing data, we hypothesize the presence of an underlying cytosine peak in L171V may have resulted from contamination by reusing cuvettes containing remnant wild type double stranded DNA during Kunkel Mutagenesis. Alternatively, the upstream process of ssDNA prep may have yielded an unfavorable concentration of double stranded DNA to single stranded DNA, resulting in plasmids of both the wild type and the mutant being replicated in the transformation. Another common source of contamination may have occurred when picking the transformed colonies.

Herein we expand our predictions and hypothesis rational through deeper structural analysis, introduced earlier in the narrative and we examine previously published results that relate closely to this set of mutations (**Table 1**).

Table 1. Catalytic efficiency and thermostability characterization of L171 variants obtained from D2D[19] database. Currently, assay results on the 171 site are available for aliphatic, sulfur, amide and basic amino acid classifications but not for aromatic, hydroxyl and acidic amino acids.[20]

Variant	Expression (mg/mL)	K_M (m_M)	k_{cat} (min^{-1})	k_{cat}/K_M ($m_M^{-1}min^{-1}$)	T_{50} (^oC)
Wild Type (L171)	1.50 ± 1.07	5.35 ± 0.87	661.5 ± 48.4	130.77 ± 28.20	39.5 ± 0.1
L171A	1.00 ± 0.00	11.09 ± 0.42	662.4 ± 6.5	59.75 ± 2.38	38.8 ± 0.5
L171M	2.40 ± 0.00	4.32 ± 0.29	887.4 ± 24.3	205.64 ± 14.76	—
L171Q	1.12 ± 0.00	8.00 ± 1.15	77.2 ± 4.6	9.65 ± 1.51	—
L171R	1.00 ± 0.00	3.36 ± 0.23	331.2 ± 6.5	98.50 ± 6.93	38.8 ± 0.3

This study's mutant set: L171G, L171V and L171W aimed to investigate changes in amino acid size on overall structural stability by mutating non-polar, small amino acids to larger volume occupying ones at one site: 171. We note the hydrophobicity trend is not consistent with the volumetric trend and we challenge which trend is dominant to one another.[20]

Based on the previously published data in **Table 1** of L171, the trend for non-polar amino acids with increasing space filling effects by volume and appears to be proportional to k_{cat} but inversely proportional to K_M. Noting from a previous study, features of overall catalytic efficiencies may not be predictive of either k_{cat} or K_M independently[10] when fitted to the Michaelis-Menten model[16]. In this model, K_M approximates how readily substrates bind to the enzyme, k_{cat} quantifies the rate at which product is formed and detaches from the enzyme, and the overall catalytic efficiency is represented as k_{cat}/K_M.

We hypothesize L171G will have similar functional outcomes to L171A but with increased K_M value and a lower k_{cat}. When comparing the kinetic and thermostability assay results from L171A

to L171 (**Table 1**), there is a reduced binding efficiency, similar turnover rate and reduced overall catalytic efficiency. We expect to see the same effects in L171G relative to L171 based on the trend previously described. Conversely, for L171V, a decreased K_M and increased k_{cat} relative to L171A but not to L171. The rationale for this is due to their increasing volumetric trend[20] from G, A, V, L, M, W for nonpolar amino acids in this study. Here, the previously published assay results for k_{cat} and K_M in **Table 1** define the quantifiable lower and upper bounds to base a hypothesis for L171G and L171V.

A reasonable hypothesis was formed for L171G and L171V as L171A and L171 are both aliphatic, nonpolar amino acids. However, there has not been a ring-containing amino acid that has been characterized at this site yet. A hypothesis for L171W is not as strongly supported due to its novelty as an aromatic structure that has yet to be characterized and assayed. Although L171M and L171W are amino acids with hydrophobic side chains, a hypothesis based on the absence of ring structure and inability to form a new hydrogen bond interaction seems not strongly supported given these limitations. The previous characterization of L171Q, might lend insights as there is the potential to form a hydrogen bond on Y166, which we modeled occurring in the L171W variant. However, whether this translates, or is lost given the ring structures interactions, is difficult to predict. Solely on the basis of increasing volumetric size for non-polar amino acids from **Table 1** which suggests space-filling affects are proportional to k_{cat} but inversely proportional to K_M, we hypothesize L171W to have an increased kinetic activity and decreased thermostability relative to L171 based on the other nonpolar amino acids: L171A, L171 regardless of its aliphatic characteristic.

By focusing in on one site of the protein, we investigate the space filling effects of nonpolar structures to determine if the trend for non-polar aliphatic amino acids with increasing space filling effects are consistent with the trends shown in **Table 1.** Moreover, to introduce novel chemical class[19,] that may be useful in protein design and observe how they compare to other mutants on the same site.[20] These mutants add to the Design to Data database to quantitatively catalog the effects of a single point mutation on BglB to potentially become the foundation for developing improved algorithms that more accurately predicts a mutations' effect based on empirically determined protein thermal stability and kinetic activity.

V. Acknowledgements

This work was supported by the University of California Davis, the National Institutes of Health (R01 GM 076324-11), the National Science Foundation (award nos. 1827246, 1805510, and 1627539), and the National Institute of Environmental Health Sciences of the National Institutes of Health (award no. P42ES004699). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, National Institute of Environmental Health Sciences, National Science Foundation, or UC Davis.

VI. References

(1) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. A Relationship between Protein Stability and Protein Function. *Proceedings of the National Academy of Sciences* **1995**, *92* (2), 452–456.

(2) Borgo, B.; Havranek, J. J. Automated Selection of Stabilizing Mutations in Designed and Natural Proteins. *Proceedings of the National Academy of Sciences* **2012**, *109* (5), 1494–1499.

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

(4) Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annual review of biochemistry* **2018**, *87*, 105–129.

(5) Soh, L. M.; 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 synthetic biology* **2017**, *6* (4), 610–618.

(6) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010**, *329* (5989), 309–313.

(7) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453* (7192), 190–195.

(8) Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *Journal of the American Chemical Society* **2012**, *134* (39), 16197–16206.

(9) Mak, W. S.; Siegel, J. B. Computational Enzyme Design: Transitioning from Catalytic Proteins to Enzymes. *Current opinion in structural biology* **2014**, *27*, 87–94.

(10) 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. Kinetic Characterization of 100 Glycoside Hydrolase Mutants Enables the Discovery of Structural Features Correlated with Kinetic Constants. *PloS one* **2016**, *11* (1), e0147596.

(11) Carlin, D. A.; Hapig-Ward, S.; Chan, B. W.; Damrau, N.; Riley, M.; Caster, R. W.; Bethards, B.; Siegel, J. B. Thermal Stability and Kinetic Constants for 129 Variants of a

Family 1 Glycoside Hydrolase Reveal That Enzyme Activity and Stability Can Be Separately Designed. *PLOS ONE* **2017**, *12* (5), e0176255. https://doi.org/10.1371/journal.pone.0176255.

(12) Siegel, J.B.; Labonte, J.W.; Gray, J.J.; Vater, A.; Dolan, E. L. Design-to-Data Course-based Undergraduate Research Experience (D2D CURE). https://www.d2dcure.com/data/BglB_kinetic_constants.php

(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.

(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. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of chemical theory and computation* **2017**, *13* (6), 3031–3048.

(15) Kunkel, T. A. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Proceedings of the National Academy of Sciences* **1985**, *82* (2), 488–492.

(16) Benchling [Biology Software]. **2020**. Retrieved from https://benchling.com.

(17) DeLano, W. L. The PyMOL Molecular Graphics System. *http://www. pymol. org* **2002**.

(18) Wiederschain, G. Y. The Proteomics Protocols Handbook. *Biochemistry* **2006**, *71* (6), 696.

(19) Roskoski R. Michaelis-menten kinetics. In: XPharm: The Comprehensive Pharmacology Reference. ; 2007. doi:10.1016/B978-008055232-3.60041-8

(20) Pommié, C., Levadoux, S., Sabatier, R., Lefranc, G. and Lefranc, M.-P. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. Journal of Molecular Recognition, 17(1):17-32 (2004). DOI:10.1002/jmr.647. PMID:14872534.