Protein engineering and kinetic characterization of a family 1 glycoside hydrolase mutant library

## Introduction

Family 1 glycoside hydrolases have broad substrate specificity for ß-glycosidic linkages and are essential to the life of every organism. They also find industrial/commercial use in the hydrolysis of plant biomass into glucose as the final step in a pathway that breaks down complex carbohydrates into simple sugars. This family of enzymes has been the subject of numerous structural and kinetic studies elucidating the catalytic mechanism. [Isorna]

BglB, the ß-glucosidase used from this study, is from *Paenibacillus polymyxa*, a motile, Gram-positive bacterium found in soils and marine sediments. An X-ray crystal structure of BglB shows that BglB follows a classical Koshland double-displacement mechanism in which Glu 164 performs a nucleophilic attack on the anomeric carbon of the substrate’s glucose moiety. The leaving group (in natural systems, another sugar) is protonated by Glu 353. Tyr 295 orients the Glu 353 for catalysis with a hydrogen bond. [Isorna] In our system, the reporter substrate 4-nitrophenyl-ß-D-glucoside is used.

**Figure 1.** Scheme showing the reaction catalyzed by BglB on 4-nitrophenol-ß-D-glucoside.

**Figure 2**. Multiple sequence alignment of 1,558 sequences annotated as family 1 glycoside hyrolases in the Pfam database, numbered according to the BglB reference sequence.

Here, we report data for 104 single mutants of BglB comprising a set of computationally-designed mutations in the active site and a systematic alanine scan of the protein-ligand interface (composed of residues within 12 Å of the ligand).

## Results

### Computationally-directed engineering of BglB

A macromolecular model of BglB constructed based on PDB entry 2JIE was engineered using a graphical front-end to Rosetta. Hypothesis-driven mutations were designed and scored using Foldit, with the requirement that the overall protein energy remain the same or improve with the mutation. No other restrictions were placed on engineered mutations.

### Protein yields

Previous studies using random mutagenesis have reported that about 90% of random mutants produce insoluble protein [cite]. In contrast, our study, in which mutants were generated within the context of a physically-realistic scoring function, greater than 95% of designed structures expressed active, soluble protein. The average yield in our study was 0.7 mg/mL.

### Kinetic characterization of mutants

Michaelis-Menten kinetic constants for 104 mutants comprising a systematic alanine scan of the protein-ligand interface, together with a set of computationally-designed mutants, were obtained to under 25% error using a colorimetric assay. (For mutants with error > 25% in kcat or KM, kcat/KM was determined via a linear fit.) Ten biological replicates of the wild type enzyme had an average yield of 0.5 mg/mL, an average kcat of 880 M/min, an average KM of 5 mM, and kcat/KM of 171,000 M/min. Our limit of detection for kcat/KM was 10 M/min, and the highest kcat/KM observed was 564,000 M/min (mutant R240A). An additional constant, *ki*, is reported for 4 mutants displaying substrate inhibition.

**Figure 3.** Heatmap of kinetic constants. Data are scaled relative to the wild type enzyme and ordered by sequence position.

## Discussion

### Effect of mutations on catalysis

Many of the mutations, particularly those in which a conserved catalytic residue is mutated to alanine, have the readily predictable effect of no detectable enzymatic activity, with overall catalytic efficiencies below 10 M/min compared to the wild type value of 171,000 M/min. Some mutations, however, reveal surprises in how the structure of this enzyme determines its function.

For example, the mutation E222A results in a 15,000-fold decrease in *kcat*, yet glutamate appears in this position in less than 5% of sequences in the of the Pfam database’s entry for family 1 glycoside hydrolases. It is interesting to note that the KM for this mutant is 10-fold lower than wild type, indicating greater affinity between the substrate and active site pocket. Why Glu 222 proved to be essential for catalytic hydrolysis of 4-nitrophenyl-ß-D-glucoside is unclear, yet this finding lends credence to the view that enzymes must trade stability for function [Bloom].

As another example, the mutant R240A, which is predicted to stabilize an alternate conformation of Glu 222 which brings the carboxylic acid moiety to within 4.5 Å of the leaving group from an original distance of 6.5 Å, has a *kcat*which is 100-fold faster than WT, and a *kcat*/*KM* 10-fold greater. While a crystal structure will be necessary to confirm this hypothesis, our molecular modeling is consistent with our experimental results in establishing a critical role for Glu 222. It is unclear and perhaps counterintuitive why the interaction between Glu 222 and the nitro group of 4-nitrophenol stabilizes this species as a leaving group and increases the rate of catalysis.

**Figure 4.** The active site of wild type BglB and two mutants, E222A and R240A. The protein with all side chains is shown in white. The substrate, 4-nitrophenol-ß-D-glucoside, is shown in grey. Catalytic residues (navy) and mutant residues (gold) are shown as thick sticks.

### Variety of effects observed

The variety of differences we see between mutants is very interesting. From the systematic alanine scan of the protein-ligand interface alone we learned that a single point mutation can knock out enzymatic activity (E164A), completely destabilize the protein or prevent its folding (W34A), greatly reduce kcat while leaving KM unchanged (N354A), or destabilize the protein-ligand complex while lowering the catalytic activation energy (S17A). [Need transition sentence.]

### Conclusion

In this work, over 100 computationally-designed mutants of a family 1 glucosidase are kinetically characterized to under 25% error. To the best of our knowledge this is the largest data set of mutant enzymes produced and kinetically characterized in a uniform manner. While no specific physical features were being probed, we have uncovered a significant functional importance of non-conserved amino acids for the tested substrate through this broad search. We believe this type of data set will be invaluable for the future development of computational enzyme engineering algorithms.

## Materials and Methods

### Molecular modeling and design

Molecular models of ß-glucosidase B were created from PDB ID 2JIE, Uniprot P22505, a recombinant protein crystallized with the substrate analog 2-deoxy-2-fluoro-alpha-D-glucopyranose and solved to 2.3 Å [Isorna]. A transition state model of 4-nitrophenyl-ß-D-glucoside was created in Spartan and docked into the crystal structure using functional constraints imposed by the enzyme mechanism. The Rosetta front-end Foldit was used to design *in silico* mutations.

### Mutagenesis, expression, and purification

Sequence-verified plasmids constructed via Kunkel mutagenesis (Transcriptic) were transformed into *E. coli BL21(DE3)* (Novagen). Small-scale (5 mL) cultures were grown to saturation, pelleted, and resuspended in induction medium containing 0.1 M IPTG (Sigma). Cultures were allowed to express at 18 C for 24 hours before harvesting. Proteins were purified via immobilized metal ion affinity chromatography (50 uL bed volume of Ni-NTA (HisPur) resin, wash buffer: 15 mM imidazole, 50 mM HEPES, 150 mM sodium chloride) and eluted in four bed volumes (elution buffer: 25 mM EDTA, 50 mM HEPES, 150 mM sodium chloride). Protein concentration was determined by the ratio of absorbance at 260 nm and 280 nm (BioTek Epoch) and confirmed by SDS-PAGE (Life).

### Assay and data analysis

Mutant proteins were assayed in 100 uL reactions containing 50 mM HEPES, 150 mM sodium chloride, 25 mM EDTA, and serial dilutions of substrate at concentrations from 75 mM to 0 mM. Rate of product (4-nitrophenol) formation was monitored by the absorbance at 420 nm. The observed rates were to fit to the Michaelis-Menten equation using SciPy [cite].

## References

[Do not fill out by hand]

## Supplemental information

[Just a list of what we’re including for now.]

1. Complete table of kinetic constants for all mutants
2. Diagnostic plots for each mutant
3. Score table from MSA for position 222