Appendix B

Using computational library design to alter the specificity of a xylanase

The work described here was carried out in the Mayo lab. All TAX activity screening and kinetic assays were carried out by Dr. Roberto Chica.

Abstract

Computational library design allows us to explore additional highly ranked sequences in the energy landscape near the minimum energy conformation that was predicted by a protein design calculation. Here, our goal was to use computational library design to adjust the specificity of a xylanase that exhibits a broad specificity for monosaccharide and disaccharide substrates. So far, none of our designs have exhibited an increase in specificity for the two substrates tested. However, this project is ongoing and calculations are currently being carried out on additional substrates.

Introduction

Computational library design is a special application of computational protein design discussed in Chapter I. Instead of identifying a single best sequence, the goal of computational library design is to specify a library of sequences that represents a group of top-scoring sequences from the computational design calculation and can be encoded by a single degenerate codon at each design position. This protein design procedure called Combinatorial Libraries Emphasizing and Reflecting Scored Sequences (CLEARSS) was developed in the Mayo lab and is advantageous over existing computational library design software because it ranks possible libraries via energy calculations of every member of each library and allows direct control over the size of the library.[1,2]

The relationship between protein structure and function is not completely understood. Consequently, many approximations are used in the protein design algorithm, including a fixed backbone structure, discrete sidechain rotamers, an energy function optimized for the stabilization of small proteins, and heuristic models of protein function that are necessary for incorporation into the force field. The screening of a large number of highly ranked protein sequences identified by computational library design can compensate for these limitations by allowing us to examine a larger portion of sequence space instead of a single sequence, improving our chances of finding a sequence that incorporates the function of interest into the protein.[3]

The xylanase from the thermophilic fungus *Thermoascus aurantiacus* (TAX) is a convenient system for testing computational library design procedures because it has been well studied with established mechanism, specificities, and kinetics (Table B-1, Figure B-1).[4,5] Glycosidase activity can be readily monitored using commercial

fluorogenic or chromogenic substrate analogs (Figure B-2). TAX has been recombinantly expressed in high yield in *Escherichia coli* and can be purified to homogeneity with a single affinity chromatography step (see Appendix D). Also, a high-resolution crystal structure of TAX bound to xylobiose is available at 1.7 Å resolution (PDB: 1GOR).[5]

The specificity of TAX is highly stringent in some respects and broad in others. TAX can hydrolyze *para*-nitrophenol (*p*NP)-derivatives of many pentose and hexose mono- and disaccharides, such as glucose and xylose; however it is inactive on others, such as mannose and maltose (Table B-1).[5] Upon modeling these inactive sugars into the active site of TAX, it is evident that steric clashes are the usual cause of the inactivity (Figure B-3). For example, glucopyranoside may be less active than xylopyranoside because the hydroxymethyl on C5 of glucose is expected to clash with W275. Mannopyranoside may be less active than glucopyranoside because the inverted stereochemistry at C2 causes the hydroxyl to clash with W84, N126, and E127. Maltose appears to be inactive because the $\alpha(1\rightarrow4)$ bond causes the second glucose to adopt a conformation that is almost perpendicular to the binding pocket. In general, disaccharides are more active than monosaccharides because they have an increased number of binding interactions to compensate for distortion in the transition state (TS).

In the work described here, we used CLEARSS to modify the specificity of TAX. In addition to evaluating the effectiveness of the library design calculation methods, these experiments were designed to elucidate the exact requirements for specificity in the TAX active site, especially the primary binding interactions that are critical for substrate

binding. None of the designs to increase the specificity for xylose or glucose have been successful so far; however, this project is ongoing.

Materials and Methods

Carbohydrate structures

The structures of the carbohydrate substrates were generated using the molecular modeling program BIOGRAF[6] to modify the xylobiose structure found in the crystal structure from Lo Leggio *et al*. (PDB: 1GOR).[5] The xylobiose structure was appropriately modified, hydrogen atoms were added, and the resulting structure was minimized for 50 steps. Charges were calculated using an electrostatic potential fit, methanol solvation including a dielectric constant of 33.62, and a hybrid density functional B3LYP as implemented by the Jaguar 5.5 software package and a 6-31G** basis set.[7] Rotamers were created using canonical torsions (60°, 180°, and 300°) for each rotatable bond in the region of carbohydrate that differs from xylobiose (e.g., the C5 hydroxymethyl group on glucose). The initial position of each carbohydrate structure within the TAX active site was determined by overlaying with the crystallographic xylobiose.

Phoenix calculations

Protein design calculations were carried out with Phoenix[1,2,8] using xylose- or glucose-based intermediates. For the xylose calculations, only Trp and Phe were allowed at position 275 to account for the results of the site-saturation mutagenesis, which showed

that only these two active residues at this position result in an active enzyme. Other design positions were 90 and 276. These positions were allowed to sample all rotamers of all amino acids except Pro. Residues in the immediate area of the design residues were allowed to change conformation but not identity (46, 47, 50, 83, 84, 87, 89, 130, 172, 207, 209, 239, and 267). The nucleophile of the reaction, E237, was required to be Ala to prevent steric clashes that could result from the close proximity of the nucleophile and the TS in the active site in the absence of a covalent bond.

An occlusion-based solvation potential was applied with scale factors of 0.05 for nonpolar burial, 2.5 for nonpolar exposure, and 1.0 for polar burial.[8] Other standard parameters were applied as in Lassila *et al.,* and a backbone-independent conformer library was used to represent side-chain flexibility.[9] The ligand was allowed to translate ± 0.2 Å in every direction in 0.2 Å steps and rotate ± 5° in every direction in 5° steps. As in the enzyme design calculations (Chapters II, III, IV, and Appendix A), geometric constraints were imposed to preserve important contacts between the intermediate and the active site sidechains. In this case, these contacts are wild-type ligand-binding contacts that are found in the crystal structure of TAX and are described in terms of distance only (Figure B-4).

To preserve the contact between xylose and K50, the Nζ of a Lys residue was required to be between 3.0 and 3.5 Å from O4 of xylose and between 2.5 and 3.0 Å from O3 of xylose. For the contact between H83 and xylose, the Nε of a Hid residue was required to be between 2.5 and 3.5 Å from O2 and O3 of xylose. Finally, the Nδ of Asn was required to be between 2.5 and 3.0 Å from O2 of xylose to preserve the contact between xylose and N172.

As in previous enzyme design calculations, sidechain-ligand interaction energies were biased to favor those contacts that satisfy the geometries. Sequence optimization was carried out with FASTER,[10,11] and a Monte Carlo-based algorithm[12,13] was used to sample sequences around the minimum energy conformation from FASTER (FMEC).

For the glucose calculations, hydrophobic residues (Ala, Ile, Leu, Val, Phe, Tyr, and Trp) were allowed at position 275 and nine internal rotamers of glucose were used to represent flexibility in the hydroxymethyl group. Otherwise, the glucose calculations were carried out in the same manner as the xylose calculations.

Library design calculations

Library design calculations were also carried out in Phoenix. The size of the libraries was set to 120 and the wild-type residue was required to be a member of the library at each design position.

Library construction

Construction of the libraries was carried out using splicing by overlap extension (SOE) mutagenesis to introduce degenerate codons at the design positions.[14] The site-saturation mutagenesis primers that were used are listed in Table B-2 (Integrated DNA Technologies). These primers were designed with the same rules as those for site-directed mutagenesis described in Chapter III, Materials and Methods. Mutagenesis primers for the xylose and glucose libraries were identical except for the replacement of the NNS site-saturation codon with the degenerate codon from the library design calculation (Table B-3). The mutagenesis fragments were constructed by combining 170

ng of template DNA (TAX-pET11a, described in Appendix D), 10x Thermopol buffer (New England Biolabs), 1 U Vent DNA polymerase (New England Biolabs), 0.5 mM dNTP mixture, 25 μmol forward or reverse mutagenesis primer, and 25 μmol forward or reverse flanking primer (TAX_NdeI_forward or TAX_BamHI_reverse) in a total reaction volume of 100 μL. The reactions were carried out on Mastercycler Personal Thermocycler (Eppendorf) using the temperature program described in Table B-4. The gene fragments were run on a 1% agarose gel and the bands were extracted and purified using a QIAquik Gel Extraction Kit (Qiagen).

Assembly reactions were carried out using an equimolar mixture of the forward and reverse fragments (200 ng total) and 10x Thermopol buffer (New England Biolabs), 0.5 mM dNTP mixture, 25 μmol forward and reverse flanking primers. and 1 U Vent DNA polmerase (New England Biolabs) in a total of 100 μL. The temperature program used for the PCR is described in Table B-4.

After amplification, the reactions were purified with a QIAquik PCR Purification Kit (Qiagen) then digested with *Bam*HI and *Nde*I (New England Biolabs). The digested genes were then run on a 1% agarose gel and the bands were extracted and purified using a QIAquik Gel Extraction Kit (Qiagen). The digested genes were then ligated into a similarly digested pET11a vector (Novagen). The libraries were transformed into *E. coli* BL-21 Gold (DE3) ultracompetent cells (Stratagene) and individual colonies were sequenced to confirm an adequate distribution of amino acids at each site (Agencourt).

TAX expression

Colonies were picked into 250 μL LB/ampicillin supplemented with 10% glycerol in 96-well plates and grown overnight at 37°C with shaking. These pre-cultures were used to inoculate 300 μL cultures in Overnight Express Instant TB media (Merck Biosciences), which were grown overnight at 37°C with shaking. The cells were pelleted, washed with phosphate buffered saline (PBS), and frozen on dry ice. The pellets were thawed at 30°C and resuspended with lysis buffer (1x CelLytic B (Sigma-Aldrich), PBS pH 7.4, 1 mg/mL hen egg white lysozyme (Sigma-Aldrich)). Lysis was carried out at 30°C for 30 min with shaking and the lysate was centrifuged for 15 min at 4°C at 3000 × g.

TAX activity assays

200 mM stocks of 4-methylumbelliferyl-β-D-glucopyranoside (MUG) and 4-methylumbelliferyl-β-D-xylopyranoside (MUX) were made in *N,N*-dimethylformamide (DMF). For the assays, the stocks were diluted to 10 mM MUG or MUX in citrate-phosphate buffer, pH 5.0 with a total DMF concentration of 5%. For end-point screening assays, 20 μL of cell lysate supernatant was added to 80 μL of the buffer with substrate in black 96-well microtiter plates with clear bottoms (Greiner). The plates were incubated for 30 to 60 min at 40°C and the fluorescence intensity was measured at λ_{em} = 445 nm (λ_{ex} = 360 nm) using a Safire^2 microplate reader (Tecan). Variants that exhibited end-point fluorescence intensity at least 1.5 times that of the median fluorescence value of the plate were considered to be active.

For kinetic analysis of selected variants, protein expression, purification, and concentration determination were carried out as for the TAX-based designs described in Chapter IV, Materials and Methods. Dilution series of MUX and MUG were made by serial two fold dilutions into DMF starting with a 250 mM stock solution in DMF. Final substrate concentrations ranged from 98 μM to 12.5 mM. In a clear bottom 96-well plate, 180 μL 50 mM citrate-phosphate buffer, pH 5.0 was combined with 10 μL substrate stock in DMF. The reaction was initiated by the addition of 10 μL TAX to a final concentration 30 μM. The release of 4-methylumbelliferone was monitored by an increase in fluorescence intensity at 445 nm with a λ_{ex} of 360 nm. The conversion factor for arbitrary fluorescence units (AFU) to concentration of 4-methylumbelliferone was determined to be 120 ± 23 AFU/μM by a standard curve.

Results and Discussion

The goal of our calculations here was to increase the specificity of TAX for both xylose and glucose individually. The design positions for these calculations were determined based on an overlay of the xylobiose and cellobiose-bound crystal structures of two xylanases (Figure B-5).[5,15] The positions with the most deviation in their conformations were identified as Q90, W275, and R276. To determine the tolerance of TAX for mutations in the active site, site-saturation mutagenesis was performed at these three sites. TAX activity assays on MUX indicated that in the case of xylose, only Phe and Trp are tolerated at position 275. At position 90, Gln (wild-type), Ser, and Arg are the residues that confer the most activity, but most other residues are also tolerated.

Position 276 shows a high tolerance for mutation with the greatest activity resulting from Arg (wild-type), Val, Leu, and Gly.

The results of the site-saturation mutagenesis suggest that our xylose libraries should be limited to Phe and Trp at position 275, and all residues should be sampled at the other two positions. Because no degenerate codon encodes just Phe and Trp, two separate calculations were carried out: one with Phe only at position 275 (Xyl1-Phe) and another with Trp only at 275 (Xyl2-Trp). In the library design calculations, the size of these two libraries was reduced to 60 so that the total size of the combined xylose libraries would be 120. Unfortunately, the signal for the assays of the MUG substrates was too low to get any meaningful site-saturation data. For the glucose calculations, we allowed only hydrophobic residues at position 275 to allow room for the C5 hydroxymethyl group and to help shield the active site from the solvent.

The results of the library design calculations are shown in Table B-3. In each of the libraries, a majority of the library members have calculated energies that are favorable. The two xylose libraries are identical except for the residue at position 275. Thus, the combination of these two libraries results in a full-sized library with 50% Trp and 50% Phe at position 275. It is evident from these results that because of the genetic code, the degenerate codon that is selected after the library design can encode for more amino acids than were represented in the initial design calculation. These cases test our assumptions about the restrictions on the identity of a given position and allow us to determine if, for example, it was reasonable to require only hydrophobic residues at position 275 in the glucose calculation.

The diversity in the xylose libraries at position 90 includes Arg, Asn, and Gln, which were found to be highly active from site-saturation mutagenesis. In all of the libraries, the high level of diversity at position 276 agrees with the tolerance of this site as seen during the site-saturation mutagenesis.

The designed combinatorial libraries were screened in crude cell lysate in 96-well plates, and the initial rates of the hydrolysis of MUG and MUX were compared. Some trends were observed from the results of this initial screening: (1) the W275F mutation eliminates MUG hydrolysis activity in variants with this mutation. This mutation also tends to decrease the activity for MUX somewhat, resulting in an increase in specificity for MUX. (2) Mutants with the R276V mutation (without the W275F mutation), show decreased activity for both substrates, but the overall specificity for MUG is increased. (3) The R276L mutation in the absence of W275F decreases the glucose hydrolysis rate while maintaining a high xylose hydrolysis rate, resulting in an increase in specificity for MUX.

Based on initial screening results, three point mutants were chosen for further kinetic characterization: W275F, R276V, and R276L. The kinetic constants resulting from these assays are shown in Table B-5. Specificity was determined by the ratio of k_{cat}/K_m of MUX and MUG. For wild-type TAX, the ratio of k_{cat}/K_m values of MUX to those for MUG is about 9. The R276V mutation decreased k_{cat}/K_m for both MUX and MUG. The R276V and R276L mutants were expected to increase the specificity of TAX for xylose over glucose. Instead, a slight decrease in MUX specificity was observed (2.9 and 2.2, respectively), but these changes in specificity are not large enough to be considered significant. In contrast, the W275F mutation, which was expected to decrease

the specificity for MUX over MUG, caused a slight increase in the specificity for xylose to about 10. The observed changes in specificity were not large enough to be considered significant. The flexibility of the TAX active site near positions 275 and 276 may have contributed to the inability to design a Kemp elimination enzyme in the natural binding pocket, as described in Chapter III. The inherent flexibility at these positions may allow the binding pocket to adjust to fit a variety of substrates regardless of the identity at these positions, making it impossible to change the specificity for glucose and xylose by solely manipulating the residues here.

Conclusions

So far, we have not been able to significantly alter the specificity of TAX using computational library design; however, xylose and glucose are the only substrates for which TAX has been redesigned. The variation in the structures of xylose and glucose may be too subtle and correspond to a region of the protein that is inherently flexible, allowing rearrangement upon binding an alternate substrate. Additional substrates with more substantial differences in structure from xylobiose (e.g., maltose or mannose) may prove to be better substrates for a specificity switch. Cellobiose may make a better substrate for a change in TAX specificity because it has an additional glucose molecule with a hydroxymethyl group that must also be accommodated. In addition, the higher specific activity that TAX shows for many disaccharides should make screening with these substrates more consistent than with monosaccharides, which exhibit low signal-to-noise ratios.

Other protein design strategies may also be employed to change the specificity of TAX. "Negative design" is one such strategy, where in addition to designing for increased binding to one specific substrate to shift specificity, a separate calculation is carried out with the goal of decreasing binding to another, alternate substrate. The best sequence in this type of calculation would have a low energy when bound to the target substrate and a high energy when bound to the alternate substrate.

Because of the limited number of substrates tested up to this point, we cannot draw any conclusions about our computational library design procedure. However, we are encouraged by the correlation between the site-saturation mutagenesis results and the sequences predicted in the library design calculation. As a larger number of substrates with more diverse structures are used to design combinatorial TAX libraries, we will be better able to assess the effectiveness of our library design methods.

References

1. Allen, B. D. *Development and validation of optimization methods for the design of protein sequences and combinatorial libraries*. California Institute of Technology: Pasadena, CA, **2009**.

2. Nisthal, A.; Allen, B. D., (*Manuscript in preparation*). **2009**.

3. Treynor, T. P.; Vizcarra, C.; Nedelcu, D.; Mayo, S. L., Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 48-53.

4. Lo Leggio, L.; Kalogiannis, S.; Bhat, M. K.; Pickersgill, R. W., High resolution structure and sequence of *T. aurantiacus* xylanase I: implications for the evolution of thermostability in family 10 xylanases and enzymes with (β)α-barrel architecture. *Proteins* **1999**, *36*, 295-306.

5. Lo Leggio, L.; Kalogiannis, S.; Eckert, K.; Teixeira, S. C.; Bhat, M. K.; Andrei, C.; Pickersgill, R. W.; Larsen, S., Substrate specificity and subsite mobility in *T. aurantiacus* xylanase 10A. *FEBS Lett.* **2001**, *509*, 303-308.

6. BIOGRAF version 3.21, Molecular Simulations, Inc., Burlington, MA, **1992**.

7. Jaguar 5.5, Schrodinger, L.L.C., Portland, OR, **1991-2003**.

8. Chica, R.; Allen, B. D., (*Manuscript in preparation*). **2009**.

9. Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L., Combinatorial methods for small-molecule placement in computational enzyme design. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16710-16715.

10. Desmet, J.; Spriet, J.; Lasters, I., Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins* **2002**, *48*, 31-43.

11. Allen, B. D.; Mayo, S. L., Dramatic performance enhancements for the FASTER optimization algorithm. *J. Comput. Chem.* **2006**, *27*, 1071-1075.

12. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H., Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087-1092.

13. Voigt, C. A.; Gordon, D. B.; Mayo, S. L., Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **2000**, *299*, 789-803.

14. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R., Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **1989**, *77*, 51-59.

15. Notenboom, V.; Birsan, C.; Warren, R. A.; Withers, S. G.; Rose, D. R., Exploring the cellulose/xylan specificity of the β-1,4-glycanase cex from *Cellulomonas fimi* through crystallography and mutation. *Biochemistry* **1998**, *37*, 4751-4758.

Table B-1. Specific activity of TAX on 2.5 mM *p*NP-glycosides. A dash indicates that no activity was detected. The data in this table were taken from Lo Leggio *et al*.[5]

substrate	activity (U/mg)
pNP-β-D-xylopyranoside	60
pNP-β-D-glucopyranoside	12.7
pNP-β-D-mannopyranoside	-
pNP-β-D-cellobioside	1290
pNP-β-D-lactoside	281
pNP-β-D-maltoside	-

Table B-2. Mutagenesis primers for site-saturation mutagenesis libraries. The degenerate codon is indicated in red. TAX_NdeI_forward and TAX_BamHI_reverse are the flanking primers used for the mutagenesis.

name	sequence
Q90X_forward	5'-CTGGTTTGGCACAGC**NNS**CTGCCGTCTTGGGTG-3'
Q90X_reverse	5'-CACCCAAGACGGCAG**SNN**GCTGTGCCAAACCAG-3'
W275X_forward	5'-GCCGATCCTGATTCT**NNS**CGCGCATCCACTACCC-3'
W275X_reverse	5'-GGGTAGTGGATGCGCG**SNN**AGAATCAGGATCGGC-3'
R276X_forward	5'- CGATCCTGATTCTTGG**NNS**GCATCCACTACCCCG-3'
R276X_reverse	5'- CGGGGTAGTGGATGC**SNN**CCAAGAATCAGGATCG-3'
TAX_NdeI_forward	5'-GAAGGAGATATACATATGGCAGAAGCG-3'
TAX_BamHI_reverse	5'- GTTAGCAGCCGGATCCCTAATGGTG-3'

Table B-3. Designed TAX libraries. The non-standard bases in the degenerate codons represent equimolar mixtures of bases: R = A, G; M = A, C; K = G, T; S = G, C; D = G, A, T; B = G, T, C; N = A, T, G, C.

library	residue #	wild-type	sampled amino acids	amino acids in library	degenerate codon
Xyl1-Phe	90	Q	ACDEFGHIKLMNQRSTVWY	HKNQRS	MRM
	275	W	F	F	TTC
	276	R	ACDEFGHIKLMNQRSTVWY	CFGILMRSVW	NKS
Xyl2-Trp	90	Q	ACDEFGHIKLMNQRSTVWY	HKNQRS	MRM
	275	W	W	W	TGG
	276	R	ACDEFGHIKLMNQRSTVWY	CFGILMRSVW	NKS
Glc	90	Q	ACDEFGHIKLMNQRSTVWY	HIKLMNQRS	MDS
	275	W	AFILVWVY	ACFGILMPSTVW	NBK
	276	R	ACDEFGHIKLMNQRSTWVY	ACFGILMPRSTVW	NBS

Table B-4. Thermocycler temperature programs for mutagenesis reactions.

SOE mutagenesis		
Temp (°C)	**Time (min)**	**# cycles**
94	5	
94	1	15x
55	1	
72	1	
72	2	
4	hold	

Table B-5. Kinetic constants for TAX variants with MUX or MUG.

	MUX			MUG				
TAX variant	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}M^{-1}$)	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($s^{-1}M^{-1}$)	MUX/MUG k_{cat}/K_m	MUG/MUX k_{cat}/K_m
wild-type	7.5E-03	3.3	2.26	1.4E-03	0.26	8.77	8.8	0.11
W275F	4.7E-03	2.6	1.8	1.6E-03	0.62	2.88	2.9	0.35
R276V	9.9E-03	2.6	3.82	9.1E-03	1.7	2.19	2.2	0.46
R276L	2.6E-03	2.0	1.28	3.5E-04	0.13	10.02	10.0	0.10

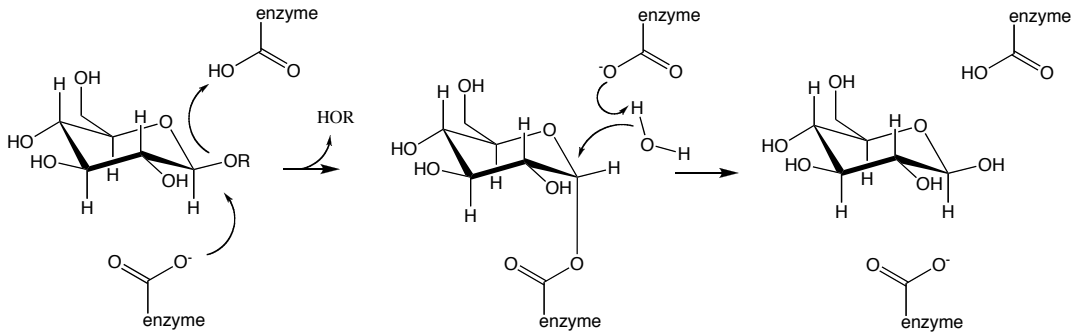

Figure B-1. Mechanism of retaining glycosidases.

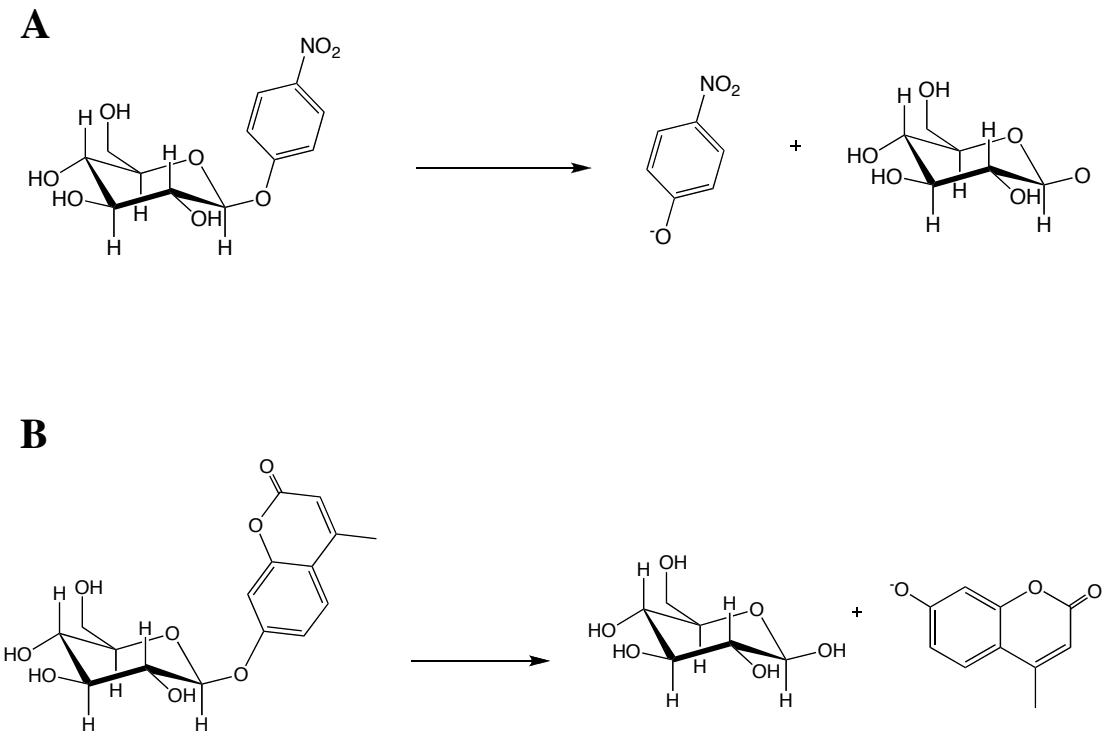

Figure B-2. Xylanase activity assays. **(A)** The hydrolysis of p-NP-glucopyranoside releases pNP, which absorbs with a λ_{max} of 410 nm. **(B)** The hydrolysis of 4-methylumbelliferyl-glucopyranoside releases 4-methylumbelliferyl, which is fluorescent with a λ_{ex} = 360 nm and a λ_{em} = 445 nm.

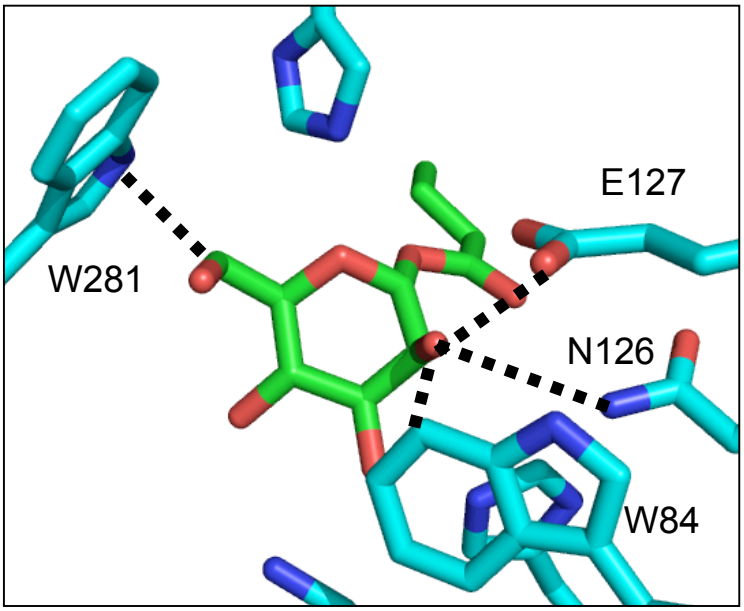

Figure B-3. Predicted clashes of mannose in the TAX active site. The modeled mannose-acyl intermediate structure is shown in green. Steric clashes are indicated with dotted lines.

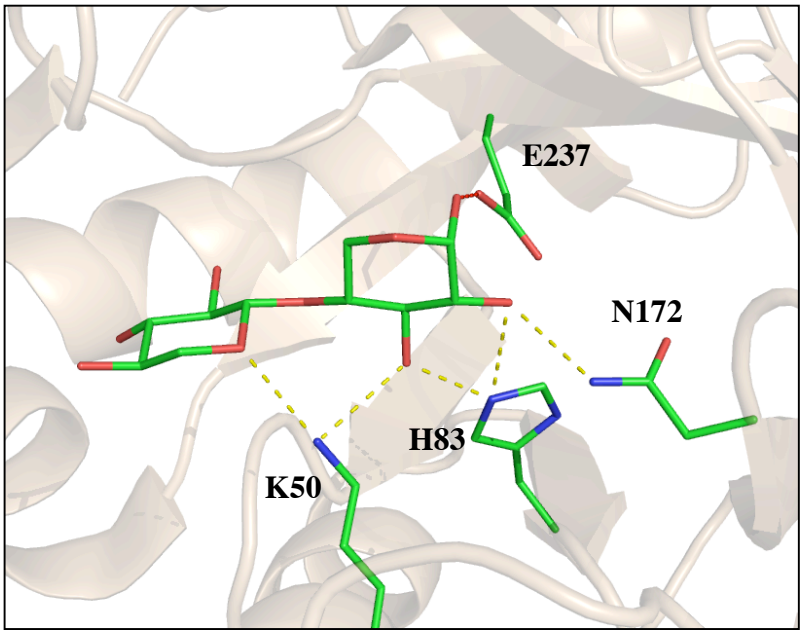

Figure B-4. Wild-type hydrogen bonds preserved in TAX calculations. The hydrogen bonds between K50, H83, N172, and the ligand are indicated with yellow dotted lines.

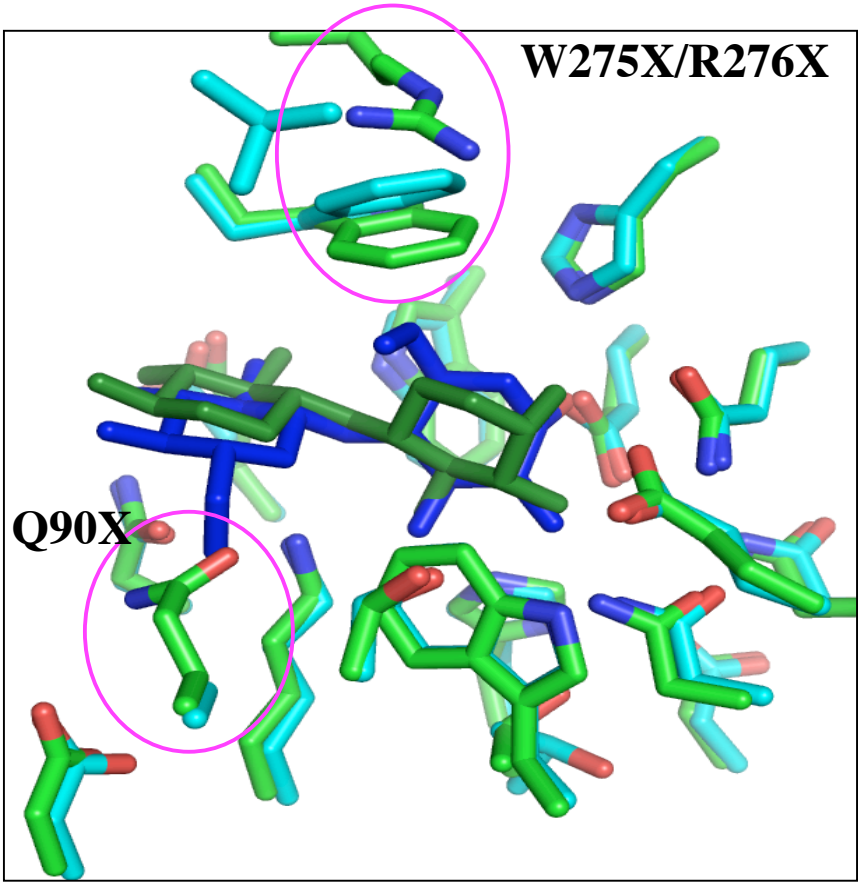

Figure B-5. Site-saturation mutagenesis positions in TAX. The *Cellulomonas fimi* xylanase is shown in cyan with the cellobiose shown in blue.[15] The TAX structure and xylobiose are shown in green.[5] Sites of site-saturation mutagenesis are indicated.