Supporting Information

for the manuscript entitled:

Exploiting topological constraints to reveal buried sequence motifs in the membrane-bound N-linked oligosaccharyl transferases

Marcie B. Jaffee and Barbara Imperiali

Department of Biology and Department of Chemistry, Massachusetts Institute of Technology,

77 Massachusetts Avenue, Cambridge, MA, 02139, USA.

S1: Identities of STT3 sequences used in all bioinformatic analysis.

Name	Description
gi|1322489|emb|CAA96722.1|	STT3 [Saccharomyces cerevisiae]
gi|11497941|ref|NP_069165.1|	transmembrane oligosaccharyl transferase, putative [Archaeoglobus fulgidus DSM 4304]
gi|15669720|ref|NP_248533.1|	putative transmembrane oligosaccharyl transferase [Methanocaldococcus jannaschii DSM 2661]
gi|87045854|gb|ABD17750.1|	STT3 [Methanococcus voltae PS]
gi|134045200|ref|YP_001096686.1|	oligosaccharyl transferase, STT3 subunit [Methanococcus maripaludis C5]
gi|150012189|gb|ABR54641.1|	Oligosaccharyl transferase STT3 subunit [Methanococcus vannielii SB]
gi|261402368|ref|YP_003246592.1|	Oligosaccharyl transferase STT3 subunit [Methanocaldococcus vulcanius M7]
gi|281485606|ref|NP_001164010.1|	STT3, subunit of the oligosaccharyltransferase complex, homolog B [Rattus norvegicus]
gi|148226196|ref|NP_001083986.1|	STT3, subunit of the oligosaccharyltransferase complex, homolog A [Xenopus laevis]
gi|211965698|gb|EEB00894.1|	oligosaccharyl transferase STT3, putative [Toxoplasma gondii ME49]
gi|121904152|gb|EAY09105.1|	Oligosaccharyl transferase STT3 subunit family protein [Trichomonas vaginalis G3]
gi|86153042|ref|ZP_01071247.1|	Oligosaccharyl transferase STT3 subunit superfamily [Campylobacter jejuni subsp. jejuni HB93-13]
gi|308159114|gb|EFO61662.1|	Oligosaccharyl transferase STT3 subunit [Giardia lamblia P15]
gi|127799903|gb|AAH48348.2|	STT3, subunit of the oligosaccharyltransferase complex, homolog A (S. cerevisiae) [Homo sapiens]
gi|55154464|gb|AAH85313.1|	STT3, subunit of the oligosaccharyltransferase complex, homolog A (S. cerevisiae) [Mus musculus]
gi|30851502|gb|AAH52433.1|	STT3, subunit of the oligosaccharyltransferase complex, homolog B (S. cerevisiae) [Mus musculus]
gi|238881972|gb|EEQ45610.1|	oligosaccharyl transferase STT3 subunit [Candida albicans WO-1]
gi|18976528|ref|NP_577885.1|	oligosaccharyl transferase stt3 subunit related protein [Pyrococcus furiosus DSM 3638]
gi|18419993|ref|NP_568380.1|	STT3A (STAUROSPORIN AND TEMPERATURE SENSITIVE 3-LIKE A); oligosaccharyl transferase [Arabidopsis thaliana]
gi|529357|gb|AAC24442.1|	Hypothetical protein T12A2.2 [Caenorhabditis elegans]
gi|253510784|gb|EES89443.1|	oligosaccharyl transferase pglB [Helicobacter canadensis MIT 98-5491]
gi|17738187|ref|NP_524494.1|	oligosaccharyl transferase 3 [Drosophila melanogaster]
gi|34556499|ref|NP_906314.1|	oligosaccharyltransferase [Wolinella succinogenes DSM 1740]
gi|307721432|ref|YP_003892572.1|	Oligosaccharyl transferase STT3 subunit [Sulfurimonas autotrophica DSM 16294]
gi|224373660|ref|YP_002608032.1|	oligosaccharyl transferase, STT3 subunit [Nautilia profundicola AmH]
gi|219868319|gb|ACL48654.1|	Oligosaccharyl transferase STT3 subunit [Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774]
gi|325066002|gb|ADY74009.1|	Oligosaccharyl transferase STT3 subunit [Desulfurobacterium thermolithotrophum DSM 11699]
gi|317114826|gb|ADU97316.1|	Oligosaccharyl transferase STT3 subunit [Thermovibrio ammonificans HB-1]

S2: Python code for programs developed to a) parse and b) organize and group sequence data according to predicted topological constraints.

a)

```
#Program 1: TMH.py
#extracts topology prediction data from TMHMM output
#takes in topology results and extracts the gi number and ranges in the sequences predicted to be TMhelix (or inside, or #outside, by changing 'TMhelix' to appropriate
S
import sys

if len(sys.argv) == 1:
	print "Error: Must input a file name"; sys.exit()
else:
	fn = sys.argv[1]
fh = open(fn)
lines = fh.readlines()
output = sys.argv[2]
sys.stdout = open(output, 'a')
seqs = []
r = ''
for line in lines:
		string = line
		if len(string) > 1:
			if string[2] == '_':
				m = 1
				splat = string.split()
				j = splat[0]
				j = j.split('_')
				gi = j[1]
				if splat[2] == 'outside':
					a = splat[3]
					a = int(a)
					a = a - 2
					if a < 1:
						a = 1
					b = splat[4]
					b = int(b)
					b = b + 2
					print gi, a, b
```

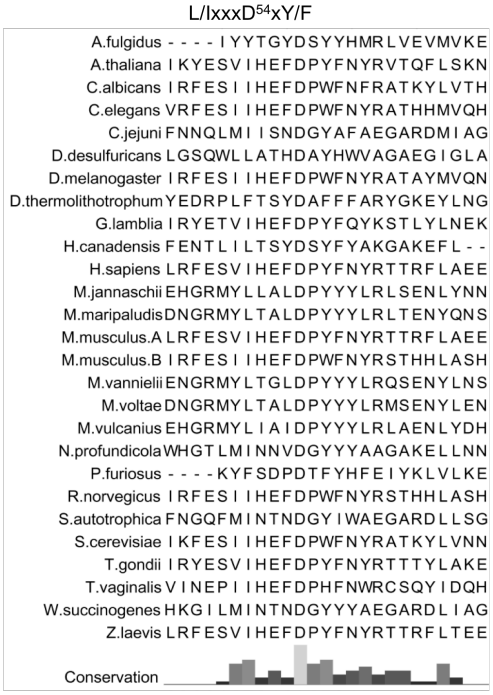
---

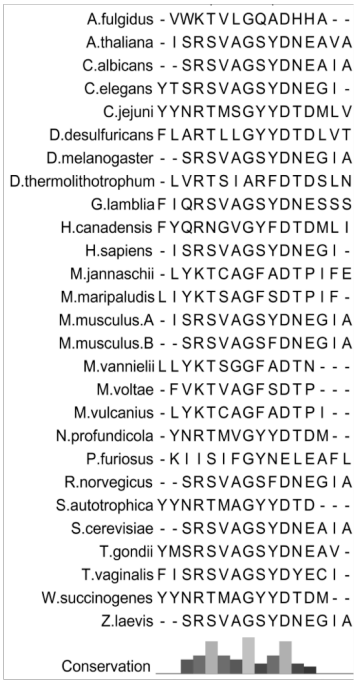
b)

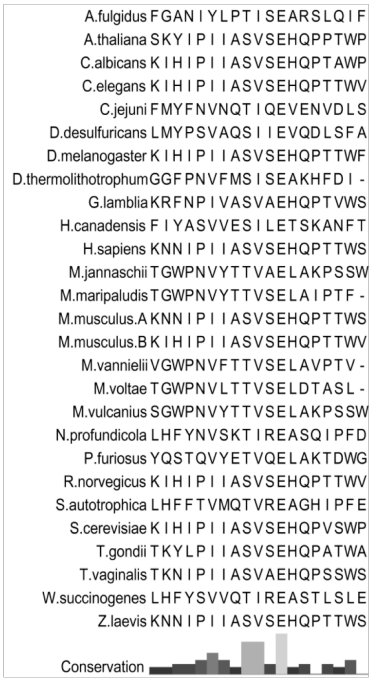
```
#Program 2: Extract.py
#first input- takes in list of fasta sequences and removes newline
#second input- file output from TMHMM program with gi and its range for a specific topological section
#output- prints the gi, the range, and the sequence corresponding to that range

import sys

if len(sys.argv) == 1:
	print "Error: Must input a file name"; sys.exit()
```

```
else:
	fn = sys.argv[1]
		seq = sys.argv[2]
fh = open(fn)
lines = fh.readlines()

output = sys.argv[3]
sys.stdout = open(output, 'a')
seqs = []
gi = ''
for line in lines:
		if line[0] == '>':
				seqs.append(gi)
				r = ''
				line = line.split('|')
				gi = line[1]
				gi = gi + '_'
		else:
				line = line.rstrip("\n")
				line = str(line)
				gi+=line
seqs = seqs[1:]
fh = open(seq)
linez = fh.readlines()
for i in linez:
		line = i.split( )
		gi = line[0]
		range1 = line[1]
		if range1 < 0:
				range1 = 1
		range2 = line[2]
		range1, range2 = int(range1), int(range2)
		range1 = range1 - 1
		for j in seqs:
				k = j.split('_')
				if gi == k[0]:
						sec = k[1]
						#range2 = len(sec)
						sec1 = sec[range1:range2]
						range1, range2 = range1 +1, range2 +1
						range1, range2 = str(range1), str(range2)
						print '>' + gi, '[' + range1 + ':' + range2 + ']'
						print sec1
```

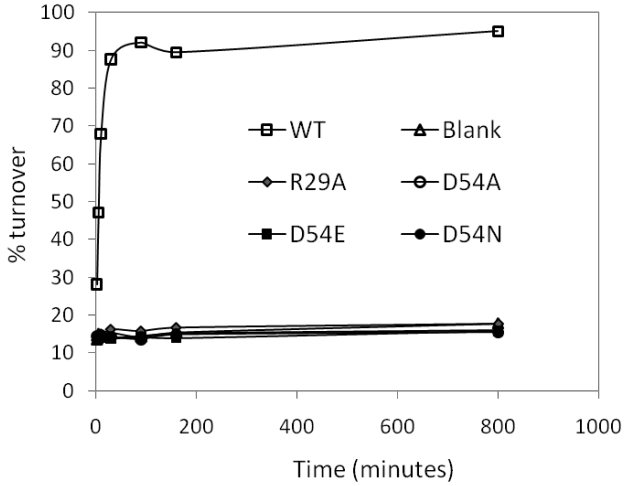
S3: Full alignments for discovered regions of OTase conservation using all 28 sequences listed in Table S1. Alignment conservation calculation uses the MAAFT method of multiple sequence alignment (Liingstone, C. D., Barton, G. J. (1993) *CABIOS* 9: 745-756).

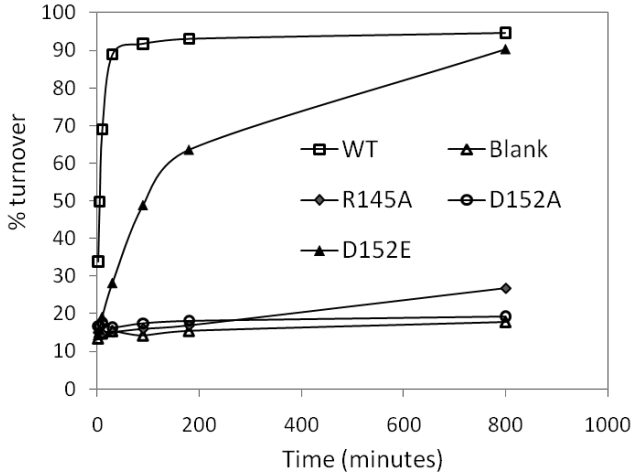


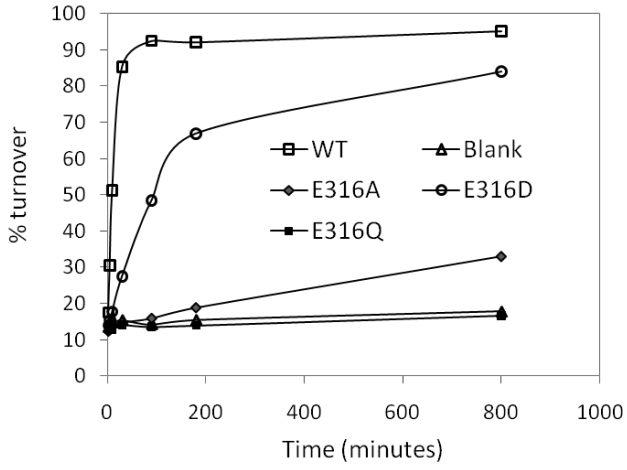


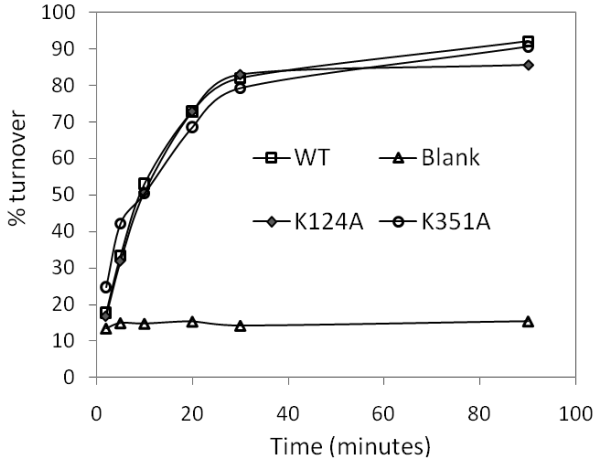


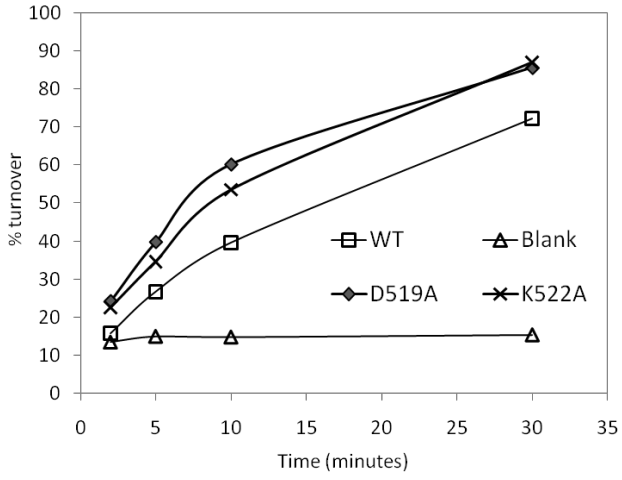
S4: PglB activity assays performed on PglB mutants. Shown are assay results for loop mutants with extended (overnight) time points, negative control mutants K124A and K351A (showing WT-levels of activity), and D^{519}xxK and D^{553}xxK mutant assays.

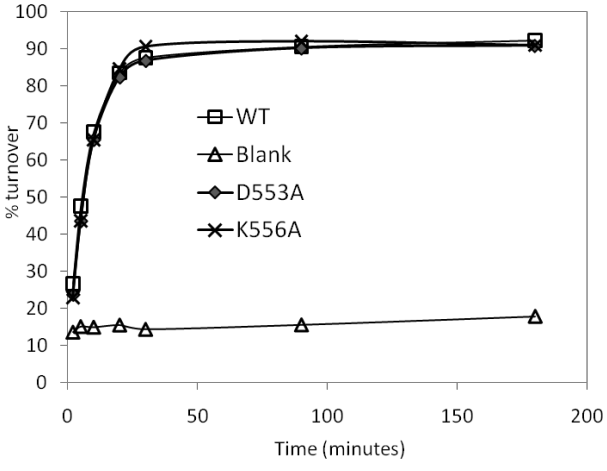




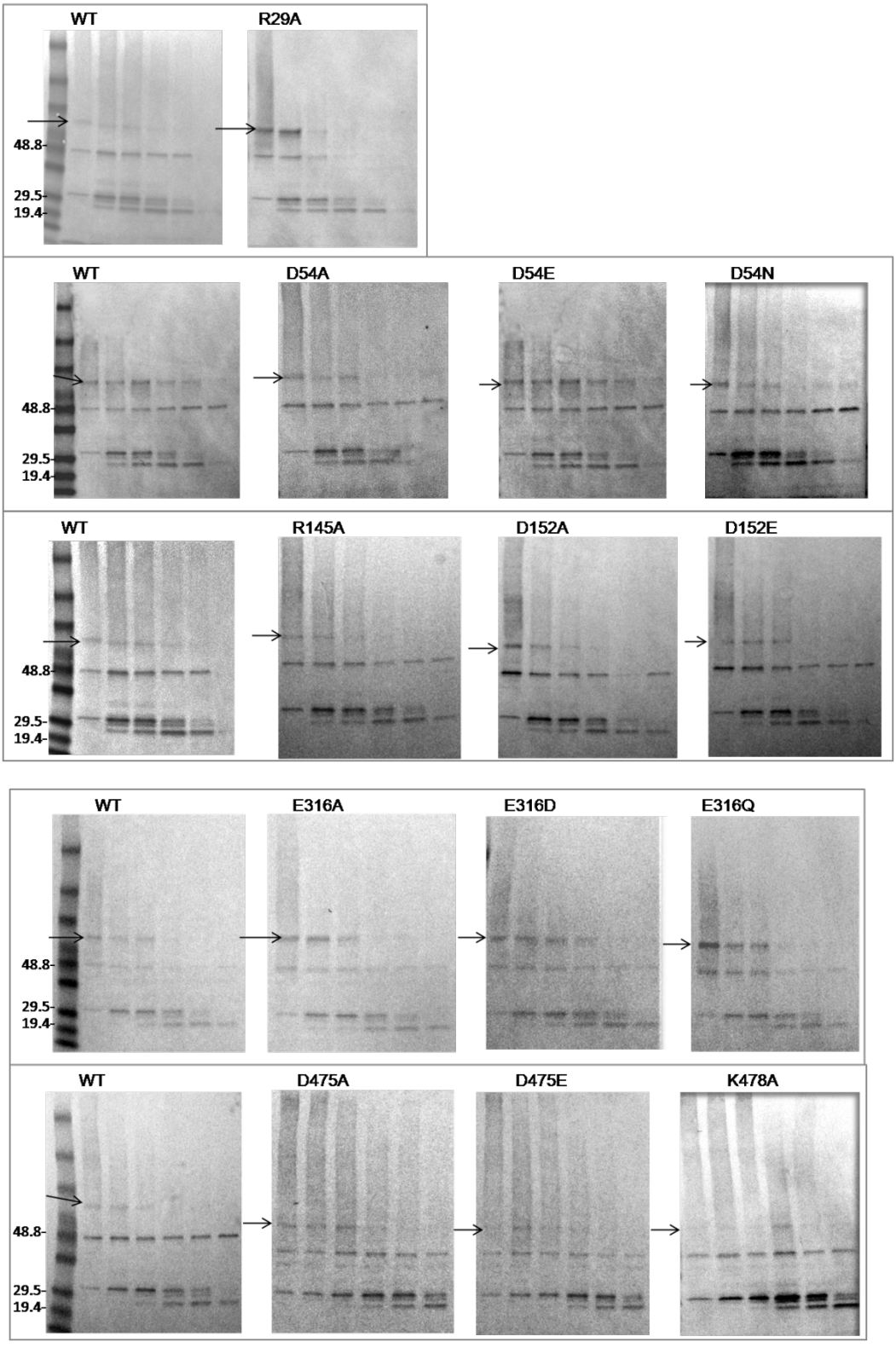




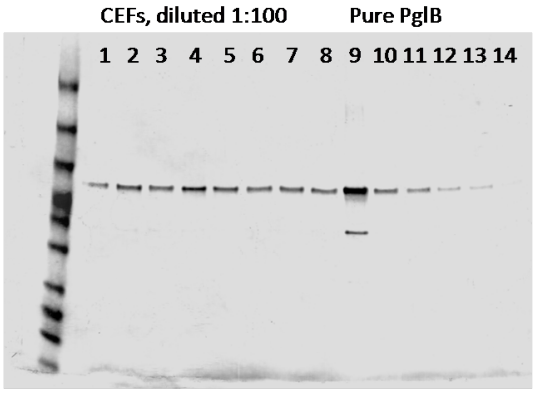


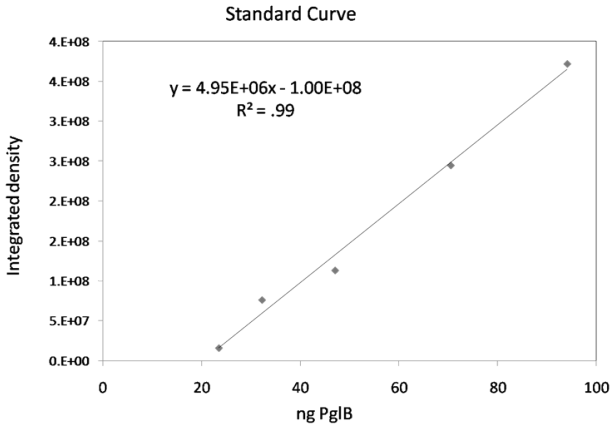


S5: PglB mutants maintain wild-type tertiary structure. Limited proteolysis of mutants assayed in sets, each accompanied by wild-type PglB for comparison. Anti-His western blot analysis was performed on fractions of the of the reaction quenched at 0, 5, 10, 30, 60, and 180 minutes. Arrows indicate the location of the full-length PglB band at the start of the reaction. **Numbers to the left of wild-type blots indicate molecular weight (kDa) of select protein standards.**

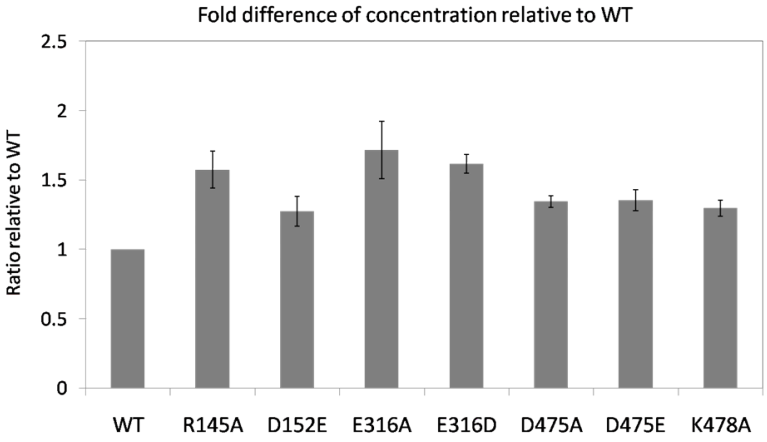


S6: Quantitative western blotting used to determine relative levels of mutant enzymes which showed partial activity. Average fold differences in activity were used to correct for variations when measuring relative rate data at varying substrate concentrations. As indicated in the legend below the blot, lanes 1-8 show mutant CEFs and lanes 9-14 show various levels of pure, quantified PglB. The intensities of the varied amounts of pure protein are used to derive a standard curve (upper right). This curve was used to estimate the quantities of protein in the CEFs shown in the blot. Fold differences in enzyme levels are shown in the bar graph on the bottom right. Error bars indicate the standard error from averages of multiple western blots.





1. WT
2. R145A
3. D152E
4. E316A
5. E316D
6. D475A
7. D475E
8. K478A
9. 470 (ng)
10. 94.1
11. 70.5
12. 47.05
13. 32.25
14. 23.5



S7: Tables showing percent of wild-type rate for mutants at varied concentrations of polyprenyldiphosphate-disaccharide substrate. Left table contains numbers derived directly from measured rates. Right table contains numbers that have been corrected for concentration variation as measured by quantitative western blot.

% of WT rate, directly from measured rates

	0.01 µM	0.1 µM	1 µM
WT	100	100	100
R145A	0.15	0.24	0.37
D152E	4.2	7.4	9.7
E316A	0.3	0.8	1.0
E316D	4.6	18.2	27.0
D475A	18.1	28.3	44.5
D475E	33.5	34.4	48.4
K478A	45.4	40.5	36.4

% of WT rate, corrected for concentration differences

	0.01 µM	0.1 µM	1.0 µM
WT	100	100	100
R145A	0.13	0.15	0.23
D152E	3.2	5.8	7.6
E316A	0.2	0.4	0.6
E316D	3	11	17
D475A	10	21	33
D475E	17	25	36
K478A	25	31	28