Supporting Information

for

Characterization of a New Glycosynthase Cloned by Using Chemical Complementation

Haiyan Tao, Pamela Peralta-Yahya, John Decatur, and Virginia W. Cornish*

Experimental Section

Unless otherwise noted, reagents were obtained from Sigma-Aldrich and used without further purification. Restriction enzymes were purchased from New England Bio-labs. The dNTPs used in the Polymerase Chain Reaction (PCR) were purchased from GE Healthcare Life Sciences. Oligonucleotides were purchased from Invitrogen. QuikChange® site-directed mutagenesis kit was purchased from Stratagene. The ρ-Nitrophenyl-β-d-Cellotetraoside was purchased from TRC Biomedical Research Chemicals. Lacβ1? 4Celβ1? PNP was prepared using Cel7B:E197A glycosynthase. Restriction digests were carried out as recommended by New England Biolabs. Cel5A:E307G enzyme was purified using a HisTrap™ Kit from GE Healthcare Life Sciences. The Cel5A:E307G catalyzed glycosyl transfer reactions were monitored using an Orion fluoride selective electrode and an accumet AR25 pH meter, and the yields were determined by reversed phase HPLC using a Vydac analytical C_{18} column. Nuclear Magnetic Resonance (NMR) spectra were recorded on Bruker Fourier Transform NMR spectrometers 500 or 800 MHz. Mass spectra (MS) were recorded at the Columbia University Department of Chemistry Mass Spectral laboratory.
Construction of the shuttle vectors for the LEU2 enrichment assay. First, an 858-bp DNA containing two SfiI sites and a stuffle were inserted to plasmid p426-Met25 (ATCC) and p426Gal1 (ATCC) between the SpeI and EcoRI sites to generate plasmid pHT2150 and pHT2151, respectively. The Cel5A gene was amplified from pJF-Ac[1] by using primer VWC1623 (5'-GCA TAC GTC GGC CCC CGG GGC CATG TAC GAC GCT AGC CTG-3') and VWC1624 (5'-GCA TTG CTG GGC CGG CAG GGC CTTA TTA GTG GTG GTG GTG GTG-3'), and was inserted into the two SfiI sites of pHT2150 and pHT2151 to generate pHT2187 and pHT2188, respectively. The E307G mutant was generated from both plasmids by QuikChange® by using primers VWC1611 (5'-GGT ATT CCT GTA ATA ATC GGT GGTGT GGA GCA GTA GAT AAG A-3') and VWC1612 (5'-TCT TAT TTA CTG CTC CAC AAC CAC CGA TTA TTA CAG GAA TAC C-3'). The E307S mutant was generated from this plasmid by QuikChange® using primers VWC1613 (5'-GGG GTA TTC CTG TAA TAA TCG GTGT TGT GGA GTA GAT AAG A-3') and VWC1614 (5'-TTG TTC TTA TCT ACT GCT CCA CAA GAA CCG ATT ATT ACA GGA ATA CCA C-3'). CelG gene was amplified from plasmid pCT714 using primer VWC1625 (5'-GCA TAC GTC GGC CCC CGG GGC CATG GGA CAG CAA CAA CGA TG-3') and VWC1648 (5'-GCA TTG CTG GGC CGG CAG GGC CTT A GTG GTT GGT GGT GGT GGT GGT GAG CTT GTG GTG TTC CGT AGT AC-3'), and was inserted into the two SfiI sites of pHT2150 and pHT2151 to generate pHT2190 and pHT2191, respectively.[2] The E345G mutant was generated from both plasmids by QuikChange® using primers VWC1617 (5'-CTC CAA TCC TAA TCG GTG GTGTGG GCG GAC ATA TG-3') and VWC1618 (5'-CAT ATG TCC GCC CCA ACC GAT TAG GAT TGG AG-3'), and the E345S mutant was generated from this plasmid by QuikChange® by using primers VWC1615 (5'-CGC TCC AAT CCT AAT CGG TTC TTG GGG CGG ACA TAT GGA-3') and VWC1616 (5'-TCC ATA TGT CCG CCC CAA GAA CCG ATT AGG ATT GGA GCG-3'). Cel5N gene was amplified from the genome DNA of C. cellulolyticum by using primer VWC1667 (5'-GCA TAC GTC GGC CCC CGG GGC CAT GGA TAC AAA TAA TGA TGA CTG GCT GC-3') and VWC1650 (5'-GCA TTG CTG GGC CGG CAG GGC CTT AGT GGT GGT GGT GGT GGA TAA TTG GTC CAT TCT CTDGCT AG-3'), and was inserted into the two SfiI sites of pHT2150 and pHT2151 to generate pHT2197 and pHT2200, respectively. The Cel5N:E345S mutant was generated from both plasmids by QuikChange® using primers VWC1694 (5'-GCT CCT CTT CCT ATA GGT TCT TGG GGA GGA TTC ATG GAT GGT G-3').
and VWC1695 (5'-CAC CAT CCA TGA ATC CTC CCC AAG AAC CTA TAA GAA GAG GAG C-3').

**LEU2 enrichment assay.** The plasmids for *LEU2* Enrichment Assay were prepared by using QIAprep® miniprep or maxiprep kits. The concentration of plasmids was determined by UV absorption at 260 nm. For each assay, the molar ratio of the plasmid encoding glycosidase to the plasmid encoding Gly or Ser mutant was controlled roughly at 4:1. Then the plasmid mix was transformed to yeast strain V1019Y by using high efficiency yeast transformation protocol.[3] 100 µL of resuspended cells were plated onto synthetic complete (SC) plate containing 2% glucose, lacking uracil, histidine and tryptophan. The plates were incubated at 30°C for 3 days. 5 mL of sterile distilled water was added to the SC plate, and the plate was shaken by hand until all the colonies were suspended in water. The cell suspension was collected into a 15 mL falcon tube. The combined cell suspension was then centrifuged at room temperature using a Sorvall RT7 Plus centrifuge at 2000 rpm for 5 min. The cell pellet was resuspended in 2.0 mL of 20% glycerol. 5 µL of the cell suspension was added into a 15 mL Falcon tube containing 3 mL SC media containing 2% galactose, 2% raffinose, 10 µM Dex-Cel and 10 µM Mtx-Lac-F, with or without methionine, lacking uracil, histidine, tryptophan and leucine. The tubes were incubated in a rotary shaker at 200 rpm and 30°C. On the 6th and the 9th days, 100 µL of cell culture was plated onto non-selective synthetic complete plate containing 2% glucose, lacking uracil, histidine, and tryptophan. The resulted colonies on each plate were collected for plasmid preparation using EZNA kit (Omega Bio-Tek). The resulted plasmids were the PCR amplified using universal primers VWC1051 (for constructs derived from pHT2150, 5'-CGT GTA ATA CAG GGT CGT C-3'), VWC1162 (for constructs derived from pHT2151, 5'-CTG GCC CCA CAA ACC TTC-3'), and VWC1052 (5'-GGG ACC TAG ACT TCA GGT TG-3'). The PCR products were then subject to restriction analysis. Restriction enzymes HphI and PstI were used for analysis of Cel5A Variants, HphI and PvuII for CelG variants, and HphI for Cel5N variants.

**Cel5A:E307G purification.** The *E. coli* expression vector of Cel5A:E307G was constructed from pJF-Ac[1]. The Cel5A:E307G enzyme was expressed in TG1 cells as reported and purified using HisTrap® Kit as recommended.

**Cel5A:E307G glycosynthase reactions:** All reactants were dissolved in pH 7 sodium phosphate buffer containing 25 mM sodium phosphate and 100 mM NaCl. In a ty-
pical reaction, 200 µL of 50 mM donor substrate, 250 µL of 40 mM PNPC and 117 µL of 85 µM Cel5A:E307G were added to in 433 µL sodium phosphate buffer to make the final volume 1 mL. The reaction was monitored using a fluoride ion electrode as reported.[3] The products were then purified by reserved-phase HPLC on a Vydac semi-prep C18 column.

**The configuration determination for the tetrasaccharide products from the Cel5A glycosynthase catalyzed reactions.** The products of the Cel5A:E307G catalyzed reaction were purified by using HPLC, and their structures were determined by mass spectrometry and NMR spectra (1H NMR, COSY, TOCSY, HSQC and/or HMBC), TOCSY spectrum are especially useful since they provide the H-H couplings within the ring system. The configurations of the glycosidic bond were determined based on the chemical shift of the ring protons. In addition, the spectra from Lacβ1?4Celβ1?PNP and commercial available p-Nitrophenyl-β-D-Cellotetraoside were used for comparison.

1. Lacβ1?3Celβ1?PNP

The tetrasaccharide product isolated from the Cel5A catalyzed Lac-F and PNPC reaction is partially contaminated with the starting material PNPC, as indicated by the small anomer peak at 4.42 ppm in the proton NMR; their 1H NMR spectra are compared in Figure S6. The presence of PNPC was further confirmed by HPLC coimmigration and mass spectrometry. Since the PNPC contamination is minor and does not affect the determination of the product configurations, the 1D and 2D NMR spectra were taken without further purification. First, the newly formed glycoside bond was determined to be a β linkage based on the 8 Hz coupling constant of the anomeric proton of the C ring (Figures S2 and 4). The product configurations were then determined based on the assignment of the ring protons and carbons, and the spectra derived from Lacβ1?4Celβ1?PNP were used as references (Figures S7-10). The protons present in the same ring were partially linked and assigned using TOCSY (Figure S4) and COSY spectra (Figure S3). Specifically, the anomeric peak at 4.74 ppm was assigned to the anomer proton of the C ring, which is significantly shifted toward downfield when compared to that from Lacβ1?4Celβ1?PNP. The peak with the downfield chemical shift at 84 ppm on the HSQC spectra (Figure S5) was assigned to the 3rd carbon of B ring, since it does not correlate with any protons at position 2, but correlate well with the putative B3 proton. All together, these evidences
indicated that the tetrasaccharide product bears a beta-1,3 linkage. MS (FAB\(^+\)) \textit{m/z} 788.13 ([M+H]\(^+\), C\(_{30}\)H\(_{46}\)NO\(_{23}\) requires 788.2).

2. Lacβ1?3Lacβ1?3Celβ1?PNP

The other product from the Cel5A catalyzed Lac-F and PNPC reaction has six hexose units with all β linkages, as indicated by proton NMR (Figures S11-13). Similar to that from Lacβ1?3Celβ1?PNP, the anomic peak at 4.70 ppm indicates a β-1,3 linkage. We assumed that the second lactose unit is attached to the tetrasaccharide product by a beta 1,3-linkage, as the axial hydroxyl group of the galactose unit is not likely a substrate for glycosynthases derived from cellulase. In addition, we did not observe significant ring proton change from the TOCSY spectrum except for the overall shift of one galactose unit, so it is likely that the second Lactose was attached to the product using the same β-1,3 linkage.

3. The tetrasaccharide fraction from the reaction of Cel-F and PNPC.

Based on the proton NMR and mass spectra, the tetrasaccharide fraction is a mixture. The products in the mixture have β linkages since the all anomic protons show an 8 Hz coupling constant. Commercially available Celβ1?4Celβ1?PNP (p-Nitrophenyl-β-D-Cellotetraoside) was used to prepare standard NMR spectra (Figures S14-17). As shown in supplemental figure 18 and 19, the proton and COSY spectra from the tetrasaccharide mixture contain the all peaks from Celβ1?4Celβ1?PNP. The major part of other fraction was determined as Celβ1?3Celβ1?PNP. Similar to Lacβ1?3Celβ1?PNP, the anomic peak at 4.67 ppm was assigned to the anomic proton of the C ring, which is correlated with the B3 carbon peak at 84 ppm in the HMBC spectrum (Figure S20). Beside the β-1,3 and the β-1,4 product, this mixture also contain a small fraction of unidentified product, as shown in the proton and COSY spectra (Figures S18 and 19). MS (ESI\(^+\)) \textit{m/z} 811.66 ([M+Na]\(^+\), C\(_{30}\)H\(_{45}\)NO\(_{23}\)Na requires 810.2).

References:


Figure S1. Preliminary results of the LEU2 enrichment assays. A mixture of plasmids encoding a roughly 1:4 ratio of the glycosidase nucleophile mutant to the wild-type glycosidase was transformed *en masse* into yeast three-hybrid strain V1019Y. After selection, the plasmids were extracted, and the ORF were amplified using PCR and analyzed using restriction digestion. For each assay, the first lane indicate the plasmids mixture before selection, the second lane indicate the plasmids mixture after six days of selection, and the third lane indicate the plasmids mixture after nine days of selection. The lanes with black triangle are 100 bps DNA ladder (New England Biolab). #1: Cel5A:E307G, under the control of MET25 promoter; #2: Cel5A:E307S, under the control of MET25 promoter; #3: Cel5A:E307G, under the control of Gal1 promoter; #4: Cel5A:E307S, under the control of Gal1 promoter; #5: CelG:E345G, under the control of MET25 promoter; #6: CelG:E345S, under the control of MET25 promoter; #7: CelG:E345G, under the control of Gal1 promoter; #8: CelG:E345S, under the control of Gal1 promoter; #9: Cel5N:E345S, under the control of MET25 promoter; #10: Cel5N:E345S, under the control of Gal1 promoter. The bands that were used to compare the glycosynthase enrichment were highlighted. Since all the glycosynthase variants involved in this research removed the HphI site located at the mutation site of their wild-type, for all of the assays, the upper bands indicate the glycosynthase mutants, and the lower bands indicate the wild-type glycosidases.
Figure S2. $^1$H NMR spectrum of Lacβ1?3Celβ1?PNP

Figure S3. COSY spectrum of Lacβ1?3Celβ1?PNP.
Figure S4. TOCSY spectrum of Lacβ1? 3Celβ1? PNP.

Figure S5. HSQC spectrum of Lacβ1? 3Celβ1? PNP.
Figure S6. The NMR comparison between PNPC and Lacβ1?3Celβ1?PNP.

Figure S7. $^1$H NMR spectrum of Lacβ1?4Celβ1?PNP.
Figure S8. COSY spectrum of Lacβ1-4Celβ1-PNP.

Figure S9. TOCSY spectrum of Lacβ1-4Celβ1-PNP.
Figure S10. HSQC spectrum of Lacβ1?4Celβ1?PNP.

Figure S11. The ^1H NMR spectrum of the hexasaccharide from the reaction of Lac-F and PNPC.
Figure S12. COSY spectrum of the hexasaccharide from the reaction of Lac-F and PNPC.

Figure S13. TOCSY spectrum of the hexasaccharide from the reaction of Lac-F and PNPC.
**Figure S14.** $^1$H NMR spectrum of $p$-Nitrophenyl-$\beta$-d-cellotetraoside.

**Figure S15.** COSY spectrum of $p$-Nitrophenyl-$\beta$-d-cellotetraoside.
Figure S16. $^1$H NMR spectrum of the tetrasaccharide fraction from the reaction of Cel-F and PNPC.

Figure S17. COSY spectrum of the tetrasaccharide fraction from the reaction of Cel-F and PNPC.
Figure S18. $^1$H NMR Comparison of the tetrasaccharide fraction from the reaction of Cel-F and PNPC and $p$-nitrophenyl-$\beta$-$D$-cellotetraoside.

Figure S19. COSY comparison of the tetrasaccharide fraction from the reaction of Cel-F and PNPC and $p$-nitrophenyl-$\beta$-$D$-cellotetraoside.
Figure S20. HMBC spectrum of the tetrasaccharide fraction from the reaction of Cel-F and PNPC.

Figure S21. HSQC-TOCSY spectrum of the tetrasaccharide fraction from the reaction of Cel-F and PNPC (70 ms mix).
Figure S22. HSQC-TOCSY spectrum of the tetrasaccharide fraction from the reaction of Cel-F and PNPC (20 ms mix).