J|A|C|S ARTICLES

Published on Web 07/12/2010

Fluorescence Activated Cell Sorting as a General Ultra-High-Throughput Screening Method for Directed Evolution of Glycosyltransferases

Guangyu Yang,[†,§,‡] Jamie R. Rich,[†] Michel Gilbert,[#] Warren W. Wakarchuk,[#] Yan Feng,[§,‡] and Stephen G. Withers*[,†]

Centre for High-throughput Biology (CHiBi) and Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, P. R. China, Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, Jilin University, 2519 Jiefang Road, Changchun 130023, P. R. China, and National Research Council of Canada, Institute for Biological Sciences, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada

Received May 14, 2010; E-mail: withers@chem.ubc.ca

Abstract: Glycosyltransferases (GTs) offer very attractive approaches to the synthesis of complex oligosaccharides. However, the limited number of available GTs, together with their instability and strict substrate specificity, have severely hampered the broad application of these enzymes. Previous attempts to broaden the range of substrate scope and to increase the activity of GTs via protein engineering have met with limited success, partially because of the lack of effective high-throughput screening methods. Recently, we reported an ultra-high-throughput screening method for sialyltransferases based on fluorescence-activated cell sorting (Aharoni et al. *Nat. Methods* **2006**, *3*, 609−614). Here, we considerably improve this method via the introduction of a two-color screening protocol to minimize the probability of false positive mutants and demonstrate its generality through directed evolution of a neutral sugar transferase, β-1,3-galactosyltransferase CgtB. A variant with broader substrate tolerance than the wild-type enzyme and 300-fold higher activity was identified rapidly from a library of $>10^7$ CgtB mutants. Importantly, the variant effected much more efficient synthesis of G_{M1a} and asialo G_{M1} oligosaccharides, the building blocks of important therapeutic glycosphingolipids, than did the parent enzyme. This work not only establishes a new methodology for the directed evolution of galactosyltransferases, but also suggests a powerful strategy for the screening of almost all GT activities, thereby facilitating the engineering of glycosyltransferases.

Introduction

The chemical synthesis of complex oligosaccharides is often an extremely challenging process because of the need for selective, labor-intensive protection/deprotection steps.[1,2] Enzymatic synthesis by glycosyltransferases (GTs), on the other hand, offers significant advantages over purely chemical methods due to the exquisite control conferred over the stereo- and regiochemistry of the reaction products.[3−5] However, broad application of GTs is hampered by the limited availability of enzymes and by their often strict substrate specificities. In addition, enzymes isolated directly from nature are not always suitable for synthetic processes that employ high concentrations of substrate and possibly also organic solvents. Therefore, access to GTs tailored for the efficient synthesis of specific oligosaccharides and glycoconjugates would be of great interest. However, while the field of protein engineering has undergone explosive growth over the past 10 years, there have been relatively few success stories with respect to the directed evolution of GTs compared with other important enzyme families such as lipases, proteases, and glycosidases. This is mainly due to the fact that typical screens for GTs are based on microtiter plate assays and thus require significant amounts of costly reagents, and have limited throughput.[6−9] Indeed, most previous directed evolution studies involving GTs screened no more than 1000 colonies. Screening of larger libraries would allow the exploration of a much larger protein sequence space,

[†] University of British Columbia.
[§] Shanghai Jiao Tong University.
[‡] Jilin University.
[#] National Research Council of Canada, Institute for Biological Sciences.

(1) Wu, D.; Fujio, M.; Wong, C. H. *Bioorg. Med. Chem.* **2008**, *16*, 1073–1083.
(2) Hanson, S.; Best, M.; Bryan, M. C.; Wong, C. H. *Trends. Biochem. Sci.* **2004**, *29*, 656–663.
(3) Macmillan, D.; Daines, A. M. *Curr. Med. Chem.* **2003**, *10*, 2733–2773.
(4) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. *Annu. Rev. Biochem.* **2008**, *77*, 521–555.
(5) Weijers, C. A.; Franssen, M. C.; Visser, G. M. *Biotechnol Adv.* **2008**, *26*, 436–456.
(6) Park, S. H.; Park, H. Y.; Sohng, J. K.; Lee, H. C.; Liou, K.; Yoon, Y. J.; Kim, B. G. *Biotechnol. Bioeng.* **2009**, *102*, 988–994.
(7) Persson, M.; Palcic, M. M. *Anal. Biochem.* **2008**, *378*, 1–7.
(8) Williams, G. J.; Goff, R. D.; Zhang, C. S.; Thorson, J. S. *Chem. Biol.* **2008**, *15*, 393–401.
(9) Williams, G. J.; Zhang, C.; Thorson, J. S. *Nat. Chem. Biol.* **2007**, *3*, 657–662.

 10.1021/ja104167y

thereby improving the chances of identifying useful mutations.[10,11] The development of high-throughput, low-cost screens for the interrogation of large GT libraries is therefore of great interest.

Fluorescence-activated cell sorting (FACS) is a specialized application of flow cytometry.[12,13] In a FACS instrument, multiple fluorescent parameters (size, structure, and fluorescent signals) of individual sample particles (cells, microbeads, or emulsions) can be analyzed by a focused laser beam at rates of approximately 10^7/h. A charge can be applied to the particles of interest, which are deflected into a collection tube by a charged plate. Originally developed for the analysis of protein−ligand interactions, FACS has recently emerged as a powerful tool for screening enzyme libraries due to its unique sensitivity and its ability to analyze as many as 10^8 mutants/day.[14−16] The major challenge in FACS screening is how to maintain the link between genotype and phenotype because most enzyme substrates do not remain associated with the cells. Several different FACS screening approaches have been developed, largely based on the attachment of products to cell surfaces or entrapment of the product inside the cells, and these have been applied to the directed evolution of proteases,[17−20] peroxidases,[21] esterases,[22] glutathione transferases,[23−25] and nucleoside kinases[26] in order to improve their expression level, stability, ligand binding, activity, or substrate specificity. We have recently reported a novel FACS-based, high-throughput screening method, which makes use of differences in physical properties between the substrate and product, for the directed evolution of the sialytransferase CstII from *Campylobacter jejuni*.[27] Libraries of mutant sialyltransferase (ST) genes were expressed in *Escherichia coli* cells that are naturally permeable to an exogenous glycosyl acceptor (Bodipy-lactose, **1**) (Figure 1). Active STs catalyzed sialylation of **1**, entrapping the charged fluorescent product within the cell and thereby co-localizing the genotype and phenotype (Figure 2a). Using FACS, the library was screened at a rate of more than 10^7 events/h and an ST variant with up to 400-fold improved activity for **1** was thus identified. However, improvements were specific to **1**, as enhanced activity had arisen from the evolution of a binding site for the substrate-appended fluorophore, as demonstrated crystallographically.[27] An approach that avoids the coincident evolution of dye binding sites is therefore necessary.

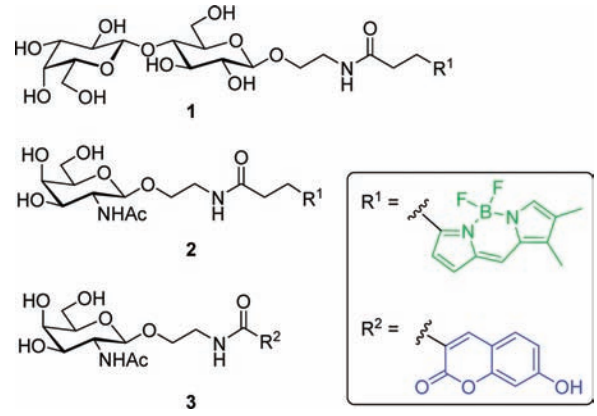


Figure 1. Fluorescently labeled lactose (**1**) and differentially labeled *N*-acetyl-D-galactosamine (**2** and **3**) derivatives used in screens for glycosyltransferase activity.

Here, we present a method to evolve improved GT activity through simultaneous use of two selection substrates bearing the same sugar but chemically distinct fluorophores, thereby minimizing the probability of selecting for a dye-binding site (Figure 2b). Significantly, we have applied this strategy to a β-1,3-galactosyltransferase (CgtB), which transfers the neutral sugar galactose, thereby demonstrating that cellular entrapment is not restricted to charged products and highlighting the generality of this ultra-high-throughput screen for GTs. Screening of a library of $>10^7$ CgtB mutants against fluorescently labeled derivatives of *N*-acetylgalactosamine (**2** and **3**) identified a range of improved mutants, with up to 300-fold enhanced catalytic activities and improved ability to synthesize G_{M1a} and asialo G_{M1} oligosaccharides, the building blocks of important therapeutic glycosphingolipids.

Experimental Section

Generation of Random-Mutagenesis Library. DNA encoding amino acids 1−272 of CgtB was subjected to error-prone PCR using the Gene Morph II Mutagenesis kit (Stratagene) with 0.03 ng or 0.06 ng of pCW-CgtBΔC30-MalE as template, following the manufacturer's protocol. The primers used were CgtB-F, 5′-GAAAGGGAGCTCACATATGTTTAAAATTTCAATCATCTTACCAAC-3′ and CgtB-R, 5′-GAAGGTCGGAATTCCGGTTTTATTTTATATATTTGAATATATAGC-3′. The PCR product was digested and ligated to pCW-MalE-C vector using NdeI and EcoRI restriction sites and the ligation mixture was electroporated to *E. cloni* 10G elite cells (Lucigen) according to the manufacturer's protocol. The transformed cells were grown overnight at 37 °C in LB medium supplemented with ampicillin (100 mg/mL) and the library plasmid DNA was extracted. Several individual clones from each library were sequenced and shown to have an average mutational frequency of ∼2 and ∼5 mutations per gene in the 0.03 and 0.06 ng template libraries, respectively. The plasmid DNA from the two libraries was mixed in a 1:1 ratio and was transformed into *E. coli* JM107* for screening.[28] These were the

(10) Aharoni, A.; Amitai, G.; Bernath, K.; Magdassi, S.; Tawfik, D. S. *Chem. Biol.* **2005**, *12*, 1281–1289.
(11) Dietrich, J. A.; McKee, A. E.; Keasling, J. D. *Annu. Rev. Biochem.* **2010**, *79*, 563–590.
(12) Givan, A. L. *Flow Cytometry: First Principles*; Wiley-Liss: NY, 1992.
(13) Shapiro, H. M. *Practical Flow Cytometry*, 4th ed.; Wiley-Liss: NY, 2004.
(14) Becker, S.; Schmoldt, H. U.; Adams, T. M.; Wilhelm, S.; Kolmar, H. *Curr. Opin. Biotechnol.* **2004**, *15*, 323–329.
(15) Farinas, E. T. *Comb. Chem. High Throughput Screening* **2006**, *9*, 321–328.
(16) Yang, G.; Withers, S. G. *ChemBioChem* **2009**, *10*, 2704–2715.
(17) Olsen, M. J.; Stephens, D.; Griffiths, D.; Daugherty, P.; Georgiou, G.; Iverson, B. L. *Nat. Biotechnol.* **2000**, *18*, 1071–1074.
(18) Varadarajan, N.; Gam, J.; Olsen, M. J.; Georgiou, G.; Iverson, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6855–6860.
(19) Varadarajan, N.; Rodriguez, S.; Hwang, B. Y.; Georgiou, G.; Iverson, B. L. *Nat. Chem. Biol.* **2008**, *4*, 290–294.
(20) Varadarajan, N.; Georgiou, G.; Iverson, B. L. *Angew. Chem., Int. Ed.* **2008**, *47*, 7861–7863.
(21) Lipovsek, D.; Antipov, E.; Armstrong, K. A.; Olsen, M. J.; Klibanov, A. M.; Tidor, B.; Wittrup, K. D. *Chem. Biol.* **2007**, *14*, 1176–1185.
(22) Becker, S.; Höbenreich, H.; Vogel, A.; Knorr, J.; Wilhelm, S.; Rosenau, F.; Jaeger, K. E.; Reetz, M. T.; Kolmar, H. *Angew. Chem., Int. Ed.* **2008**, *47*, 5085–5088.
(23) Kawarasaki, Y.; Griswold, K. E.; Stevenson, J. D.; Selzer, T.; Benkovic, S. J.; Iverson, B. L.; Georgiou, G. *Nucleic. Acids. Res.* **2003**, *31*, e126.
(24) Griswold, K. E.; Kawarasaki, Y.; Ghoneim, N.; Benkovic, S. J.; Iverson, B. L.; Georgiou, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10082–10087.
(25) Griswold, K. E.; Aiyappan, N. S.; Iverson, B. L.; Georgiou, G. *J. Mol. Biol.* **2006**, *364*, 400–410.
(26) Liu, L.; Li, Y.; Liotta, D.; Lutz, S. *Nucleic Acids Res.* **2009**, *37*, 4472–4481.
(27) Aharoni, A.; Thieme, K.; Chiu, C. P.; Buchini, S.; Lairson, L. L.; Chen, H.; Strynadka, N. C.; Wakarchuk, W. W.; Withers, S. G. *Nat. Methods.* **2006**, *3*, 609–614.
(28) Antoine, T.; Heyraud, A.; Bosso, C.; Samain, E. *Angew. Chem., Int. Ed.* **2005**, *44*, 1350–1352.

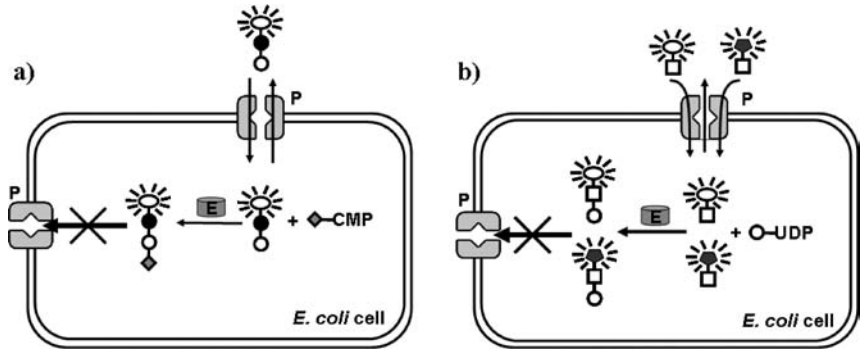


Figure 2. Schematic view of the fluorescent product entrapment method employed to screen mutant GT libraries. Fluorescently labeled acceptors such as **1**, **2**, or **3** are transported into the cell by a transmembrane sugar transport protein (permease). After an incubation period during which the acceptor may be modified enzymatically, unreacted acceptor is removed in a washing step. *E. coli* cells containing catalytically active GTs retain the fluorescent product inside the cell as it is no longer a substrate for the permease. Catalytically active enzymes can be identified and isolated by fluorescence activated cell sorting: (a) in the original method, a single fluorescent substrate was used to screen the library of mutant sialyltransferases;[27] (b) in this improved method, two selection substrates were employed simultaneously in order to avoid the evolution of dye binding sites. E, glycosyltransferase; P, permease.

same engineered cells used for the original FACS screening work in which both the *lacZ* β-galactosidase gene and the sialic acid aldolase gene (*nanA*) have been knocked out. Knockout of the *nanA* gene was not essential for this work, nor was it deleterious.

Generation of DNA Shuffling Library. To construct the shuffling library, the plasmid DNA isolated from the third round of FACS enrichment was amplified by PCR with CgtB-F and CgtB-R primers. The amplified genes were purified and mixed with the wild-type CgtB genes in a 50:1 ratio. The gene mixture was digested using Benzonase (Novagen) and those DNA fragments smaller than 100 bp were collected and reassembled using the following PCR profile: 95 °C for 1 min; 45 cycles of 94 °C for 30s, 50 °C for 30 s, 72 °C for 1 min; and 72 °C for 5 min. Analysis of the reassembly product by agarose gel electrophoresis revealed a smear encompassing the ~800 bp target length. A portion of the assembly reaction mixture was used as template for PCR amplification with CgtB-F and CgtB-R primers. The resulting genes were purified, digested, and ligated into pCW-MalE-C vector and transformed into *E. coli* JM107* competent cells for screening.

Site-Directed Mutagenesis. Site-directed mutagenesis was performed using the Quikchange methodology (Stratagene) following the manufacturer's protocol. Mutations were confirmed by DNA sequencing. The primers used were

CgtB_K151E_F: AATTATTGCAGAGAAAAATTTATATTG-GACTATGTGG,

CgtB_K151E_R: CAATATAAATTTTTCTCTG-CAATAATTTTTTTTAC,

CgtB_L227G_F: GAATGAAGTGCGTGTTAAAAATAATAT-TCAAGAGTTGC,

CgtB_L227G_R: GAATATTATTTTTAACACGCACTTCAT-TCTTAGTATTAC,

CgtB_E234D_F: AAATAATATTCAAGATTTGCAGTTGGTTT-TAAACTA

TTTAAGG,CgtB_E234D_R:GTTTAAAACCAACTGCAAATCT-TGAATATTATTTTTAACAAGCAC.

Flow Cytometric Screening. Flow cytometric screening of CgtB for glycosyltransferase activity was carried out essentially as described before,[27] except that two fluorescent substrates were used simultaneously. Briefly, plasmid DNA (pCW-CgtBΔC30-MalE) encoding the CgtB libraries was transformed into JM107* competent cells and used to directly inoculate (without plating on agar) LB media supplemented with ampicillin (100 mg/mL) and grown overnight at 37 °C. The cells were then diluted 1/50 in M9 mineral cultured media and grown at 37 °C with vigorous shaking. When the absorbance at 600 nm (A600) reached 0.5−0.7, IPTG was added to a final concentration of 1 mM, and the cultures were transferred to 20 °C and grown overnight. Cells were spun down (1 mL) and resuspended in 50 μL of M9 media supplemented with 0.2−0.5 mM concentrations of **2** and **3**. Following 10−30 min of incubation, cells were spun down and the excess acceptor sugars were removed by washing the cells with LB media and phosphate-buffered saline (PBS, pH 7.4). The sample was placed on ice and taken to the FACS for further analysis and sorting.

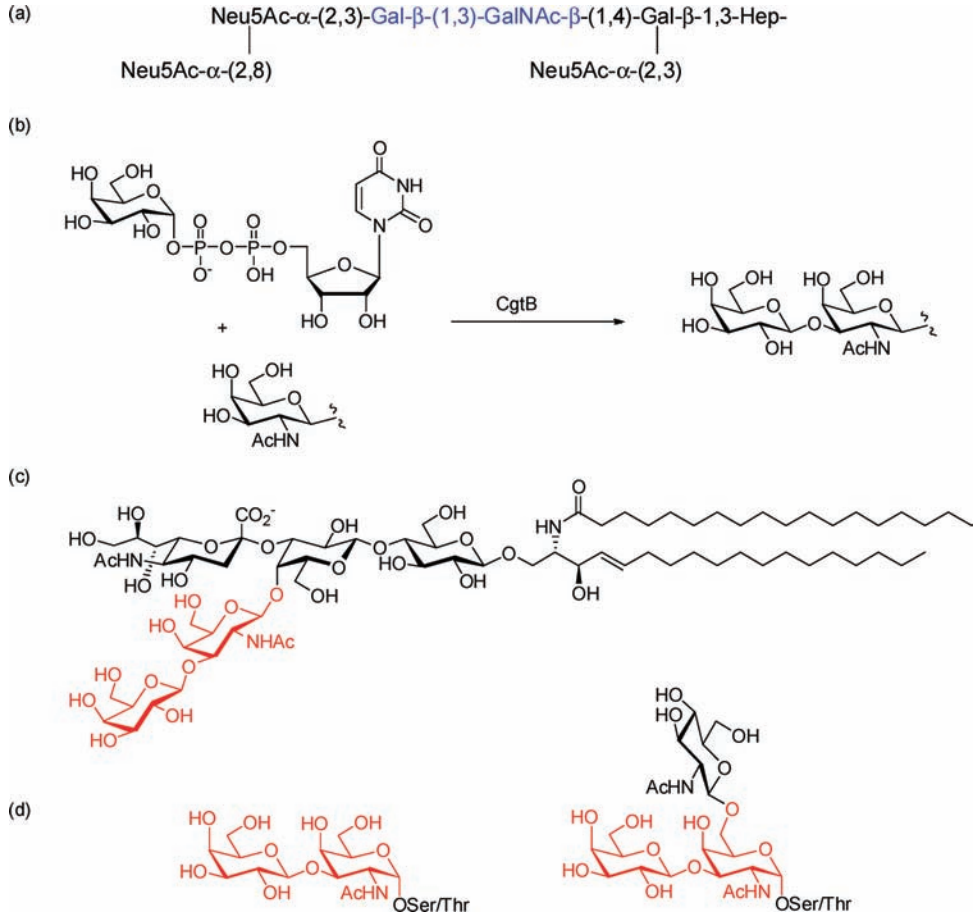
The cells were diluted with PBS as needed to obtain a flow cytometric event rate of ~4000/s in the FACSAria flow cytometer (Becton-Dickinson) using PBS as sheath fluid. The threshold for event detection was set to forward and side scattering. The average sort rate was ~4000 events/s, using a 70 μm nozzle, exciting argon ion (488 nm), and 405 nm lasers, and measuring emissions passing the 530 ± 20 nm (FITC) band-pass filter for the bodipy emission, and the 450 nm (violet 1) filter for the coumarin emission. Cells were sorted into Eppendorf tubes containing 0.5−1 mL of LB medium. Pools of sorted positives were plated on LB agar plates supplemented with ampicillin (100 mg/mL). Colonies were allowed to grow overnight at 37 °C and the cells were removed from the agar plates for the next round of FACS enrichment. FACS data were processed using the Flowjo Software (Tree Star).

Protein Expression and Purification. Expression and purification of wild-type CgtB and its mutants were carried out as previously described.[29] Briefly, the *E. coli* cells containing CgtB variants were inoculated in 2YT medium supplemented with ampicillin (100 mg/mL) at 37 °C overnight. The cells were transferred into fresh medium and grown at 37 °C. When A600 reached 0.6−1.0, IPTG was added to a final concentration of 1 mM, and the cultures were transferred to 20 °C and grown overnight. The cells were harvested and suspended in 10% (w/v) Buffer A (which contains 20 mM HEPES, pH 7.0, 200 mM NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA, and 10% glycerol). After ultrasonic cell disintegration, the cell suspension was centrifuged at 12 000 rpm for 30 min. The supernatant was diluted as needed in Buffer A and applied to a 20 mL column of amylose resin (New England Biolabs). After sample application, the column was washed with 2 column volumes to elute unbound proteins. The bound protein was eluted by washing the column with buffer B (Buffer A and 20 mM maltose). The fractions containing the target protein were collected and dialyzed overnight against 100 vol of 50 mM sodium acetate, pH 6.0, and 20% glycerol at 4 °C. The purity of all of the variants of CgtB was greater than 90% (see the Supporting Information). The protein concentration was determined according to the Bradford method, and bovine serum albumin (BSA) was used as the standard.

Kinetic Analysis. Kinetic analysis of the CgtB variants was performed at 37 °C in 20 mM HEPES buffer, pH 7.5, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 50 mM KCl, 0.25 mM NADH, 0.7 mM

(29) Bernatchez, S.; Gilbert, M.; Blanchard, M. C.; Karwaski, M. F.; Li, J.; DeFrees, S.; Wakarchuk, W. W. *Glycobiology* **2007**, *17*, 1333–1343.

Scheme 1. (a) Partial Structure of Lipooligosaccharide Outer Core from *C. jejuni* O:19 Strain OH4384.[33] (b) Reaction Catalyzed by CgtB Glycosyltransferases from *C. jejuni*. (c) Structure of Representative Ganglioside (Ganglioside G_{M1a}) with Galβ1,3GalNAcβ Subunit Highlighted in Red. (d) Minimal Core 1 and Core 2 Glycoprotein *O*-Glycan Structures with Galβ1,3GalNAcα Subunit Highlighted in Red



PEP, and 0.1% (w/v) BSA. Release of UDP was coupled to the oxidation of NADH ($\lambda = 340$ nm, $\varepsilon = 6.22$ mM^{-1} cm^{-1}) using the enzymes pyruvate kinase (PK) and lactate dehydrogenase (LDH).[30] Background hydrolysis of UDP-Gal in the absence of acceptor sugar was measured for the different acceptors and subtracted from the enzyme transfer rate. Absorbance measurements were obtained using a UV-2550 spectrophotometer (Shimadzu). Grafit 4.0 (Erithacus software) was used to calculate kinetic parameters and associated errors by direct fit of the initial rate to the appropriate equations.

Specific activities of the CgtB variants using asialo G_{M2}-FCHASE as acceptor were determined as described previously.[29]

Synthesis. See the Supporting Information.

Results and Discussion

The CgtB galactosyltransferases (CAZy family GT2)[31] are a group of related enzymes from different strains of *C. jejuni* that catalyze the biosynthesis of the Galβ1,3GalNAc linkage within their respective structurally distinct lipooligosaccharides (Scheme 1a,b)[32] To our knowledge, these are the only glycosyltransferases identified to date that utilize both α- and β-configured *N*-acetylgalactosamine as an acceptor, thus, suggesting simultaneous promise for synthesis of the Galβ1,3GalNAc subunit common to **both** the core 1 and 2 *O*-glycans as well as the ganglioside glycolipids (Scheme 1c,d).[29] Previous improvements in the *C. jejuni* OH4384 CgtB variant involving a C-terminal deletion and fusion to maltose binding protein (CgtBΔC30-MalE) provided a starting point for directed evolution in this study.[29]

FACS-based screening for CgtB activity in *E. coli* (Figure 1b) required that (i) fluorescent acceptor substrates reach the cytoplasm, (ii) glycosylation of the acceptor result in product entrapment, and (iii) unreacted acceptor be washed out of the cell. To this end, we synthesized bodipy- and coumarin-labeled fluorescent GalNAc acceptors **2** and **3**, with the expectation that they would be freely transported into and out of *E. coli*. The expectation then was that upon galactosylation of **2** and **3** by plasmid-encoded CgtB, the product disaccharides would be entrapped within the cell, allowing unreacted acceptors to be washed away without loss of fluorescent product (Figure 2b). In this scenario the fluorescent intensity of the cells should correlate with CgtB activity.

To test the feasibility of this strategy, *E. coli* cells expressing, respectively, native CgtB, its improved variant CgtBΔC30-MalE, and empty pCW-MalE-C plasmid were incubated with both fluorescent GalNAc acceptors **2** and **3** at concentrations that optimize the dynamic range accessible while minimizing selection for low K_m values. An *E. coli* strain JM107*, which

(30) Gosselin, S.; Alhussaini, M.; Streiff, M. B.; Takabayashi, K.; Palcic, M. M. *Anal. Biochem.* **1994**, *220*, 92–97.
(31) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233–238.
(32) Gilbert, M.; Brisson, J. R.; Karwaski, M. F.; Michniewicz, J.; Cunningham, A. M.; Wu, Y. Y.; Young, N. M.; Wakarchuk, W. W. *J. Biol. Chem.* **2000**, *275*, 3896–3906.
(33) Aspinall, G. O.; McDonald, A. G.; Pang, H.; Kurjanczyk, L. A.; Penner, J. L. *Biochemistry.* **1994**, *33*, 241–249.

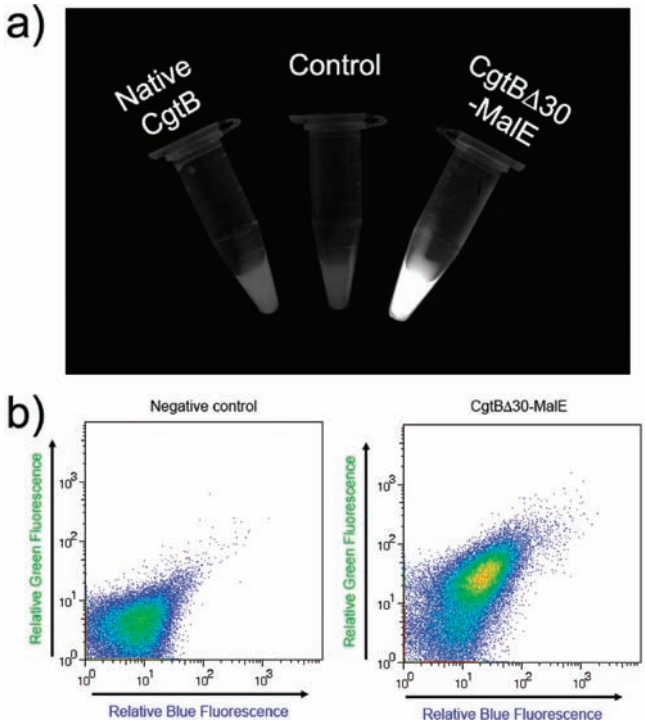


Figure 3. (a) Visualization of fluorescence entrapment within *E. coli* cells expressing native CgtB, the CgtBΔC30-MalE variant, or empty plasmid (control). (b) Flow cytometry profile of whole cell fluorescence in cells expressing empty plasmid (left, negative control) and CgtBΔC30-MalE (right) after 15 min incubation with bodipy-GalNAc (**2**, 0.25 mM) and coumarin-GalNAc (**3**, 0.25 mM), followed by a washing step.

lacks β-galactosidase (lacZ), was used to minimize the *in vivo* degradation of substrate and product.[28] After extensive washing, the fluorescence intensities of the cells were visualized under ultraviolet light. As shown in Figure 3a, the cells expressing empty plasmid were only weakly fluorescent. However, cells expressing CgtBΔC30-MalE showed a strong fluorescence signal, while an intermediate level of fluorescence intensity was seen for the cells expressing native CgtB (Figure 3a). The flow cytometry profile showed that CgtBΔC30-MalE cells have both substantially higher green and blue fluorescence corresponding to entrapment of both bodipy and coumarin disaccharide conjugates, respectively (Figure 3b). The fluorescence signals are stable for more than 2 h at 4 °C, providing an ample window for the FACS analysis and sorting.

While the mechanism of transport and entrapment of the fluorescent substrate/product has not been clearly established, several pieces of evidence point to mediation by lactose permease. No fluorescence entrapment was obtained when using *E. coli* Tuner strain (Novagen) in which the lactose permease gene (*lacY*) is knocked out, suggesting that the fluorescent acceptor could not get into the cells without the *lacY* gene (see the Supporting Information for details). In addition, when regular *E. coli* JM107* cells (which contain the *lacY* gene) were incubated with the fluorescent acceptors together with high concentrations of lactose, transport of the fluorescent acceptor into the cells was inhibited (see the Supporting Information). Since we have independently demonstrated that lactose does not appreciably inhibit CgtB-catalyzed galactosyl transfer to Bodipy-GalNAc, this result suggests that the lactose competitively inhibits entry of Bodipy-GalNAc via the lactose permease, but that the galactosylated product does not get out. Consequently, this screening system can be used with GTs that use

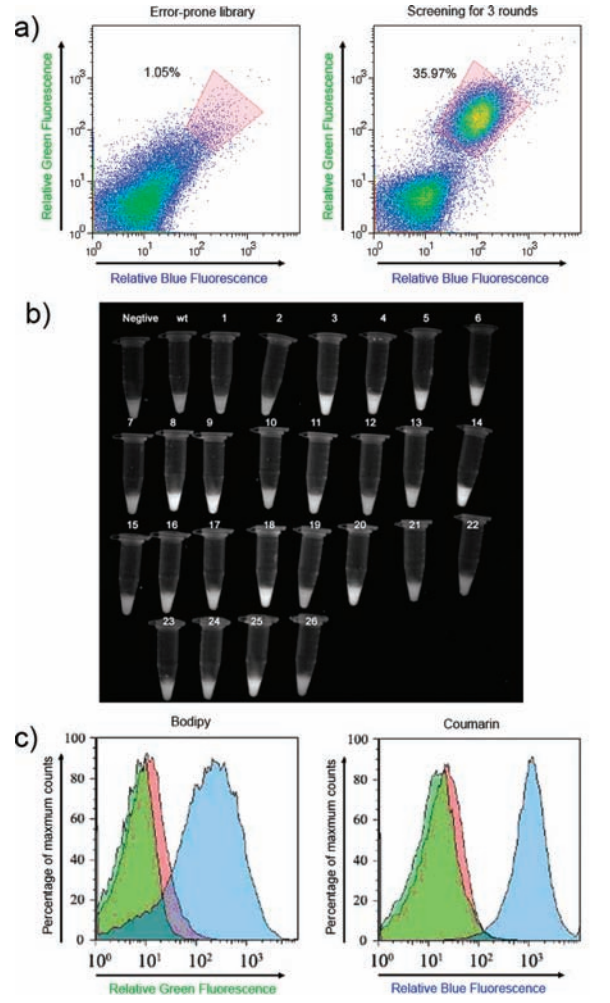


Figure 4. (a) FACS analysis of the error-prone PCR library of CgtBΔC30-MalE and enrichment of the *E. coli* population with significantly higher fluorescence entrapment after three rounds of screening. The shaded area (pink) defines the gate employed to sort cells in the top 0.3%−1.0% of both green (bodipy) and blue (coumarin) fluorescence intensity. A highly fluorescent population was evolved after 3 rounds of enrichment. (b). Fluorescence intensity (under UV irradiation) of the 26 mutants E1-E26 randomly selected from the enriched population obtained after three rounds of sorting the EP-PCR library. WT: the parent enzyme, CgtBΔC30-MalE. (c) FACS histogram analysis of the green and blue fluorescence of CgtBΔC30-MalE (pink), its randomized library (green), and the S42 mutant (blue) in *E. coli* cells.

galactose derivatives as acceptors. Indeed, in addition to the β-1,3-galactosyltransferase CgtB reported in this work, we have shown that this *E. coli*-based screening system can be readily applied to sialyltransferases, α-1,4-galactosyltransferases, and an α-1,3-galactosyltransferase (unpublished work). Importantly, it should be possible to also use other sugar transport proteins to further extend the application of this strategy.

Having established a valid reporter of galactosyltransferase activity, DNA encoding amino acids 1−272 of CgtB was subjected to error-prone PCR (EP-PCR) and cloned into the pCW-MalE-C vector to generate a library of CgtBΔC30-MalE variants with ∼2−5 mutations per gene. The library was transformed into *E. coli* JM107* to generate $\sim 2 \times 10^7$ colonies. Three iterative rounds of sorting were performed by FACS to enrich those cells in the top 0.3−1.0% of both green and blue fluorescence intensity (Figure 4a). After each round of sorting, the "positive" cells were collected into growth medium and plated on agar prior to the next round of enrichment. Twenty

Table 1. Specific Activities of CgtBΔC30-MalE and Selected Mutants Using Benzyl Glycosides of *N*-Acetylgalactosamine as Acceptors

enzyme	mutations	specific activity (mU/mg)	
		5 mM GalNAc-α-OBn[a]	5 mM GalNAc-β-OBn[a]
CgtBΔC30-MalE	NONE	178	18
E8[b]	E234D	796	104
E14[b]	N26K, C40S, K50R, K151E, L227G	1110	111
E20[b]	K130I, L70I	593	51
S42[c]	N26K, L68I, K151E, L227G, E234D	2510	274
HS1[d]	K151E	301	55
HS2[d]	K151E, E234D	1205	143
HS3[d]	L227G	590	65

[a] Errors in the data were less than 8%. [b] Error-prone PCR library. [c] DNA shuffling library. [d] Hot spot mutants constructed using site directed mutagenesis.

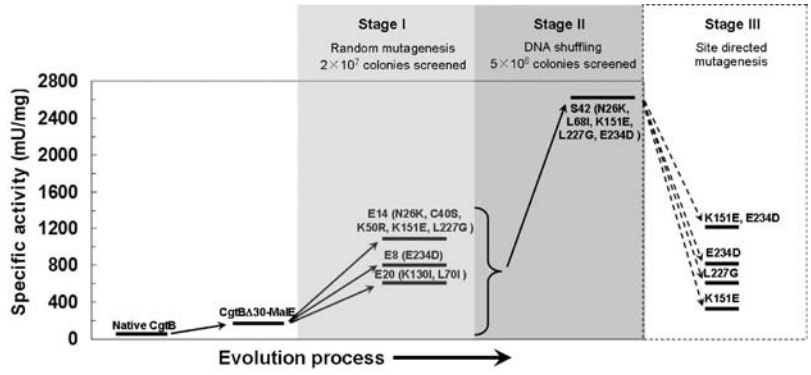


Figure 5. Evolution pathway of β-1,3-galactosyltransferase CgtB from *C. jejuni* OH4384. In an earlier report, CgtBΔC30-MalE was constructed and showed a much higher activity than the native CgtB.[29] In our current work, the CgtBΔC30-MalE gene was randomized in stage I using the Gene Morph II Mutagenesis kit, ∼2 × 10^7 colonies were screened, and several improved mutants were obtained. In stage II, the positive mutants from stage I were shuffled together with the parental CgtBΔC30-MalE gene, 5 × 10^6 colonies were screened, and the best mutant in this study, S42, was isolated. In stage III, specific mutants were constructed by site-directed mutagenesis and their effects on the activity improvement were determined and analyzed. The specific activities were measured using GalNAc-α-OBn as acceptor.

six clones were randomly selected after the third round of sorting and about 40% of them showed much higher activities than the wild type CgtBΔC30-MalE, as indicated by the fluorescent intensity of the cells under UV irradiation in the presence of **2** and **3** (Figure 4b). Indeed, the specific activities of these mutants for transfer of galactose to simple glycosides of *N*-acetylgalactosamine showed significant improvements over that of the parent enzyme CgtBΔC30-MalE (*vide infra*). Eleven of the CgtB genes carrying apparent beneficial mutations were sequenced and five of these were found to be of identical composition (equivalent to the mutant E14 in Table 1, see the Supporting Information). This finding suggests that these mutants have the highest activities in the entire EP-PCR library and have therefore been substantially enriched during the FACS screening.

In an effort to improve upon the already significant gains in glycosyltransferase activity obtained via the EP-PCR approach, CgtB genes on plasmids from the enriched population (Figure 4a) were backcrossed with the parental CgtB gene by DNA shuffling, giving rise to a second library consisting of ∼5 × 10^6 colonies. After three rounds of FACS as above, a total of 48 positive clones were assessed for improvements in activity relative to CgtBΔC30-MalE, as suggested by the fluorescence intensity of the cells under UV irradiation in the presence of **2** and **3**. From within this group, the CgtB genes from eight clones with improved fluorescence retention evident to the naked eye were sequenced. Among these proteins, the most active enzyme S42 (Figure 4c), contained five amino acid mutations. As previously observed after FACS sorting of the EP-PCR library, the enriched pool contained multiple enzymes bearing the same mutations. In this case, three of the eight sequenced CgtB genes were identical to the most active mutant S42, further substantiating the capacity of this screening method to effectively enrich the most active colony in a large population. Inspection of the mutations observed in positive clones from the EP-PCR library and the shuffling library identified three "hot spots", Lys151Glu (11 times), Leu227Gly (9 times), and Glu234Asp (8 times), sites at which mutations appear particularly beneficial to activity.

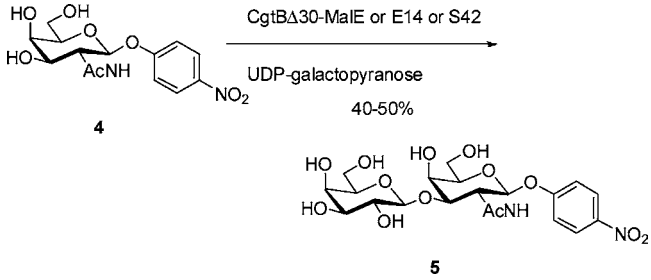
To investigate effects of each hot-spot mutation on catalytic activity, the individual mutants K151E, L227G, and K151E/E234D were constructed by site-directed mutagenesis and the kinetic parameters of each were determined along with those for the wild-type CgtB and some of the most active variants from the EP-PCR and DNA shuffling libraries, which included the E234D single mutant. The enzymes were expressed and purified to greater than 90% homogeneity as described in the Supporting Information. Specific activities of these mutant enzymes with UDP galactopyranose donor and the α- and β-linked benzyl glycosides of GalNAc (GalNAc-OBn) as acceptors were determined and the data obtained are presented in Table 1. In accordance with their fluorescence profiles, all of the variants were significantly more active than the parent CgtBΔC30-MalE. Mutant S42 showed the highest activity, an increase of 14-fold over the parent, irrespective of the anomeric configuration of the acceptor. Insight into the origin of this rate increase is obtained from data on the three individual hot-spot mutations (K151E, L227G, E234D). Activities of these three mutants were 1.7-, 3.3-, and 4.5-fold higher than wild-type when using αGalNAc-OBn acceptor, and 3.0-, 3.6-, and 5.8-fold higher for βGalNAc-OBn. Indeed, similar activity enhancements were seen for each substrate with each mutant, clearly showing that we have eliminated false positive mutants arising from the

Table 2. Kinetic Parameters for the Transfer of Galactose from UDP-Galactopyranose to G_{M2} Oligosaccharide by Selected CgtB Mutants

	substrate					
	UDPGal[a,d]			G_{M2} oligosaccharide,[d]		
enzyme	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M[c]	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M[c]
CgtBΔC30-MalE	54	0.22	250	460	23	20
S42	34	0.03	1160	91	0.3	320
K151E	150	0.24	630	1080	26	42
E234D	65	0.07	900	120	1.4	88
L227G	55	0.11	490	140	2.3	61
K151E/E234D	170	0.08	2200	410	2.8	150

[a] Kinetic measurements at 5 mM G_{M2} oligosaccharide and variable UDP-Gal concentration. [b] Kinetic measurements at 10 mM UDP-Gal and variable G_{M2} oligosaccharide concentrations. [c] Measured in min^{-1} mM^{-1}. [d] The errors in the k_{cat} and K_M data are ~5%, and the errors in the k_{cat}/K_M data are less than 10%.

Scheme 2. Galactosylation of **4** Catalyzed by Native Enzyme or Mutants, Including E14 and S42, Yields the Same Product



generation of dye binding sites. Our strategy of simultaneous screening against differentially labeled substrates therefore appears to have been successful. It is also of interest that the effects of the hot spot mutations appear to be independent of each other, thus non-interacting (Figure 5). On the other hand, the contributions of the N26K and L68I mutations to the rate enhancement seen in S42 seem to be very small, if not negligible, since K151E, L227G, and E234D already account for most of the activity increase.

To assess the effects of these mutations on glycosylation of a physiologically relevant substrate, kinetic parameters for the transfer of galactose to the G_{M2} oligosaccharide were determined (see Supporting Information for structures of acceptor and product), and these are shown in Table 2. The specificity constant (k_{cat}/K_M) of the double mutant K151E/E234D for G_{M2} is greater than that of either single mutant K151E or E234D but lower than that of the S42 mutant, which contains all three "hotspots" (K151E, E234D, L227G), indicating that the effects of these mutations are essentially additive for acceptor kinetics. The influence of the K151E mutation is exerted via an improvement in k_{cat}, while L227G and E234D affect activity via decreased K_M values. Indeed the K_M value of S42 for G_{M2} is some 80-fold lower than that of the parent enzyme. With the tetrasaccharide acceptor, a 16-fold improvement in k_{cat}/K_M was observed for S42 over the parent enzyme.

To confirm that the observed rate enhancements are of practical use, we sought to establish that the mutants generated the same reaction products as the parent enzyme. The galactosylation of *p*-nitrophenyl 2-acetamido-β-D-galactopyranoside **4** was compared using CgtBΔC30-MalE, and mutants E8, E14, E20, and S42 (Scheme 2). Analysis of the reactions by thin-layer chromatography provided initial confirmation that the same products were formed in each case. This was then confirmed for the products formed by CgtBΔC30-MalE, E14, and S42 through their isolation and analysis by ^{1}H NMR spectroscopy.

Scheme 3. Some Useful Transformations Catalyzed by CgtB Galactosyltransferase Mutant S42

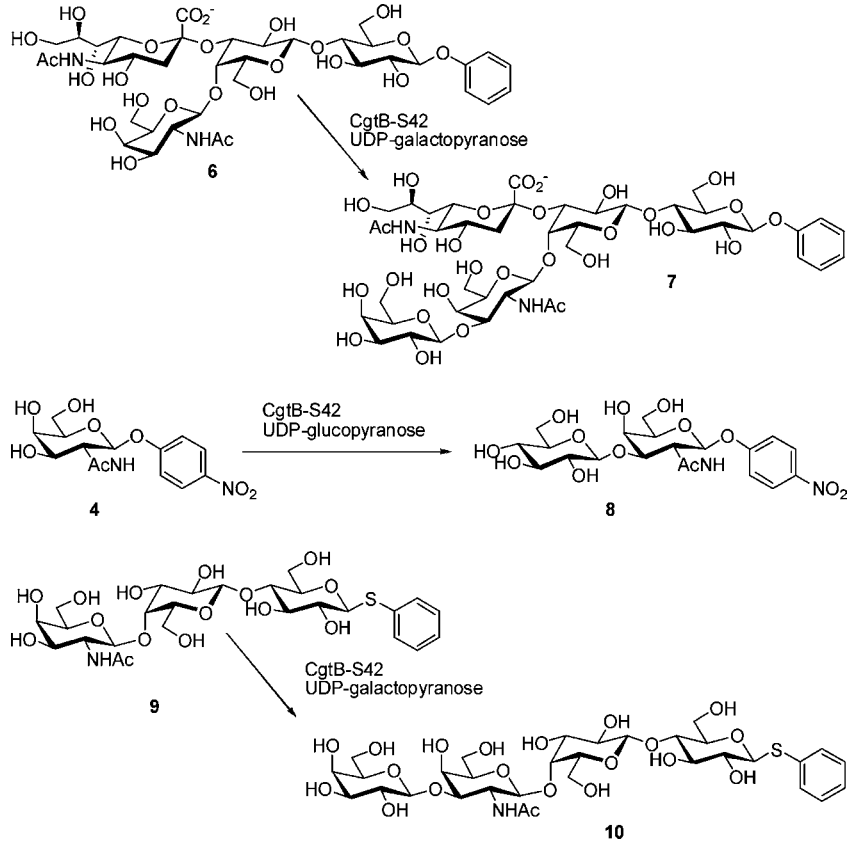


Table 3. The Specific Activities (mU/mg) of CgtBΔC30-MalE and Mutant S42 with Different Donor and Acceptor Substrates

acceptor	donor	CgtBΔC30-MalE spec. act.	S42 spec. act	enhancement (fold)
Asialo G_{M2}-FCHASE[a]	UDP-galactose	100	2800	28
GalNAc-α-OBn	UDP-glucose	<0.13	38.0	>298

[a] For structures of FCHASE oligosaccharides, see the Supporting Information.

Similarly, the pentasaccharide product formed upon galactosylation of G_{M2} oligosaccharide with CgtB mutant S42 had an identical ^{1}H NMR spectrum to that obtained from an authentic sample of G_{M1a} oligosaccharide.

The synthetic potential of the best mutant identified, S42, was demonstrated by synthesis of G_{M1a} phenyl glycoside (**7**, G_{M1}OPh) from its precursor tetrasaccharide G_{M2}OPh **6** as shown in Scheme 3a, and further explored by testing its donor sugar specificity using acceptor **4**. In keeping with the specificity of wild-type CgtB, neither UDP-GlcNAc nor UDP-GalNAc served as donors. However, surprisingly UDP-Glc served as a substrate for CgtB S42. Specific activity was enhanced at least 300-fold over the parent enzyme, the activity of which was only barely detectable with UDP-Glc donor (Scheme 3b, Table 3). This tolerance of S42 toward structural modifications at C-4 of the galactose donor suggests its potential in the chemoenzymatic synthesis of C-4 modified variants of ganglioside G_{M1} or *O*-glycans containing the T-antigen structural subunit. Moreover, S42 should be an excellent starting point for the evolution of novel glycosyltransferases with new substrate profiles.

Another important class of synthetic targets is that of the asialo G_{M1a}-type oligosaccharides such as **10**. While these compounds can be obtained by sialidase-catalyzed degradation of the more complex G_{M1a} oligosaccharides (Scheme 1c), this reaction is typically only performed on an analytical scale.[34] Further, in addition to the fact that such a degradative approach is inherently inefficient, the commercial ganglioside preparations that are needed are isolated from natural sources such as bovine brain and canine blood, and are, thus, expensive and potential vectors for cross-species contamination by prions or other infectious agents, and, therefore, are unattractive.

Since asialo G_{M2}-type oligosaccharides (such as **9**) are available from lactosides via a single GalNAc transferase-catalyzed step, transfer of a Gal residue to asialo G_{M2}-type oligosaccharides by *Campylobacter* CgtB represents a viable approach toward the assembly of asialo G_{M1a} structures. Mutant S42 transfers galactose to the asialo G_{M2} oligosaccharide **9** 30-fold faster than do either the parental enzyme CgtBΔC30-MalE, or the other CgtB variants found in previous work.[29] Thus, use of S42, in conjunction with improved GalNAc transferases, would make this route practical and economically viable.[35,36]

Conclusion

The use of ultra-high-throughput screening methods such as FACS sorting allows the interrogation of mutant libraries of a size that would be inaccessible by plate-based assays on either cost or time considerations. This study generalizes our previous approach,[27] both by demonstrating its application to neutral sugar transferases and by showing that simultaneous use of chemically and spectroscopically distinct fluorescent tags on the same sugar acceptor removes the selective bias toward evolution of dye binding sites. It is probable that, by the use of acceptors with both different sugar moieties and different fluorescent dyes, one could create positive and negative selection pressures for the isolation of highly active GT mutants with tailored alterations in substrate specificity. The merit of this methodology lies not only in its thousand times greater throughput than previous microtiter assays, but also in its use of *E. coli* cells as microreactors. In such a small system, the volumes of reagent needed for screening are millions-fold less than that required for robotic screening, resulting in huge savings in cost and availability of reagents.[37]

This work paves new avenues for the engineering of almost all GTs. Given the available information on sequence, structure, and function of sugar transport proteins (http://www.membranetransport.org/; http://www.tcdb.org/),[38] it may be possible to develop a repertoire of engineered *E. coli* strains that can be used as a tool box for the directed evolution of many GTs. This technique will greatly increase the rate at which novel engineered GTs are identified, thereby contributing to the continued development of glycoscience.

Acknowledgment. We thank Amir Aharoni for helpful discussions, Hongming Chen for providing activated bodipy dye, and Marie-France Karwaski for technical help. We acknowledge the China Scholarship Council for a fellowship to G.Y., the Michael Smith Foundation for Health Research for a fellowship to J.R.R., the Canada Research Chairs Program for a Tier 1 chair to S.G.W. and the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research for grant support.

Supporting Information Available: Structures, syntheses, sequencing results, purification of mutants, and spectra as mentioned in text. This material is available free of charge via the Internet at http://pubs.acs.org.

JA104167Y

(34) Iwamori, M.; Kaido, T.; Iwamori, Y.; Ohta, Y.; Tsukamoto, K.; Kozaki, S. *J. Biochem.* **2005**, *138*, 327–334.
(35) Gilbert, M.; Karwaski, M. F.; Bernatchez, S.; Young, N. M.; Taboada, E.; Michniewicz, J.; Cunningham, A. M.; Wakarchuk, W. W. *J. Biol. Chem.* **2002**, *277*, 327–337.
(36) Yamashiro, S.; Haraguchi, M.; Furukawa, K.; Takamiya, K.; Yamamoto, A.; Nagata, Y.; Lloyd, K. O.; Shiku, H.; Furukawa, K. *J. Biol. Chem.* **1995**, *270*, 6149–6155.
(37) Agresti, J. J.; Antipov, E.; Abate, A. R.; Ahn, K.; Rowat, A. C.; Baret, J. C.; Marquez, M.; Klibanov, A. M.; Griffiths, A. D.; Weitz, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4004–4009.
(38) Saier, M. H. *Mol. Microbiol.* **2000**, *35*, 699–710.