BIOCHEMISTRY

Directed evolution of an α1,3-fucosyltransferase using a single-cell ultrahigh-throughput screening method

Yumeng Tan[1]*, Yong Zhang[1]*, Yunbin Han[1,2], Hao Liu[1], Haifeng Chen[1], Fuqiang Ma[1,3], Stephen G. Withers[4], Yan Feng[1], Guangyu Yang[1]†

Fucosylated glycoconjugates are involved in a variety of physiological and pathological processes. However, economical production of fucosylated drugs and prebiotic supplements has been hampered by the poor catalytic efficiency of fucosyltransferases. Here, we developed a fluorescence-activated cell sorting system that enables the ultrahigh-throughput screening ($>10^7$ mutants/hour) of such enzymes and designed a companion strategy to assess the screening performance of the system. After three rounds of directed evolution, a mutant M32 of the α1,3-FucT from *Helicobacter pylori* was identified with 6- and 14-fold increases in catalytic efficiency (k_{cat}/K_m) for the synthesis of Lewis x and 3′-fucosyllactose, respectively. The structure of the M32 mutant revealed that the S45F mutation generates a clamp-like structure that appears to improve binding of the galactopyranose ring of the acceptor substrate. Moreover, molecular dynamic simulations reveal that helix α5, is more mobile in the M32 mutant, possibly explaining its high fucosylation activity.

INTRODUCTION

Fucosylation is often the final step in the biosynthesis of a number of oligosaccharides and glycoconjugates that are essential for a wide range of biological processes in eukaryotes, including angiogenesis, fertilization, cell adhesion, inflammation, and tumor metastasis (*1*, *2*). Given their bioactivity and utility, the synthesis of fucose-containing glycans such as cancer-related antigens (*3*, *4*) and human milk oligosaccharides (*5*, *6*) is of substantial applied interest (Fig. 1A). Natural fucosylation reactions are performed by fucosyltransferases (FucTs; Fig. 1B), which catalyze the transfer of fucose from guanosine diphosphate β-L-fucose (GDP-Fuc) to various oligosaccharides, glycoproteins, or glycolipids. Furthermore, several retaining α-L-fucosidases are known to also have transfucosylation activity (Fig. 1B), and this has been exploited for the synthesis of fucose-containing glycans (*7*, *8*). Although both FucTs and fucosidases have been biochemically characterized in considerable detail, their broader applications in biotransformation or metabolic engineering has, to date, been hindered by the low catalytic activity of these enzymes. Thus, obtaining more efficient enzymes that are tailored for the synthesis of specific oligosaccharide and glycoconjugates would be of great scientific and industrial interest.

Directed evolution has long been a key strategy to generate enzymes with desired properties (e.g., high catalytic activity) through consecutive rounds of diversification and selection (*9–11*). One of the crucial factors in directed evolution strategies is the development of suitable screening or selection techniques, which would considerably increase the chance of obtaining mutants with desired properties from large mutagenesis libraries and reduces the time and cost (*12*). Recently, fluorescence-activated cell sorting (FACS) has shown great promise for significantly advancing screening capabilities for glyco-

syltransferases (GTs) (*12–14*), and one of the important strategies involving FACS relies on the exquisite substrate specificity of lactose permeases (LacY) that are located in the cell membrane. The key to the screening method is to design appropriate fluorescent substrates that can be accessed by GTs inside the cell after transport by the LacY "gate" but for which the glycosylation products are not eligible for specific transmembrane transporter of sugars and are therefore accumulated within the cell (*14*).

In theory, any enzymes catalyzing a modification that interferes with recognition of transmembrane sugar transporters should be amenable to screening by similar high-throughput strategies. However, the only such attempts reported to date have targeted the C-3 of the galactose moiety of the lactose(Galβ1,4Glc)/LacNAc(Galβ1,-4GlcNAc), for example, use of α2,3-sialyltransferase (*12*) and β1,3-galactosyltransferase (*14*). Fucosylation enzymes can indeed modify LacNAc or lactose substrates at various positions. For example, an α1,3-FucT (FutA) can add a fucose to the C₃-OH position of the GlcNAc group of the disaccharide LacNAc (*15–17*); an α1,2-FucT (FutC) catalyzes the fucose transfer to the C₂-OH position of galactose (*18*, *19*), while an α-L-Fucosidase (FucD) can add the fucose to the C₃-OH position of galactose (*8*). We reasoned that the galactoside transporter, LacY, could translocate a range of disaccharides or their analogs into *Escherichia coli* cells, and the fucosylated modifications by enzymes may well alter LacY recognition, making it possible to develop new screening methods for the engineering of these important enzymes.

In this study, we designed and synthesized several fluorescently labeled substrate derivatives for fucosylation reactions. These fucosylation enzymes generate fluorescent oligosaccharide products that are readily retained in cells, thus facilitating development of a new FACS-based fucosylation enzyme directed evolution system. Moreover, we developed a competitive allele-specific TaqMan polymerase chain reaction (PCR) method to enable the rapid quantitative assessment of single-cell high-throughput screening efficiency; we are unaware of other instances in which this technology was used for enzyme screening. As a model, we used this new established FACS platform for the directed evolution of FutA and identified a range of improved mutants by three rounds of directed evolution. Our solution of the crystal structure of the best-performing mutant

[1]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China. [2]State Key Laboratory Cultivation Base for TCM Quality and Efficacy, College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China. [3]CAS Key Laboratory of Biomedical Diagnostics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, China. [4]Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada.
*These authors contributed equally to this work.
†Corresponding author. Email: yanggy@sjtu.edu.cn
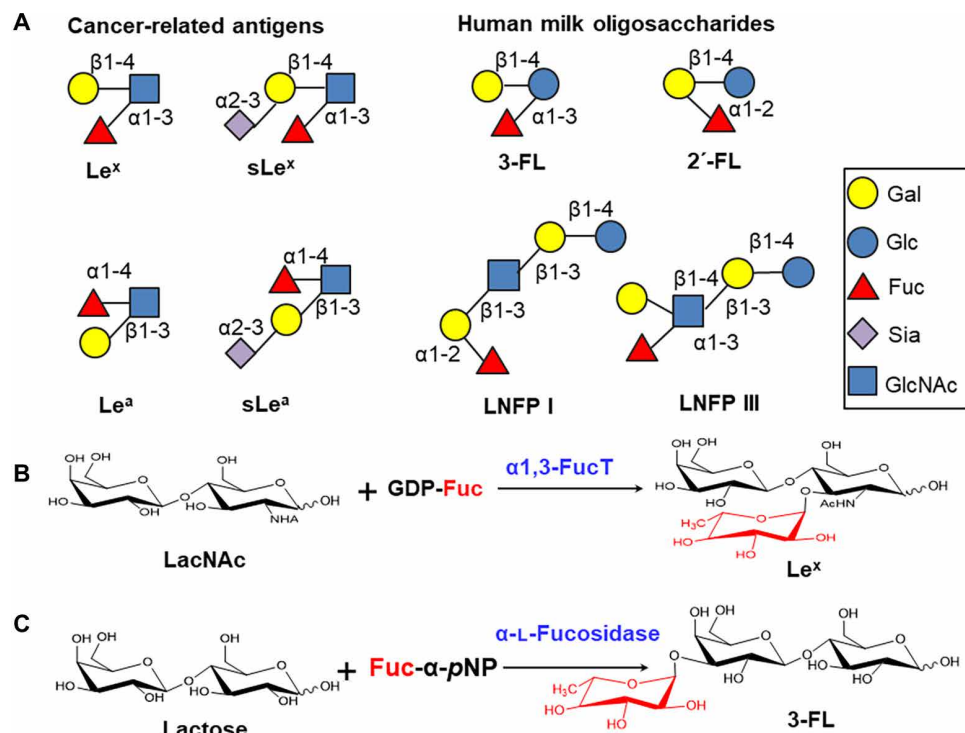
Fig. 1. Fucosylation reactions catalyzed by fucosylation enzymes. (A) Structures of various disease-associated and human milk oligosaccharides. (**B**) Schematic depiction of fucosylation reactions catalyzed by FucTs and fucosidases. Typically, an FutA catalyzes the L-fucose transfer from a GDP-Fuc donor substrate to a LacNAc acceptor substrate via an α1,3-linkage, forming the Le[x]. α-L-Fucosidases can also transfer a fucose moiety from the para-nitrophenyl α-L-fucopyranoside (Fuc-α-pNP) to lactose, forming a 3-fucosyllactose.

in complex with the donor substrate GDP-Fuc provided valuable insights into the mechanism through which fucosyl-transferring enzymes recognize their substrates.

RESULTS

LacY substrate specificity for various fucosylated products

To test whether the fucosylation modification can entrap enough fluorescent signal in the cell by affecting LacY substrate specificity, we chose FutA from *Helicobacter pylori* strain NCTC 11639 (GenBank accession number AAB81031.1) (15) for glycosyl modification. Since LacNAc is a natural substrate of FutA, we designed and synthesized fluorescently labeled derivatives of LacNAc: Compound 1 is the LacNAc conjugated with a bodipy fluorophore; compound 2 is the LacNAc conjugated with a coumarin fluorophore (Fig. 2A). These LacNAc derivatives with two chemically distinct fluorophores enable screening strategies that minimize the probability of selection for an altered dye-binding site rather than altered activity for an enzyme variant.

An *E. coli* strain JM107*, which has galactosidase (LacZ) knocked out (to avoid lactose hydrolysis) (12) but expresses a GDP-fucose synthetase (FKP), a bifunctional enzyme that has both fucokinase and fucose-1-phosphate guanylyltransferase activities, was used to test the ability of cells to retain the fucosylated products. Theoretically, acceptor substrates are capable of entering and exiting a cell freely via the plasma membrane–localized LacY transporter, and once inside the cell, they should be subjected to fucosylation by corresponding enzymes to generate trisaccharide products. Given that these newly formed trisaccharide products are not substrates for the LacY transporter, which primarily transports disaccharides (20, 21),

the new trisaccharides will be, in effect, entrapped inside of cells, thereby allowing unreacted substrates to be washed away without significant loss of fluorescence intensity (Fig. 2B).

Pursuing this strategy, we incubated FutA-expressing cells FutA[(+)] and empty vector pUC18 control cells FutA[(−)] with fucose and fluorescently labeled 1 and 2. After several rounds of washing with LB medium and phosphate-buffered saline (PBS) buffer, we observed relatively strong fluorescence signals for the FutA[(+)] samples under ultraviolet light but not for the FutA[(−)] control (Fig. 2C). Furthermore, FACS analysis demonstrated that the cell expressing the recombinant enzyme FutA had stronger green (bodipy conjugates) and blue fluorescence (coumarin conjugates) signals than the control cell (Fig. 2D).

In addition, to explore whether this trisaccharide product entrapment strategy can be more broadly applied to other fucosylation enzymes, including FutC from *H. pylori* (19) and FucD from *Anaerolinea thermophila* (22), we also synthesized diverse fluorescently labeled lactose derivatives as transfer acceptors (fig. S1A) and tested the fluorescence retaining capacity of cells expressing these two enzymes. As with FutA, these cells expressing the recombinant FutC or FucD enzyme had more intense fluorescence signals than did the empty vector control cells (fig. S1, C and D). These assays indicate that fluorescently labeled disaccharide acceptor substrates can be taken up into cells via the lactose permease LacY, but the fucosylation enzymes product, fluorescent trisaccharides, are transported much less efficiently, leading to product accumulation inside cells.

Note that there are also other sugar transmembrane transporters in *E. coli* cells (23), i.e., sugar efflux pumps such as SetA, which is capable of transporting a range of sugars and sugar analogs (24). To
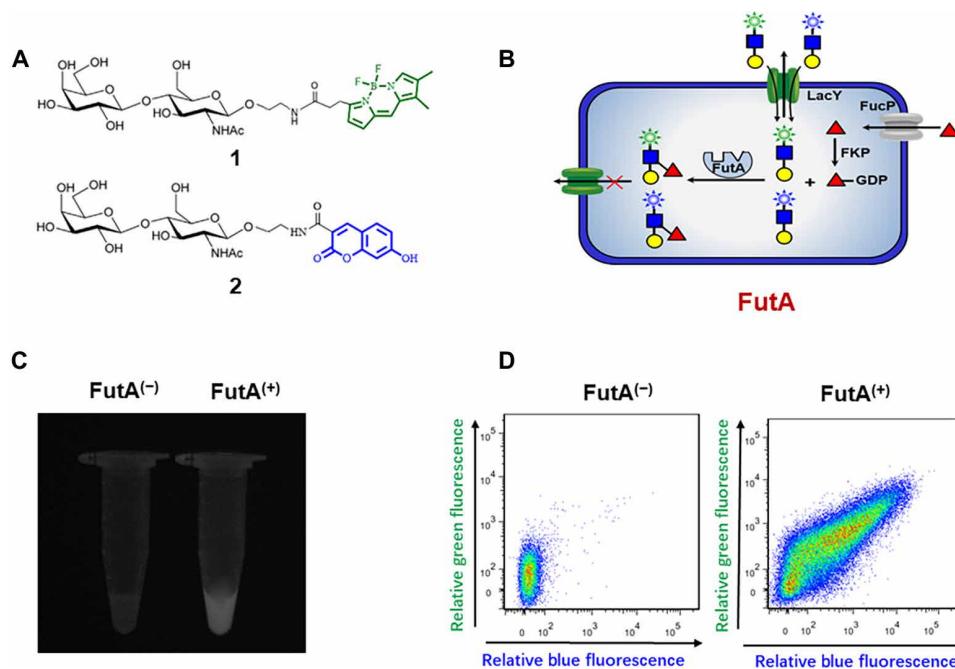
Fig. 2. Scheme for the product entrapment strategy and the established cell-based fucosylation assay for FACS screening. (**A**) Two kinds of fluorescently labeled LacNAc derivatives (1 and 2) were designed and synthesized for cell entrapment analysis. (**B**) Fluorescently labeled acceptor substrates are transported into the cell via LacY; fucose enters into the cell via a fucosyl transporter (FucP) and was converted into GDP-fucose donor substrate by GDP-fucose synthase (FKP). After incubation and washing, *E. coli* cells expressing fucosylation enzymes accumulate fluorescent trisaccharide enzyme products, as the LacY transport rate for such products is significantly reduced compared to their disaccharide substrate form. Thus, the fluorescence intensity accumulation inside cells carries information about the catalytic activity of the fucosylation enzymes being assayed/screened. These cells with FutA activity can be further isolated using FACS. (**C**) Visualization of fluorescence entrapment within FutA$^{(+)}$ and FutA$^{(-)}$ cells under an ultraviolet light. (**D**) Flow cytometry profiles of FutA$^{(+)}$ and FutA$^{(-)}$ cell fluorescence after 30-min incubation with 1.5 mM fucose, 0.5 mM bodipy-LacNAc, and coumarin-LacNAc, followed by a washing step. Green and blue signals represent cells retaining bodipy and coumarin fluorescently labeled oligo-saccharides, respectively.

address whether other transporters interfere with fluorescent fucosylation product entrapment, *E. coli* cells with empty vector control or the cells with plasmids for FutA, FutC, or FucD, were reacted with substrates, followed by washing with PBS buffer and with fresh M9 minimal medium. Given that we did not conduct LB medium wash, only the fluorescence outside the cell was washed away. These cells, which retained most of the fluorescently labeled di- and trisaccharide compounds (in some cases), were incubated in fresh M9 medium at 37°C for 30 min, and the fluorescence leakage in the culture supernatants was analyzed. We did observe some fluorescence leakage in the supernatants for all of the cell types; however, the intensity of the fluorescence leakage in the supernatant for the FutA, FutC, and FucD cells was only 15, 8.6, and 20.3%, respectively, as strong as the intensity for the control cells, indicating that the oligosaccharide products can indeed be entrapped efficiently inside cells (fig. S2A). There might be other transporters, which recognize and transport sugar products, but their efficiency is apparently much slower than that of LacY permease. Our results thereby established that cells expressing the fucosylation enzymes could retain their fluorescent trisaccharide products for a sufficiently long time to enable FACS-based screening following directed evolution targeted at improving fucosylation enzyme activity (fig. S2B).

To boost the screening efficiency, we further optimized a series of parameters for the screening, including fluorescence signal stability, induction time, reaction time, and substrate concentration. The fluorescence signals of the FutA$^{(+)}$ cells were stable for more than 1 hour at room temperature, providing a more-than-ample window for FACS-based screening (fig. S3A). For the expression of recombinant fucosylation enzymes, we found that the cells induced expression at around 18 hours produced the strongest fluorescence signals, suggesting that the protein expression achieved by this point is highly amenable to these entrapment assays (fig. S3B). The fucosylation reaction was essentially linear within 1 hour while almost saturated at about 1.5 hours. To distinguish between positive and negative mutants in subsequent experiments, we therefore set the reaction time to 0.5 hours (fig. S3C). Last, we found that the optimal substrate concentrations for the screening system are 1.5 mM donor fucose and 0.5 mM acceptors (fig. S3D).

Evaluation of the FACS-based fucosylation activity screening system

We next determined whether the cells expressing the recombinant fucosylation enzymes could be used with FACS. The first and most obvious test was to see whether FutA$^{(+)}$ cells could be sorted from FutA$^{(-)}$ control cells via the FACS system. Therefore, we analyzed FACS efficiency by preparing cell mixtures with 1:10, 1:100, 1:1000, and 1:10,000 FutA$^{(+)}$/FutA$^{(-)}$ cell ratios. After a substrate incubation, the cells were harvested, washed, and analyzed via FACS; the top 0.5% of cells with the strongest fluorescence intensities were sorted to minimize possible false positives. To examine FACS enrichment efficiency, we plated the sorted cells and verified the presence of the FutA gene by colony PCR. Notably, a single round of high-stringency FACS sorting archived a population with 100% FutA$^{(+)}$ cells, even for the 1:10,000 mixture (table S1). Considering that previous direct

evolution studies reported fold increases between 40 and 330 for a single round of FACS (25–27), this system showing >10,000-fold enrichment ability highlights its extremely excellent performance.

As this novel screening system is so efficient in identifying FutA$^{(+)}$ cells that it rapidly saturated the conventional enrichment test, we further developed a more stringent and quantitative approach to evaluating its enrichment performance. Instead of using completely inactive FutA$^{(-)}$ control cells, we adopted *E. coli* cells with markedly reduced FutA protein expression than normal FutA$^{(+)}$ cells to test FACS efficiency (Fig. 3A). We created the cells expressing FutA at a relatively low level with the robust ribosomal binding site (RBS)– ATG spacing technique. Briefly, the 13th base (cytosine) of the pUC18 plasmid polyclonal area, which is positioned at the back of the ribosome binding site, was deleted to generate the FutA$^{(+)}$-RBS cells. High-performance liquid chromatography (HPLC) analysis of cell extracts showed that the catalytic activity of enzyme in the FutA$^{(+)}$-RBS cells decreased around three times in comparison with the normal FutA$^{(+)}$ cells. Compared with distinguishing between FutA$^{(+)}$ and completely inactive FutA$^{(-)}$, this high/low activity screening is obviously more relevant to a real directed evolution experiment.

We prepared mixtures of FutA$^{(+)}$ and FutA$^{(+)}$-RBS cells and performed one round of FACS (Fig. 3B). The enriched cell populations were plated and grown as colonies. We then used a technique named competitive allele-specific TaqMan PCR (with TaqMan probes) to distinguish between the FutA$^{(+)}$ and FutA$^{(+)}$-RBS colonies (exploiting the aforementioned deletion of the 13th base of the pUC18 polyclonal area). Encouragingly, after this single FACS round, we observed an 18-fold enrichment for FutA$^{(+)}$ cells (Fig. 3C), indicating that this new FACS system can effectively offer the performance required to enable sensitive screens for improved fucosylation activity.

Directed evolution of FutA

Targeting increased catalytic efficiency toward acceptor LacNAc substrate, we performed three rounds of directed evolution for FutA. In the first round, an error-prone PCR was used to generate a random mutagenesis library; the second round used an ordered recombination mutagenesis (ORM) to accumulate optimal mutations; and the third round applied a combinatorial active-site saturation testing (CAST) (28) approach to specifically mutagenize 16 selected residues on the basis of structural analysis.

The first round of directed evolution introduced random mutations in FutA encoding sequence by the error-prone PCR and generated a library of 4×10^6 unique colonies with two to five mutations per *FutA* gene variant. We used three FACS sorting iterations to successively enrich the top 0.5% of cells in terms of both green (bodipy conjugates) and blue (coumarin conjugates) fluorescence intensity. After each of the three sortings, the *FutA* gene of each "positive" cell was amplified using the FutA-F and FutA-R primers (table S4), and the PCR product was introduced into a new host cell. To confirm that this error-prone PCR-based round was indeed producing/identifying FutA variants with improved catalytic activity, we randomly selected 20 clones after the third round of FACS. Assays using LacNAc as the substrate showed that 40% of these mutant variants were at least 30% more efficient than wild-type FutA in transferring fucose onto LacNAc. Among these FutA variants, three mutants M11 (Y199N/V368A/ D407N), M12 (S45F), and M13 (E340D) exhibited roughly twofold improvement over the activity of the wild-type enzyme (table S2).

The next directed evolution round used the ORM to iteratively accumulate optimal mutations at four selected amino acid sites. The top four sites were selected on the basis of initial identification from the error-prone PCR round and on follow-up experiments that tested and ranked the individual contribution of those single point mutations (fig. S4). Considering that the single mutants S45F and E340D are already available, thus, we constructed other three single mutants Y199N, V368A, and D407N (designated M21, M22, and M23) to evaluate individual contributions. Assays showed that the single mutation E340D resulted in more pronounced improvement in catalytic activity than did other beneficial mutations, and therefore, the ORM path for this mutant was initiated. A new mutant, designated M24, was constructed by a combination of E340D and V368A. The activity of M24 further improved 2.3-fold against the wild type. By introducing the S45F mutation into M24, we generated the M25 variant. Adding Y199N mutation into the M25 further created a mutant named M26 (S45F/Y199N/E340D/V368A). Assay analysis revealed that the M26 obtained an overall 2.74-fold increase of specific activity against that of the wild type (table S2). Thus, the M26 was chosen as the template for the third round of directed evolution.

It is known that amino acid residues near substrate-binding pockets often have particularly profound effects on enzyme catalysis (28). The mutagenesis component of the third round of directed evolution exploited this direction. This is a semi-rational method guided by (i) the crystal structure of wild-type FutA [Protein Data Bank (PDB): 2NZY] (29) and (ii) HotSpot Wizard 2.0 for automated identification of hotspots and smart libraries for protein engineering (30). In that, we selected all of 51 residues within 8-Å distance from the boundary surface of LacNAc and GDP-Fuc for target mutagenesis (fig. S5A). Moreover, the HotSpot Wizard algorithm provided a substrate-binding residue prediction based on a comprehensive sequence/ structure alignment with the reported complex structures of GT acceptor. Sixteen predicted hotspot residues (V30, W33, E37, E38, K40, E41, N44, V46, G72, P74, L75, Y92, D127, H131, K223, and N226) were picked out for site-directed mutagenesis (fig. S5B). Upon applying the CAST strategy on FutA, the following amino acid clusters were defined V30/W33, E37/E38/K40/E41, N44/S45/V46, G72/ P74/L75, D127/R128/H131, K223/N226, and Y92. Seven small libraries, A to G, respectively, were then created separately using complete saturation (i.e., simultaneous randomization of each cluster). These libraries were combined into a pooled library (library 2) comprising about 2×10^4 variants that was screened using the established FACS system to identify mutants with improved activity for both synthesized fluorescent substrates 1 and 2. Subsequently, we randomly selected and characterized 20 enriched variants by spectrophotometric and HPLC analysis. A mutant named M32 (S45F/ D127N/R128E/H131I/Y199N/E340D/V368A) was identified as the most active FutA variant, exhibiting a more than 4.7-fold improvement in catalytic activity for LacNAc over wild-type FutA (table S3). In addition, an obviously improved K223E mutation found in M31 variant was also introduced into the M32 mutant. However, the resulting mutant, having eight residue substitutions, exhibited reduced enzyme activity compared to its parent (table S2) thus, the K223E mutation was not used further in efforts to improve catalytic efficiency.

Characterization and kinetic analysis of the directed evolution mutants

Before further characterization, given that the directed evolution process is known to sometimes change the substrate and/or product profiles of target enzymes, we checked the reaction product(s) generated by the best mutant M32. After incubation of M32 with GDP-Fuc
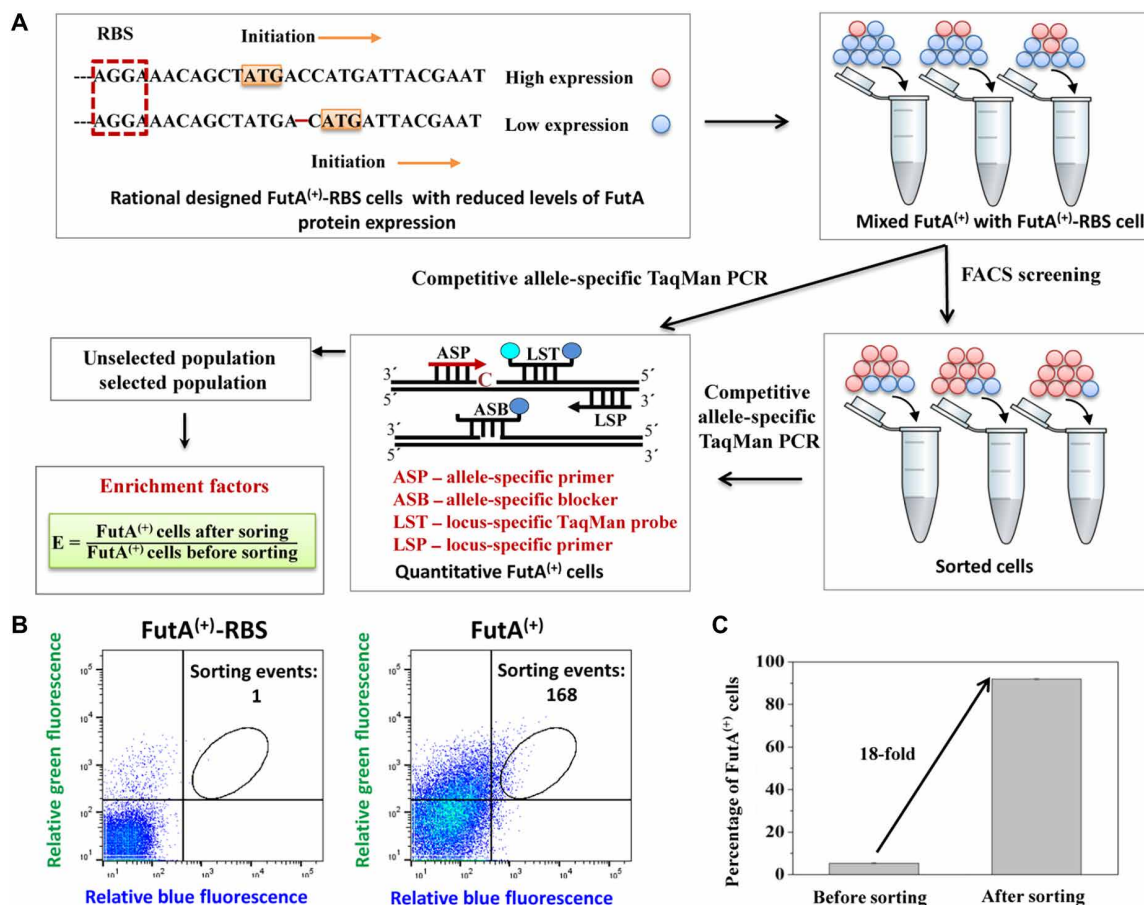
Fig. 3. Analysis of FACS-based screening efficiency by competitive allele-specific TaqMan PCR. (A) The RBS-ATG spacing technique was used to create two populations of cells: (i) normal FutA[(+)] cells and (ii) FutA[(+)] cells with a weakened FutA activity resulting from reduced FutA expression [FutA[(+)]-RBS cells]. Cell mixtures of FutA[(+)] and FutA[(+)]-RBS were prepared and applied to one round of FACS sorting. The unsorted and sorted variant pools were quantified using competitive allele-specific TaqMan PCR, and then, enrichment factors were calculated according to FutA[(+)] cell ratios before and after sorting. **(B)** Flow cytometric screening of FutA[(+)] and FutA[(+)]-RBS cells. **(C)** Percentage of FutA[(+)] cells increased after sorting.

and LacNAc, HPLC–electrospray ionization mass spectrometry analysis revealed a single product peak of mass/charge ratio of 522.19 [M-H] eluting at 3.6 min, identical to the Le[x] reference standard. Moreover, HPLC–electrospray ionization tandem mass spectrometry analysis revealed that the Le[x] standard and the M32 product had identical fragmentation patterns (fig. S6), demonstrating that M32 catalyzes the same reaction as does wild-type FutA.

To comprehend the specific activity improvement for the obtained beneficial FutA mutants, the kinetic parameters for LacNAc and GDP-Fuc substrates were first determined with the wild type, M26 and M32 mutants. Since FutA has a promiscuous activity to lactose, we also examined kinetic parameters for lactose as the acceptor substrate. It showed that a 4.1-fold improvement in the k_{cat}/K_m value of the mutant M26 arose from increases in the k_{cat} values, with little effect on the K_m values. Further mutations (D127N/R128E/H131I) led to progressive increases of catalytic activities, with the k_{cat}/K_m value of the best mutant M32 being sixfold higher than that of the wild type for LacNAc. The increase in k_{cat}/K_m for M32 was derived from a 10.8-fold decrease in the K_m value for lactose compared with the wild-type enzyme, thus substantially increased binding affinity, while the k_{cat} value for the GDP-Fuc donor increased 6.2-fold with little effect on the K_m value (Table 1).

Structural analysis of the best mutant M32

To gain a better structural understanding of mechanisms underlying the activity improvements of the best mutant M32, we solved its crystal structure in complex with GDP-Fuc at a resolution of 3.12 Å (PDB code 5ZOI) using a hanging-drop vapor diffusion method. Crystallization of the mutant was also dependent on deletion of the C-terminal 115 residues as reported for the wild type (*31*). There were no gross structural changes in the M32 mutant compared to the wild-type enzyme, with an root mean square deviation of 0.82 Å over 351 residues (fig. S8A). As a typical member of the GT-B family, FutA consists of an N-terminal domain (NTD) and C-terminal domain (CTD) with similar Rossmann folds, encompassing residues 20 to 150 and 160 to 320 to bind acceptor and donor substrates, respectively (fig. S8B). Upon inspection of the mutations in the M32 structure, we noted that four of seven mutated residues (S45F, D127N, R128E, and H131I) are located in two helices (α2 and α5) in the NTD binding acceptor substrate (Fig. 4A). All of these residues are within 8 Å of the known catalytic center. Meanwhile, the Y199N mutation is located in the CTD substrate-binding domain, while the other mutation of E340D is located on the protein surface.

We further docked the acceptor substrate LacNAc to the wild-type FutA and the M32 mutant. The results revealed that both the

Table 1. Kinetic parameters for wild-type FutA and selected beneficial mutants. The kinetic assays were performed in three independent replicates, and the fitting curves for the kinetic parameters are presented in fig. S7.

Enzyme	LacNAc*			GDP-Fuc†			Lactose*			GDP-Fuc†		
	$k$cat (s^{-1})	K_M (mM)	$k$cat/K_M (s^{-1} mM^{-1})	$k$cat (s^{-1})	K_M (mM)	$k$cat/K_M (s^{-1} mM^{-1})	$k$cat (s^{-1})	K_M (mM)	$k$cat/K_M (s^{-1} mM^{-1})	$k$cat (s^{-1})	K_M (mM)	$k$cat/K_M (s^{-1} mM^{-1})
Wild type	7.91 ± 0.18	0.52 ± 0.05	15.28 ± 1.13	16.28 ± 0.62	0.09 ± 0.01	180.89 ± 13.40	1.24 ± 0.05	99.60 ± 0.08	0.01 ± 0.005	0.05 ± 0.01	0.02 ± 0.01	2.53 ± 0.28
M26	30.60 ± 0.04	0.49 ± 0.02	62.45 ± 1.74	25.97 ± 0.57	0.08 ± 0.01	324.63 ± 17.34	1.27 ± 0.02	24.30 ± 1.00	0.05 ± 0.001	0.22 ± 0.01	0.03 ± 0.01	7.35 ± 0.39
M32	34.27 ± 0.70	0.37 ± 0.04	92.62 ± 8.23	30.03 ± 0.04	0.08 ± 0.01	375.38 ± 13.81	1.26 ± 0.03	9.20 ± 1.00	0.14 ± 0.01	0.31 ± 0.01	0.02 ± 0.01	15.59 ± 1.31

*Kinetic measurements at a non-limiting concentration of the fucosyl donor substrate GDP-Fuc (0.4 mM) but variable concentrations of the acceptors LacNAc (0.02 to 8 mM) or lactose (4 to 20 mM). †Kinetic measurements at non-limiting concentrations of the acceptors LacNAc or lactose (both 0.4 mM) but a variable concentration of the fucosyl donor substrate GDP-Fuc (0.02 to 8 mM).

shape and the orientation of LacNAc are complementary to a deep pocket in the FutA N-terminal domain; an aromatic cluster that consists of W33 and W34 surrounds the LacNAc. The substitution of S45F on helix α2 in the M32 structure creates a new clamp-like structure with W33 and W34 at the bottom of the acceptor substrate–binding pocket (Fig. 4B). This local structure could improve binding affinity of the protein for the galactopyranose ring of the acceptor substrates by enhancing both stacking and hydrophobic interactions, which explains much lower acceptor K_m values of M32 for both LacNAc and lactose substrates than were seen with wild-type FutA.

Analysis of the complex structure indicated that there is a considerable distance between the donor and the acceptor substrates in the FutA (approximately 10 Å), suggesting that there must be a substantial conformational change during catalysis to bring the NTD and CTD together. Notably, the helix α5 having the triple mutations D127N/R128E/H131I seems to play a hinge role between the N- and C-terminal domains (Fig. 4A and fig. S8B). These D127N/R128E/H131I mutations showed similar side-chain configurations to wild type but a markedly altered local electrostatic potential environment on the surface of hinge helix α5 (Fig. 4C). The subtle structural changes may affect the interdomain motion of M32 and potentially produce a more favorable conformation for catalysis, contributing to increased catalytic efficiency of the mutants and higher k_{cat} values.

We further used the FutA-LacNAc and M32-LacNAc complexes as initial models to perform 100-ns molecular dynamic (MD) simulations to explore the molecular mechanism of catalysis by FucTs (fig. S8C). Three independent trajectories were simulated for each of the two systems. The conformer ensemble consisted of 20,000 snapshots sampled for each 5-ps time step and showed that FucTs could adopt either a relatively open or close form in which the hinge helix α5 shifts about 2.6 Å as a consequence of the conformational changes (fig. S8D). MD analysis revealed that a long connecting loop region (residues 132 to 168) of helix α5 undergoes a marked fluctuation with a range of around 2 Å (Fig. 4D), suggesting that this highly flexible loop allows motion of the helix α5. The improved catalytic parameters of the M32 mutant with three mutations in the hinge helix α5 thus emphasize the importance of conformational dynamics in FucTs for efficient catalysis.

In addition, the protein surface residue Y199N is positioned near C-terminal domain (the donor binding domain in GT-B family) and is at the protein surface (Fig. 4A), and this residue may somehow improve donor substrate recognition. Alternatively, it may influence the overall efficiency via some long-distance residue interaction. Given that our determined M32 structure occurs as a homodimer, it is reasonable to speculate that E340D and V368A mutations—located at the homodimer interface—may facilitate its dimerization, perhaps promoting folding in a way that can increase its overall FucT activity.

DISCUSSION

It is now widely appreciated that directed evolution technologies can effectively improve the catalytic activity of many types of enzyme (32, 33), but directed evolution of FucTs is still in its very early days. Foundational work by Choi *et al.* (31) identified mutants of 1,3-FucTs with increased activity by use of a 96-well plate screening system based on color change of a pH indicator. This approach required high concentrations of extremely expensive natural GDP-fucose substrates, resulting in high screening costs and limited throughput. Technological advances over the last decade have helped the FACS-based screening system to becoming a powerful tool for directed evolution in general and for GTs in particular. To date, FACS platforms have been used to screen two GTs: an α2,3-sialyltransferase (12) and a β1,3-galactosyltransferase (14). Thus, our present study extends the scope to include screening for FucT and transfucosylation activity.

It bears emphasizing that the key to our successful screening strategy was the entrapment of fucosylated products inside cells based on the differential substrate specificity of the lactose permease LacY or other sugar transporters. Many studies in recent years have characterized the substrate specificity of LacY for various *E. coli* sugars (20, 34), and several very interesting and biotechnologically relevant findings have been reported (21, 35). Olsen and Brooker (35) investigated the sugar specificity properties of LacY and found that the relative importance of —OH groups around the galactose ring is OH-$_3$ > OH-$_4$ > OH-$_6$ > OH-$_2$ > OH-$_1$, while Abramson reported the first crystal structure of lactose permease from *E. coli* and found that six residues play major roles in substrate recognition (20).

Our present study extends this knowledge. First, somewhat similar findings from other reports show that the terminal sugar moiety of the lactose disaccharide is critical for LacY recognition, we show that fucosylation of the C_3-OH positions of the GalNAc moiety of LacNAc interferes with LacY transporting efficiency. Notably, we found that fucosylation of the C_2-OH or C_3-OH position of the galactose moiety also interferes with the translocation of corresponding trisaccharide products; we are unaware of other studies reporting
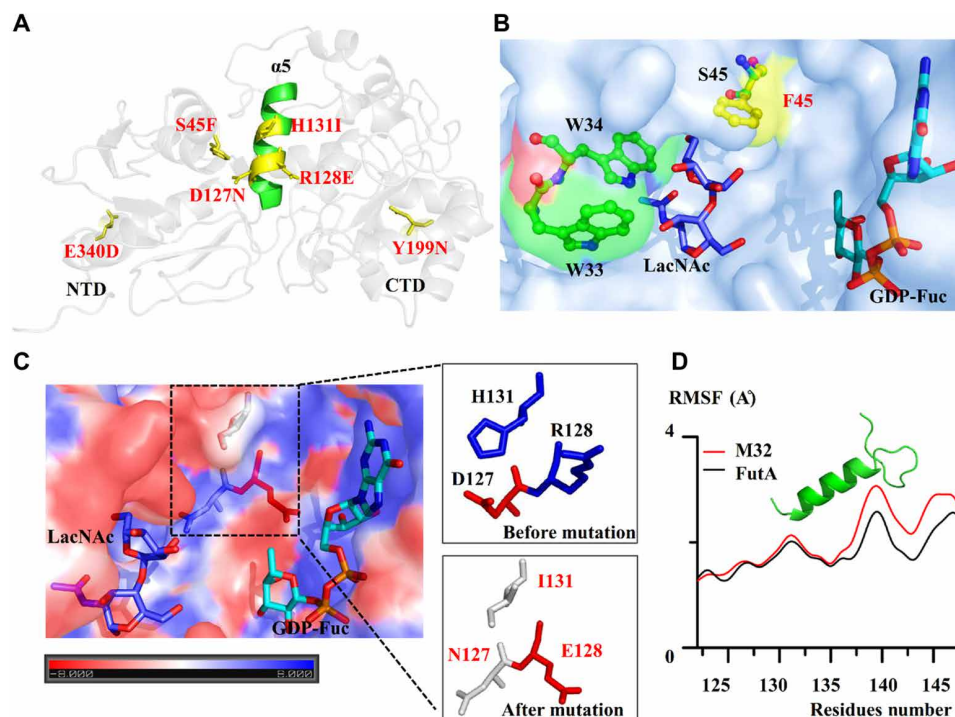
Fig. 4. Structural insight into the improved catalytic activity of the best M32 mutant. (**A**) Backbone diagram of the M32 mutant (PDB code 5ZOI) with mutations accumulations during directed evolution. Mutated residues are depicted in yellow sticks. Helix α5 having the triple mutations D127N/R128E/H131I located between the NTD and CTD is colored in green. (**B**) Enhanced interaction toward the LacNAc acceptor in the M32 mutant. The S45F mutation of M32 resulted in a new clamp-like structure with W33 and W34 at the bottom of the substrate-binding pocket. Key aromatic residues and S45 are shown in green sticks, and substituted residue F45 was represented in yellow stick. (**C**) Local electrostatic surface of M32 active pocket (red, electronegative; blue, electropositive; contoured from −8 to 8 kT/e). These D127N/R128E/H131I mutations showed a changed local electrostatic potential environment on the surface of hinge helix α5. (**D**) Root mean square fluctuation (RMSF) of wild-type FutA and M32 mutant residues from 122 to 148 region backbones in 100 ns constrained MD simulation. The segment of 122 to 148 residues are shown in cartoon.

that modifications to the inner sugar moiety of a disaccharide can alter substrate binding to sugar permeases or transporters. Thus, our work suggests that it is feasible to use modification(s) on the inner sugar moiety to enable entrapment of enzymatic products, a strategy that may further extend the scope of this screening approach.

The evaluation of the performance of various FACS-based screening systems has typically been based on "model screening" strategies that attempt to distinguish inactive "mock" cells from enzyme-active cells at various ratios (*32*, *33*). However, it is now becoming clear that such low resolution and binary (on/off) criteria are not immediately applicable/translatable to actually identifying cells that contain kinetically attractive positive mutants in directed evolution screens. As the goal of most screens is to identify highly active mutants, we show that an idealized approach for evaluating FACS-based screening systems would include the ability to differentiate among cells that contain moderately versus highly active mutant enzymes. Our study addressed this issue: The extremely strong performance of our FACS-based screening system (~100% of cells after one round of FACS were positive) drove us to find alternatives to typical model screening evaluations. We successfully used the RBS-ATG spacing technique to generate a diversity of differentially active FutA mutant variant cells based on differential transcription, and this enabled robust characterization of our system's performance. Complementing the RBS-ATG spacing technique, we also used competitive allele-specific TaqMan PCR with custom TaqMan probes to quickly quantify the efficiency of our FACS-based screening system. These alternative evaluation concepts

can be generalized. That is, a researcher needing to carefully evaluate their high-efficiency FACS screening methodology should consider generating a population of cells with differentially active overall enzyme activity (based on transcription activity) using the easy-to-use RBS-ATG spacing technique.

The FutA enzymes, which catalyze the transfer of fucose from GDP-Fuc to the C_3-OH of glucose/GlcNAc moieties, are used industrially not only for the synthesis of Le[x] in the production of anti-inflammatory drugs and antitumor vaccines but also for the large-scale synthesis of 3-fucosyllactose for prebiotics (*36*). Although several FutAs have been cloned and characterized from mammals and bacteria, most of these suffer from poor catalytic efficiency, preventing their use for the large-scale synthesis of Le[x] and 3-fucosyllactose (table S3) (*37*). Using our FACS-based high-throughput screening system with FutA, we found that just three rounds of directed evolution led to the identification of FutA mutants with 6-fold and 14-fold increases in catalytic efficiency (k_{cat}/K_m) for the synthesis of Le[x] and 3-fucosyllactose, respectively, which is the highest-to-date reported activity for Le[x] production by biocatalysis (table S3).

Our crystallographic analysis of the best mutant M32 revealed that four of seven beneficial mutations (S45F, D127N, R128E, and H131I) lie within 10 Å of the active site catalytic base and are distributed on two helices in the NTD (Fig. 4A). A new clamp-like local structure formed by the S45F substitution and the other two aromatic residues, W33 and W34, appears to enhance acceptor substrate binding and facilitate catalysis by this mutant. This is

consistent with a previous study showing that the S46F mutation in FutA from *H. pylori* 26695 increased catalytic activities (*31*). Moreover, crystal structures of the GT-B family enzymes, WaaG (PDB: 2IV7; GT4) (*38*) and GumK (PDB: 2HY7; GT70) (*39*), revealed similar surface-exposed aromatic hydrophobic residues (tryptophan and phenylalanine) on their N-terminal domains. As reported earlier (*40*), GT-B superfamily members undergo conformational changes involving movement of the NTD and CTD upon substrate binding. Notably, the three mutations D127N, R128E, and H131I clustered on a helix hinge between the NTD and CTD, likely altering the interdomain structure and influencing domain motions to providing a more reactive conformation. This likely extends to other GT-B enzymes. Through analysis of MD simulations, we proposed a dynamic model in which the NTD and CTD of GT-B enzymes undergo dynamic hinge-bending motions upon binding their respective acceptor and donor, thereby initiating domain closure (fig. S8E). Since interdomain motion appears to be essential for catalysis by GT-B type GTs, it is tempting to speculate that the hinge region could be a new hotspot for the rational design of catalytic properties into GT-B enzymes. Overall, this and other key mutated residues identified in our study highlighted important sites for engineering catalytic activities of GT-B type enzymes, which should also motivate hypothesis-driven studies into the basis of FucT substrate recognition, thereby facilitating future rational engineering efforts.

Since LacY transports disaccharides in and out of *E. coli* cells, we envision that our method should be suitable for use in improving other GTs by designing appropriate fluorescently labeled disaccharide substrates. In addition, we anticipate that this dual-color fluorescence method can be modified to facilitate the engineering of substrate specificity for enzymes, provided that the method's selection criteria are switched from seeking efficiency to seeking novel catalytic functionality. For example, different types of substrates could be conjugated with green fluorescent bodipy and blue fluorescent coumarin for the directed evolution of enzyme specificity toward different substrates. As enzymatic fucosylation represents an important target for the production of both pharmaceuticals and human milk oligosaccharides, our new FACS system is likely to have application in the engineering of these enzymes. Moreover, the techniques (e.g., cellular entrapment and alternatives to model screening for evaluation) used in our study should find use more broadly in future engineering efforts seeking to improve the catalytic performance of other enzyme classes.

MATERIALS AND METHODS
Generation of random mutagenesis library
The DNA sequence encoding amino acids 1 to 421 of FutA was subjected to error-prone PCR using 0 to 0.3 mM Mn^{2+} to control the extent of mutation with 0.03 ng of pUC18-FutA as the template. The primers used for this step can be found in table S4. The PCR products were digested and ligated into the pUC18 vector using the *Hind*III and *Eco*RI restriction sites, and the ligation mixture was electroporated to *E. coli* 10G ELITE cells (Lucigen, USA) 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 verified by sequencing; this indicated average mutational frequencies of ~2, ~4, ~6, and ~7

mutations per gene in the 0, 0.1, 0.2, and 0.3 mM Mn^{2+} libraries, respectively. The plasmid DNA from the four libraries was mixed in a 1:1:1:1 ratio and was then transformed into *E. coli* JM107* cells for screening.

Generation of ORM
ORM was performed to rapidly determine the optimal combination of these beneficial point mutations, similar to iterative site-directed mutagenesis. All primers are shown in table S4. The whole-plasmid PCR method was performed using Proofast Super-Fidelity DNA polymerase with the M26 plasmid as the template DNA. The PCR products were digested with *Dpn* I to remove the parent plasmids and loaded with GelRed (Shanghai Life iLab Biotech) to check DNA quality on 1% agarose gel. The resulting DNA mixture was transformed into *E. coli* JM107* competent cells.

Docking simulation of LacNAc
The crystal structure of FutA (PDB codes: 2NZY and 5ZOI) was used for acceptor LacNAc docking analysis. The coordinates of the acceptor LacNAc were generated and energetically optimized with the CHARMm force field in Discovery Studio 3.5. The software was also used for the docking of LacNAc into the binary structure of FutA with donor GDP-fucose. The active site was assigned around the region of Oe2 for Glu^{95}. Among the 50 docking poses generated from docking simulations, the one with the minimum docking energy value E/d [<0.97 kcal/(mol Å)] was selected.

Generation of the combinatorial active-site saturation test libraries
The mutagenesis of CAST was performed using primer pairs (table S4). The target amino acid position was coded by NNK (sense strand), where N = A, G, C, or T and K = G or T. The whole-plasmid PCR method was performed using Proofast Super-Fidelity DNA polymerase with the plasmid DNA isolated from the second round of FACS enrichment as the template DNA using the following PCR thermocycling profile: 95°C for 3 min, 30 cycles of 95°C for 15 s, 55°C for 15 s, 72°C for 5 min, and 72°C for 10 min. The PCR products were digested with *Dpn*I to remove the parent plasmid and then were cleaned using a DNA purification kit (Shanghai Life iLab Biotech, China). The resulting DNA mixture was transformed into *E. coli* JM107* cells for screening.

Screening via flow cytometry
Flow cytometric screening of FutA activity was carried out essentially as described before (*15*, *17*), except that the two fluorescent substrates were used. Briefly, we transformed *E. coli* JM107 LacZ⁻ (*12*, *14*) cells with pACKC18-FKP recombinant plasmid encoding for GDP-fucose synthase and then named it as JM107*. Plasmid DNA (pUC18-FutA) encoding the FutA libraries was transformed into JM107* competent cells and used to directly inoculate LB media supplemented with ampicillin and chloramphenicol (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 OD_{600} (optical density at 600 nm) reached 0.5 to 0.7, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cultures were transferred to 20°C and grew overnight. Cells were spun down (2 ml) and resuspended in 50 μl of M9 media supplemented with 1.5 mM fucose (RayGood Biotech, Shanghai, China), 0.5 mM concentrations of fluorescently

labeled substrates. Following 30 min of incubation, cells were spun down, and the excess acceptor sugars were removed by washing the cells with LB media and PBS (pH 7.4).

The cells were diluted with PBS to obtain a flow cytometric event rate of ~6000/s in a FACSAria II flow cytometer (Becton Dickinson) using PBS as the sheath fluid. The threshold for event detection was set to forward and side scattering. The average sort rate was ~6000 events/s, using a 85-μm nozzle, exciting argon ion (508 nm) and 405-nm lasers, and measuring emissions passing the 515-nm (fluorescein isothiocyanate) band-pass filter for the bodipy emission and the 450-nm (violet 1) filter for the coumarin emission. The detection threshold was set at forward scatter (FSC) > 10,000 and side scatter (SSC) > 10,000. Cells were sorted into 1.5-ml tubes and centrifuged at 14,000g for 30 min. Genes from positive cells were amplified using the FutA-F and FutA-R primers (table S4) with Proofast Super-Fidelity DNA polymerase (ATG Biotech). The 50 μl of reaction mixtures contained ~11 μl of water, 25 μl of Proofast buffer (2×), 1 μl of 10 mM dNTPs (deoxynucleoside triphosphates), 10 μl of sorted cells, 20 μM primers mix (1 μl each), and 1 μl of Proofast Super-Fidelity DNA polymerase using the following PCR thermocycling profile: 95°C for 3 min; 40 cycles of 95°C for 15 s, 55°C for 15 s, 72°C for 90 s, and 72°C for 10 min. The PCR product was digested and ligated into the pUC18 vector using HindIII and EcoRI restriction sites, and the ligation mixture was electroporated into E. coli 10G elite cells (Lucigen) according to the manufacturer's protocol. Extracted plasmids were then electroporated into E. coli JM107* cells for the next round of FACS enrichment. FACS data were processed using FlowJo software 10 (TreeStar).

Plus/minus reference model screening
The E. coli FutA[(+)] cells expressing FuctA were mixed with FutA[(−)] cells harboring empty plasmid pUC18 at ratios of 1:10, 1:100, 1:1000, and 1:10,000. The cell mixtures were reacted with two fluorogenic substrates at 37°C for 30 min, and the excess acceptor sugars were removed by washing the cells with LB media and PBS (pH 7.4). Next, the samples were sorted by flow cytometry, and single E. coli cells with the top 0.5% of blue fluorescence intensity were collected into 1.5-ml tubes containing 0.5 ml of super optimal broth with catabolite repression (SOC) medium. After growing the collected cells on agar plates, negative and positive colonies were identified by individual bacterial colony PCR with the universal primers M13F(−47)/M13R(−48). The enrichment factors were calculated on the basis of the positive ratio values from before and after sorting.

Competitive allele-specific TaqMan PCR to examine FACS-based screening efficiency
The deletion mutant FutA[(+)]-RBS cells was constructed using a whole-plasmid PCR method with 0.03 ng of pUC18-FutA as the template. The primers used here can be found in table S4. The E. coli JM107* cells expressing FutA were mixed with the deletion mutants [FutA[(+)]-RBS] at ratios of 1:10, 1:100, and 1:1000. These mixtures were incubated with both substrates 1 and 2 at 37°C for 30 min. The samples were then sorted by flow cytometry, and the E. coli cells with the top 0.5% of blue fluorescence were collected into 1.5-ml tubes. The unsorted and sorted variant pools were then quantified using the following competitive allele-specific TaqMan PCR to detect point mutations. Very briefly, the competitive allele-specific TaqMan PCR principle is based on the design of locus-specific primers, an allele-specific primer (ASP), an allele-specific "blocker" oligo, and locus-specific TaqMan probes (Fig. 3A and table S4). The

E. coli JM107* cells expressing FutA were mixed with the deletion mutant cells at various ratios (0.1, 0.5, 1, 2.5, 5, 10, 25, and 100%) standard sample, with each of sample having three replicates. The 50 μl of reaction mixtures contained 25 μl of TaqMan Master Mix (2×), 0.6 μl of 10 mM ASP, 0.4 μl of 10 mM TaqMan probe, 0.6 μl of 10 mM primer, 2.4 μl of 30 μM blocker oligo, and 6 μl of mixed standard sample. The thermocycling program used here was 95°C for 10 s and then five cycles with 92°C for 15 s and 58°C for 1 min, followed by additional 40 cycles with 92°C for 15 s and 60°C for 1 min. A standard curve was constructed by fitting this equation to the log (N_0) and Ct values for a known series of N_0.

Protein expression and purification
The futA gene sequence was subcloned into the pET21a vector (Novagen) and was transformed into E. coli BL21 (DE3) pLysS cells. Transformants were inoculated in LB medium supplemented with ampicillin (100 mg/ml) and grown at 37°C overnight. The cells were then transferred into fresh medium and grown at 37°C; when the OD_{600} reached 0.8 to 1.0, IPTG was added to a final concentration of 0.5 mM and induced expression overnight at 20°C. The cells were harvested and suspended in binding buffer [25 mM tris-HCl (pH 7.5), 100 mM NaCl, and 20 mM imidazole] and lysed by sonication on ice. The recombinant proteins were affinity-purified using a Ni-NTA (Ni^{2+}-nitrilotriacetate) column (Smart-lifesciences, Changzhou, China) and eluted with buffer [200 mM imidazole, 500 mM NaCl, 100 mM tris-HCl (pH 8.5), and 10 mM 2-mercaptoethanol]. Proteins were further purified by gel filtration using a Superdex-75 column (GE Healthcare) equilibrated against buffer [25 mM tris-HCl (pH 8.0), 100 mM NaCl, and 1 mM dithiothreitol]. Last, the pure proteins were concentrated with a 30-kDa Vivaspin-20 concentrator (GE Healthcare) to ~10 mg/ml.

Enzymatic activity and kinetic analysis
FutA produces GDP as a side product, the production of GDP was thus coupled with a pyruvate kinase/lactate dehydrogenase assay to enable spectrophotometric monitoring of the consumption of NADH (nicotinamide adenine dinucleotide) (26); here, the excitation was at 340 nm, and the emission was at 460 nm for measuring FutA activity. The activity was measured at 37°C for 5 min in a final volume of 0.1 ml containing 100 mM tris-HCl (pH 7.5), 1 mM MnCl$_2$, 1 mM phosphoenolpyruvate, 50 μM NADH, 13.5 U of pyruvate kinase, 30 U of lactate dehydrogenase, 4 mM LacNAc or lactose, 400 μM GDP-Fuc (Carbosynth Co. Ltd., Berkshire, UK), and appropriate amounts of purified FutA variants. The assay was initiated upon the addition of the purified FutA, and the decrease in the fluorescence emission at 460 nm was monitored. Kinetic analyses of the FutA variants were carried out for the fucosyl donor substrate GDP-Fuc with the acceptor substrates LacNAc or lactose, as described previously (29). The kinetic parameters were obtained by fitting initial velocity data to the Michaelis-Menten equation using GraphPad Prism 5.0 (GraphPad Inc.).

Crystallization
Crystallization experiments were conducted in 48-well plates by the hanging-drop vapor diffusion method at 293 K, and each hanging drop was prepared by mixing 1.0 μl each of protein solution and reservoir solution. Initial crystallization trials yielded some small crystals from the polyethylene glycol (PEG)/Ion crystallization screen (Hampton Research). Crystal quality was improved by the microseed matrix screening method (41), and the seed stock was made

from the initial crystals. Ultimately, diffraction-quality crystals were grown in hanging drops at 21°C by mixing 1 μl of protein [16 mg/ml in 25 mM tris-HCl (pH 7.5) and 100 mM NaCl] with an equal volume of 0.1 M MES monohydrate (pH 6.0), 0.05 M CaCl$_2$, and 45% PEG-200. Crystals belonged to the space group P 3 2 1, with unit cell dimensions a = 121.4 Å, b = 121.4 Å, c = 32.4 Å, β = 90°. Complex crystals were obtained by cocrystallization with GDP-Fucose. The GDP-Fucose was dissolved to a final concentration of 2 mM in the protein solution for 2 hours at 4°C before setting up hanging drop crystallization experiments as described above.

Data collection and structure determination

For x-ray diffraction experiments, crystals were fished out from the crystallization drop using a nylon loop, soaked briefly in a cryo-protectant solution of the crystallization solution supplemented with 30% (v/v) ethylene glycol and flash-cooled in liquid nitrogen. X-ray diffraction datasets were collected on beamline BL17U and BL19U at the Shanghai Synchrotron Research Facility. All diffraction data were indexed, integrated, and scaled using HKL-2000.

Initial phases for each structure were determined by molecular replacement. The structure of apo-M32 was solved using the program BALBES with the Auto-Rickshaw pipeline. The structure was completed with alternating rounds of manual model building with Coot and refinement with REFMAC5 in CCP4suite. The structure of the M32-substrate complex was determined by molecular replacement with the program MOLREP using the apo-M32 as a search model. The structure of GDP-Fucose was built with Coot Ligand Builder, and restraints were created using PRODRG. Iterative model building was performed with Coot software, and refinement was carried out with REFMAC5 in CCP4suit. The final models were evaluated by PROCHECK (*42*). Data collection and refinement statistics are provided in table S5. PyMOL (DeLano Scientific; www.pymol.org) was used to produce molecular graphics renditions.

MD simulation

The simulations were conducted using the AMBER14 package with a ff12SB force field. The initial structure was taken from our docking results and then solvated in the truncated octahedron boxes of 10 Å, with counter ions added to neutralize the systems. A cutoff of 8 Å was used for the van der Waals and short-range electrostatic interactions. Partial mesh Ewald and SHAKE algorithms were adopted to calculate long-range electrostatic interactions and constrain the lengths of bonds with hydrogen atoms (*43*). An integration time step was set to 2 fs. The system was minimized with 500 steps of the steepest descent method and 3500 steps of conjugate gradient algorithm and then heated to a finite temperature of 298 K in NVT (constant particle number, volume and temperature) ensemble. After 10-ps simulations for equilibration, the simulations were processed in the NPT (constant particle number, pressure, and temperature) ensemble at 298 K.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/10/eaaw8451/DC1

Fig. S1. Scheme for the fluorescent product entrapment strategy and the cell-based FutC and FucD fucosylation assays using FACS.
Fig. S2. Analysis of fluorescence retention in various cells.
Fig. S3. Optimization of the FACS-based system.
Fig. S4. Site-directed mutagenesis and ordered recombination of the mutation site.
Fig. S5. Rational selection of candidates for "best hit" from C$_α$ of catalytic key residue and clustering of α helices on substrate-binding sites for CAST and SSM.

Fig. S6. LC-MS analyses of the Le^X from LacNAc catalyzed by FutA variants.
Fig. S7. Steady-state kinetics of wild type, M26, and M32 measured using various substrates.
Fig. S8. Structural insight into the improved activity of the best M32 mutant.
Table S1. Model screening of FutA$^{(+)}$ cells.
Table S2. Specific activities of FutA and selected mutants using LacNAc and lactose as acceptors.
Table S3. Activity comparison between the best mutant in the present study and previously reported FutA enzymes.
Table S4. Primers used in this study.
Table S5. Data collection and refinement statistics.
References (*44*, *45*)

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. B. Ma, J. L. Simalagrant, D. E. Taylor, Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **16**, 158R–184R (2006).
2. E. Miyoshi, K. Moriwaki, T. Nakagawa, Biological function of fucosylation in cancer biology. *J. Biochem.* **143**, 725–729 (2008).
3. M. N. Christiansen, J. Chik, L. Lee, M. Anugraham, J. L. Abrahams, N. H. Packer, Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525–546 (2014).
4. E. Miyoshi, K. Moriwaki, N. Terao, C.-C. Tan, M. Terao, T. Nakagawa, H. Matsumoto, S. Shinzaki, Y. Kamada, Fucosylation is a promising target for cancer diagnosis and therapy. *Biomolecules* **2**, 34–45 (2012).
5. L. Bode, Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
6. C. Kunz, S. Rudloff, W. Baier, N. Klein, S. Strobel, Oligosaccharides in human milk: Structural, functional, and metabolic aspects. *Annu. Rev. Nutr.* **20**, 699–722 (2000).
7. B. Ma, J. L. Simala-Grant, D. E. Taylor, Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **12**, 158–184 (2006).
8. J. R. Díaz, R. J. Carbajo, A. P. Lucena, V. Monedero, M. J. Yebraa, Synthesis of fucosyl-*N*-Acetylglucosamine disaccharides by transfucosylation using α-L-Fucosidases from *Lactobacillus casei*. *Appl. Environ. Microb.* **79**, 3847–3850 (2013).
9. F. H. Arnold, Design by directed evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).
10. U. T. Bornscheuer, M. Pohl, Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5**, 137–143 (2001).
11. M. S. Packer, D. R. Liu, Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
12. A. Aharoni, K. Thieme, C. P. C. Chiu, S. Buchini, L. L. Lairson, H. Chen, N. C. J. Strynadka, W. W. Wakarchuk, S. G. Withers, High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat. Methods* **3**, 609–614 (2006).
13. G. Yang, S. G. Withers, Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *Chembiochem* **10**, 2704–2715 (2010).
14. G. Yang, J. R. Rich, G. Michel, W. W. Wakarchuk, Y. Feng, S. G. Withers, Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases. *J. Am. Chem. Soc.* **132**, 10570–10577 (2010).
15. Z. Ge, N. W. Chan, M. M. Palcic, D. E. Taylor, Cloning and heterologous expression of an α1,3-fucosyltransferase gene from the gastric pathogen *Helicobacter pylori*. *J. Biol. Chem.* **272**, 21357–21363 (1997).
16. S. L. Martin, M. R. Edbrooke, T. C. Hodgman, D. H. van den Eijnden, M. I. Bird, Lewis X biosynthesis in *Helicobacter pylori*. Molecular cloning of an α(1,3)-fucosyltransferase gene. *J. Biol. Chem.* **272**, 21349–21356 (1997).
17. D. A. Rasko, G. Wang, M. M. Palcic, D. E. Taylor, Cloning and Characterization of the α(1,3/4) Fucosyltransferase of *Helicobacter pylori*. *J. Biol. Chem.* **275**, 4988–4994 (2000).
18. L. Engels, L. Elling, WbgL: A novel bacterial α1,2-fucosyltransferase for the synthesis of 2'-fucosyllactose. *Glycobiology* **24**, 170–178 (2014).
19. D. B. Stein, Y. N. Lin, C. H. Lin, Characterization of *Helicobacter pylori* α1,2-fucosyltransferase for enzymatic synthesis of tumor-associated antigens. *Adv. Synth. Catal.* **350**, 2313–2321 (2010).
20. J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
21. I. Smirnova, V. Kasho, X. Jiang, H. M. Chen, S. G. Withers, H. R. Kaback, Oversized galactosides as a probe for conformational dynamics in LacY. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4146–4151 (2018).
22. M. Lezyk, C. Jers, L. Kjaerulff, C. H. Gotfredsen, M. D. Mikkelsen, J. D. Mikkelsen, Novel α-L-fucosidases from a soil metagenome for production of fucosylated human milk oligosaccharides. *PLOS ONE* **11**, e0147438 (2016).
23. M. H. Saier Jr., Families of transmembrane sugar transport proteins. *Mol. Microbiol.* **35**, 699–710 (2010).

24. Y. Sun, C. K. Vanderpool, Regulation and function of *Escherichia coli* sugar efflux transporter A (SetA) during glucose-phosphate stress. *J. Bacteriol.* **193**, 143–153 (2011).

25. A. Amir, A. Gil, B. Kalia, M. Shlomo, S. D. Tawfik, High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **12**, 1281–1289 (2005).

26. E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly, A. D. Griffiths, High-throughput screening of enzyme libraries: In vitro evolution of a β-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.* **12**, 1291–1300 (2005).

27. O. Raluca, P. Radivoje, N. Jovana, F. Rainer, Ultra-high-throughput screening method for the directed evolution of glucose oxidase. *Chem. Biol.* **21**, 414–421 (2014).

28. M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* **117**, 4192–4196 (2005).

29. H.-Y. Sun, S.-W. Lin, T.-P. Ko, J.-F. Pan, C.-L. Liu, C.-N. Lin, A. H.-J. Wang, C.-H. Lin, Structure and mechanism of *Helicobacter pylori* fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J. Biol. Chem.* **282**, 9973–9982 (2007).

30. A. Pavelka, E. Chovancova, J. Damborsky, HotSpot Wizard: A web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* **37**, W376–W383 (2009).

31. Y. H. Choi, J. H. Kim, B. S. Park, B.-G. Kim, Solubilization and iterative saturation mutagenesis of α1,3-fucosyltransferase from *Helicobacter pylori* to enhance its catalytic efficiency. *Biotechnol. Bioeng.* **113**, 1666–1675 (2016).

32. A. D. Griffiths, S. D. Tawfik, Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.* **22**, 24–35 (2014).

33. N. Varadarajan, J. Gam, M. J. Olsen, G. Georgiou, B. L. Iverson, Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6855–6860 (2005).

34. H. R. Kaback, L. Guan, It takes two to tango: The dance of the permease. *J. Gen. Physiol.* **151**, 878–886 (2019).

35. S. G. Olsen, R. J. Brooker, Analysis of the structural specificity of the lactose permease toward sugars. *J. Biol. Chem.* **264**, 15982–15987 (1989).

36. C. Dumon, E. Samain, B. Priem, Assessment of the two *Helicobacter pylori* α-1,3-oligosaccharides by metabolically engineered *Escherichia coli*. *Biotechnol. Prog.* **20**, 412–419 (2010).

37. J. Yu, J. Shin, M. Park, E. Seydametova, S.-M. Jung, J.-H. Seo, D.-H. Kweon, Engineering of α-1,3-fucosyltransferases for production of 3-fucosyllactose in *Escherichia coli*. *Metab. Eng.* **48**, 269–278 (2018).

38. C. Martinez-Fleites, M. Proctor, S. Roberts, D. N. Bolam, H. J. Gilbert, G. J. Davies, Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chem. Biol.* **13**, 1143–1152 (2006).

39. M. Barreras, S. R. Salinas, P. L. Abdian, M. A. Kampel, L. Ielpi, Structure and mechanism of GumK, a membrane associated glucuronosyltransferase. *J. Biol. Chem.* **283**, 25027–25035 (2008).

40. D. Albesa-Jové, M. E. Guerin, The conformational plasticity of glycosyltransferases. *Curr. Opin. Struct. Biol.* **40**, 23–32 (2016).

41. A. D'Arcy, T. Bergfors, S. W. Cowanjacob, M. Marsh, Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallogr. F Struct. Biol. Commun.* **F70**, 1117–1126 (2014).

42. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *PROCHECK*: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).

43. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

44. S. Rabbani, V. Miksa, B. Wipf, B. Ernst, Molecular cloning and functional expression of a novel *Helicobacter pylori* α-1,4 fucosyltransferase. *Glycobiology* **15**, 1076–1083 (2005).

45. R. A. Alm, L.-S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. Dejonge, G. Carmel, P. J. Tummino, A. Caruso, M. U. Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, T. J. Trust, Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180 (1999).

Citation: Y. Tan, Y. Zhang, Y. Han, H. Liu, H. Chen, F. Ma, S. G. Withers, Y. Feng, G. Yang, Directed evolution of an α1,3-fucosyltransferase using a single-cell ultrahigh-throughput screening method. *Sci. Adv.* **5**, eaaw8451 (2019).