Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry

Lan Ban1,2,10, Nicholas Pettit1,3,4,10, Lei Li5,7,10, Andreea D Stuparu1,2, Li Cai3,4, Wenlan Chen3,4, Wanyi Guan3,6, Weiqing Han7,8, Peng George Wang7,8* & Milan Mrksich1,2,9*

Glycosyltransferases catalyze the reaction between an activated sugar donor and an acceptor to form a new glycosidic linkage. Glycosyltransferases are responsible for the assembly of oligosaccharides in vivo and are also important for the in vitro synthesis of these biomolecules. However, the functional identification and characterization of new glycosyltransferases is difficult and tedious. This paper describes an approach that combines arrays of reactions on an immobilized array of acceptors with an analysis by mass spectrometry to screen putative glycosyltransferases. A total of 14,280 combinations of a glycosyltransferase, an acceptor and a donor in four buffer conditions were screened, leading to the identification and characterization of four new glycosyltransferases. This work is notable because it provides a label-free method for the rapid functional annotation of putative enzymes.

RESULTS

Reaction screening
We analyzed the genomes of several microorganisms to identify putative glycosyltransferases to include in the screen. In this way, we emphasize that the SAMDI method can be used in a nonbiased screen to identify unanticipated transferase activities, although we note that this approach can also be used to analyze glycosyltransferases that are taken from a single species or can be directed at a mechanistic question. Some fraction of these enzymes may not catalyze glycosylation reactions—they could, for example, have hydrolase activity or act on acceptors that use other functional groups as the nucleophile—and thus would not be active in the present screen. To speed up the analysis, we used a high-throughput screening platform based on metal plates that have an array of 384 gold-coated islands in the standard microtiter plate geometry15. We prepared 24 oligosaccharide acceptors (Supplementary Methods, Supplementary Table 1 and Supplementary Figs. 1–4) and immobilized individual acceptors on the gold features by either forming desorption-ionization mass spectrometry (in a technique referred to as SAMDI)16–18. When the monolayer is irradiated with the laser, the alkanethiols are desorbed from the gold substrate but undergo little fragmentation, providing spectra that directly reveal the masses of the ligand-substituted alkanethiols (or alkyl disulfides).

In this way, the spectra show clear peaks for both the substrate and the products of an enzyme-mediated conversion and are valuable for efficiently identifying enzyme activities.

We used the SAMDI assay to evaluate nearly 60,000 reactions comprising unique combinations of a putative glycosyltransferase, a nucleotide donor, an immobilized acceptor and a buffer composition. Forty-four reactions showed new peaks that corresponded to individual glycosylation products and gave a functional validation of four new glycosyltransferases, including an enzyme with a previously unknown specificity in forming a glycosidic linkage.

1Howard Hughes Medical Institute, Northwestern University, Evanston, Illinois, USA. 2Department of Chemistry, Northwestern University, Evanston, Illinois, USA. 3Department of Chemistry, Ohio State University, Columbus, Ohio, USA. 4Department of Biochemistry, Ohio State University, Columbus, Ohio, USA. 5National Glycoengineering Research Center, Shandong University, Shandong, China. 6State Key Laboratory of Microbial Technology, Shandong University, Shandong, China. 7State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, China. 8College of Pharmacy, Nankai University, Tianjin, China. 9Department of Biomedical Engineering, Northwestern University, Evanston, Illinois, USA. 10These authors contributed equally to this work. *e-mail: pwang@nankai.edu or milan.mrksich@northwestern.edu
monolayers from carbohydrate-terminated alkanethiol reagents or performing immobilization reactions of a thiol-functionalized carbohydrate with a maleimide-terminated monolayer or an azido-functionalized carbohydrate with an alkyne-terminated monolayer (Supplementary Fig. 5). Our use of multiple strategies for immobilizing the acceptors owes to the availability of, or efficient access to, the carbohydrate reagents. The choice of chemistries used to link the acceptor to the monolayer should not have a major influence on the glycosylation reaction, as glycosyltransferases generally have selectivity for the terminal carbohydrate unit of the acceptor. We therefore expect that the activities that are discovered in the screen will not be compromised by the choice of immobilization reaction; in support of this idea, we discuss later the use of a homogeneous format to verify the activities of the glycosyltransferases found in this work.

An example of the screening assay is shown in Figure 1 and starts with a monolayer to which the acceptor lactose is immobilized. A SAMDI spectrum of the monolayer showed a peak at a mass to charge ratio (m/z) of 1,296, which corresponds to the mixed disulfide with a single lactose group. The peak at m/z 693 represents the tri(ethylene glycol) disulfide. The monolayer was treated with bovine α1,3-galactosyltransferase (GGTA1) and analyzed by SAMDI to reveal a peak at m/z 1,458, which corresponds to the mixed disulfide containing the trisaccharide that resulted from the enzymatic galactosylation of lactose. The absence of a peak at m/z 1,296 indicated that the enzymatic reaction was essentially complete. We then applied this assay to screen 85 glycosyltransferases (Supplementary Results and Supplementary Table 4), including 76 putative bacterial enzymes that had not been previously characterized, with the goal of identifying new glycosyltransferase activities.

Because traditional methods for protein expression and purification are tedious and time consuming and can compromise protein activity, we used an in vitro expression system to rapidly prepare the glycosyltransferases, which we then assayed in an unpurified form (Supplementary Fig. 7). Each individual protein was first mixed with one of the seven sugar donors dissolved in one of four buffers and then applied to individual gold islands presenting the sugar acceptor (Fig. 1). The four buffers were selected because of their common use in enzyme assays and included different divalent ions and pH values (Methods).

From the 57,120 reactions tested in the screen, 44 had new glycosylation products (Supplementary Table 5). Included in these hits were glycosylation activities for four previously uncharacterized enzymes (Fig. 2). Two of these enzymes, BF0009 (GT80; GTs are numbered as listed in Supplementary Table 4) and BF0614 (GT84), are from Bacteroides fragilis and catalyze the GalNAcylation of β-glucose and cellobiose and the galactosylation of N-acetylglucosamine (GlcNAc), respectively (Fig. 2a–c). The other...
two enzymes, HD0466 (GT24) and AAF28363.1 (GT09), are from *Haemophilus ducreyi* and transfer GlcNAc from UDP-GlcNAc to lactose and to GlcNAc, respectively (Fig. 2d,e). We also found that two known galactosyltransferases could use donors that have not been reported previously. The glycosyltransferases GGTa1 (GT02) and LgtC (GT06), in addition to using the UDP-galactose (UDP-Gal) donor, were also found to glycosylate their substrates with UDP-glucose (UDP-Glc) (for GGTa1) or with UDP-Glc and guanosine diphosphate mannose (GDP-Man) (for LgtC) (Fig. 2f and Supplementary Fig. 8).

To elaborate on one example, the protein expressed from the gene *BF0009* (GT80) was mixed with the donor UDP-GalNAc and applied to a spot presenting β-glucose. The SAMDI spectrum revealed a new peak at $m/z$ 1,337 (Fig. 2a), which is 203 Da greater than the peak for the acceptor-terminated alkyl disulfide ($m/z$ 1,134) and is consistent with the addition of GalNAc to the acceptor. To estimate the yield with which the glycosyltransferases were expressed in the active form, we included nine known glycosyltransferases in the screen and found that six were active (Supplementary Table 5). We believe that this fraction provides a fair estimate of the yield for the expression of the other putative enzymes examined in this work, although we did not determine whether the inactivity arose from improper folding of the enzymes, a requirement for cofactors or regulatory domains or an inaccessibility of the immobilized substrates.

### Identification of the glycosidic linkages in the products

The SAMDI screen is effective at identifying combinations of glycosyltransferases, donors and acceptor substrates that generate new glycosidic linkages, but the use of mass spectrometry does not provide information on the regiochemical and stereochemical structure of the linkage. To characterize the products, we expressed the four newly identified enzymes in *Escherichia coli* BL21(DE3) cells and used them in preparative glycosylation reactions to generate milligram quantities of the products, which we then characterized using one-dimensional and two-dimensional NMR (Fig. 2, Supplementary Table 2 and Supplementary Results). We found that *BF0009* (GT80) catalyzes the formation of a β1,3 linkage between GalNAc and glucose; *BF0614* (GT84) joins galactose and GlcNAc through a β1,4 bond; *HD0466* (GT24) joins GlcNAc and lactose through a β1,3 linkage; and AAF28363.1 (GT09) creates a β1,4 linkage between two GlcNAc residues.

### Biochemical characterization of the new enzymes

We selected three glycosyltransferases that gave new activities in the screen and characterized their kinetic parameters. These experiments used a pulldown format wherein reactions were performed in solution, quenched and then applied to a monolayer to allow the substrate and product to undergo immobilization before analysis by SAMDI. This method avoids perturbations that may arise from presentation of the ligands at the surface. We used an azido-modified oligosaccharide as the acceptor and a monolayer presenting a terminal alkyne group to selectively immobilize the substrate and product of the reaction (Supplementary Table 3 and Supplementary Fig. 9). Previous work has shown that the SAMDI method provides quantitative information for the rates of enzyme-catalyzed reactions, including reactions of glycosyltransferases. In one example, we characterized the BF0009-mediated transfer of UDP-GalNAc to an azidogluucose substrate by performing a series of reactions, with the glucose azide and UDP-GalNAc being present at several concentrations and for several reaction times (Supplementary Fig. 9c). The yields were determined by integrating the areas of the mass peaks for the product and the acceptor substrate (Supplementary Fig. 9b and Supplementary Fig. 10). Kinetic parameters were determined from double-reciprocal plots (Supplementary Fig. 11) and are summarized in Table 1. We also performed this experiment using a radiolabeled assay and found that the results agreed with our determination of kinetic parameters using the SAMDI pulldown assay (Supplementary Table 6). These experiments revealed several features of the glycosyltransferases. The value of the Michaelis constant ($K_m$) for the UDP-Gal donor is an order of magnitude lower with the human enzyme, reflecting a stronger interaction between the human homologue and the donor. In another example, HD0466 from *H. ducreyi* and the closely related β1,3N-acetylglucosaminyltransferase from *Neisseria meningitidis* (LgtA) each interact with the donor with similar affinities and perform the glycosylation reaction with similar maximum velocities, $V_{max}$.

Many glycosyltransferases have conserved folds and structural motifs that bind, and require, divalent metal ions for activity. Therefore, we also investigated the metal-dependent activities of the...
Table 1 Kinetic parameters for glycosylation reactions

| Parameter | BF0009 | BF0614 | HD0466 |
|-----------|--------|--------|--------|
| K_M (mM)  | 4.51 (0.071) | 4.92 (0.075) | 2.89 (0.032) |
| K_M (mM)  | 6.16 (0.047) | 5.62 (0.068) | 3.52 (0.041) |
| K_M (mM)  | 2.88 (0.039) | 0.446 (0.023) | 0.366 (0.044) |
| V_max (nmol min⁻¹) | 0.25 (0.013) | 0.22 (0.011) | 0.19 (0.0095) |

The example we report here resulted in the discovery of four bacterial proteins that have specific glycosyltransferase activities. Although the biological roles of these enzymes are unknown, they provide new catalysts that can be used to generate oligosaccharides with new structures. For example, the protein BF0009 from B. fragilis is particularly interesting because the enzyme catalyzes a reaction that yields a linkage that has not previously been observed in bacterial systems. Further, this enzyme only recognizes UDP-GalNAc as the donor substrate among the seven donors screened, making it more selective than related enzymes, including bovine milk galactosyltransferase, which accepts UDP-Gal, UDP-Glc and UDP-GalNAc. We also note that carbohydrate-associated antigens are related to the pathogenesis of B. fragilis, and future work may reveal whether the newly identified GalNAc–β1,3Glc linkage exists in the bacterium and whether it is relevant to pathogenesis.

In summary, this work describes an effective strategy for combining carbohydrate arrays and mass spectrometry for the functional annotation of enzyme families. The method is noteworthy because it combines the immobilized arrays with a true label-free detection method that allows for direct readout of the biochemical activities on the array. This method for the high-throughput characterization of glycosyltransferase activity offers a new opportunity for the identification of new and interesting glycosyltransferases from both bacterial and eukaryotic sources. We believe it will also be key for the functional annotation of other enzyme families.

**METHODS**

**In vitro expression of putative glycosyltransferases.** The sources of the materials used, including bacterial and eukaryotic cell lines, are summarized in Supplementary Table 5. The sources of the materials used, including bacteria strains, bacterial and eukaryotic cell lines, and other genetic experiments, can be found in Supplementary Methods. The in vitro expression of the putative glycosyltransferases was performed using the Expressway Cell-Free E. coli Expression System (Invitrogen), as described by the vendor. In brief, 2 μg of plasmid was used for each 100 μl volume of the reaction, and the reactions were incubated at 37 °C for 4 h. Insoluble particulates were removed by centrifugation, and the supernatant containing the soluble protein was used immediately in glycosylation reactions. A vector provided by the manufacturer (harboring a nonglycosyltransferase gene) was used as a positive control for protein expression, and the empty vector pMCSG7 was used as a negative control and gave no detectable endogenous glycosyltransferase activity. The expression of the target enzymes was confirmed by western blot using mouse histidine-specific antibody and mouse IgG-specific conjugated horseradish peroxidase secondary antibody (Supplementary Fig. 7). The films were developed using the enhanced chemiluminescence with the western blot detection reagent.

**Reaction screening.** Reactions were performed on metal plates that had a 24 × 12 array of gold islands modified with monolayers. The preparation of the oligosaccharide-terminated monolayers is discussed in detail in the Supplementary Methods. Each sugar nucleotide donor was dissolved in one of the four buffer systems (0.75 mM, 4 μl) and combined with protein (2 μl) in a 384-well plate. The buffers used in this work were as follows: Tris-HCl (30 mM, pH 8.0) and MnCl₂ (10 mM); sodium cacodylate (100 mM, pH 6.0) and MnCl₂ (10 mM); Tris-HCl (50 mM, pH 7.5) and CaCl₂ (10 mM); or Hepes (50 mM, pH 7.5). The resulting reaction mixtures were then transferred to individual gold islands on a 384-well metal plate presenting one of the 24 sugar acceptors. Liquid handling was performed by a Tecan Freedom EVO 200 robot. We estimate that the enzyme was present at concentrations in the low nanomolar range. The plates were kept in humidified chambers at 37 °C for 4 h and then rinsed with water and ethanol. The plates were dried under nitrogen, treated with a solution of 2,4,6-trihydroxyacetophenone (THAP) matrix (5 mg ml⁻¹, 0.5 ml per plate) and analyzed by SAMDI mass spectrometry. A 355-nm neodymium-doped yttrium aluminum garnet (Nd:YAG) laser was used as the desorption/ionization source with an accelerating voltage of 20 kV and an extraction delay time of 50 ns. All spectra were acquired automatically using the positive reflector mode. The combinations of glycosyltransferase, donor and acceptor substrate that gave a glycosylation reaction are summarized in Supplementary Table 5. The new enzymes were expressed again in E. coli and were used to synthesize oligosaccharide products at a preparative scale. See the Supplementary Methods for the details of enzyme expression, obtaining carbohydrate preparation and mass characterization. The new glycosyltransferase LgtA1 were also expressed in E. coli, and their activities with UDP-Glc and GDP-Man (for LgtC) or with UDP-Glc (for GGTAT) were confirmed in a SAMDI assay.

**Determination of the kinetic parameters of glycosyltransferases using SAMDI.** To obtain kinetic constants for the new glycosyltransferases, reactions were performed in solution, and then the substrate and product were immobilized to a monolayer.
by way of a Click reaction\(^2\). Reactions were initiated in 384-well plates and quenched by the addition of EDTA and then applied to a monolayer and analyzed by SAMDI, according to the protocols listed here. Each reaction contained one of the purified enzymes, the sugar donor, the acceptor, Tris-\(\text{HCl} (50 \text{ mM}, \text{pH} 8.0)\) and \(\text{MnCl}_2 (10 \text{ mM})\) in a total volume of 10 \(\mu\text{L}\). The enzymes were used at the following concentrations: \(\text{BF0009}, 0.14 \text{ mg ml}^{-1}; \text{BF0614}, 0.3 \text{ mg ml}^{-1}; \) and \(\text{HD0466}, 0.3 \text{ mg ml}^{-1}\). The concentrations of the donors ranged from 100 \(\mu\text{M}\) to 5 \(\text{mM}\), and the concentrations of the acceptors ranged from 50 \(\mu\text{M}\) to 5 \(\text{mM}\). For each set of concentrations, the reactions were carried out for durations of 2–30 min at intervals of 2–3 min and terminated by adding a mixture of cold ethanol and EDTA (10 \(\text{mM}, 20 \mu\text{L}\)). Each reaction mixture was then applied to an individual gold circle of the array (in a volume of \(2 \mu\text{L}\)) that was modified with the alkylene-terminated monolayer. An aequorin solution (1 \(\mu\text{L}\) per reaction) containing \(\text{CulBr} (2 \text{ mM})\) and triethylamine (0.5 \(\text{mM}\)) was applied to each circle, and the reactions were incubated at 25 °C for between 30 min and 6 h, depending on the concentration of the azaido sugars. The completion of the reactions was monitored by SAMDI. The slide was then rinsed with water and then ethanol and then was dried under nitrogen. For quantification, the extent of glycosylation (\(R\)) was determined from the peak intensities for the product (\(I_\text{p}\)) and the acceptor substrate (\(I_\text{a}\)) on the SAMDI spectra using the formula \(R = \frac{I_\text{p}}{I_\text{a} + I_\text{p}}\). We confirmed that the measured ratio reflected the actual ratio of the two azaido sugars in solution by performing a calibration experiment, which is discussed in detail in the Supplementary Methods and Supplementary Figure 10. The yield of the glycosylation, calculated from the equation given in Supplementary Figure 9b, was plotted against the reaction time. The linear region of the plot was fitted to obtain the slope, which represented the initial velocity \(v_0\). Double-reciprocal plots of the initial velocities are shown in Supplementary Figure 11. For these plots, the donors were the variable substrates, and the acceptors were the constant substrates. The data were fit to equation (1), which has previously been used to describe bisubstrate enzyme kinetics\(^{19,20}\).

\[
\frac{1}{v_0} = \frac{K_A + K_B + [B]}{v_{\text{max}} [A] [B]} \times [A] [B]
\]

In this equation, \([A]\) is the concentration of the acceptor, \([B]\) is the concentration of the donor, \(v_0\) is the maximum velocity, \(K_A\) and \(K_B\) represent the cognate Michaelis constants for substrates \(A\) and \(B\), respectively, and \(K_{\text{cat}}\) is the dissociation constant of the substrate \(A\). The 1:1 metal activity studies were carried out in a cuvet and the details can be found in the Methods section. For the study of 1,2-fucosyltransferase of \(B.\ fragilis\), we confirmed that the measured ratio reflected the actual ratio of the two azaido sugars in solution by performing a calibration experiment, which is discussed in detail in the Supplementary Methods and Supplementary Figure 10. The yield of the glycosylation, calculated from the equation given in Supplementary Figure 9b, was plotted against the reaction time. The linear region of the plot was fitted to obtain the slope, which represented the initial velocity \(v_0\). Double-reciprocal plots of the initial velocities are shown in Supplementary Figure 11. For these plots, the donors were the variable substrates, and the acceptors were the constant substrates. The data were fit to equation (1), which has previously been used to describe bisubstrate enzyme kinetics\(^{19,20}\).

\[
\frac{1}{v_0} = \frac{K_A + K_B + [B]}{v_{\text{max}} [A] [B]} \times [A] [B]
\]

In this equation, \([A]\) is the concentration of the acceptor, \([B]\) is the concentration of the donor, \(v_0\) is the maximum velocity, \(K_A\) and \(K_B\) represent the cognate Michaelis constants for substrates \(A\) and \(B\), respectively, and \(K_{\text{cat}}\) is the dissociation constant of the substrate \(A\). The metal activity studies were carried out in a cuvet and the details can be found in the Methods section. For the study of 1,2-fucosyltransferase of \(B.\ fragilis\), we confirmed that the measured ratio reflected the actual ratio of the two azaido sugars in solution by performing a calibration experiment, which is discussed in detail in the Supplementary Methods and Supplementary Figure 10. The yield of the glycosylation, calculated from the equation given in Supplementary Figure 9b, was plotted against the reaction time. The linear region of the plot was fitted to obtain the slope, which represented the initial velocity \(v_0\). Double-reciprocal plots of the initial velocities are shown in Supplementary Figure 11. For these plots, the donors were the variable substrates, and the acceptors were the constant substrates. The data were fit to equation (1), which has previously been used to describe bisubstrate enzyme kinetics\(^{19,20}\).

\[
\frac{1}{v_0} = \frac{K_A + K_B + [B]}{v_{\text{max}} [A] [B]} \times [A] [B]
\]

In this equation, \([A]\) is the concentration of the acceptor, \([B]\) is the concentration of the donor, \(v_0\) is the maximum velocity, \(K_A\) and \(K_B\) represent the cognate Michaelis constants for substrates \(A\) and \(B\), respectively, and \(K_{\text{cat}}\) is the dissociation constant of the substrate \(A\). The metal activity studies were carried out in a cuvet and the details can be found in the Methods section. For the study of 1,2-fucosyltransferase of \(B.\ fragilis\), we confirmed that the measured ratio reflected the actual ratio of the two azaido sugars in solution by performing a calibration experiment, which is discussed in detail in the Supplementary Methods and Supplementary Figure 10. The yield of the glycosylation, calculated from the equation given in Supplementary Figure 9b, was plotted against the reaction time. The linear region of the plot was fitted to obtain the slope, which represented the initial velocity \(v_0\). Double-reciprocal plots of the initial velocities are shown in Supplementary Figure 11. For these plots, the donors were the variable substrates, and the acceptors were the constant substrates. The data were fit to equation (1), which has previously been used to describe bisubstrate enzyme kinetics\(^{19,20}\).

\[
\frac{1}{v_0} = \frac{K_A + K_B + [B]}{v_{\text{max}} [A] [B]} \times [A] [B]
\]

In this equation, \([A]\) is the concentration of the acceptor, \([B]\) is the concentration of the donor, \(v_0\) is the maximum velocity, \(K_A\) and \(K_B\) represent the cognate Michaelis constants for substrates \(A\) and \(B\), respectively, and \(K_{\text{cat}}\) is the dissociation constant of the substrate \(A\). The metal activity studies were carried out in a cuvet and the details can be found in the Methods section. For the study of 1,2-fucosyltransferase of \(B.\ fragilis\), we confirmed that the measured ratio reflected the actual ratio of the two azaido sugars in solution by performing a calibration experiment, which is discussed in detail in the Supplementary Methods and Supplementary Figure 10. The yield of the glycosylation, calculated from the equation given in Supplementary Figure 9b, was plotted against the reaction time. The linear region of the plot was fitted to obtain the slope, which represented the initial velocity \(v_0\). Double-reciprocal plots of the initial velocities are shown in Supplementary Figure 11. For these plots, the donors were the variable substrates, and the acceptors were the constant substrates. The data were fit to equation (1), which has previously been used to describe bisubstrate enzyme kinetics\(^{19,20}\).

\[
\frac{1}{v_0} = \frac{K_A + K_B + [B]}{v_{\text{max}} [A] [B]} \times [A] [B]
\]

In this equation, \([A]\) is the concentration of the acceptor, \([B]\) is the concentration of the donor, \(v_0\) is the maximum velocity, \(K_A\) and \(K_B\) represent the cognate Michaelis constants for substrates \(A\) and \(B\), respectively, and \(K_{\text{cat}}\) is the dissociation constant of the substrate \(A\). The metal activity studies were carried out in a cuvet and the details can be found in the Methods section. For the study of 1,2-fucosyltransferase of \(B.\ fragilis\), we confirmed that the measured ratio reflected the actual ratio of the two azaido sugars in solution by performing a calibration experiment, which is discussed in detail in the Supplementary Methods and Supplementary Figure 10. The yield of the glycosylation, calculated from the equation given in Supplementary Figure 9b, was plotted against the reaction time. The linear region of the plot was fitted to obtain the slope, which represented the initial velocity \(v_0\). Double-reciprocal plots of the initial velocities are shown in Supplementary Figure 11. For these plots, the donors were the variable substrates, and the acceptors were the constant substrates. The data were fit to equation (1), which has previously been used to describe bisubstrate enzyme kinetics\(^{19,20}\).