Engineering Orthogonal Polypeptide GalNAc-Transferase and UDP-Sugar Pairs

Junwon Choi,[†,∥,1] Lauren J. S. Wagner,[††,1] Suzanne B. P. E. Timmermans,[†††,∥∥] Stacy A. Malaker,[†] Benjamin Schumann,[†,∥∥∥] Melissa A. Gray,[†] Marjoke F. Debets,[†,∥∥∥∥] Megumi Takashima,[††††,∥∥∥∥∥] Jase Gehring,[†††††,∥∥∥∥∥∥] and Carolyn R. Bertozzi*[,†,§]

[†]Department of Chemistry and [§]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, United States

[††]Department of Chemistry, University of California, Berkeley, CA 94720, United States

[†††]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

[††††]Department of Nutritional Sciences, University of California, Berkeley, CA 94720, United States

[†††††]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, United States

Abstract

O-GalNAc glycans constitute a major part of the human glycome. They are difficult to study because of the complex interplay of 20 distinct glycosyltransferase isoenzymes that initiate this form of glycosylation, the polypeptide GalNAc transferases (GalNAc-Ts). Despite proven disease relevance, correlating the activity of individual GalNAc-Ts with biological function remains challenging due to a lack of tools to probe their substrate specificity in a complex biological environment. Here, we develop a "bump–hole" chemical reporter system for studying GalNAc-T activity in vitro. Individual GalNAc-Ts were rationally engineered to contain an enlarged active site (hole) and probed with a newly-synthesized collection of 20 (bumped) UDP-GalNAc analogs to identify enzyme–substrate pairs that retain peptide specificities but are otherwise completely orthogonal to native enzyme–substrate pairs. The approach was applicable to multiple GalNAc-T isoenzymes, including GalNAc-T1 and -T2 that prefer non-glycosylated peptide substrates and GalNAcT-10 that prefers a pre-glycosylated peptide substrate. A detailed investigation of enzyme kinetics and specificities revealed

the robustness of the approach to faithfully report on GalNAc-T activity and paves the way for studying substrate specificities in living systems.

Introduction

Glycosylation with O-linked α-GalNAc (historically called mucin-type O-glycosylation) is one of the most abundant forms of post-translational modification in higher eukaryotes[1] and is essential for normal embryonic development,[2] immune cell function,[3] and metabolic homeostasis.[4] Aberrations in O-GalNAc glycosylation are associated with tumor progression.[5] Accordingly, the enzymes that initiate O-GalNAc glycosylation, the polypeptide *N*-acetylgalactosaminyl transferases (GalNAc-Ts), are differentially regulated, and their expression has been associated with various disease states.[6–9] The human genome encodes 20 GalNAc-T paralogs, constituting one of the largest and arguably most complex glycosyltransferase families.[10] GalNAc-Ts are membrane-associated, type II Golgi-resident enzymes with adjacent catalytic and lectin domains that face the Golgi lumen. Some isoenzymes seem to share functional redundancies, as reflected in subtle animal knockout phenotypes,[11,12] whereas others are essential for viability or specific organ functions.[2,13]

O-GalNAc glycans frequently decorate proteins in clusters called mucin glycodomains.[14] These densely glycosylated sequences are built by sequential action of "early" GalNAc-Ts (e.g., GalNAc-T1 and -T2) that glycosylate unmodified or sparsely glycosylated peptide sequences, and "late" GalNAc-Ts (e.g., GalNAc-T10) that act adjacent to existing O-GalNAc glycans.[10,15] The substrate specificities of the various isoenzymes collectively create an O-GalNAc glycoproteome with discrete biological consequences.

Despite their participation in many facets of human biology, our understanding of GalNAc-T substrate selection is limited. GalNAc-Ts transfer a GalNAc residue from the nucleotide sugar donor UDP-GalNAc to the hydroxyl groups of Ser, Thr, and possibly Tyr residues of proteins in the secretory pathway and supposedly the nucleus.[16,17] The protein substrates and specific glycosylation sites modified by individual GalNAc-Ts are not well described, hampering our understanding of their

biological function and disease causality. Classical genetic approaches toward understanding GalNAc-T function are complicated by protein substrate redundancies and the existence of overlapping protein target sites.[12,18] More targeted genetic studies have validated roles for GalNAc-T2 glycosylation of ApoCIII and ANGPTL3 in high-density lipoprotein metabolism,[9] GalNAc-T3 glycosylation of FGF23 in familial tumoral calcinosis,[19] and GalNAc-T11 glycosylation of Notch in heterotaxy, associated with congenital heart disease.[13] Despite these forays, the primary source of information about specific amino acid preferences of individual GalNAc-Ts is still the performance of in vitro assays with peptide substrates.[20–22]

Clausen and co-workers made a major breakthrough to discern GalNAc-T substrate specificity using cell lines engineered to generate truncated O-GalNAc glycoproteins with limited glycan elaborations.[23] Glycoproteomics analyses of these so-called SimpleCells – with and without a single GalNAc-T knockout – identified isoenzyme-specific glycosylated proteins and sites.[18,24,25] However, given the interdependencies and substrate redundancies of various GalNAc-Ts, this loss of function approach may not generalize across the enzyme family. Furthermore, late GalNAc-Ts have not been addressed by the SimpleCell approach yet. So far, no method is available to probe the function of an individual GalNAc-T isoenzyme in the presence of others without the drawbacks associated with genetic knockouts.

We envisioned a complementary approach in which individual GalNAc-T isoenzymes are engineered to acquire a gain of function – the ability to mark their specific protein substrates and glycosites with a chemical handle. This has been achieved with other enzyme families, including kinases,[26,27] acetyltransferases,[28] methyltransferases,[29] and ADP-ribosyltransferases,[30] using the “bump–hole” engineering strategy.[31] Key to this strategy is the identification of gatekeeper residues that can be mutated to alter the active site so as to accommodate a complementary substrate. Critically, the “bumped” substrate should not be recognized by native members of the enzyme family. There are examples wherein the active sites of individual glycosyltransferases were mutated to accommodate modified nucleotide-sugars bearing chemical handles.[32–34] However, to the best of our knowledge,

glycosyltransferase enzyme families have not yet been subjects of orthogonal enzyme–substrate engineering.

Here, we report a strategy for engineering orthogonal GalNAc-T and UDP-sugar pairs that can, in principle, be applied across the enzyme family (Scheme 1). We identified a combination of active site point mutations that caused loss of function with UDP-GalNAc but gain of function with synthetic UDP-GalNAc analogs that contain a chemical handle. We found orthogonal enzyme–substrate pairs that retained peptide glycosylation site preferences and catalytic parameters comparable to their native counterparts. The generality of the strategy was demonstrated with bump–hole pairs for the early GalNAc-Ts, -T1 and -T2, and the late GalNAc-T, -T10.

Scheme 1. A Bump–Hole Approach[a]

[a]Mutagenesis of key gatekeeper residues in the active site of a GalNAc-T introduces a "hole" in the engineered GalNAc-T that accommodates an enlarged UDP-GalNAc analog modified with a "bump" (orange circle) and chemical handle (orange diamond). The *N*-acyl side chain of UDP-GalNAc contains a methyl group (red Me) that is modified on the UDP-GalNAc analog to an *R*-group (red R), representing the bump and chemical handle. Monosaccharides are represented with colored boxes: GalNAc (yellow) and GalNAc analog (orange).

Results and Discussion

To identify gatekeeper residues we might target for mutagenesis, we analyzed the available crystal structures of the catalytic and lectin domains of GalNAc-T1,[35] -T2,[36–38] and -T10,[39] some of which contained bound GalNAc or UDP-GalNAc in the active site. We considered different positions around GalNAc when selecting a location in the active site for mutagenesis. Notably, we previously found that virtually all GalNAc-T isoenzymes can use UDP-*N*-azidoacetylgalactosamine (UDP-GalNAz) as a substrate both in vitro and in living cells or organisms, although with reduced efficiency compared to UDP-GalNAc.[40–42] Thus, we anticipated that larger *N*-acyl substituents would be required to avoid recognition by wild-type GalNAc-T isoenzymes, and these would have to be accommodated by an engineered active site hole. We noted a high degree of conservation of three hydrophobic residues within 5 Å of the methyl group of GalNAc that create a binding pocket (Figure 1). We hypothesized that some combination of I253, L310, and F361 could serve as active site gatekeeper residues across the enzyme family (Figures 1 and S1).

Figure 1. Identification of gatekeeper residues. (A) Residues within 5Å of GalNAc methyl carbon for GalNAc-T2 (PDB ID: 4D0T). Five of seven amino acids in close proximity to the GalNAc methyl contain side chains; of these, H359 and D224 coordinate Mn^{2+}, and 3 are promising hydrophobic residues. (B) Space-filling model of gatekeeper residues within 5Å of GalNAc methyl in GalNAc-T2 (PDB ID: 4D0T). (C) Amino acid sequences of human GalNAc-T1–GalNAc-T20 surrounding potential gatekeeper residues demonstrate a high degree of conservation, with 13 isoenzymes containing I253/L310 and 18 containing I253. Only GalNAc-T8 and -T18 have dramatically different

residues at position 253 and around 310. GalNAc-Ts used in this study are boxed. Clustal Omega was used to generate a multiple sequence alignment of the amino acid sequences corresponding to the full-length genes of human GalNAc-T1–GalNAc-T20 (Figure S2; Table S1). GalNAc-Ts are ordered based on homology, and GalNAc-Ts that predominantly prefer GalNAc-peptides are denoted with an asterisk.[10]

To test these hypotheses, we initially focused on GalNAc-T2, an early isoenzyme that recognizes unglycosylated peptide substrates and is one of the best characterized of the family.[36–38] We generated four single point mutants (I253A, L310A, F361A, and F361S) and one double mutant (I253A/L310A) of a soluble, FLAG epitope-tagged construct secreted from mammalian cells.[43] Preliminary screens with purified proteins and the known peptide substrate EA2-biotin (PTTDSTTPAPTTKK(biotin))[42] showed minimal activity from either F361 mutant with UDP-GalNAc (**1**) or UDP-GalNAz (**2**),[42,44] in line with mutagenesis data that were published thereafter.[38] The remaining three mutants (I253A, L310A, and I253A/L310A) and wild-type GalNAc-T2 were investigated further.

We next designed a quantitative enzymatic assay with the sensitivity and flexibility necessary to accommodate chemically diverse UDP-GalNAc analog libraries and measure kinetic parameters of mutated GalNAc-Ts. Widely-used glycosyltransferase assays rely on the detection of UDP as a side product of the glycosylation reaction.[45] Although we used this method for initial screens (Figures S3 and S4), we deemed it of limited utility, as background hydrolysis of UDP-sugars under these conditions produces UDP that limits sensitivity. The assay had to be independent of the nature of the chemical handle, precluding the use of our previously developed azido-ELISA.[46] We therefore used the chromophore 2,4-dinitrophenyl-5-L-alanine amide, which was installed on peptide substrates using Marfey's reagent and is convenient to monitor by UV detection on an HPLC. We prepared an artificial, labeled peptide via solid-phase peptide synthesis as an optimal GalNAc-T2 substrate (Peptide-1) based on previously identified amino acid preferences (Figure 2A).[21]

Figure 2. Synthesis of a peptide substrate and a panel of UDP-GalNAc analogs. (A) Synthetic route for Peptide-1. Blue T indicates the Thr glycosylation site used by GalNAc-T2. (B) Synthetic routes for UDP-GalNAc analogs. Route 1 was used to synthesize UDP-sugars ((*S*)-**3**, (*R*)-**3**, *(S)*-**4**, (*R*)-**4**, (*S*)-**5**, (*R*)-**5**, (*S*)-**6**, (*R*)-**6**, (*S*)-**8**, (*R*)-**8**, (*S*)-**12**, (*R*)-**12**, (*S*)-**14**, (*R*)-**14**), and Route 2 was used to synthesize UDP-sugars (**2**, **7**, **9**, **10**, **11**, **13**). (C) A panel of UDP-GalNAc derivatives with azide or alkyne chemical handles. Compounds **1** and **2** are the natural substrate UDP-GalNAc and known analog UDP-GalNAz, respectively. Reagents and conditions: a) NEt(*i*-Pr)$_2$, DMF, r.t.; b) R-COOH, COMU®, NEt(*i*-Pr)$_2$, DMF, 0 °C to r.t.; c) *N,N'*-dimethyl-1,3-diaminopropane, THF, r.t.; d) *i*-Pr$_2$NPO(OAll)$_2$, 1*H*-tetrazole, CH_2Cl_2, 0 °C, then *m*-CPBA, −78 °C; e) Pd(PPh$_3$)$_4$, sodium *p*-toluenesulfinate, THF/MeOH, r.t.; f) i. Uridine 5′-monophosphomorpholidate 4-morpholine-*N,N*′-dicyclohexylcarboxamidine salt, 1-methylimidazole hydrochloride, NEt_3, DMF, r.t.; ii. MeOH/water/NEt_3, r.t.; g) HEPES buffer (pH = 8.0), 0 °C to r.t.

We next synthesized a collection of UDP-GalNAc analogs, chemically diversified at the C2-acylamide moiety and including bioorthogonal azide or alkyne groups. Synthesis of nucleotide sugars is nontrivial, and both chemical and enzymatic routes have been developed.[46–50] As UDP-GalNAc biosynthetic enzymes are unlikely to accommodate all envisaged modifications,[47,51] we chose to chemically synthesize 20 bumped UDP-GalNAc analogs by two routes adapted from literature procedures (Figure 2B).[46,49,52] Installation of the acylamide side chain into a UDP-galactosamine scaffold via active ester chemistry proceeded smoothly to provide compounds **2**, **7**, **9**, **10**, **11**, and **13** (Figure 2B, Route 2). For UDP-GalNAc analogs containing chiral acylamide α-carbon atoms, an alternate route was selected to minimize the risk of acylamide epimerization during installation. Thus, acylamides were introduced at an early stage in the syntheses of compounds ((*S*)-**3**, (*R*)-**3**, *(S)*-**4**, (*R*)-**4**, (*S*)-**5**, (*R*)-**5**, (*S*)-**6**, (*R*)-**6**, (*S*)-**8**, (*R*)-**8**, (*S*)-**12**, (*R*)-**12**, (*S*)-**14**, (*R*)-**14**) (Figure 2B, Route 1), using

COMU® as a highly efficient coupling reagent that precluded epimerization. Together, these two routes and an optimized purification procedure enabled us to generate a structurally diverse panel of synthetic UDP-GalNAc analogs (Figure 2C).

With purified enzymes, synthetic peptide substrate, and UDP-GalNAc analogs in hand, we assessed the viability of GalNAc-T bump–hole engineering in a combinatorial fashion, using an HPLC-based enzymatic assay (Figure 3A). Wild-type GalNAc-T2 efficiently transferred a GalNAc residue from UDP-GalNAc (**1**) to the acceptor peptide, and the single mutants T2(I253A) and T2(L310A) still utilized UDP-GalNAc as a donor substrate, although with reduced activity (Figure 3B). In contrast, UDP-GalNAc was a poor substrate for the double mutant enzyme T2(I253A/L310A), which we felt might allow for the development of an orthogonal GalNAc-T and UDP-sugar pair.

Figure 3. Screening GalNAc-T2 for an orthogonal enzyme–substrate pair. (A) Scheme for glycosylation reaction with Peptide-1, GalNAc-T2, and UDP-GalNAc or UDP-GalNAc analog to form glycosylated Peptide-1. Blue T indicates the Thr glycosylation site used by GalNAc-T2. (B) Glycopeptide formation by wild-type and mutant GalNAc-T2. UDP-GalNAc and Peptide-1 were incubated with GalNAc-T2 at 37 °C for 1 h, and the reaction was quenched by the addition of aqueous EDTA (150 mM, pH = 8.0). The percent conversion to glycopeptide product was quantified by HPLC separation and peak integration. All data represent the mean of technical triplicates, and the error

bars represent the standard deviation. (C) Bump–hole pair optimization for GalNAc-T2. Glycosylation by wild-type and double mutant GalNAc-T2 was compared for UDP-GalNAc (**1**) and UDP-GalNAc analogs with Peptide-1. Reactions were performed and quantified as in B. Heat map (blue shading) shows percent glycosylated Peptide-1 formed by wild-type or double mutant GalNAc-T2 with UDP-GalNAc or analogs. Red values represent the mean of technical triplicates.

We investigated how enlargement of the GalNAc side chain alters substrate activity for the wild-type and double mutant enzymes (Figures 3C and S5 for selected UDP-GalNAc analogs and Figure S3 for additional analogs). UDP-GalNAz (**2**) was a better substrate for wild-type GalNAc-T2 than for the double mutant T2(I253A/L310A). Strikingly, however, some UDP-GalNAc analogs with longer or branched *N*-acyl chains were better substrates for T2(I253A/L310A) than for wild-type GalNAc-T2. These included UDP-GalNAc analogs with additional alkyl substituents α to the amide as in (*S*)-**3** and (*S*)-**4**, longer side chains as in **7**, **11** ("UDP-GalNAlk"),[53] and **13**, or both branches and length as in (*S*)-**12**, (*R*)-**12**, (*S*)-**14**, and (*R*)-**14**. For analogs **11**–**14**, we observed that as the size of the alkyne-containing bump increased, wild-type enzyme activity plummeted. In contrast, the double mutant enzyme maintained activity with these larger alkyne-functionalized UDP-GalNAc analogs. In the extreme case of (*R*)-**14**, glycopeptide formation by the wild-type enzyme was undetectable, while the double mutant still produced glycopeptide product.

An orthogonal enzyme–substrate pair should ideally retain the catalytic efficiency of the native enzyme–substrate pair to appropriately emulate biological function. Michaelis–Menten kinetic analysis revealed that both native and bump–hole enzyme–substrate pairs had comparable kinetic parameters (Table 1). Catalytic constants (k_{cat}) of enzyme–substrate pairs T2(I253A/L310A) and (*S*)-**3**, **7**, **11**, or (*S*)-**12** differed less than two-fold from the wild-type GalNAc-T2/**1** pair. The k_{cat} of the engineered pair T2(I253A/L310A)/**13** was approximately five-fold lower. Concomitantly, this pair displayed a ten-fold lower K_m compared to the wild-type pair, suggesting that **13** benefitted from increased contacts within the active site leading to stronger substrate binding. Thus, the catalytic efficiency (k_{cat}/K_m) of T2(I253A/L310A)/**13** is two-fold higher than wild-type GalNAc-T2/**1**. In contrast, (*S*)-**12** exhibited the weakest interaction with T2(I253A/L310A) among all enzyme–substrate

pairs tested, with a ten-fold higher K_m than the wild-type pair, suggesting that the methyl branch at the acylamide α-position experiences a steric clash with the enzyme.

Table 1. Kinetic parameters of wild-type and engineered GalNAc-T2 and UDP-sugar pairs.[a]

T2/UDP-Sugar	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (mM^{-1} s^{-1})
WT-T2/**1**	0.813 ± 0.017	29.5 ± 2.4	27.5
T2(I253A/L310A)/(*S*)-**3**	0.566 ± 0.014	42.9 ± 3.8	13.2
T2(I253A/L310A)/**7**	0.609 ± 0.033	162 ± 21	3.76
T2(I253A/L310A)/**11**	0.683 ± 0.021	56.2 ± 5.8	12.1
T2(I253A/L310A)/(*S*)-**12**	0.839 ± 0.052	430 ± 46	1.95
T2(I253A/L310A)/**13**	0.158 ± 0.003	2.60 ± 0.78	60.9

[a]To determine K_m and k_{cat} values for UDP-GalNAc and UDP-GalNAc analogs, initial rates were measured by incubating wild-type or double mutant GalNAc-T2 with concentrations of UDP-sugars varying from 15.6 μM to 500 μM and with a constant concentration of acceptor peptide (Peptide-1 = 267 μM for **1**, (*S*)-**3**, **11**; Peptide-1 = 250 μM for **7**, (*S*)-**12**, **13**). The glycosylation was conducted at 37 °C, and three aliquots were taken within 15 min and quenched by the addition of aqueous EDTA (150 mM, pH = 8.0). Products were quantified by HPLC separation and peak integration. The enzymatic kinetic parameters were obtained by nonlinear regression fitting using GraphPad Prism. All data represent the mean of technical triplicates, and error depicts the standard deviation.

To enable potential application in biological systems, wild-type GalNAc-Ts must not accept bumped UDP-GalNAc analogs in the presence of UDP-GalNAc. We examined the relative selectivity of wild-type GalNAc-T2 toward UDP-GalNAc compared to selected analogs (Figure 4). Competition experiments revealed that in the presence of a 1:1 mixture of UDP-GalNAc and either UDP-GalNAc analog (*S*)-**3**, **7**, **11**, or **13**, wild-type GalNAc-T2 indeed preferentially transferred GalNAc to Peptide-1. In the reaction that included analog **13**, the glycopeptide product observed was almost exclusively that derived from UDP-GalNAc.

Figure 4. Selectivity of wild-type GalNAc-T2 for UDP-GalNAc relative to UDP-GalNAc analogs. (A) Scheme for competition experiment between UDP-GalNAc and UDP-GalNAc analog. Wild-type GalNAc-T2 was treated with Peptide-1 and an equal ratio of UDP-GalNAc and UDP-GalNAc analog in a competition experiment, and glycosylation reactions were terminated at 20–30% glycopeptide formation. (B) Selectivity of wild-type GalNAc-T2 for UDP-GalNAc (**1**) over UDP-GalNAc analog ((*S*)-**3**, **7**, **11**, or **13**) in a competition experiment. Reactions were performed as in A. UDP-sugars and Peptide-1 were incubated with GalNAc-T2 at 37 °C for 30 min, and the reaction was quenched by the addition of aqueous EDTA (150 mM, pH = 8.0). The percent conversion to glycopeptide product was quantified by HPLC separation and peak integration. Percent of Peptide-1 modified with GalNAc or GalNAc analog was measured, and the selectivity ratio is shown in blue. All data represent the mean of technical triplicates, and the error bars represent the standard deviation.

We next sought to determine if our bump–hole strategy could be generalized to study other members of the GalNAc-T family. We thus extended the approach to GalNAc-T1, an early isoenzyme that is phylogenetically distant from GalNAc-T2, and GalNAc-T10, a late isoenzyme that prefers glycosylated substrates and, as such, is particularly difficult to study in vivo.[10,15] We generated wild-type and double mutant versions of GalNAc-T1 and -T10 using the same methods and gatekeeper residues homologous to GalNAc-T2 (I238A/L295A for -T1, I266A/L321A for -T10). We designed and prepared acceptor substrates labeled with 2,4-dinitrophenyl-5-L-alanine amide in the same fashion as Peptide-1. Similar to GalNAc-T2, amino acid preferences of GalNAc-T1 have been

studied, and we used an optimized sequence for Peptide-2, which contains a single threonine for glycosylation (Figure 5A).[21] GalNAc-T10, as a late isoenzyme, required a pre-installed GalNAc directly adjacent C-terminally to the glycosylation site. To this end, MUC5AC-3, a known GalNAc-T10 glycopeptide substrate, was used as the sequence for Peptide-3 (Figure 5A).[12,54]

Gratifyingly, the bump–hole approach developed for GalNAc-T2 was directly transferrable to both early and late GalNAc-T isoenzymes. Wild-type GalNAc-T1 efficiently catalyzed glycosylation with UDP-GalNAc, whereas minimal activity was observed with UDP-GalNAc analog **13** (Figure 5B). In contrast, T1(I238A/L295A) did not utilize UDP-GalNAc as a substrate but efficiently transferred the modified GalNAc residue from analog **13** to Peptide-2 (Figures 5B and S4). Similar to GalNAc-T2 engineering, the k_{cat} of T1(I238A/L295A) with **13** was approximately five-fold lower than the k_{cat} of wild-type GalNAc-T1 with UDP-GalNAc (Figure 5C). In contrast, the K_m value was unaltered, resulting in a reduction of catalytic efficiency by less than an order of magnitude (Figure 5C). The selectivity of the late isoenzyme GalNAc-T10 showed identical trends; the preference of wild-type GalNAc-T10 for UDP-GalNAc over **13** was reversed in the double mutant T10(I266A/L321A), highlighting the universality of our bump–hole approach for multiple members of the GalNAc-T family with diverse peptide substrate preferences (Figure 5B). Mutation of I266 and L321 to alanine residues had no effect on the k_{cat} and conferred only a two-fold higher K_m with the UDP-sugar analog, despite a switch in specificity from UDP-GalNAc to **13** (Figure 5C).

GalNAc-T/UDP-Sugar	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (mM^{-1} s^{-1})
WT-T1/**1**	0.551 ± 0.008	22.8 ± 1.5	24.2
T1(I238A/L295A)/**13**	0.097 ± 0.002	22.4 ± 2.1	4.33
WT-T10/**1**	0.956 ± 0.019	5.14 ± 0.87	186
T10(I266A/L321A)/**13**	1.09 ± 0.03	12.7 ± 1.6	85.4

Figure 5. Orthogonal GalNAc-T and UDP-sugar pairs for GalNAc-T1 and GalNAc-T10. (A) Scheme for glycosylation reaction with GalNAc-T, peptide, and UDP-GalNAc or UDP-GalNAc analog to form glycosylated peptide. Glycosylation reactions with GalNAc-T1 utilized Peptide-2, and reactions with GalNAc-T10 utilized Peptide-3. Blue T indicates the Thr glycosylation site used by the GalNAc-T of interest. (B) Glycopeptide formation by wild-type or double mutant GalNAc-T1 or GalNAc-T10 with UDP-GalNAc (**1**) or **13**. Reactions were performed as in A. GalNAc-T, UDP-sugar, and peptide were incubated at 37 °C for 1 h (-T10) or 2 h (-T1), and the reaction was quenched with aqueous EDTA (150 mM, pH = 8.0). All data represent the mean of technical triplicates, and the error bars represent the standard deviation. (C) Kinetic parameters of wild-type and orthogonal GalNAc-T and UDP-sugar pairs. To determine K_m and k_{cat} values for UDP-GalNAc and UDP-GalNAc analogs, initial rates were measured by incubating wild-type or double mutant GalNAc-Ts with varying concentrations of UDP-sugars and a constant concentration of acceptor peptide. For GalNAc-T1, the concentration of UDP-sugars varied from 15.6 μM to 500 μM, and the concentration of acceptor Peptide-2 was held at 250 μM. For GalNAc-T10, the concentration of UDP-sugars varied from 15.6 μM to 250 μM, and the concentration of acceptor Peptide-3 was held at 266 μM. The glycosylation was conducted at 37 °C, and three aliquots were taken within 15 min and quenched by the addition of aqueous EDTA (150 mM, pH = 8.0). Products were quantified by HPLC separation and peak integration. The enzymatic kinetic parameters were obtained by nonlinear regression fitting using GraphPad Prism. All data represent the mean of technical triplicates, and error depicts the standard deviation.

Finally, we explored whether the glycosylation site specificity of all three GalNAc-T bump–hole pairs was altered due to the active site mutations we introduced (Figure 6). In contrast to protein substrates,

the specificities of GalNAc-T isoenzymes toward synthetic peptide substrates have been exhaustively mapped and are the basis for recent mechanistic and structural studies.[15,20,21,42,55] We performed glycosylation reactions with known peptide and/or glycopeptide substrates of GalNAc-T1, -T2 and -T10, and fragmented and manually sequenced the resulting mono-(GalNAc-T1 and -T2) or di-glycopeptides (GalNAc-T10) by tandem mass spectrometry to identify site preferences. The glycopeptide MUC5AC-3 contains multiple potential glycosylation sites, three of which are glycosylated at different frequencies by wild-type GalNAc-T2 and UDP-GalNAc.[56] We observed similar fine specificities for these sites with the T2(I253A/L310A)/**13** pair, with a preference for Thr13 over Thr9 and Thr10 (Figure 6B). A similar retention of site specificity was found using other known GalNAc-T2 substrates, the glycopeptide MUC5AC-13 and the peptide EA2 (Figure 6B).[37,56] Glycosylation site specificity was also retained for GalNAc-T1 and -T10 bump–hole pairs, using EA2 and MUC5AC-3 substrates, respectively (Figure 6B).[54,57] Of note, GalNAc-T10 contains a GalNAc binding pocket in the glycosyltransferase domain directly adjacent to the active site, conferring selectivity toward pre-glycosylated substrates.[58] The double mutant T10(I266A/L321A) maintains specificity for Thr2 of MUC5AC-3, the only Thr adjacent to an existing glycosite, and exhibits minimally altered kinetic parameters. These data indicate that bump–hole engineering retains the nature of T10(I266A/L321A) as a late isoenzyme.

Peptide	Glycosylated peptide	WT-T2/**1**	T2(I253A/L310A)/**13**
MUC5AC-3 (GTT*PSPVPTTSTTSAP)	GTT*PSPVPTTSTT*SAP	64.7%	63.2%
	GTT*PSPVPTT*STTSAP	22.7%	17.5%
	GTT*PSPVPT*TSTTSAP	12.6%	19.3%
MUC5AC-13 (GTTPSPVPTTSTT*SAP)	G(TT)*PSPVPTTSTT*SAP	95.4%	94.6%
	GTTPSPVPT*TSTT*SAP	4.6%	5.4%
EA2 (PTTDSTTPAPTTK)	PTTDSTT*PAPTTK	100%	100%
Peptide	Glycosylated peptide	WT-T1/**1**	T1(I238A/L295A)/**13**
EA2	PTTDSTT*PAPTTK	100%	100%
Peptide	Glycosylated peptide	WT-T10/**1**	T10(I266A/L321A)/**13**
MUC5AC-3	GT*T*PSPVPTTSTTSAP	100%	100%

Figure 6. Glycosylation of natural peptide substrates by wild-type and engineered GalNAc-T isoenzyme–substrate pairs. (A) Scheme for glycosylation reaction with GalNAc-T, natural peptide substrate, and UDP-GalNAc or UDP-GalNAc analog to form glycosylated peptide. Glycosylation reactions were terminated at 10–20% glycopeptide formation. (B) Percent of glycosylated peptide formed out of total glycosylated peptide formed. Reactions were performed as in A at 37 °C and quenched by the addition of aqueous EDTA (150 mM, pH = 8.0). Naturally-occurring glycopeptides MUC5AC-3 and MUC5AC-13 each contain a single GalNAc-O-Thr (T*). Red T* indicates the site of glycosylation by the GalNAc-T of interest. Glycosylation of MUC5AC-3 by GalNAc-T2 yields a major product that is glycosylated at Thr3 for wild-type GalNAc-T2/**1** and either Thr2 or Thr3 for T2(I253A/L310A)/**13**, labeled (TT)*. (C) Representative MS/MS spectrum of EA2 glycosylated by T2(1253A/L310A)/**13** upon fragmentation and sequencing. Fragmentation pattern of EA2 amino acid sequence to generate *c*-ions (blue) and *z*-ions (red) is shown.

Conclusions

Taken together, these data show that GalNAc-T bump–hole pairs retain the glycosylation site fine specificities and general kinetic parameters of the native enzyme–substrate pairs. Such an outcome is promising for application of this technology to the discovery of new protein substrates and glycosylation sites in living systems. It is particularly notable that this strategy translates successfully to GalNAc-T10, which, as a late isoenzyme, is notoriously difficult to study.

This work represents the first orthogonal bump–hole system for a glycosyltransferase family, an achievement that would not have been possible without structural data, particularly the first structure of GalNAc-T10 bound to GalNAc and UDP reported in 2006.[39] Currently, five total GalNAc-T family members have been structurally characterized with the addition of several recent crystals,[15] establishing a foundation for the identification of promising gatekeeper residues to target for mutagenesis. Further forays into structural characterization of other glycosyltransferase families are warranted to expand this approach across the larger enzyme superfamily.

It should be noted that, while structure data formed a cornerstone to this study, the ability to synthesize and rapidly screen many substrate analogs proved critical to identifying orthogonal bump–hole pairs. By exploring chemical space around the *N*-acyl position, we identified substrates with exquisite specificity for double mutant GalNAc-Ts. We envision that bump–hole pairs such as those described here could be generally applicable across the GalNAc-T family, given their well-conserved UDP-GalNAc binding site.[10,15] The bump–hole enzyme–substrate pairs we developed here set the stage for a new approach to the identification of biological substrates of GalNAc-Ts in living systems, analogous to work done with other enzyme families.[31,59,60]

Methods

Expression of GalNAc-Ts. The soluble domains of GalNAc-T2, -T1, and -T10 include both the catalytic and lectin domains, and the design of our mammalian secretion constructs were based on published constructs.[43] These truncated constructs were cloned into pFLAG-myc-CMV19, containing an N-terminal preprotrypsin leader sequence and an N-terminal FLAG tag. Mutations in GalNAc-T2 and -T1 were introduced using site-directed mutagenesis and mutations in -T10 were introduced during gene synthesis (Table S2).

Truncated GalNAc-Ts were expressed in HEK-293T cells and purified from the culture medium by FLAG affinity chromatography. Glycerol was added to purified proteins to a final concentration of 25% (v/v). Proteins were quantified by densitometry of Coomassie-stained SDS-PAGE gel bands, aliquoted and stored at −80 °C. Western blot confirmed the identity of FLAG-tagged GalNAc-Ts (Figure S6).

Representative experimental procedure for the preparation of UDP-*N*-acetyl-α-D-galactosamine derivatives (Figure 2B). Route 1: 1,3,4,6-Tetra-*O*-acetyl-2-amino-2-deoxy-β-D-galactopyranose hydrochloride and azido acids were prepared according to literature procedures. A mixture of 1,3,4,6-tetra-*O*-acetyl-2-amino-2-deoxy-β-D-galactopyranose (192 mg, 0.500 mmol), the azido acid (0.500 mmol), and Hünig's base (0.261 mL, 1.50 mmol) in DMF (4.00 mL) in a 25-mL round-bottom flask was cooled to 0 °C. COMU® was added, and the reaction mixture was stirred at 0 °C for 1 h. The solution was allowed to warm to r.t. and stirred for 3 h. The mixture was diluted by the addition of ethyl acetate (50 mL), rinsed with HCl (1 M; 2 x 10 mL), saturated aqueous $NaHCO_3$ (2 x 10 mL), and brine (20 mL), dried over Na_2SO_4, and concentrated. The product was purified by column chromatography.

Diallyl galactosyl 1-phosphates were prepared and deallylated according to a literature procedure.[61] Tri-*O*-acetylated UDP-sugars were prepared by treating sugar 1-phosphates (0.200 mmol) with the uridine 5′-monophosphomorpholidate 4-morpholine-*N*,*N*′-dicyclohexylcarboxamidine salt (224 mg, 0.326 mmol), 1-methylimidazole hydrochloride (128 mg, 1.08 mmol), and NEt_3 (55.8 μL, 0.400 mmol) in DMF (3.92 mL) in a 25-mL round-bottom flask at r.t. for 12 h.[52] The tri-*O*-acetylated UDP-sugar was purified by column chromatography on C-18 silica gel and then preparative HPLC on C-18 silica gel. The compound was dissolved in MeOH/water/NEt_3 (5 mL, 5:2:1) in a 25-mL round-bottom flask, and the reaction mixture was stirred at r.t. overnight. The product was purified by preparative HPLC on C-18 silica gel. Finally, the purified compound was passed through a Bio-Rad AG® 50W-X8 resin (Na^+ form) and lyophilized.

Route 2: *N*-Hydroxysuccinimide (NHS) esters were prepared according to a literature procedure.[62] A solution of the NHS ester (0.150 mmol) in DMF (1.08 mL) was added to a mixture of UDP-D-galactosamine disodium salt (30.5 mg, 0.0500 mmol) in HEPES buffer (0.1 M, pH = 8.0; 1.08 mL) at 0 °C. The reaction mixture was allowed to warm to r.t. and stirred overnight. Next, the mixture was purified by column chromatography on C-18 silica gel and then preparative HPLC on C-18 silica gel. Finally, the purified compound was passed through a Bio-Rad AG® 50W-X8 resin (Na^+ form) and lyophilized.

Experimental procedure for the preparation of peptides (Figure 2A). Peptide-1 and -2: Peptides were synthesized on 2-chlorotrityl chloride resin by solid phase peptide synthesis using *N*-Fmoc-protected amino acids. Each coupling step was performed with *N*-Fmoc-protected amino acid (10 equiv.), COMU® (10 equiv.), and *N*,*N*-diisopropylethylamine (20 equiv.) in DMF at r.t. for 30 min under N_2 agitation. Fmoc deprotection was conducted with 20% piperidine in DMF at r.t. for 20 min under N_2 agitation. The N-terminus was reacted with (*S*)-2-((5-fluoro-2,4-dinitrophenyl)amino)propanamide (10 equiv.) and *N*,*N*-diisopropylethylamine (10 equiv.) in DMF at r.t. overnight under N_2 agitation. Peptides were cleaved and deprotected by a mixture of trifluoroacetic acid (88% v/v), triisopropylsilane (2% v/v), 1,4-dithiothreitol (5% w/v), and water (5% v/v) at r.t. for 1.5 h. The mixture was concentrated, triturated with cold Et_2O, redissolved in water, and lyophilized. The desired peptide was purified by preparative HPLC on C-18 silica gel. **Peptide-3:** *N*-(9-Fluorenylmethyloxycarbonyl)-*O*-(2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-galactopyranosyl)-L-threonine was prepared according to literature procedures.[63–65] Peptides were synthesized on 2-chlorotrityl chloride resin by solid phase peptide synthesis using *N*-Fmoc-protected amino acids. Each coupling step was performed with *N*-Fmoc-protected amino acid (10 equiv.), COMU® (10 equiv.), and *N*,*N*-diisopropylethylamine (20 equiv.) in DMF at r.t. for 30 min under N_2 agitation. For the reaction with the glycosylated amino acid, the coupling reaction was conducted with *N*-(9-fluorenylmethyloxycarbonyl)-*O*-(2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-galactopyranosyl)-L-threonine (2 equiv.), COMU® (2 equiv.), and *N*,*N*-diisopropylethylamine (4 equiv.) in DMF at r.t. overnight under N_2 agitation. Fmoc deprotection was conducted with 20% piperidine in DMF at r.t.

for 20 min under N_2 agitation. The N-terminus was reacted with (*S*)-2-((5-fluoro-2,4-dinitrophenyl)amino)propanamide (10 equiv.) and *N*,*N*-diisopropylethylamine (10 equiv.) in DMF at r.t. overnight under N_2 agitation. Peptides were cleaved and deprotected by a mixture of trifluoroacetic acid (88% v/v), triisopropylsilane (2% v/v), 1,4-dithiothreitol (5% w/v), and water (5% v/v) at r.t. for 1.5 h. The mixture was concentrated, triturated with cold Et_2O, redissolved in water, and lyophilized. The crude product was purified by preparative HPLC on C-18 silica gel. For the deacetylation of the sugar moiety, purified glycopeptides were treated with aqueous hydrazine (5%) for 1 h. The deprotected glycopeptides were purified by preparative HPLC on C-18 silica gel.

Representative procedure for the glycosylation by GalNAc-Ts with UDP-GalNAc or UDP-GalNAc analogs (Figures 3, S5, 5A, and 5B). The glycosylation reaction was initiated by the addition of wild-type or mutant GalNAc-T (-T2 = 50.0 nM; -T1 = 160 nM; or -T10 = 120 nM) in Tris-HCl buffer (16.7 mM Tris-HCl, 100 mM NaCl, 25% glycerol, pH = 7.4; 25.0 μL) to the mixture of UDP-sugar (500 μM), and peptide (Peptide-1, Peptide-2, or Peptide-3; 100 μM) in Tris-HCl buffer (25 mM Tris-HCl, 20 mM $MnCl_2$; 25.0 μL) at 0 °C, resulting in a final reaction mixture containing GalNAc-T (-T2 = 25.0 nM; -T1 = 80.0 nM; -T10 = 60.0 nM), peptide (Peptide-1, Peptide-2, or Peptide-3; 50.0 μM), UDP-sugar (250 μM) in Tris-HCl buffer (20.8 mM Tris-HCl, 10 mM $MnCl_2$, 50 mM NaCl, 12.5% glycerol, pH = 7.4; 50.0 μL). The glycosylation was conducted at 37 °C for 1 h (-T2; -T10) or 2 h (-T1) and quenched by the addition of aqueous EDTA (150 mM, pH = 8.0; 25.0 μL). Glycopeptide formation was determined by HPLC and peak integration.

Michaelis–Menten kinetics, UDP-sugar competition experiments, and glycosylation of natural peptide substrates including mass spectrometry-based sequencing are described in the Supporting Information.

Further details can be found in the Supporting Information.

Safety statement: No unexpected or unusually high safety hazards were encountered.

Associated Content

Supporting Information

Additional experimental details, data, and figures including crystal structure modeling (Figure S1), GalNAc-T multiple sequence alignment (Figure S2), enzymatic activity screens (Figures S3, S4, and S5), GalNAc-T expression data (Figure S6), kinetics graphs, MS/MS spectra, synthetic schemes, and NMR spectra.

Author Information

Corresponding Author

*E-mail: bertozzi@stanford.edu

Author Contributions

[1]J.C. and L.W. contributed equally to this work.

Present addresses

[II]Chemical Kinomics Research Center, Korea Institute of Science and Technology (KIST), 5 Hwarangro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea

[III]Bio-Organic Chemistry Research Group, Institute for Complex Molecular Systems, Eindhoven University of Technology, PO Box 513, 5600 MB, Eindhoven, The Netherlands

[IIII]The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK, and Imperial College London, 80 Wood Lane, London W12 0BZ, UK

[IIIII]Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA

[IIIIII]Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

[IIIIIII]Department of Biology and Biological Engineering, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA

Notes

The authors declare no competing financial interest.

Acknowledgements

We thank Lawrence Tabak (National Institutes of Health, Bethesda, MD) for the kind gift of the gene for full length human GalNAc-T2 in the plasmid pCMV-NTAP. This work was supported by NIH

(R01 CA200423). J.C. was supported by the Korea Institute of Science and Technology (KIST). S.B.P.E.T. was supported by the Nora Baart Foundation and the Stichting Jo Kolk Studiefonds. S.A.M. was supported by a National Institute of General Medical Sciences F32 Postdoctoral Fellowship (F32-GM126663-01). B.S. was supported by a Feodor Lynen Fellowship by the Alexander von Humboldt Foundation. M.A.G. was supported by the National Science Foundation Graduate Research Fellowship (NSF GRFP) and the Stanford ChEM-H Chemistry/Biology Interface Predoctoral Training Program. M.F.D. was supported by a NWO Rubicon Postdoctoral Fellowship. J.C. thanks Dr. Mason Appel and Dr. Neil G. Rumachik for critical discussions.

References

(1) Vakhrushev, S. Y.; Steentoft, C.; Vester-Christensen, M. B.; Bennett, E. P.; Clausen, H.; Levery, S. B. Enhanced Mass Spectrometric Mapping of the Human GalNAc-Type O-Glycoproteome with SimpleCells. *Mol. Cell. Proteomics* **2013**, *12* (4), 932–944.

(2) Tran, D. T.; Zhang, L.; Zhang, Y.; Tian, E.; Earl, L. A.; Ten Hagen, K. G. Multiple Members of the UDP-GalNAc: Polypeptide *N*-Acetylgalactosaminyltransferase Family Are Essential for Viability in Drosophila. *J. Biol. Chem.* **2012**, *287* (8), 5243–5252.

(3) Xu, Z.; Weiss, A. Negative Regulation of CD45 by Differential Homodimerization of the Alternatively Spliced Isoforms. *Nat. Immunol.* **2002**, *3* (8), 764–771.

(4) Schjoldager, K. T.-B. G.; Vester-Christensen, M. B.; Bennett, E. P.; Levery, S. B.; Schwientek, T.; Yin, W.; Blixt, O.; Clausen, H. O-Glycosylation Modulates Proprotein Convertase Activation of Angiopoietin-like Protein 3 . *J. Biol. Chem.* **2010**, *285* (47), 36293–36303.

(5) Radhakrishnan, P.; Dabelsteen, S.; Madsen, F. B.; Francavilla, C.; Kopp, K. L.; Steentoft, C.; Vakhrushev, S. Y.; Olsen, J. V.; Hansen, L.; Bennett, E. P.; Woetmanne, A.; Yin, G.; Chen, L.; Song, H.; Bak, M.; Hlady, R. A.; Peters, S. L.; Opavsky, R.; Thode, C.; Qvortrup, K.; Schjoldager, K. T.-B. G.; Clausen, H.; Hollingsworth, M. A.; Wandall, H. H. Immature

Truncated O-Glycophenotype of Cancer Directly Induces Oncogenic Features. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (39), E4066–E4075.

(6) Kingsley, P. D.; Ten Hagen, K. G.; Maltby, K. M.; Zara, J.; Tabak, L. A. Diverse Spatial Expression Patterns of UDP-GalNAc: Polypeptide *N*-Acetylgalactosaminyl-Transferase Family Member mRNAs during Mouse Development. *Glycobiology* **2000**, *10* (12), 1317–1323.

(7) Peng, R.-Q.; Wan, H.-Y.; Li, H.-F.; Liu, M.; Li, X.; Tang, H. MicroRNA-214 Suppresses Growth and Invasiveness of Cervical Cancer Cells by Targeting UDP-*N*-Acetyl-α-D-Galactosamine:Polypeptide *N*-Acetylgalactosaminyltransferase 7. *J. Biol. Chem.* **2012**, *287* (17), 14301–14309.

(8) Lavrsen, K.; Dabelsteen, S.; Vakhrushev, S. Y.; Levann, A. M. R.; Haue, A. D.; Dylander, A.; Mandel, U.; Hansen, L.; Frödin, M.; Bennett, E. P.; Wandall, H. H. De Novo Expression of Human Polypeptide *N*-Acetylgalactosaminyltransferase 6 (GalNAc-T6) in Colon Adenocarcinoma Inhibits the Differentiation of Colonic Epithelium. *J. Biol. Chem.* **2018**, *293* (4), 1298–1314.

(9) Khetarpal, S. A.; Schjoldager, K. T.; Christoffersen, C.; Raghavan, A.; Edmondson, A. C.; Reutter, H. M.; Ahmed, B.; Ouazzani, R.; Peloso, G. M.; Vitali, C.; Zhao, W.; Somasundara, A. V. H.; Millar; J. S.; Park, Y.; Fernando, G.; Livanov, V.; Choi, S.; Noé, E.; Patel, P.; Ho, S. P.; Myocardial Infarction Exome Sequencing Study; Kirchgessner, T. G.; Wandall, H. H.; Hansen, L.; Bennett, E. P.; Vakhrushev, S. Y.; Saleheen, D.; Kathiresan, S.; Brown, C. D.; Jamra, R. A.; LeGuern, E.; Clausen, H.; Rader, D. J. Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans, Nonhuman Primates, and Rodents. *Cell Metab.* **2016**, *24* (2), 234–245.

(10) Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide GalNAc-Transferase Gene Family. *Glycobiology* **2012**, *22* (6), 736–756.

(11) Block, H.; Ley, K.; Zarbock, A. Severe Impairment of Leukocyte Recruitment in ppGalNAcT-1-Deficient Mice. *J. Immunol.* **2012**, *188* (11), 5674–5681.

(12) Ten Hagen, K. G.; Fritz, T. A.; Tabak, L. A. All in the Family: The UDP-GalNAc:Polypeptide *N*-Acetylgalactosaminyltransferases. *Glycobiology* **2003**, *13* (1), 1R–16R.

(13) Boskovski, M. T.; Yuan, S.; Pedersen, N. B.; Goth, C. K.; Makova, S.; Clausen, H.; Brueckner, M.; Khokha, M. K. The Heterotaxy Gene GALNT11 Glycosylates Notch to Orchestrate Cilia Type and Laterality. *Nature* **2013**, *504*, 456–459. https://doi.org/10.1038/nature12723.

(14) Woods, E. C.; Kai, F.; Barnes, J. M.; Pedram, K.; Pickup, M. W.; Hollander, M. J.; Weaver, V. M.; Bertozzi, C. R. A Bulky Glycocalyx Fosters Metastasis Formation by Promoting G1 Cell Cycle Progression. *elife* **2017**, *6*, e25752.

(15) de las Rivas, M.; Lira-Navarrete, E.; Gerken, T. A.; Hurtado-Guerrero, R. Polypeptide GalNAc-Ts: From Redundancy to Specificity. *Curr. Opin. Struct. Biol.* **2019**, *56*, 87–96.

(16) Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T.-B. G.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; Marcos-Silva, L.; Gupta, R.; Bennett, E. P.; Mandel, U.; Brunak, S.; Wandall, H. H.; Levery, S. B.; Clausen, H. Precision Mapping of the Human O-GalNAc Glycoproteome through SimpleCell Technology. *EMBO J.* **2013**, *32*, 1478–1488.

(17) Cejas, R. B.; Lorenz, V.; Garay, Y. C.; Irazoqui, F. J. Biosynthesis of O-*N*-Acetylgalactosamine Glycans in the Human Cell Nucleus . *J. Biol. Chem.* **2019**, *294* (9), 2997–3011.

(18) Schjoldager, K. T.; Joshi, H. J.; Kong, Y.; Goth, C. K.; King, S. L.; Wandall, H. H.; Bennett, E. P.; Vakhrushev, S. Y.; Clausen, H. Deconstruction of O-Glycosylation–GalNAc-T Isoforms Direct Distinct Subsets of the O-Glycoproteome. *EMBO Rep.* **2015**, *16* (12), 1713–1722.

(19) Kato, K.; Jeanneau, C.; Tarp, M. A.; Benet-Pagès, A.; Lorenz-Depiereux, B.; Bennett, E. P.;

Mandel, U.; Strom, T. M.; Clausen, H. Polypeptide GalNAc-Transferase T3 and Familial Tumoral Calcinosis. *J. Biol. Chem.* **2006**, *281* (27), 18370–18377.

(20) Revoredo, L.; Wang, S.; Bennett, E. P.; Clausen, H.; Moremen, K. W.; Jarvis, D. L.; Ten Hagen, K. G.; Tabak, L. A.; Gerken, T. A. Mucin-Type O-Glycosylation Is Controlled by Short- and Long-Range Glycopeptide Substrate Recognition that Varies among Members of the Polypeptide GalNAc Transferase Family. *Glycobiology* **2016**, *26* (4), 360–376.

(21) Gerken, T. A.; Raman, J.; Fritz, T. A.; Jamison, O. Identification of Common and Unique Peptide Substrate Preferences for the UDP-GalNAc:Polypeptide α-*N*-Acetylgalactosaminyltransferases T1 and T2 Derived from Oriented Random Peptide Substrates. *J. Biol. Chem.* **2006**, *281* (43), 32403–32416.

(22) Gerken, T. A.; Jamison, O.; Perrine, C. L.; Collette, J. C.; Moinova, H.; Ravi, L.; Markowitz, S. D.; Shen, W.; Patel, H.; Tabak, L. A. Emerging Paradigms for the Initiation of Mucin-Type Protein O-Glycosylation by the Polypeptide GalNAc Transferase Family of Glycosyltransferases. *J. Biol. Chem.* **2011**, *286* (16), 14493–14507.

(23) Steentoft, C.; Vakhrushev, S. Y.; Vester-Christensen, M. B.; Schjoldager, K. T. B. G.; Kong, Y.; Bennett, E. P.; Mandel, U.; Wandall, H.; Levery, S. B.; Clausen, H. Mining the O-Glycoproteome Using Zinc-Finger Nuclease-Glycoengineered SimpleCell Lines. *Nat. Methods* **2011**, *8* (11), 977–982.

(24) Hintze, J.; Ye, Z.; Narimatsu, Y.; Madsen, T. D.; Joshi, H. J.; Goth, C. K.; Linstedt, A.; Bachert, C.; Mandel, U.; Bennett, E. P.; Vakhrushev, S. Y.; Schjoldager, K. T. Probing the Contribution of Individual Polypeptide GalNAc-Transferase Isoforms to the O-Glycoproteome by Inducible Expression in Isogenic Cell Lines. *J. Biol. Chem.* **2018**, *293* (49), 19064–19077.

(25) Schjoldager, K. T.-B. G.; Vakhrushev, S. Y.; Kong, Y.; Steentoft, C.; Nudelman, A. S.; Pedersen, N. B.; Wandall, H. H.; Mandel, U.; Bennett, E. P.; Levery, S. B.; Clausen, H. Probing Isoform-Specific Functions of Polypeptide GalNAc-Transferases Using Zinc Finger Nuclease Glycoengineered SimpleCells. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (25), 9893–

9898.

(26) Fouda, A. E.; Pflum, M. K. H. A Cell-Permeable ATP Analogue for Kinase-Catalyzed Biotinylation. *Angew. Chem., Int. Ed.* **2015**, *54* (33), 9618–9621.

(27) Blethrow, J. D.; Glavy, J. S.; Morgan, D. O.; Shokat, K. M. Covalent Capture of Kinase-Specific Phosphopeptides Reveals Cdk1-Cyclin B Substrates. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (5), 1442–1447.

(28) Yang, C.; Mi, J.; Feng, Y.; Ngo, L.; Gao, T.; Yan, L.; Zheng, Y. G. Labeling Lysine Acetyltransferase Substrates with Engineered Enzymes and Functionalized Cofactor Surrogates. *J. Am. Chem. Soc.* **2013**, *135* (21), 7791–7794.

(29) Islam, K.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Expanding Cofactor Repertoire of Protein Lysine Methyltransferase. *ACS Chem. Biol.* **2011**, *6* (7), 679–684.

(30) Carter-O'Connell, I.; Jin, H.; Morgan, R. K.; David, L. L.; Cohen, M. S. Engineering the Substrate Specificity of ADP-Ribosyltransferases for Identifying Direct Protein Targets. *J. Am. Chem. Soc.* **2014**, *136* (14), 5201–5204.

(31) Islam, K. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem. Biol.* **2018**, *25* (10), 1171–1184.

(32) Rodriguez, A. C.; Yu, S. H.; Li, B.; Zegzouti, H.; Kohler, J. J. Enhanced Transfer of a Photocross-Linking *N*-Acetylglucosamine (GlcNAc) Analog by an O-GlcNAc Transferase Mutant with Converted Substrate Specificity. *J. Biol. Chem.* **2015**, *290* (37), 22638–22648.

(33) Ramakrishnan, B.; Qasba, P. K. Structure-Based Design of β1,4-Galactosyltransferase I (β4Gal-T1) with Equally Efficient *N*-Acetylgalactosaminyltransferase Activity *J. Biol. Chem.* **2002**, *277* (23), 20833–20839.

(34) Clark, P. M.; Dweck, J. F.; Mason, D. E.; Hart, C. R.; Buck, S. B.; Peters, E. C.; Agnew, B. J.; Hsieh-Wilson, L. C. Direct In-Gel Fluorescence Detection and Cellular Imaging of O-GlcNAc-Modified Proteins. *J. Am. Chem. Soc.* **2008**, *130* (35), 11576–11577.

(35) Fritz, T. A.; Hurley, J. H.; Trinh, L.-B.; Shiloach, J.; Tabak, L. A. The Beginnings of Mucin Biosynthesis: The Crystal Structure of UDP-GalNAc:Polypeptide-*N*-Acetylgalactosaminyltransferase-T1. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (43), 15307–15312.

(36) Fritz, T. A.; Raman, J.; Tabak, L. A. Dynamic Association between the Catalytic and Lectin Domains of Human UDP-GalNAc:Polypeptide α-*N*-Acetylgalactosaminyltransferase-2. *J. Biol. Chem.* **2006**, *281* (13), 8613–8619.

(37) Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañón, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide *N*-Acetylgalactosaminyltransferase 2. *Angew. Chem., Int. Ed.* **2014**, *53* (31), 8206–8210.

(38) Lira-Navarrete, E.; de las Rivas, M.; Compañón, I.; Pallarés, M. C.; Kong, Y.; Iglesias-Fernández, J.; Bernardes, G. J. L.; Peregrina, J. M.; Rovira, C.; Bernadó, P.; Bruscolini, P.; Clausen, H.; Lostao, A.; Corzana, F.; Hurtado-Guerrero, R. Dynamic Interplay between Catalytic and Lectin Domains of GalNAc-Transferases Modulates Protein O-Glycosylation. *Nat. Commun.* **2015**, *6*, 6937.

(39) Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H. Structural Basis of Carbohydrate Transfer Activity by Human UDP-GalNAc: Polypeptide α-*N*-Acetylgalactosaminyltransferase (pp-GalNAc-T10). *J. Mol. Biol.* **2006**, *359* (3), 708–727.

(40) Hang, H. C.; Yu, C.; Kato, D. L.; Bertozzi, C. R. A Metabolic Labeling Approach toward Proteomic Analysis of Mucin-Type O-Linked Glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (25), 14846–14851.

(41) Prescher, J. A.; Dube, D. H.; Bertozzi, C. R. Chemical Remodelling of Cell Surfaces in Living Animals. *Nature* **2004**, *430*, 873–877.

(42) Pratt, M. R.; Hang, H. C.; Ten Hagen, K. G.; Rarick, J.; Gerken, T. A.; Tabak, L. A.; Bertozzi, C. R. Deconvoluting the Functions of Polypeptide *N*-α-Acetylgalactosaminyltransferase Family Members by Glycopeptide Substrate Profiling. *Chem. Biol.* **2004**, *11* (7), 1009–1016.

(43) Hagen, F. K.; Ten Hagen, K. G.; Beres, T. M.; Balys, M. M.; VanWuyckhuyse, B. C.; Tabak, L. A. cDNA Cloning and Expression of a Novel UDP-*N*-Acetyl-D-Galactosamine:Polypeptide *N*-Acetylgalactosaminyltransferase. *J. Biol. Chem.* **1997**, *272* (21), 13843–13848.

(44) Wagner, L. J. S. Engineering an Orthogonal {O-Glycosyltransferase} and Donor Sugar Pair. Ph.D. Dissertation, University of California, Berkeley, 2015.

(45) Sheikh, M. O.; Halmo, S. M.; Patel, S.; Middleton, D.; Takeuchi, H.; Schafer, C. M.; West, C. M.; Haltiwanger, R. S.; Avci, F. Y.; Moremen, K. W.; Wells, L. Rapid Screening of Sugar-Nucleotide Donor Specificities of Putative Glycosyltransferases. *Glycobiology* **2017**, *27* (3), 206–212.

(46) Hang, H. C.; Yu, C.; Pratt, M. R.; Bertozzi, C. R. Probing Glycosyltransferase Activities with the Staudinger Ligation. *J. Am. Chem. Soc.* **2004**, *126* (1), 6–7.

(47) Guan, W.; Cai, L.; Wang, P. G. Highly Efficient Synthesis of UDP-GalNAc/GlcNAc Analogues with Promiscuous Recombinant Human UDP-GalNAc Pyrophosphorylase AGX1. *Chem.—Eur. J.* **2010**, *16* (45), 13343–13345.

(48) Wolf, S.; Zismann, T.; Lunau, N.; Meier, C. Reliable Synthesis of Various Nucleoside Diphosphate Glycopyranoses. *Chem.—Eur. J.* **2009**, *15* (31), 7656–7664.

(49) Heidlas, J. E.; Lees, W. J.; Whitesides, G. M. Practical Enzyme-Based Syntheses of Uridine 5′-Diphosphogalactose and Uridine 5′-Diphospho-*N*-Acetylgalactosamine on a Gram Scale. *J. Org. Chem.* **1992**, *57* (1), 152–157.

(50) Muthana, M. M.; Qu, J.; Li, Y.; Zhang, L.; Yu, H.; Ding, L.; Malekan, H.; Chen, X. Efficient One-Pot Multienzyme Synthesis of UDP-Sugars Using a Promiscuous UDP-Sugar Pyrophosphorylase from *Bifidobacterium Longum* (BLUSP). *Chem. Commun.* **2012**, *48* (21),

2728–2730.

(51) Pouilly, S.; Bourgeaux, V.; Piller, F.; Piller, V. Evaluation of Analogues of GalNAc as Substrates for Enzymes of the Mammalian GalNAc Salvage Pathway. *ACS Chem. Biol.* **2012**, *7* (4), 753–760.

(52) Tsukamoto, H.; Kahne, D. *N*-Methylimidazolium Chloride-Catalyzed Pyrophosphate Formation: Application to the Synthesis of Lipid I and NDP-Sugar Donors. *Bioorg. Med. Chem. Lett.* **2011**, *21* (17), 5050–5053.

(53) Batt, A. R.; Zaro, B. W.; Navarro, M. X.; Pratt, M. R. Metabolic Chemical Reporters of Glycans Exhibit Cell-Type-Selective Metabolism and Glycoprotein Labeling. *ChemBioChem* **2017**, *18* (13), 1177–1182.

(54) Cheng, L.; Tachibana, K.; Zhang, Y.; Guo, J.; Tachibana, K. K.; Kameyama, A.; Wang, H.; Hiruma, T.; Iwasaki, H.; Togayachi, A.; Kudo, T.; Narimatsu, H. Characterization of a Novel Human UDP-GalNAc Transferase, pp-GalNAc-T10. *FEBS Lett.* **2002**, *531* (2), 115–121.

(55) Kightlinger, W.; Lin, L.; Rosztoczy, M.; Li, W.; Delisa, M. P.; Mrksich, M.; Jewett, M. C. Design of Glycosylation Sites by Rapid Synthesis and Analysis of Glycosyltransferases. *Nat. Chem. Biol.* **2018**, *14* (6), 627–635.

(56) Raman, J.; Fritz, T. A.; Gerken, T. A.; Jamison, O.; Live, D.; Liu, M.; Tabak, L. A. The Catalytic and Lectin Domains of UDP-GalNAc:Polypeptide α-*N*-Acetylgalactosaminyltransferase Function in Concert to Direct Glycosylation Site Selection. *J. Biol. Chem.* **2008**, *283* (34), 22942–22951.

(57) Tenno, M.; Ohtsubo, K.; Hagen, F. K.; Ditto, D.; Zarbock, A.; Schaerli, P.; von Andrian, U. H.; Ley, K.; Le, D.; Tabak, L. A.; Marth, J. D. Initiation of Protein O Glycosylation by the Polypeptide GalNAcT-1 in Vascular Biology and Humoral Immunity. *Mol. Cell. Biol.* **2007**, *27* (24), 8783–8796.

(58) Perrine, C. L.; Ganguli, A.; Wu, P.; Bertozzi, C. R.; Fritz, T. A.; Raman, J.; Tabak, L. A.;

Gerken, T. A. Glycopeptide-Preferring Polypeptide GalNAc Transferase 10 (ppGalNAc T10), Involved in Mucin-Type O-Glycosylation, Has a Unique GalNAc-*O*-Ser/Thr-Binding Site in Its Catalytic Domain Not Found in ppGalNAc T1 or T2. *J. Biol. Chem.* **2009**, *284* (30), 20387–20397.

(59) Banko, M. R.; Allen, J. J.; Schaffer, B. E.; Wilker, E. W.; Tsou, P.; White, J. L.; Villén, J.; Wang, B.; Kim, S. R.; Sakamoto, K.; Gygi, S. P.; Cantley, L. C.; Yaffe, M. B.; Shokat, K. M.; Brunet, A. Molecular Cell Chemical Genetic Screen for AMPKα2 Substrates Uncovers a Network of Proteins Involved in Mitosis. *Mol. Cell* **2011**, 44 (6), 878–892.

(60) Wang, R.; Islam, K.; Liu, Y.; Zheng, W.; Tang, H.; Lailler, N.; Blum, G.; Deng, H.; Luo, M. Profiling Genome-Wide Chromatin Methylation with Engineered Posttranslation Apparatus within Living Cells. *J. Am. Chem. Soc.* **2013**, *135* (3), 1048–1056.

(61) Beahm, B. J.; Dehnert, K. W.; Derr, N. L.; Kuhn, J.; Eberhart, J. K.; Spillmann, D.; Amacher, S. L.; Bertozzi, C. R. A Visualizable Chain-Terminating Inhibitor of Glycosaminoglycan Biosynthesis in Developing Zebrafish. *Angew. Chem., Int. Ed.* **2014**, *53* (13), 3347–3352.

(62) Zhang, Z.; Hejesen, C.; Kjelstrup, M. B.; Birkedal, V.; Gothelf, K. V. A DNA-Mediated Homogeneous Binding Assay for Proteins and Small Molecules. *J. Am. Chem. Soc.* **2014**, *136* (31), 11115–11120.

(63) Plattner, C.; Höfener, M.; Sewald, N. One-Pot Azidochlorination of Glycals. *Org. Lett.* **2011**, *13* (4), 545–547.

(64) Winans, K. A.; King, D. S.; Rao, V. R.; Bertozzi, C. R. A Chemically Synthesized Version of the Insect Antibacterial Glycopeptide, Diptericin, Disrupts Bacterial Membrane Integrity. *Biochemistry* **1999**, *38* (36), 11700–11710.

(65) Miermont, A.; Barnhill, H.; Strable, E.; Lu, X.; Wall, K. A.; Wang, Q.; Finn, M. G.; Huang, X. Cowpea Mosaic Virus Capsid: A Promising Carrier for the Development of Carbohydrate Based Antitumor Vaccines. *Chem.—Eur. J.* **2008**, *14* (16), 4939–4947.

TOC Graphic

Wild type GalNAc-T
UDP-GalNAc
HO
OH
O
HO
HN
O
UDP
O
Me
Peptide
HO
UDP
HO
OH
O
HO
HN
O
UDP
O
R
UDP-GalNAc analog
= N3 ,
= Chromophore
UDP
Install R
Engineered
GalNAc-T
Mutate
Gatekeeper
Residues
Wild type GalNAc-T
Mn2+
UDP-GalNAc