ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

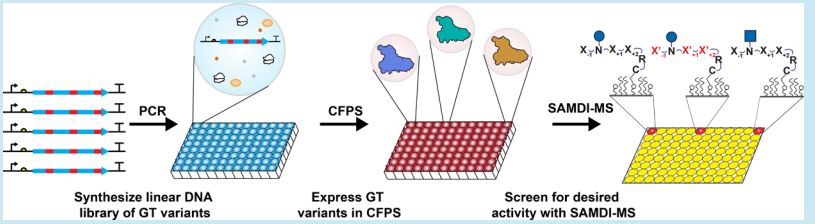
# Using High-Throughput Experiments To Screen *N*-Glycosyltransferases with Altered Specificities

Liang Lin, Weston Kightlinger, Katherine F. Warfel, Michael C. Jewett,* and Milan Mrksich*

Cite This: *ACS Synth. Biol.* 2024, 13, 1290−1302 Read Online








**ABSTRACT:** The important roles that protein glycosylation plays in modulating the activities and efficacies of protein therapeutics have motivated the development of synthetic glycosylation systems in living bacteria and in vitro. A key challenge is the lack of glycosyltransferases that can efficiently and site-specifically glycosylate desired target proteins without the need to alter primary amino acid sequences at the acceptor site. Here, we report an efficient and systematic method to screen a library of glycosyltransferases capable of modifying comprehensive sets of acceptor peptide sequences in parallel. This approach is enabled by cell-free protein synthesis and mass spectrometry of self-assembled monolayers and is used to engineer a recently discovered prokaryotic *N*-glycosyltransferase (NGT). We screened 26 pools of site-saturated NGT libraries to identify relevant residues that determine polypeptide specificity and then characterized 122 NGT mutants, using 1052 unique peptides and 52,894 unique reaction conditions. We define a panel of 14 NGTs that can modify 93% of all sequences within the canonical $X_{-1}$-N-$X_{+1}$-S/T eukaryotic glycosylation sequences as well as another panel for many noncanonical sequences (with 10 of 17 non-S/T amino acids at the $X_{+2}$ position). We then successfully applied our panel of NGTs to increase the efficiency of glycosylation for three protein therapeutics. Our work promises to significantly expand the substrates amenable to in vitro and bacterial glycoengineering.

**KEYWORDS:** *cell-free protein synthesis, synthetic biology, high-throughput, glycosyltransferase, glycosylation, therapeutics*

## INTRODUCTION

*N*-Linked protein glycosylation is the modification of asparagine side chains with oligosaccharides and is among the most common and complex post-translational modification found in nature.[1] In eukaryotes, *N*-glycans are installed at the canonical sequence motif N-X-S/T (where X ≠ P).[2] The majority of protein therapeutics are *N*-glycosylated[3] and differences in glycosylation patterns are known to have strong effects on bioactivity,[4,5] protein stability,[6] and serum half-life.[7] The introduction of additional *N*-glycosylation sites into therapeutic proteins has also been shown to improve therapeutic properties, including prolonged serum half-life.[8,9] While the use of mammalian cell lines with endogenous *N*-glycosylation pathways is the most common method to produce glycoprotein therapeutics, these constitutive systems limit the diversity of glycan structures that can be constructed[10,11] and often suffer from heterogeneity of the glycoprotein products.[3,12,13] These limitations have motivated the development of synthetic glycosylation systems in *Escherichia coli* or in vitro to install[10,14−18] or remodel[12,16] defined glycans to precisely control the structures and properties of glycoproteins.

Among these synthetic glycosylation systems, a class of bacterial cytoplasmic enzymes known as *N*-glycosyltransferases (NGTs) are important for glycoengineering because they can efficiently transfer a single glucose residue from a uracil-diphosphate-glucose (UDP-Glc) sugar donor onto certain native eukaryotic sequences.[19,20] This glucose residue can then be extended into a full-length glycan using glycosyltransferases[21] or endoglycosidase chemoenzymatic glycan remodeling.[22,23] Rigorous analyses of NGT specificities have shown that NGTs can only modify a fraction of all possible eukaryotic *N*-glycosylation sequences.[19,22] Because there is a continuously expanding set of potential therapeutic protein targets that could be optimized by glycoengineering—including proteins lacking the canonical N-X-S/T glycosylation sequences—there is a clear need to engineer or discover NGTs that enable the

**Received:** December 22, 2023
**Revised:** February 29, 2024
**Accepted:** March 1, 2024
**Published:** March 25, 2024





https://doi.org/10.1021/acssynbio.3c00769
*ACS Synth. Biol.* 2024, 13, 1290−1302

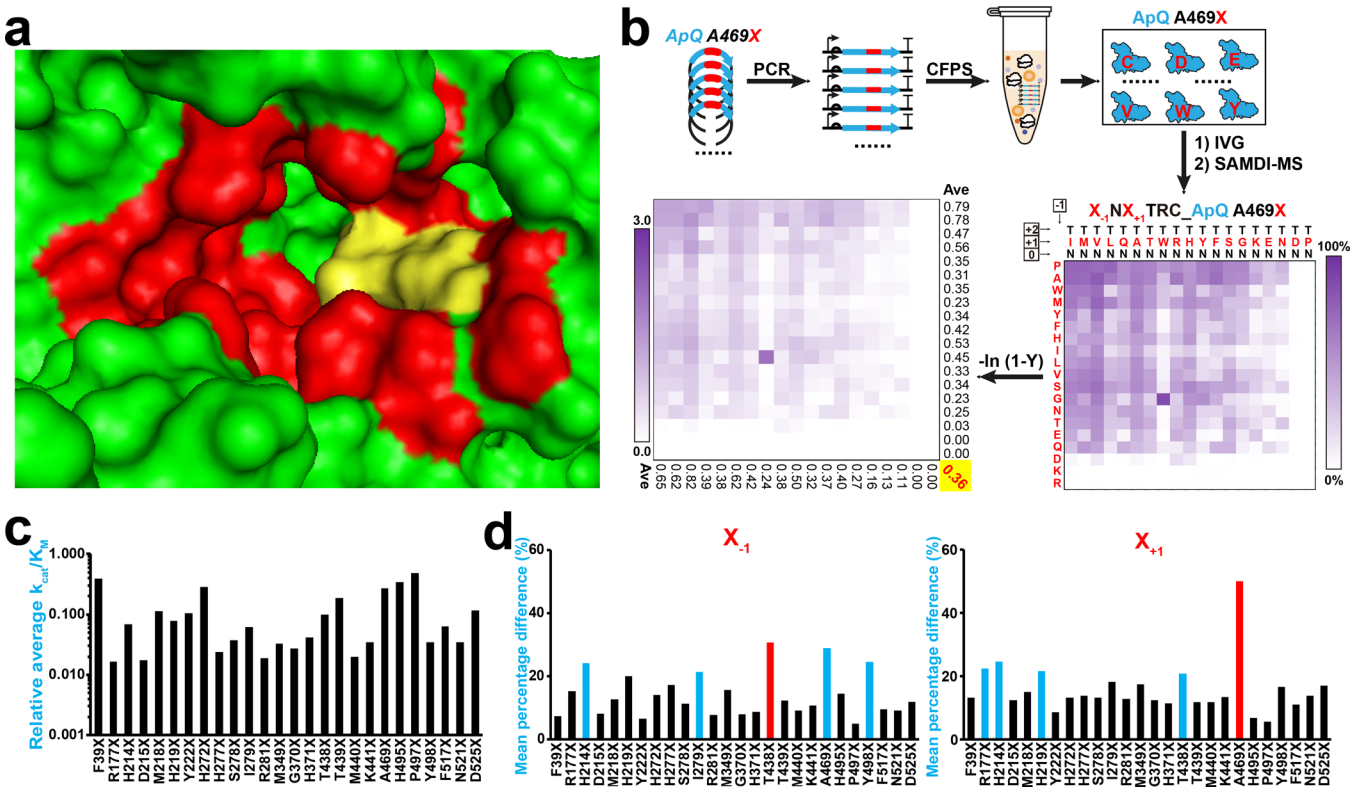



**Figure 1.** Peptide library screening to identify ApQ residues that determine acceptor peptide specificity. (a) Binding pocket of ApNGT (PDB ID: 3Q3H[31]) with 26 potential peptide binding residues (highlighted in red). The donor UDP is marked in yellow. (b) An SSVL containing equal amounts of all 19 nonwildtype amino acids was synthesized as linear DNA template for each residue (A469 as an example). These SSVLs were amplified by PCR to generate LETs and expressed in CFPS to produce protein SSVLs. The protein SSVLs were then used to modify a library of peptide substrates with the motif $X_{-1}$-N-$X_{+1}$-TRC and analyzed via SAMDI-MS. A heatmap of peptide modification is shown (bottom-right), with the same descending order of average modification for amino acids at the $X_{-1}$ and $X_{+1}$ positions, respectively, as ApQ. A new heatmap of $-\ln(1 - Y)$ (bottom-left), where $Y$ is the peptide modification, was generated. Because we use the concentration of peptides much lower than $K_M$, the average $k_{cat}/K_M$ of each SSVL can be calculated using the average $-\ln(1 - Y)$, yellow highlighted in heatmap, with the equation $k_{cat}/K_M = -\ln(1 - Y)/c/t$. The average $k_{cat}/K_M$ relative to ApQ is shown in (c). All SSVLs show decreased average $k_{cat}/K_M$. (d) Average value of $-\ln(1 - Y)$ in each row of $X_{-1}$ amino acid, and each column of $X_{+1}$ amino acid for each SSVL, is compared to that of ApQ, which chooses to have the same value of average $-\ln(1 - Y)$ of the entire library as the SSVL, to show the percentage difference using the equation $2 \times |\text{Ave(X)} - \text{Ave(ApQ)}|/(\text{Ave(X)} + \text{Ave(ApQ)})$. Mean percentage difference of the $X_{-1}$ (left) and $X_{+1}$ (right) positions are presented using the average of the percentage differences for all rows and columns, respectively. Values higher than 20 and 30% are highlighted in blue and red bars, respectively. Given the focus on a rapid screen, all experiments were completed with $n = 1$.

modification of an expanded set of acceptor sequences. The broad goal of this work is to develop a synthetic biology workflow to discover and identify a repertoire of NGTs capable of glycosylating any sequence of interest, alleviating the need to alter the primary amino acid sequences of target proteins for naturally occurring glycosylation sites or introducing new glycosylation sites.

Approaches based on traditional directed evolution are effective for engineering enzyme activity toward a single substrate but are not yet suited to developing enzymes that display multiple functions (i.e., distinct peptide specificities). Wells and collaborators addressed this limitation by developing a method to engineer peptide ligases by identifying modified substrates from the *E. coli* proteome using a liquid-chromatography tandem mass spectrometry (LC-MS/MS) proteomics approach.[24] While proteomic identification works well for ligases and proteases,[25] it is difficult to apply to glycosylation enzymes because enrichment methods for glycopeptides are not generalizable and differences in substrate peptide length strongly affect glycosylation efficiency.[19] Furthermore, proteomic identification provides some information on substrate preferences but does not directly measure activity. We have developed a general and versatile assay based on self-assembled monolayers for matrix-assisted laser desorption/ionization mass spectrometry (SAMDI-MS), which can rapidly and quantitatively measure enzymatic specificities and activities on a large number of substrates without the need to purify enzymes or substrates.[19,26,27] We have combined this method with cell-free protein synthesis (CFPS)[28] of enzymes to create the GlycoSCORES workflow, which we used to analyze the specificity of several NGTs.[19,22]

Here, we report the use of the GlycoSCORES workflow with high-throughput CFPS reactions from PCR-derived linear expression templates (LET-CFPS)[29,30] to develop a panel of NGTs that expands the range of sequences that can be directly glycosylated. Our parallel workflow to develop this panel relies on two key steps. First, we screened acceptor sequence specificity on pools of 26 site-saturated variant libraries (SSVLs) of the parent NGT. Each of the SSVLs is comprised of 19 mutants at a specific residue that was targeted for mutagenesis based on inspection of the NGT crystal structure and expected interactions with the substrate peptide. By screening these SSVLs on substrate peptide libraries, we separately identified residues that determine specificity at the


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

$X_{-2}$, $X_{-1}$, $X_{+1}$, $X_{+2}$, and $X_{+3}$ positions of the substrate peptide, relative to the glycosylated asparagine. Second, we generated and rigorously characterized precise, single or double mutants that collectively expand the set of canonical (N-X-S/T where X ≠ P) and noncanonical (N-X-Z where X ≠ P and Z ≠ S/T) peptide sequences that can be efficiently modified compared to the parent NGT alone. We discovered 13 NGT mutants, in addition to the parent NGT, which significantly increase the fraction of all $X_{-1}$ and $X_{+1}$ canonical sequence combinations (684 in total) that can be modified with high efficiency—in yields of approximately 45−65%. Another panel of NGTs allow for modification of a variety of sequences with the $X_{+2}$ positions besides S/T (10 of 17 amino acids, e.g., Ala, Asp, Met and Val). Moreover, we demonstrated the utility of our NGT mutant panel by increasing the modification efficiency of approved therapeutic proteins, compared to the parent NGT, without modifying their amino acid sequences. We expect that this method will be helpful in the development of additional enzymes with altered specificities, and we anticipate that the NGT mutants discovered here will significantly expand the application areas for bacterial and in vitro glycoengineering.

## RESULTS

**Identifying Specificity-Determining Residues by Site-Saturated Library Screening.** To develop a panel of mutant NGT enzymes capable of efficiently modifying a broad range of defined protein glycosylation sites, we first set out to identify the residues that directly determine substrate specificity. While a crystal structure of the NGT complex with the substrate could provide this information, the known crystal structure of NGT[31] from *Actinobacillus pleuropneumoniae* (ApNGT) does not provide the location of the acceptor peptide, only showing the uracil diphosphate (UDP) portion of the UDP-Glc sugar donor. Therefore, it was necessary for us to first identify residues that determine specificity by directly screening enzyme mutants. To that end, we selected 26 residues surrounding the UDP-binding pocket of ApNGT for mutagenesis (Figure 1a). We then ordered fully saturated libraries for each of these residues as linear DNA, using a previously reported Q469A mutant of ApNGT (we refer to this parent mutant as ApQ)[20,32] as a starting point because it has much higher activity than wildtype ApNGT for its peptide substrates. Each of these SSVLs contained DNA encoding enzymes with an approximately equal mixture of the 19 non-wildtype amino acids (indicated by an "X") at one of the 26 targeted residues. In this way, we test each library as a pool, rather than individual clones, to identify residues having the greatest impact on activity and peptide specificity.

We performed PCR on each of these SSVLs and directly used the resulting linear expression templates (LETs) to drive expression of protein SSVLs in CFPS (Figure 1b). All 26 SSVLs were expressed at similar levels compared to ApQ (Supplementary Figure 1). All 26 protein SSVLs as well as the parent ApQ were used directly in glycosylation assays of each peptide in a 361-member substrate library with the motif $X_{-1}$-N-$X_{+1}$-TRC where $X_{-1}$ and $X_{+1}$ are one of the 19 amino acids (Cys excluded). After in vitro glycosylation (IVG), peptides and glycopeptides were covalently pulled down onto maleimide-functionalized self-assembled monolayers by reaction with the *C*-terminal cysteine and then analyzed with SAMDI-MS for peptide modification (modification heatmaps shown in Supplementary Figure 2).

To compare the differences between ApQ and each of the SSVLs, we first calculated the average activity across the entire peptide library. We used concentrations of peptide substrates that were generally 10-fold lower than the $K_M$ for most peptide and NGT combinations[19,20] and were therefore able to compare the approximate $k_{cat}/K_M$ of each reaction using the equation $-\ln(1 - Y) = k_{cat}/(K_M{*}c{*}t)$, where $c$ is enzyme concentration, $t$ is reaction time, and $Y$ is yield of modification. We converted each modification data point within each heatmap for the 26 protein SSVLs to generate heatmaps showing $-\ln(1 - Y)$ (Figure 1b) and then obtained an approximated average $k_{cat}/K_M$ value for each SSVL across all 361 peptides. While this quantification method does not provide an exact $k_{cat}/K_M$ value for each enzyme−substrate combination (doing so would require tens of separate measurements for each substrate), it did allow us to present and compare one approximate value of average $k_{cat}/K_M$ for each SSVL based on 361 data points for each peptide substrate in the library. To enable comparison, the average $k_{cat}/K_M$ of each SSVL was then normalized to that of ApQ (Figure 1c, see Online Methods for more details). As might be expected, no SSVL showed greater activity—measured as an average across all 361 peptides—than the parent ApQ. We observed that the R177X, D215X, R281X, and M440X SSVLs had the poorest average activities (less than 2% relative to ApQ), indicating that no individual mutant in these SSVLs provides activity close to that of ApQ and that these residues may likely be important to catalysis or substrate (peptide receptor or sugar donor) binding of ApQ.

Next, to identify the residues that strongly influence specificity at each position of the substrate peptide, we quantitatively compared the differences in substrate specificity for each of the SSVLs with that of ApQ. We began by measuring the modification of our $X_{-1}$-N-$X_{+1}$-TRC peptide library at different concentrations of ApQ in order to generate a series of heatmaps for ApQ with various levels of average $-\ln(1 - Y)$. In this way, we could select the appropriate heatmap in order to compare the peptide selectivity difference of ApQ and each of the 26 protein SSVLs using heatmaps with the same value of average $-\ln(1 - Y)$ (Supplementary Figure 3). We then calculated the percentage difference of each $X_{-1}$ amino acid (each row of the heatmap) for each SSVL compared to ApQ from the average $-\ln(1 - Y)$ value for all 19 peptides within that $X_{-1}$ amino acid, using the equation $2 * |\mathrm{Ave(X)} - \mathrm{Ave(ApQ)}|/(\mathrm{Ave(X)} + \mathrm{Ave\ (ApQ)})$, where Ave(X) and Ave(ApQ) are the average $-\ln(1 - Y)$ for each SSVL and ApQ, respectively. The average of all 19 percentage differences in $X_{-1}$ amino acid rows gave the mean percentage difference for the $X_{-1}$ position (Figure 1d). We performed a similar analysis to determine the mean percentage difference for the $X_{+1}$ position for each SSVL (Figure 1d). The mean percentage difference heatmap of ApQ and all SSVLs compared to each other is shown in Supplementary Figure 4. Based on these data, we concluded that the residues playing the strongest role in determining specificity of the enzyme for the $X_{-1}$ position of the acceptor peptide are, in order from strongest to weakest: T438, A469, Y498, H214, and I279. Similarly, we found that, for the $X_{+1}$ position, residues A469, H214, R177, H219, and T438 have the greatest impact on specificity. We found that residue 469 plays a relatively strong role in determining enzyme specificity for both the $X_{-1}$ and $X_{+1}$ positions, as well as the UDP sugar donor as reported previously.[31,32]


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

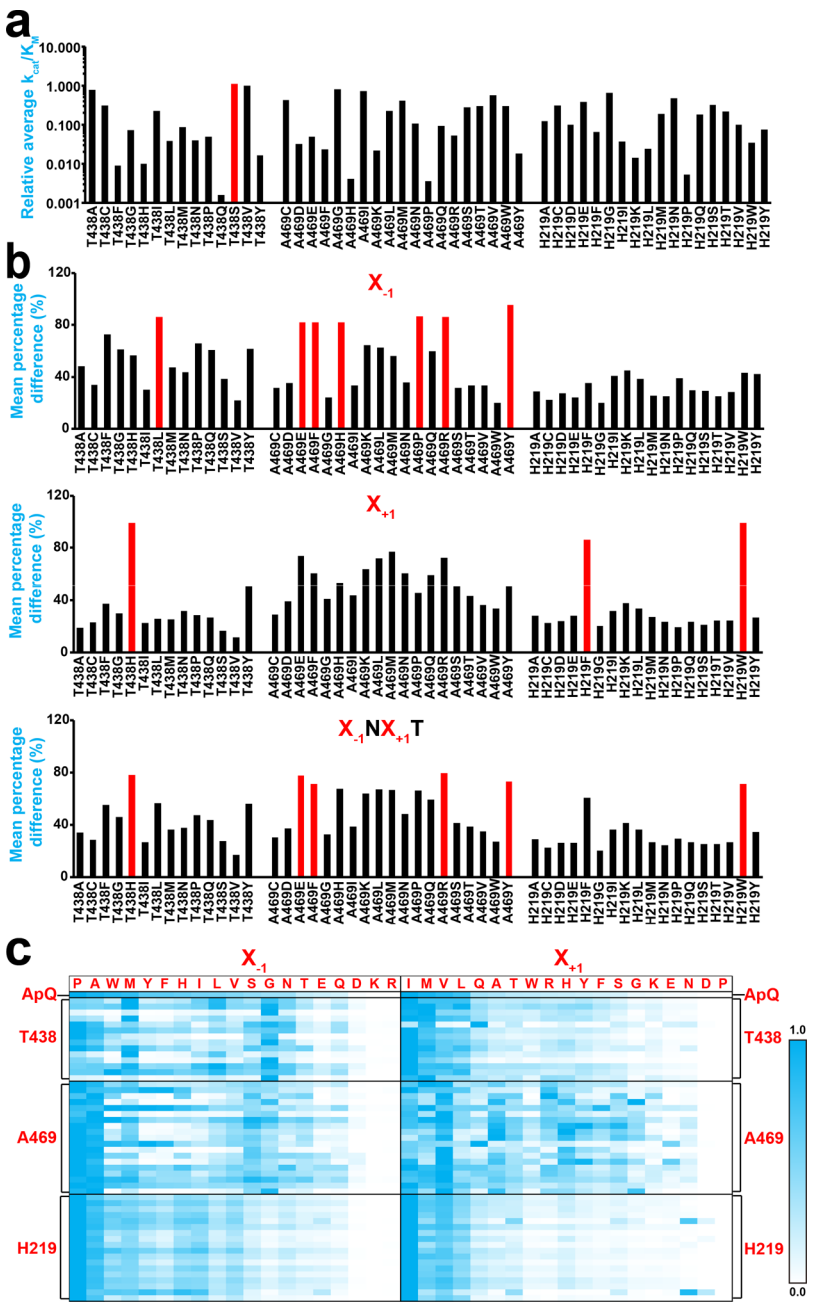



**Figure 2.** Screening individual ApQ mutants with unique specificities for the $X_{-1}$and $X_{+1}$acceptor peptide positions. (a) Relative average $k_{cat}/K_M$ of individual mutants, from T438, A469, and H219, compared to ApQ against the $X_{-1}$-N-$X_{+1}$-TRC peptide library. Only T438S shows a slight increase in relative activity (1.1-fold), while T438D/E/K/R/W and H219R show poor activities that are less than 0.001-fold of ApQ (H219R was not screened with entire library and T438D/E/K/R/W were screened but showed poor modification, see Supplementary Figure 9). (b) Mean percentage differences of $X_{-1}$ (upper), $X_{+1}$ (middle), and entire library (lower) for each mutant compared to ApQ. The mean percentage difference of the entire library is the average of the value of $X_{-1}$ and $X_{+1}$. A value higher than 75% is highlighted in red bars. (c) Heatmap of the relative selectivity for amino acids at the $X_{-1}$ and $X_{+1}$ positions of ApQ and all individual mutants of the three selected residues (T438, A469, and H219). Relative specificity is defined as the ratio of the average $-\ln(1 - Y)$ at each amino acid lane to the maximum value of all 19 $X_{-1}$ or $X_{+1}$ lanes. The amino acids at the $X_{-1}$ and $X_{+1}$ positions are organized in the same order as modification heatmaps, and the order of individual mutants at each residue is the same as (a) and (b). The heatmap with numerical values is also shown in Supplementary Figure 12. These data show that T438 mutants exhibit large specificity differences for the $X_{-1}$ position, H219F/W for $X_{+1}$, and A469 mutants for both $X_{-1}$ and $X_{+1}$. All experiments were completed with $n = 1$.

Having identified the residues that play the strongest roles in determining the polypeptide specificity of ApQ for the $X_{-1}$ and $X_{+1}$ positions of the acceptor peptide, we next pursued analogous experiments to identify residues that affect specificity for the $X_{+2}$, $X_{-2}$, and $X_{+3}$ positions. Here, we chose six representative $X_{-1}$-N-$X_{+1}$-TRC peptide sequences preferred by ApQ and produced new peptide libraries that substituted the Thr with the other 18 amino acids at the $X_{+2}$ position, adding 19 $X_{-2}$ amino acids, or inserting 19 $X_{+3}$ amino acids (again, with Cys excluded in all peptides), respectively. We then screened the specificity of select SSVLs for each position that the earlier experiments suggested to have an


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

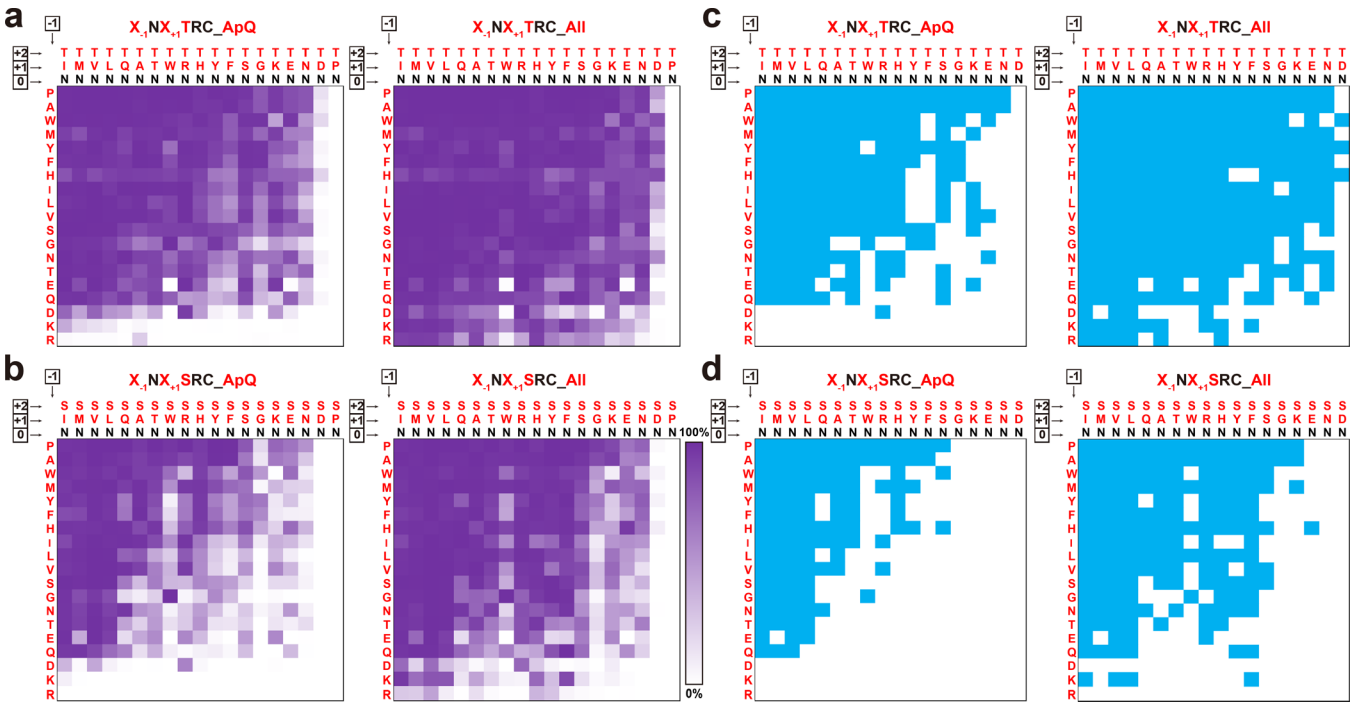



**Figure 3.** Expanded set of peptide sequences eligible for glycosylation by engineered NGTs. (a, b) Modification heatmaps of the peptide libraries $X_{-1}$-N-$X_{+1}$-T (a) or $X_{-1}$-N-$X_{+1}$-S (b) for ApQ (left) and the maximum modification from 14 selected NGTs (ApQ, H219F, H219W, T438S, T439E, A469G, A469I, H495D, H219F-T438S, H219F-H495D, H219W-T438S, H219W-H495D, A469G-H495D, and A469I-H495D) (right). All NGTs were tested using the same condition: 0.545 $\mu$M NGT produced in LET-CFPS, 30 °C for 3 h. There are improvements for the peptides inefficiently modified by ApQ. Specifically, peptides with Asn, Asp at the $X_{+1}$ position, and Lys, Arg at $X_{-1}$ are modified at higher efficiencies by the panel of 14 NGTs compared to ApQ alone. Heatmaps annotated with numerical values and the optimal NGT for each peptide substrate are shown in Supplementary Figure 19. All experiments were completed with $n$ = 1. (c, d) Comparison of all peptide substrates within the canonical glycosylation motif ($X_{-1}$-N-$X_{+1}$-T, $X_{+1} \neq$ P in (c)) or $X_{-1}$-N-$X_{+1}$-S ($X_{+1} \neq$ P in (d)) modified with more than 80% efficiency in (a) and (b) by ApQ (left) and the maximum value from 14 selected NGTs (right). The percentage of peptides with more than 80% modification (highlighted in blue) increased from 56 to 80% for the $X_{-1}$-N-$X_{+1}$-T library and from 33 to 51% for the $X_{-1}$-N-$X_{+1}$-S library.

important role in specificity. We found that D215 and R177 were important in determining specificity for the $X_{+2}$ position, H277 for $X_{-2}$, and H214 for $X_{+3}$ (Supplementary Figures 5–7). These observations identify those residues that interact with and determine specificity for the acceptor peptide (shown in Supplementary Figure 8); however, they do not rigorously establish that the interactions are direct. Of note, position 215 was recently identified as important for glycosylating non-canonical sequons in an investigation into the catalytic mechanism of ApNGT.[33]

**Screening Individual NGT Mutants with Unique Substrate Specificities.** After identifying the specificity-determining residues, we sought to screen the individual mutants at these residues to understand which peptide sequences were preferred as substrates. Based on our analyses, we decided to deconvolute the activities of each mutant within three of the 26 SSVLs—H219, T438, and A469—using the $X_{-1}$-N-$X_{+1}$-TRC peptide library (Supplementary Figures 9–11). We first isolated individual mutants from the SSVLs by circularization of the linear DNA and transformation of the resulting plasmids (see Online Methods). We found that each individual variant was expressed at similar levels (Supplementary Figure 1). Only T438S showed an increase in average glycosylation activity over ApQ (Figure 2a), while T438D/E/K/R/W and H219R had the poorest activity (less than 0.1% relative to ApQ). The activities of mutants at T438, likely an important residue for peptide binding according to our screen, varied significantly over the 19 mutants. We also analyzed the peptide selectivity for these individual mutants (Figure 2b). Most of T438 mutants exhibited altered specificities for the $X_{-1}$ position, with little effect on the $X_{+1}$ position; however, T438H showed altered $X_{+1}$ specificity and small changes in $X_{-1}$ specificity. Most A469 mutants showed different preferences for both the $X_{-1}$ and $X_{+1}$ positions. Mutations of A469 to F/H/P/R/Y had a stronger effect on $X_{-1}$ specificity while to G/I/N/S had a stronger effect on $X_{+1}$. Of the H219 mutants, only H219F/W strongly affected the peptide selectivity at the $X_{+1}$ position.

The differential selectivity of each mutant on amino acids at the $X_{-1}$ or $X_{+1}$ position allows them to be used for unique purposes (e.g., site-specific modification) (Figure 2c). For example, most of the T438 mutants preferred Met or Gly over ApQ-preferred Pro and Ala at $X_{-1}$; most A469 mutants preferred other amino acids, like Val, Ala, His, over ApQ-preferred Ile and Met at $X_{+1}$; H219F and H219W had very similar peptide specificities and exhibited significant increases in their preference for peptides with Asn and Asp at $X_{+1}$, which is not preferred by ApQ. We also performed a pairwise comparison of the specificity differences between all individual NGT mutants at each residue and found that many mutants possess unique preferences (Supplementary Figure 13).

**Expanding the Set of Sequences Eligible for Glycosylation by Selected NGT Mutants.** To identify a panel of mutant NGTs that can modify comprehensive sets of acceptor sequences, we first selected six NGTs (T438S, A469G, A469I, H219F and H219W, as well as the parent


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

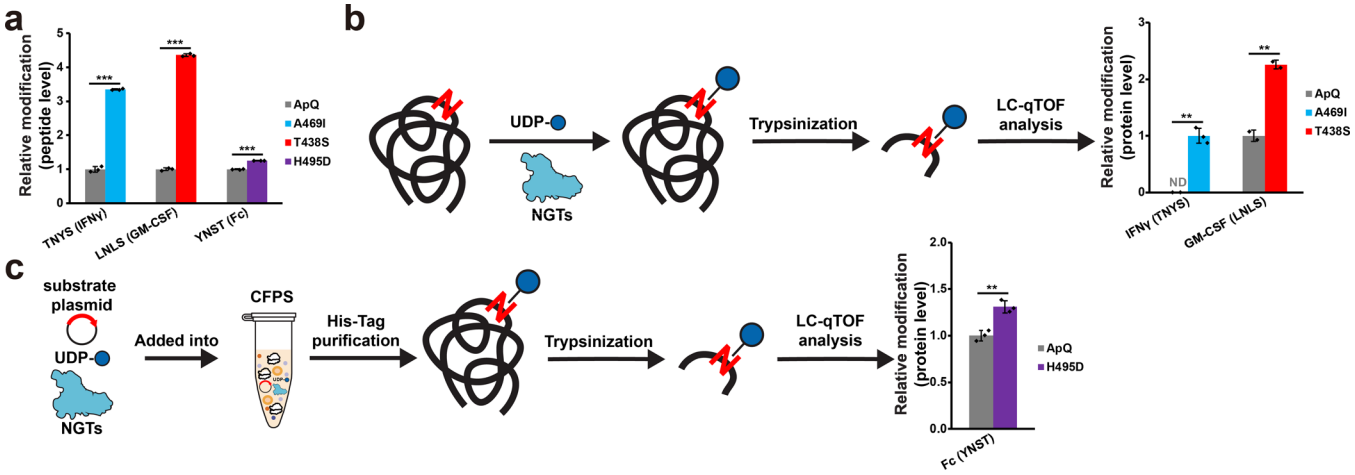



**Figure 4.** Selected NGT mutants enabling superior modification of therapeutic proteins. (a) Peptides with the sequences from approved therapeutic proteins are confirmed to have greater modification when glycosylated by purified mutants compared to ApQ. Folds of modification to ApQ are presented for each peptide. All experiments were completed with $n = 3$ IVG reactions. Experimental conditions: 1 $\mu$M (for TNYS), 0.05 $\mu$M (for LNLS), or 0.2 $\mu$M (for YNST) purified NGT, 30 °C for 3 h. (b) Purified approved therapeutic proteins exhibit enhanced modifications when glycosylated by selected mutants compared to ApQ. Folds of modification to ApQ are presented for each protein glycosylation site. After IVG, the solutions were dialyzed, trypsinized, and analyzed with LC-qTOF. ApQ showed no detectable modification of IFN$\gamma$ (marked as "ND"). All experiments were completed with $n = 2$ or $n = 3$ individual IVG reactions (as indicated in graph). Experimental condition: 5 $\mu$M purified NGT, 5 mM UDP-Glc, 30 °C for 12 h. (c) Fc showed increased modification for H495D over ApQ in CFPS expression. Folds of modification to ApQ are presented. Fc, with a 6xHis tag, was expressed by LET-CFPS supplemented with purified NGT and UDP-Glc. After the CFPS reaction, Fc was purified with magnetic beads, dialyzed, trypsinized, and analyzed with LC-qTOF. All experiments were completed with $n = 3$ individual CFPS reactions. Experimental condition: 2 $\mu$M purified NGT, 5 mM UDP-Glc, 30 °C LET-CFPS for 6 h. All protein modification efficiencies can be found in Supplementary Table 7. All $p$ values were from the two-tailed $t$-test with $p < 0.01$ (**) or $p < 0.001$ (***).

ApQ) arising from our initial screens. This panel of NGTs combined to provide the highest activity for the broadest range of peptides in the initial $X_{-1}$-N-$X_{+1}$-TRC substrate library (based on the calculated appropriate $k_{cat}/K_M$ for each peptide-NGT combination). We then screened the activities of these NGTs under identical conditions (0.545 $\mu$M NGT for 3 h at 30 °C) across a total of 684 peptide sequences of the form $X_{-1}$-N-$X_{+1}$-TRC and $X_{-1}$-N-$X_{+1}$-SRC (Supplementary Figures 14 and 15). These six NGTs all displayed less activity with Ser than with Thr at the $X_{+2}$ position. The five ApQ mutants added to the panel significantly expanded the set of sequences within the $X_{-1}$-N-$X_{+1}$-S/T motif that can be efficiently glycosylated (where the modification was greater than 80%) by 17% (118 of 684 peptides).

However, we noticed that even with these five additional ApQ mutants, the glycosylation of peptide substrates with Lys or Arg at the $X_{-1}$ position remained challenging. To address this gap and further expand the permissible substrate scope, we tested enzymes with mutations to residues that are nearby the hypothesized $X_{-1}$ interaction site (Supplementary Figure 8) and exhibited high activity in SSVL screens—T439, H495, and P497 (details described in Supplementary Figure 16). We found that T439E and some individual H495 mutants (particularly H495D) showed significantly increased preferences for peptides with Lys or Arg at the $X_{-1}$ position (heatmaps in modification of the full $X_{-1}$-N-$X_{+1}$-TRC substrate library by representative mutants are shown in Supplementary Figure 17). To further expand the set of preferred sequences, we also generated and screened double mutants that combined two single mutations identified above. Specifically, we combined H495D with mutations that provided unique specificities at the $X_{+1}$ position (A469G, A469I, and H219F/W). We also combined the H219F/W with mutations that provided unique specificities at $X_{-1}$ (T438S and H495D) (Supplementary Figure 18).

Finally, we assembled and tested a panel of 14 selected NGTs (ApQ, H219F, H219W, T438S, T439E, A469G, A469I, H495D, H219F-T438S, H219F-H495D, H219W-T438S, H219W-H495D, A469G-H495D, and A469I-H495D) with the entire or partial $X_{-1}$-N-$X_{+1}$-TRC and $X_{-1}$-N-$X_{+1}$-SRC peptide libraries under identical reaction conditions (Supplementary Figures 14–16, 18). Our goal was to demonstrate that this NGT panel could enable the glycosylation of a diverse range of peptides. We observed that these enzymes did significantly increase the maximum modification efficiency for 260 of the 684 canonical $X_{-1}$-N-$X_{+1}$-S/T glycosylation sequences over ApQ alone (Figure 3a,b and Supplementary Tables 5 and 6). Specifically, we increased the percentage of peptides with modification greater than 80% from the 45 to 66% (Figure 3c,d) and the percentage of peptides with modification greater than 5% from 80 to 93% (Supplementary Figure 19).

**Glycosylation of Approved Therapeutic Proteins Using NGT Mutants.** As a proof of principle, we next demonstrated the utility of specific mutants from the NGT mutant panel developed above to glycosylate three model therapeutic proteins: interferon-gamma (IFN$\gamma$), granulocyte–macrophage colony-stimulating factor (GM-CSF), and the constant region (Fc) of a human immunoglobulin antibody (IgG1). At the peptide level, we found that the purified A469I mutant glycosylated the sequence TNYS found in IFN$\gamma$ more efficiently than did ApQ (Figure 4a). Similarly, T438S glycosylated the LNLS sequence from GM-CSF more efficiently than did ApQ, and H495D glycosylated the YNST sequence from Fc more efficiently than did ApQ (Figure 4a). We then confirmed these relative activities at the protein level using purified IFN$\gamma$ and GM-CSF as substrates. After


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290–1302

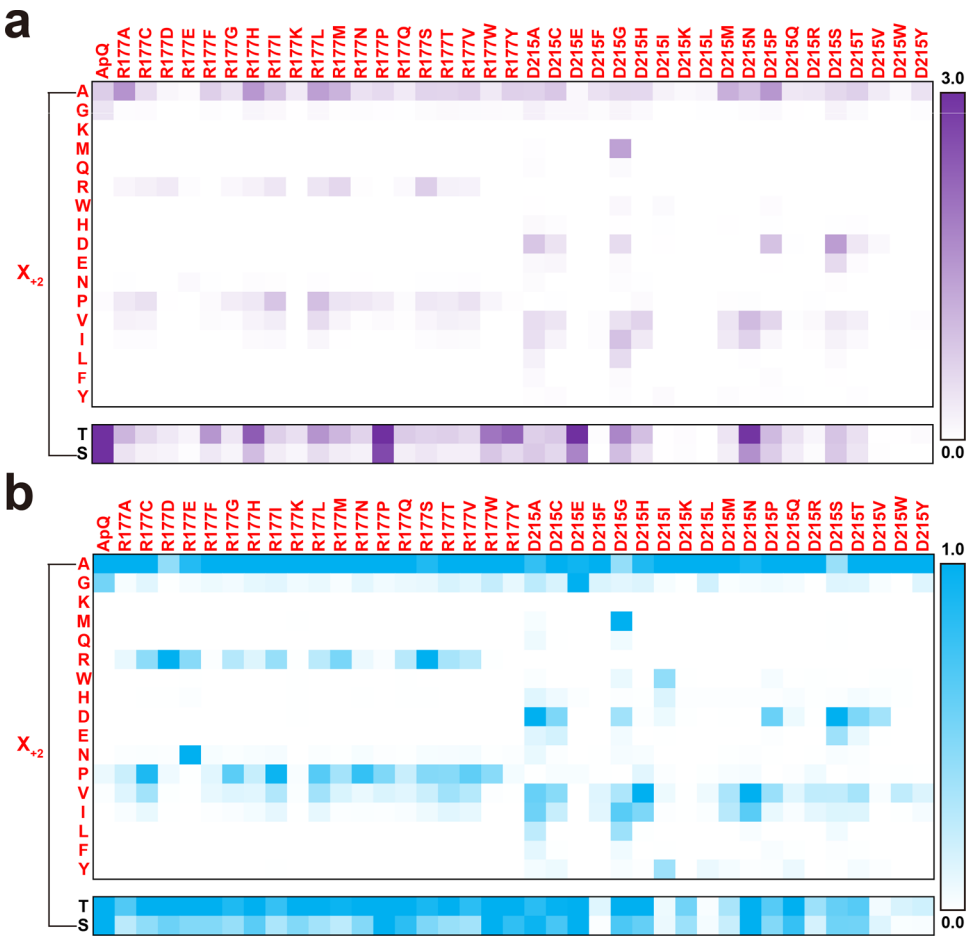



**Figure 5.** Screening NGT mutants for expanded specificity at the $X_{+2}$ position. (a) Average $-\ln(1 - Y)$ heatmap for non-S/T amino acids at $X_{+2}$ with all R177 and D215 individual mutants (values for T/S shown below). Six $X_{-1}$-N-$X_{+1}$-T peptide sequences preferred by ApQ were substituted with 18 amino acids (Cys excluded) at $X_{+2}$ and screened against all individual R177 and D215 mutants. All heatmaps show results from $n = 1$ experiment. All reactions were performed with 0.545 $\mu$M NGTs produced in LET-CFPS, 30 °C for 12 h. $X_{+2}$ amino acid lanes are arranged in the same descending order as ApQ shown in Supplementary Figure 5. The modification heatmaps for $X_{+2}$ with ApQ and all individual mutants are shown in Supplementary Figure 22. (b) Relative amino acid selectivity of $X_{+2}$ based on data in (a), divided by the maximum value of all $X_{+2}$ (except S/T) for each mutant. The relative selectivity for S/T is also shown below over the same maximum value and may be higher than 100%. When Asn is present at $X_{+2}$, the modification may come from the second Asn at NRC, rather than N-X-N if the modification for W-N-I/V-N-RC is more preferred than A-N-I/V-N-RC.

glycosylation, the target protein was digested with trypsin and analyzed by liquid chromatography-quadrupole time-of-flight (LC-qTOF) mass spectrometry (Figure 4b). The relative modification, using the percent area of integrated extracted ion chromatograms, showed that the NGT mutant achieved more efficient glycosylation than did ApQ (Figure 4b). We also used $MS^2$ to confirm the identity of the targeted peptides (Supplementary Figure 20). Notably, the glycosylation of sequences within folded intact proteins is less efficient than that of the corresponding sequences as free peptides. Using Fc as an example, we showed that the modification of proteins could be improved by supplementing NGTs at the beginning of the CFPS reaction to simultaneously express and glycosylate substrate proteins prior to folding. By adding 2 $\mu$M ApQ before rather than after the Fc expression in the CFPS reaction, we found that the glycosylation efficiency of Fc was increased from 15 to 46% (Supplementary Figure 21 and Table 7). Supplementing with the H495D mutant (2 $\mu$M) rather than ApQ pre-CFPS increased the modification of Fc to 60%, a 30% improvement (Figure 4c and Supplementary Table 7). By supplementing with a larger amount of of H495D (5 $\mu$M), 80% modification of Fc could be achieved (Supplementary Figure 21). Taken together, our data show that native amino acid sequences having the canonical glycosylation sequence N-X-S/T can be modified more efficiently with the mutant NGTs identified using our high-throughput experimentation.

**Expanding $X_{+2}$ Selectivity beyond Canonical Glycosylation Motifs.** Thus far, the enzymatic *N*-glycosylation of a target protein has required a canonical N-X-S/T glycosylation motif that is either naturally present or introduced by altering primary amino acid sequences. This requirement for a Ser or Thr residue presents a significant constraint to generalize enzymes for protein glycosylation. However, the introduction of new glycosylation sites without modifying primary amino acid sequences could also be enabled by engineered NGTs having a preference for sequences beyond the canonical glycosylation motif N-X-S/T. Therefore, the development of NGTs that do not require Ser or Thr at the $X_{+2}$ position would expand the range of sequences and proteins eligible for glycosylation. Based on our previous specificity screens and hypothesized peptide-interaction residues (Supplementary Figures 5 and 8), R177 and D215 were identified as the


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

specificity-determining residues for the $X_{+2}$ position. We then sought to discover mutants that can glycosylate peptides without Ser or Thr at the $X_{+2}$ position by screening all individual mutants of R177 and D215 across the peptide library of the form $(X_{-1}NX_{+1})X_{+2}RC$ (Supplementary Figure 22). As expected and also recently reported,[33] we found that most mutants tolerated S/T at the $X_{+2}$ position. However, we also found that R177 individual mutants tolerated A/R/P/V, D215 individual mutants tolerated A/D/E/V/I/L, and ApQ tolerated A/G at the $X_{+2}$ position (Figure 5a). D215G exhibited the broadest promiscuity for $X_{+2}$ amino acids and could modify sequences with A/M/D/V/I/L at relatively high efficiency, as well as G/Q/W/E/N/F/Y at medium efficiency. Interestingly, we found that D215F/I/L/V lost their preference for peptides with S/T at $X_{+2}$ (Figure 5b). These findings can be used to guide the choice of an NGT mutant to target a sequence with a given $X_{+2}$ amino acid (Figure 5a) or to selectively target sequences with one $X_{+2}$ amino acid over another (Figure 5b).

Finally, we sought to explore which noncanonical sequences might be targeted with our newly discovered R177 and D215 mutants that exhibited expanded specificity at the $X_{+2}$ position. In our previous SSVL screens, D215X exhibited little change in $X_{-1}$ and $X_{+1}$ specificity. Therefore, the $X_{-1}$-N-$X_{+1}$-TRC screen of D215X (Supplementary Figure 2) can be used to approximate the specificity of all D215 mutants for $X_{-1}$ and $X_{+1}$ combinations. However, we also observed that R177X significantly altered the enzyme selectivity for the $X_{-1}$ position. Therefore, we screened all R177 mutants with $X_{-1}$-N-$X_{+1}$-TRC to determine which $X_{-1}$ and $X_{+1}$ combinations could be used with non-S/T amino acids at the $X_{+2}$ position (Supplementary Figure 23). Overall, we found that 10 non-S/T amino acids (A/G/M/R/D/E/P/V/I/L) at the $X_{+2}$ position can be modified at relatively high efficiency on the peptide level. However, we note that the modification of noncanonical sequences remains less efficient than that of the canonical sequences. Further engineering or evolution of NGTs targeting noncanonical peptide sequences will be required to achieve highly efficient modification of noncanonical sequences in therapeutic proteins.

## DISCUSSION

In this work, we present a systematic method to build enzymes for synthetic glycosylation pathways. We identify enzyme residues that determine specificity for each amino acid position of peptide substrates, and we use these sites as a starting point to develop a panel of specificity-distinct NGTs capable of modifying unique sets of substrate sequences. Our high-throughput GlycoSCORES characterization technique enabled the screening of 123 individual NGTs through 52,894 independent reactions. To our knowledge, this is the most detailed glycosyltransferase engineering and characterization effort completed to date, surpassing the state-of-the-art[19] by nearly 4-fold. With minor adaptations to the workflow, this method of developing an enzymatic repertoire for modification of an entire substrate library should be applicable to other glycosyltransferases,[19] proteases,[26] phosphatases,[34] deacetylases,[35] and other enzymes.[36,37]

Two key features make this work important. First, the rigorous characterization enabled the development of a panel of 14 NGTs that significantly expand the set of sequences available for glycosylation by bacterial enzymes. Of the 684 peptides within the canonical eukaryotic glycosylation motif ($X_{-1}$-N-$X_{+1}$-S/T where $X_{+1} \neq$ P) that we surveyed, 260 peptides were found to be modified with significantly higher efficiency by one of the 13 NGT mutants compared to ApQ (Supplementary Tables 5 and 6). These variants increase the percentage of sequences that can be glycosylated by NGTs with good efficiency (more than 80% modification with ∼0.5 $\mu$M NGT and 3 h reaction) from 45 to 66%. This expanded panel of NGTs advances the rational glycosylation of a sequence of interest by identifying the optimal NGT from the heatmap reported in Supplementary Figure 19. We also note the prospect for developing double (or greater) mutants that have tailored specificities. We generally observed that double mutants have specificities that are consistent with those of each of the corresponding single mutants and therefore are more specific than the single mutants. Taking the T438H-A469H double mutant as an example, both T438H and A469H prefer Pro, Ala, Ser, and Gly at the X-1 position and we find a similar preference for the double mutant; similarly, Lys, Arg, Asp, Glu, and Ile decrease activity in the single mutants and the double mutant. For the X+1 position, Ile, Met promote activity for all three mutants; Pro, Asp, Lys, Trp decrease activity for all three mutants (Supplementary Figure 24).

Our strategy was applied to increase the modification of the therapeutic proteins IFN$\gamma$, GM-CSF, and Fc using the A469I, T438S, and H495D mutants of ApQ, respectively. We also developed NGTs that can glycosylate, or even prefer, sequences outside of the canonical N-X-S/T motif with non-S/T amino acids at the $X_{+2}$ position. This discovery widens the scope of glycoengineering, allowing researchers to investigate how glycans can be used to improve the properties of a more diverse set of proteins without the need to modify their native amino acid sequences. Notably, many of the mutants discovered in this work possess quite distinct substrate specificities, which may enable the site-specific control of glycosylation structures at multiple sequences within a single protein by sequential modification[22] to enable the precise engineering of synergistic glycan interactions.

Second, this work is important because it highlights the benefits of high-throughput experimentation. While typical-directed evolution workflows lead to enzymes capable of performing a single reaction, our approach can be used to develop enzymes having multiple properties. Indeed, the parallel approach to activity monitoring enabled by GlycoSCORES (combination of CFPS and SAMDI-MS) results in variants having unique specificities—which enable a pattern of activity on many different substrates—and affords new possibilities for glycoengineering.

While this work focuses on the initiation step of glycosylation, many enzymatic and chemoenzymatic technologies have been developed to elaborate the monosaccharide installed by NGTs into human-like or otherwise useful glycans. For example, chemoenzymatic methods using endoglycosidases and chemically synthesized oxazoline donors can be used to install full-length human glycans.[22,23] These full-length human glycans may provide increased serum half-life for proteins[38] or provide routes to tune other therapeutic effects through the installation of a homogeneous *N*-glycan on the Fc region of human IgG.[4,12] Biosynthetic approaches to extend the glucose installed by NGTs to diverse and useful glycans, including polysialic acid, have also been developed.[21] Notably, the reducing end sugar of human *N*-glycans is *N*-acetylglucosamine (GlcNAc), rather than the glucose installed by NGTs. The effect of this difference on glycoprotein immunogenicity


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

and other properties is unknown and will need to be assessed for each application. A two-step method using ApQ to install *N*-glucosamine (GlcN) and the acetyltransferase GlmA has already been developed,[32] and we are currently working on discovering NGT mutants that can more efficiently transfer GlcN and even GlcNAc. For example, we found that A469I and T438S can modify some peptides with GlcN more efficiently than ApQ (Supplementary Figure 25). Several other highly active mutants discovered in this work can also modify peptides with GlcN (Supplementary Figure 25).

Despite increasing the range of sequences that can be glycosylated, we do recognize that the protein structure surrounding a targeted glycosylation sequence can affect the efficiency of modification. NGTs normally act post-translationally on folded proteins and therefore sites that are buried or conformationally locked into secondary or tertiary structure may not be available for modification by NGTs. Thus, some targets may require the use of other existing glycosylation methods that employ oligosaccharyltransferases (OSTs). While OSTs are complex integral membrane proteins and require lipid-linked oligosaccharide (LLO) substrates, they are capable of cotranslational modification on unfolded sites.[39−41] Despite recent efforts to engineer or discover OSTs with expanded specificities,[42,43] it is often still necessary to install a glycosylation tag (GlycTag) by extending or otherwise altering the primary amino acid sequence of the target protein in order to achieve glycosylation.[17,44] Therefore, a comprehensive engineering of OSTs similar to the work performed here for NGTs is also urgently needed to expand the set of sequences permissible to modification with OSTs.

In summary, we demonstrate the application of a high-throughput experimentation strategy to engineer glycosyltransferases by using LET-CFPS and SAMDI-MS for parallelized generation and characterization of many enzyme mutants on a broad range of substrates. Using this method, we have developed a panel of rigorously characterized, readily expressed, fully soluble *N*-glycosylation enzymes with unique activities that will serve as a valuable resource for the glycoengineering community. We expect that this panel of NGTs will be especially useful in the bacterial or in vitro glycoengineering of protein therapeutics because it alleviates the need to alter primary amino acid sequences to achieve glycosylation for many protein therapeutics. Ultimately, our approach is poised to facilitate basic understanding in glycoscience and enable new applications in glycoengineering.

## METHODS

**Peptide Library Synthesis and SAMDI Screening.** All peptide libraries were synthesized with *N*-acetyl and *C*-amide, as described previously.[19] The average concentration of each peptide library was also determined as before,[19] as well as the calculation of average relative ionization factors (RIFs). SAMDI plates were also prepared and used for peptide screening as before.[19]

For peptide screening, 50 $\mu$M peptide was reacted with the indicated concentration of NGT, purified or produced in LET-CFPS, and 2.5 mM UDP-Glc in 100 mM HEPES (pH 8.0) and 500 mM NaCl at 30 °C for indicated time. Screenings for UDP-GlcN modification of peptides were completed similarly using 50 $\mu$M peptide with the 1.09 $\mu$M NGT produced in LET-CFPS and 2.5 mM UDP-GlcN (custom synthesized at Chemily Glycoscience) in 100 mM HEPES (pH 8.0) and 500 mM NaCl at 30 °C for 12 h. After the IVG reaction, TCEP-resin (Pierce) was added and incubated at 37 °C for 1 h. Two microliter solutions of these reduced IVGs were added to the islands of a 384-well maleimide-functionalized SAMDI plate and incubated at room temperature for 0.5 h. Because the reaction is not quenched during the 1 h TCEP reduction and 0.5 h SAMDI incubation steps, we approximated this time as an additional 1 h of reaction for approximate $k_{cat}/K_M$ calculations. The SAMDI plate was then washed with water, ethanol, water, and ethanol before being dried with nitrogen flow. Ten mg/mL 2′,4′,6′-trihydroxyacetophenone monohydrate (THAP; Sigma-Aldrich) in acetone was applied onto SAMDI plate as the matrix. The plate was then analyzed with a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry using an AB SCIEX 5800 TOF/TOF instrument. Spectra was processed with Applied Biosystems SciEx Time of Flight Series Explorer Software version 4.1.0. All peptide library screenings were completed with $n$ = 1. Modification efficiencies were determined using spectral peak ratios adjusted by RIF[19] (Supplementary Table 3) except the data for UDP-GlcN, which show the relative intensity of glycopeptide (product peak) to all peptides (substrate and product peaks).

**Synthesis of Linear and Plasmid SSVLs.** Linear DNA for 26 SSVLs of ApQ were synthesized by Twist Bioscience within a linearized form of the pJL1 cell-free expression backbone[45] opened within the kanamycin resistance gene such that recircularization with SapI restriction enzyme sites installed at the 5′ and 3′ ends would result in pJL1.ApQ (Supplementary Note 1). Plasmid forms of SSVLs were generated by circularizing these linear libraries using the Golden Gate assembly method with the SapI restriction enzyme.[46] Thirty nanograms (6 $\mu$L) of each linear library was incubated with 1 $\mu$L each of 10,000 U/mL SapI restriction enzyme, 10 mM ATP, 10× CutSmart buffer, and 2,000,000 T4 ligase (all products from New England Biolabs). Circularization reactions were carried out with 30 cycles of 1 min at 37 °C and 1 min at 16 °C followed by 5 min at 65 °C. Completed circularization reactions were then transformed into DH5$\alpha$ electrocompetent cells and plated on LB (KAN+). All plates produced more than 100 colonies (ensuring 5-fold coverage of the library). After overnight growth, these LB (KAN+) plates were washed with 5 mL of LB media and miniprepped to generate plasmid libraries.

**Isolating Individual NGT Mutants from SSVLs.** The plasmid SSVLs of selected residues, R177, D215, H219, T438, A469 and H495, were transformed into DH5$\alpha$ high efficiency chemically competent cells (New England Biolabs) by heat shock followed by incubation on LB agar plate (KAN+). More than 50 clones were picked from each SSVL transformation, cultured in LB (KAN+) media, miniprepped, and sequenced to isolate all 19 individual mutants.

**Construction of Single and Double Mutants.** Single mutants of ApQ were generated using single-site PCR mutagenesis of a pJL1.ApQ template, as previously reported.[47] Briefly, 25 $\mu$L PCR reactions were performed which contained 12.5 $\mu$L Q5 hot start high-fidelity 2× master mix (New England Biolabs), 1 ng template, 500 nM primer pair. The primers and Tm temperatures for these PCRs are listed in Supplementary Table 8. The PCR was initiated at 98 °C for 30 s; followed by 15 cycles of 98 °C for 10 s, $Tm_2$ for 30 s and 72 °C for 2 min; finished at $Tm_1$ for 1 min and 72 °C for 4 min. After the PCR, 2.5 $\mu$L 10× CutSmart buffer and 0.5 $\mu$L DpnI (New England Biolabs) was added and incubated at 37 °C for


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

2 h. Agarose gel electrophoresis was used to confirm the production of full-length PCR product (~3.5 kb). The PCR solutions (after DpnI treatment) were transformed in DH5α high efficiency chemically competent cells by heat shock and applied on LB agar plate (KAN+). Two clones for each plate were picked, cultured, miniprepped, and sequenced. Double mutants were generated similarly except single mutants were used as the initial plasmid template.

**LET-CFPS.** CFPS reactions were performed using crude lysate derived from *E. coli* strain BL21 Star (DE3), as described previously[22,39,48−50] using linear DNA expression templates produced by PCR rather than plasmids.[29] Crude lysates for CFPS were prepared by growth, harvest, and lysis of BL21 Star (DE3) *E. coli* cells, as described previously[48] using a total energy input of 640 J for lysis of 1 mL cell suspensions. LET-CFPS reactions were performed at 50 μL in 2.0 mL centrifuge tubes containing 1.2 mM ATP (pH 7.2), 0.85 mM GTP, UTP, and CTP (pH 7.2); 34 μg/mL folinic acid; 171 μg/mL of *E. coli* tRNA mixture; 2 mM of 20 standard amino acids; 0.33 mM nicotinamide adenine dinucleotide (NAD); 0.27 mM coenzyme-A (CoA); 1.5 mM spermidine; 1 mM putrescine; 4 mM sodium oxalate; 130 mM potassium glutamate; 10 mM ammonium glutamate; 8 mM magnesium glutamate; 57 mM HEPES (pH 7.2); 33 mM phosphoenolpyruvate (PEP, pH 7); 20% v/v NGT linear template; and 27% v/v of BL21 crude extracts.[51−53] NGT linear template was generated in a PCR reaction and used directly without purification. The 60 μL PCR reactions contained 30 μL Q5 hot start high-fidelity 2× master mix, 1.2 ng template (linear SSVLs synthesized by Twist or individual mutant plasmids), and 500 nM primer pair (ccacctctgacttgagcgtc and gcagtttcatttgatgctcgatg). The PCR was initiated at 98 °C for 30 s; followed by 36 circles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 1 min; and finished at 72 °C for 2 min. All reagents used in CFPS were purchased in Sigma-Aldrich except *E. coli* total tRNA mixture from strain MRE600 and PEP (Roche Applied Science). All CFPS reactions were carried out at 22 °C for 20 h. After the reaction, 1:1 v/v of 2× Roche complete protease inhibitor cocktail and 5 mM EDTA were added to the CFPS solutions, and the solutions were flash-frozen in liquid nitrogen and stocked at −80 °C for future use.

**Approximate $k_{cat}/K_M$ Calculation for Peptide Library $X_{-1}$-N-$X_{+1}$-TRC.** According to previous studies, the $K_M$ of optimized peptides of the form $X_{-2}$-$X_{-1}$-N-$X_{+1}$-T-$X_{+3}$-RC for ApNGT are greater than 0.5 mM,[19] and longer peptide substrates generally exhibit lower $K_M$ values than shorter peptides.[20] Previous reports have also found that the $K_M$ of ApQ differs by approximately 1.5-fold compared to ApNGT.[20] Based on these findings, we used a concentration of peptides (50 μM) that is much smaller than the $K_M$ for the NGT variants used in this study. Thus, we use equation $k_{cat}/K_M = -\ln(1 - Y)/c/t$ to approximate the value of $k_{cat}/K_M$ in which $Y$ is the modification for peptides, $c$ is the concentration of enzyme used, and $t$ is the reaction time for glycosylation. Heat maps showing values of $-\ln(1 - Y)$, which is in direct proportion to $k_{cat}/K_M$ of the peptides, $c$(NGT), and $t$, were generated from modification heat maps.

While approximate, this calculation allowed us to compare the activity of all peptide-enzyme combinations with more than 10-fold fewer reactions than would have been required to rigorously determine $k_{cat}/K_M$ values. We also note that average $k_{cat}/K_M$ values from a complete library of 361 conditions will be much more accurate than individual $k_{cat}/K_M$ approximations. We used the average $k_{cat}/K_M$ to compare the activity between each mutant or SSVLs across different enzyme concentrations and reaction times. Because the apparent average $k_{cat}/K_M$ was affected by the value of average $-\ln(1 - Y)$ (Supplementary Figure 3), the relative average $k_{cat}/K_M$ of mutants were compared to an ApQ screen that yielded the same value of average $-\ln(1 - Y)$. The optimal NGTs chosen for glycosylation of the whole set of canonical eukaryotic glycosylation sequences ($X_{-1}$-N-$X_{+1}$-T/S-RC) (Supplementary Figures 14 and 15) were determined by calculating the approximate $k_{cat}/K_M$ for each peptide-NGT combination and choosing the NGT mutant that provided the highest value. The selected NGTs (including ApQ) were screened with the same conditions. Specifically, 0.545 μM NGT was produced in LET-CFPS and combined with 2.5 mM UDP-Glc and 50 μM peptide before incubation at 30 °C for 3 h.

**Analysis of the Mean Percentage Differences between SSVLs or Single Mutants and ApQ.** To serve as references to compare the specificities of SSVLs or single mutants to the specificity of ApQ, several $X_{-1}$-N-$X_{+1}$-TRC heatmaps were analyzed after IVGs with various amounts of ApQ, yielding heatmaps with various averages of $-\ln(1 - Y)$, where $Y$ is modification. Using linear interpolation between two of these ApQ reference heatmaps, a theoretical ApQ heatmap with the same average $-\ln(1 - Y)$ as the measured heatmap for a given SSVL or single mutant was generated (reference ApQ heatmaps and description of calculation process found in Supplementary Figure 3). We then calculated the percentage difference between the average of the $-\ln(1 - Y)$ values for all 19 peptides with a given $X_{-1}$ amino acid lane in the theoretical ApQ heatmap (defined as Ave(ApQ)) and the average of the $-\ln(1 - Y)$ values for all 19 peptides with a given $X_{-1}$ amino acid lane in the measured mutant heatmap (defined as Ave(X)) using the equation 2*|Ave(X) − Ave(ApQ)|/(Ave(X) + Ave(ApQ)). The average of percentage differences for all 19 $X_{-1}$ amino acid rows gives the mean percentage difference of $X_{-1}$. The $X_{+1}$ mean percentage difference values were calculated similarly. The mean percentage difference for the whole $X_{-1}$-N-$X_{+1}$-TRC library is the average of the mean percentage differences for all $X_{-1}$ and $X_{+1}$ lanes. This calculation method was used to generate mean percentage differences shown in Figures 1 and 2 and Supplementary Figure 3.

**Combinatorial Comparison of SSVL and Single Mutant Specificities by Mean Percentage Differences.** To compare the specificities between individual mutants or SSVLs to each other, we calculated the mean percentage difference between any two mutants from mean percentage differences of each one to ApQ. The numeric value, not the absolute value, of 2*(Ave(X) − Ave(ApQ))/(Ave(X) + Ave(ApQ)) for each $X_{-1}$ or $X_{+1}$ lane was calculated as above for each mutant and defined as $PD_1$ for mutant 1 and $PD_2$ for mutant 2. The percentage difference between mutant 1 and 2 at each $X_{-1}$ or $X_{+1}$ lane was then calculated using the equation |$PD_1$ − $PD_2$|/(1 − $PD_1$*$PD_2$). Using this method, we calculated the mean percentage differences between all SSVLs or isolated mutants at each residue for $X_{-1}$, $X_{+1}$ or the entire library, respectively. This calculation is based on the assumption that the percentage difference for each $X_{-1}$ and $X_{+1}$ lane between two NGTs remains unchanged when determined from heatmaps with different average values of $-\ln(1 - Y)$. This calculation method was used to generate mean percentage differences in Supplementary Figures 4 and 13.


https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

**ApQ and Mutant Plasmid Construction, Expression in *E. coli*, and Purification.** ApQ mutant constructs were generated in the pET21b vector for in vivo expression and purification. Mutagenesis was performed the same way as described previously for in vitro constructs in the pJL1 vector. Primers and Tm used are listed in Supplementary Table 8. NGTs were purified as described previously with minor modifications.[19] Briefly, BL21 Star (DE3) chemically competent cells were transformed with pET21b.ApQ or mutant plasmids by heat shock. An overnight culture was inoculated in LB (CARB+) media. Fresh LB (CARB+) was inoculated at initial OD600 = 0.08, and the cells were grown at 37 °C at 250 rpm to 0.6−0.8 OD and induced with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) for 6 h at 30 °C. The cells were pelleted by centrifugation at 5000×*g* for 10 min at 4 °C, resuspended in Buffer 3 (20 mM Tris−HCl and 250 mM NaCl, pH 8.0), pelleted again by centrifugation at 8000×*g* for 10 min at 4 °C, and flash frozen at −80 °C. The pellets were then thawed and resuspended in 5 mL Buffer 3 per gram wet pellet weight and supplemented with 1 mg/mL lysozyme (Sigma), 1 μL benzonase (Millipore), and 1× Halt protease inhibitor (Thermo Fisher Scientific). Cells were then lysed by single-pass homogenization at 21,000 psig (Avestin) and centrifuged at 13,000×*g* for 20 min at 4 °C. Imidazole was added to the supernatant to a final concentration of 20 mM. The supernatant was applied to 1 mL Ni-NTA agarose resin (Qiagen) equilibrated with Buffer 3 with 20 mM imidazole. Following a 1 h incubation, the resin was washed once with five column volumes of Buffer 3 with 20 mM imidazole, washed twice with five column volumes of Buffer 3 with 30 mM imidazole, and once with five column volumes of Buffer 3 with 40 mM imidazole. The protein was eluted in one column volume of Buffer 3 with 500 mM Imidazole. The elution was dialyzed against 50 mM HEPES, 200 mM NaCl, pH 7.0, and flash frozen at −80 °C. Protein concentration was quantified with a NanoDrop UV−vis spectrophotometer (Thermo Fisher) using the following parameters, molecular weight: 71,502.50 Da, Extinction coefficient: 63,260 $M^{-1}$ $cm^{-1}$.

**In Vitro Protein Glycosylation and LC-qTOF Analysis of Tryptic Glycopeptides.** Ten micrometers IFNγ (Millipore) or GM-CSF (R&D Systems) was reacted with 5 μM purified ApQ or mutants and 5 mM UDP-Glc in Buffer 1, 50 mM $NaH_2PO_4$ (pH 8) and 300 mM NaCl. The reactions were carried out at 30 °C for 12 h. After the reaction, 10 μL solutions were diluted to 30 μL and dialyzed with Pierce 96-well microdialysis plate (3.5k MWCO) against 1:4 diluted Buffer 1 for 8 h at room temperature. Dialyzed solutions were added with 1 μL 0.5 mg/mL Trypsin (Pierce) in 1 mM HCl and incubated at 37 °C for 16 h. One microliter of 0.25 mM DTT was added to the reaction before resting it on ice for 1 h. The tryptic glycopeptides were analyzed as described previously.[22] Briefly, 5−10 μL of trypsinized samples were injected into a Bruker Elute UPLC system equipped with an ACQUITY UPLC Peptide BEH C18 Column, 300 Å, 1.7 μm, 2.1 mm × 100 mm (186003686 from Waters Corporation) with a 10 mm guard column (186004629 Waters Corporation) coupled to an Impact-II UHR TOF Mass Spectrometer (Bruker Daltonics, Inc.). The chromatographic separation method used 100% water with 0.1% formic acid as solvent A and 100% acetonitrile with 0.1% formic acid as solvent B. Chromatography was completed using 100% A for 1 min and a gradient of 0−50% B for 4 min. The flow rate was kept at 0.5 mL/min. Mass spectra in a range of 100−3000 Da were collected in 8 Hz. External calibration was performed for all spectra. We used MS/MS to monitor the target peptides and glycopeptides with collision energy of 50 eV (spectra shown in Supplementary Figure 20). Bruker Compass Data Analysis software version 4.1 was used to analyze the data. The targeted peaks in extracted ion chromatograms of targeted peptide and glycopeptide masses were integrated to calculate the modification using %Area, Area(P)/(Area(S) + Area(P)). Results are listed in Supplementary Table 7.

**LET-CFPS Protein Expression and Glycosylation.** pJL1.Fc was expressed in LET-CFPS the same way as NGTs (see above) with the addition of 2 or 5 μM purified ApQ or H495D mutant and 5 mM UDP-Glc. After 6 h CFPS incubation at 30 °C, 70 μL Buffer 1 with 5 mM imidazole was added into 50 μL CFPS solutions. The reactions were centrifuged with 12,000×*g* for 15 min at 4 °C, and supernatants were mixed with 30 μL His Dynabeads (Invitrogen) for a 10 min incubation. The beads were washed thrice with 120 μL Buffer 1 with 5 mM imidazole and eluted with 80 μL Buffer 1 with 500 mM imidazole. Elution solutions were dialyzed with Pierce 96-well microdialysis plate (3.5k MWCO) against 1:4 diluted Buffer 1 for 8 h at room temperature. One microliter of 0.5 mg/mL Trypsin (Pierce) in 1 mM HCl was added for 40 μL dialyzed solutions and incubated at 37 °C for 16 h. One microliter of 0.25 mM DTT was added to the reaction before resting it on ice for 1 h. LC-qTOF analysis was performed as described previously.

When noted that the glycosylation of Fc was completed after CFPS (postfolding), Fc was expressed in LET-CFPS for 30 °C and 20 h incubation, then centrifuged and supplemented with 2 μM purified ApQ and 5 mM UDP-Glc. This IVG reaction was then incubated at 30 °C for 6 h. The purification, dialysis, trypsinization, and LC-qTOF analysis of these reactions were performed as above.

***t*-Test and Data Analysis.** Two-tailed Student's *t*-tests and resulting *p* values were calculated in Microsoft Excel 2016. For all peptide library screens, only $n = 1$ experiment was used. For peptide IVGs of target protein sequences, $n = 3$ independent reactions were performed. For therapeutic protein modifications, $n = 2$ or $n = 3$ independent reactions were carried out as noted in dot plots. In the data analysis of $n > 1$, the average is presented, and standard deviations (s.d.) are shown as error bars. All heatmaps were generated in Microsoft Excel 2016.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

All data generated in this study are included in this published article (and its supplementary files) and are available upon request.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.3c00769.

Strains, plasmids, DNA sequences, peptide sequences, reaction kinetics data, mass spectrometry ionization parameters, protein expression data, and in vitro glycosylation reaction screening results for many single, double, and pooled mutant libraries (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Michael C. Jewett** − *Center for Synthetic Biology and Department of Chemical and Biological Engineering,*



https://doi.org/10.1021/acssynbio.3c00769
ACS Synth. Biol. 2024, 13, 1290−1302

*Northwestern University, Evanston, Illinois 60208, United States; Department of Bioengineering, Stanford University, Stanford, California 94305, United States;* orcid.org/0000-0003-2948-6211; Phone: (+1) 847 467 5007; Email: m-jewett@northwestern.edu; Fax: (+1) 847 491 4928

**Milan Mrksich** − *Department of Biomedical Engineering, Center for Synthetic Biology, Department of Chemistry, and Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-4964-796X; Phone: (+1) 847 467 0472; Email: milan.mrksich@northwestern.edu; Fax: (+1) 847 491 4928

### Authors

**Liang Lin** − *Department of Biomedical Engineering, Center for Synthetic Biology, and Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States*

**Weston Kightlinger** − *Center for Synthetic Biology and Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States*

**Katherine F. Warfel** − *Center for Synthetic Biology and Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-7780-6294

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.3c00769

### Author Contributions

L.L. and W.K. designed, performed, and analyzed experiments. K.W. expressed and purified NGT mutants and performed some other experiments. M.C.J. and M.M. directed the study and interpreted the data. L.L., W.K., K.W., M.C.J., and M.M. wrote the manuscript.

### Notes

The authors declare the following competing financial interest(s): M.M. is Founder and Chair of SAMDI Tech, Inc., which performs SAMDI-MS for high throughput screening in drug discovery programs of pharmaceutical companies. M.C.J. has a financial interest in Resilience, Gauntlet Bio, Synolo Therapeutics, and Stemloop, Inc.. M.C.J.s interests are reviewed and managed by Northwestern University and Stanford University in accordance with their conflict of interest policies. All other authors declare no conflicts of interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Defense Threat Reduction Agency (HDTRA1-21-1-0038, HDTRA1-20-1-0004), DARPA (W911NF-23-2-0039), and the National Science Foundation Graduate Research Fellowship program (DGE-1324585). We acknowledge Andrew Ott for assistance with LC-qTOF instrumentation. This work made use of IMSERC at Northwestern University, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the State of Illinois and International Institute for Nanotechnology (IIN).

## ■ REFERENCES

(1) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **2011**, *1*, 90.

(2) Schwarz, F.; Aebi, M. Mechanisms and principles of n-linked protein glycosylation. *Curr. Opin. Struc. Biol.* **2011**, *21*, 576−582.

(3) Sethuraman, N.; Stadheim, T. A. Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* **2006**, *17*, 341−346.

(4) Li, T.; et al. Modulating igg effector function by fc glycan engineering. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 3485−3490.

(5) Murakami, M.; et al. Chemical synthesis of erythropoietin glycoforms for insights into the relationship between glycosylation pattern and bioactivity. *Sci. Adv.* **2016**, *2*, No. e1500678.

(6) Mimura, Y.; et al. The influence of glycosylation on the thermal stability and effector function expression of human igg1-fc: Properties of a series of truncated glycoforms. *Mol. Immunol.* **2000**, *37*, 697−706.

(7) Wissing, S.; et al. Expression of glycoproteins with excellent glycosylation profile and serum half-life in cap-go cells. *BMC Proc.* **2015**, *9*, P12.

(8) Elliott, S.; et al. Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat. Biotechnol.* **2003**, *21*, 414−421.

(9) Perlman, S.; et al. Glycosylation of an n-terminal extension prolongs the half-life and increases the in vivo activity of follicle stimulating hormone. *J. Clin. Endocrinol. Metab.* **2003**, *88*, 3227−3235.

(10) Valderrama-Rincon, J. D.; et al. An engineered eukaryotic protein glycosylation pathway in escherichia coli. *Nat. Chem. Biol.* **2012**, *8*, 434−436.

(11) Keys, T. G.; Aebi, M. Engineering protein glycosylation in prokaryotes. *Curr. Opin. Syst. Biol.* **2017**, *5*, 23−31.

(12) Lin, C.-W.; et al. A common glycan structure on immunoglobulin g for enhancement of effector functions. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10611−10616.

(13) Wang, L.-X.; Amin, M. N. Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions. *Chem. Biol.* **2014**, *21*, 51−66.

(14) Jaroentomeechai, T.; et al. Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun.* **2018**, *9*, No. 2686.

(15) Schwarz, F.; et al. A combined method for producing homogeneous glycoproteins with eukaryotic n-glycosylation. *Nat. Chem. Biol.* **2010**, *6*, 264−266.

(16) Guarino, C.; DeLisa, M. P. A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology* **2012**, *22*, 596−601.

(17) Schoborg, J. A.; et al. A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnol. Bioeng.* **2018**, *115*, 739−750, DOI: 10.1002/bit.26502.

(18) Wacker, M.; et al. N-linked glycosylation in campylobacter jejuni and its functional transfer into E. *coli. Science* **2002**, *298*, 1790−1793.

(19) Kightlinger, W.; et al. Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat. Chem. Biol.* **2018**, *14*, 627−635.

(20) Song, Q.; et al. Production of homogeneous glycoprotein with multisite modifications by an engineered n-glycosyltransferase mutant. *J. Biol. Chem.* **2017**, *292*, 8856−8863.

(21) Keys, T. G.; et al. A biosynthetic route for polysialylating proteins in escherichia coli. *Metab. Eng.* **2017**, *44*, 293−301.

(22) Lin, L.; Kightlinger, W.; Hockenberry, A. J.; Jewett, M. C.; Mrksich, M. Sequential glycosylation of proteins with substrate-specific n-glycosyltransferases. *ACS Cent. Sci.* **2020**, *6*, 144−154.

(23) Lomino, J. V.; et al. A two-step enzymatic glycosylation of polypeptides with complex n-glycans. *Bioorg. Med. Chem.* **2013**, *21*, 2262−2270.

(24) Weeks, A. M.; Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol.* **2018**, *14*, 50−57.

(25) Schilling, O.; Overall, C. M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.* **2008**, *26*, 685−694.
(26) Wood, S. E.; et al. A bottom-up proteomic approach to identify substrate specificity of outer-membrane protease ompt. *Angew. Chem., Int. Ed.* **2017**, *129*, 16758−16762.
(27) Gurard-Levin, Z. A.; Kim, J.; Mrksich, M. Combining mass spectrometry and peptide arrays to profile the specificities of histone deacetylases. *ChemBioChem.* **2009**, *10*, 2159−2161.
(28) Silverman, A. D.; Karim, A. S.; Jewett, M. C. Cell-free gene expression systems: An expanded repertoire of applications. *Nature Reviews Genetics.* **2020**, *21*, 151−170.
(29) Schinn, S. M.; Broadbent, A.; Bradley, W. T.; Bundy, B. C. Protein synthesis directly from pcr: Progress and applications of cell-free protein synthesis with linear DNA. *New Biotechnol.* **2016**, *33*, 480−487.
(30) Hunt, A. C.; Vögeli, B.; Hassan, A. O.; et al. A rapid cell-free expression and screening platform for antibody discovery. *Nat. Commun.* 2023, *14*, 3897.
(31) Kawai, F.; et al. Structural insights into the glycosyltransferase activity of the actinobacillus pleuropneumoniae hmw1c-like protein. *J. Biol. Chem.* **2011**, *286*, 38546−38557.
(32) Xu, Y.; et al. A novel enzymatic method for synthesis of glycopeptides carrying natural eukaryotic n-glycans. *Chem. Commun.* **2017**, *53*, 9075−9077.
(33) Hao, Z.; Guo, Q.; Feng, Y.; Zhang, Z.; Li, T.; Tian, Z.; Zheng, J.; Da, L.-T.; Peng, W. Investigation of the catalytic mechanism of a soluble N-glycosyltransferase allows synthesis of N-glycans at noncanonical sequons. *JACS Au.* **2023**, *3*, 2144−2155.
(34) Szymczak, L. C.; Huang, C. F.; Berns, E. J.; Mrksich, M. Combining SAMDI mass spectrometry and peptide arrays to profile phosphatase activities. *Methods Enzymol.* **2018**, *607*, 389−403.
(35) Kuo, H. Y.; DeLuca, T. A.; Miller, W. M.; Mrksich, M. Profiling deacetylase activities in cell lysates with peptide arrays and SAMDI mass spectrometry. *Anal. Chem.* **2013**, *85*, 10635−10642.
(36) Kornacki, J. R.; Stuparu, A. D.; Mrksich, M. Acetyltransferase p300/cbp associated factor (pcaf) regulates crosstalk-dependent acetylation of histone h3 by distal site recognition. *ACS Chem. Biol.* **2015**, *10*, 157−164.
(37) Houseman, B. T.; Huh, J. H.; Kron, S. J.; Mrksich, M. Peptide chips for the quantitative evaluation of protein kinase activity. *Nat. Biotechnol.* **2002**, *20*, 270−274.
(38) Kontermann, R. E. Strategies for extended serum half-life of protein therapeutics. *Curr. Opin. Biotechnol.* **2011**, *22*, 868−876.
(39) Hershewe, J. M.; Kightlinger, W.; Jewett, M. C. Cell-free systems for accelerating glycoprotein expression and biomanufacturing. *Journal of Industrial Microbiology and Biotechnology.* **2020**, *47*, 977−991.
(40) Wild, R.; et al. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic n-glycosylation. *Science* **2018**, *359*, 545−550, DOI: 10.1126/science.aar5140.
(41) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011**, *474*, 350−355.
(42) Ollis, A. A.; et al. Substitute sweeteners: Diverse bacterial oligosaccharyltransferases with unique n-glycosylation site preferences. *Sci. Rep.* **2015**, *5*, No. 15237.
(43) Ollis, A. A.; Zhang, S.; Fisher, A. C.; DeLisa, M. P. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat. Chem. Biol.* **2014**, *10*, 816−822.
(44) Fisher, A. C.; et al. Production of secretory and extracellular n-linked glycoproteins in escherichia coli. *Appl. Environ. Microbiol.* **2011**, *77*, 871−881.
(45) Martin, R. W.; Des Soye, B. J.; Kwon, Y. C.; Kay, J.; Davis, R. G.; Thomas, P. M.; Marcum, R. D.; Majewska, N. I.; Chen, C. X.; Weiss, M. G.; Stoddart, A. E.; Amiram, M.; Charna, A. K.; Isaacs, F. J.; Kelleher, N. L.; Hong, S. H.; Jewett, M. C. Cell-free protein synthesis from genomically recoded bacteria enables multi-site incorporation of non-canonical amino acids. *Nat. Commun.* **2018**, *9*, No. 1203.
(46) Engler, C.; Kandzia, R.; Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **2008**, *3*, No. e3647.
(47) Liu, H.; Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **2008**, *8*, 91.
(48) Kwon, Y.-C.; Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* **2015**, *5*, No. 8663.
(49) Stark, J. C.; Jaroentomeecha, T.; Warfel, K.F.; Hershewe, J.M.; DeLisa, M.P.; Jewett, M.C Rapid biosynthesis of glycoprotein therapeutics and vaccines from freeze-dried bacterial cell lysates. *Nat. Protoc.* **2023**, *18*, 1−25.
(50) Stark, J. C.; Jaroentomeechai, T.; Moeller, T. D.; Hershewe, J. M.; Warfel, K. F.; Moricz, B. S.; Martini, A. M.; Dubner, R. S.; Hsu, K. J.; Stevenson, T. C.; Jones, B. D.; DeLisa, M. P.; Jewett, M. C. On-demand, biomanufacturing of protective conjugate vaccines. *Sci. Adv.* **2021**, *7*, No. EABE9444, DOI: 10.1126/sciadv.abe9444.
(51) Jewett, M. C.; Swartz, J. R. Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient protein synthesis. *Biotechnol. Bioeng.* **2004**, *86*, 19−26.
(52) Jewett, M. C.; Swartz, J. R. Substrate replenishment extends protein synthesis with an *in vitro* system designed to mimic the cytoplasm. *Biotechnol. Bioeng.* **2004**, *87*, 465−472.
(53) Jewett, M. C.; Calhoun, K. A.; Voloshin, A.; Wuu, J. J.; Swartz, J. R. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* **2008**, *4*, No. 220.


https://doi.org/10.1021/acssynbio.3c00769
*ACS Synth. Biol.* 2024, 13, 1290−1302