

High-Throughput Synthesis and Analysis of Intact Glycoproteins Using SAMDI-MS

José-Marc Techner,^{†,||,#} Weston Kightlinger,^{‡,||,#} Liang Lin,^{§,||} Jasmine Hersheve,^{‡,||} Ashvita Ramesh,^{§,⊥} Matthew P. DeLisa,[∇] Michael C. Jewett,^{*,‡,||} and Milan Mrksich^{*,†,§,||}

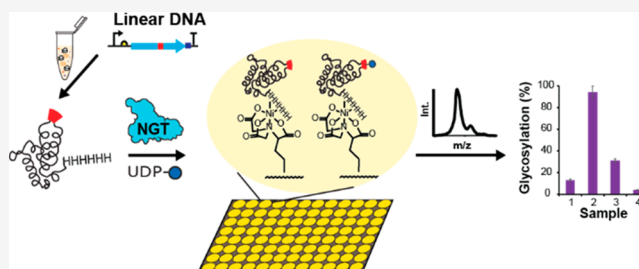
[†]Department of Chemistry, [‡]Department of Chemical and Biological Engineering, [§]Department of Biomedical Engineering, and ^{||}Center for Synthetic Biology, Northwestern University, Evanston, Illinois 60208, United States

[⊥]Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, United States

[∇]Department of Microbiology, Nancy E. and Peter C. Meinig School of Biomedical Engineering, Biochemistry, Molecular and Cell Biology, and Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, United States

Supporting Information

ABSTRACT: High-throughput quantification of the post-translational modification of many individual protein samples is challenging with current label-based methods. This paper demonstrates an efficient method that addresses this gap by combining *Escherichia coli*-based cell-free protein synthesis (CFPS) and self-assembled monolayers for matrix-assisted laser desorption/ionization mass spectrometry (SAMDI-MS) to analyze intact proteins. This high-throughput approach begins with polyhistidine-tagged protein substrates expressed from linear DNA templates by CFPS. Here, we synthesized an 87-member library of the *E. coli* Immunity Protein 7 (Im7) containing an acceptor sequence optimized for glycosylation by the *Actinobacillus pleuropneumoniae* N-glycosyltransferase (NGT) at every possible position along the protein backbone. These protein substrates were individually treated with NGT and then selectively immobilized to self-assembled monolayers presenting nickel-nitrilotriacetic acid (Ni-NTA) complexes before final analysis by SAMDI-MS to quantify the conversion of substrate to glycoprotein. This method offers new opportunities for rapid synthesis and quantitative evaluation of intact glycoproteins.



Proteins undergo a vast array of post-translational modifications (PTMs) that regulate protein function and underlie signaling pathways in the cell.^{1,2} Hence, assays that measure PTMs in protein samples are required to facilitate biological engineering and fundamental studies of cell function.^{1,3} However, the in vitro assays that are commonly used to determine the type and extent of PTMs on individual protein samples often lack generalizability or are low throughput. For example, the use of immunological probes in Western blot or enzyme-linked immunosorbent assays (ELISA) provide throughput but often lack specificity or are not available for all modifications. Methods that combine liquid chromatography and mass spectrometry (LC-MS) are quite general and provide quantitative data but have the limitation of low throughput.

The limitations of PTM assays are particularly evident for glycosylation which is the attachment of sugar moieties (glycans) to amino acid side-chains by glycosyltransferases (GTs).⁴ Glycan structures are highly diverse and analytical probes with sufficient specificity are often not available, and the design rules governing the installation and efficiency of glycosylation at a given site remain unclear for many glycosylation systems.⁵ Thus, there remains an urgent need

for analytical methods that enable rapid and quantitative assessment of glycosylation efficiency for hundreds of protein samples. Here we report a strategy to address the broader need for high-throughput PTM assays and illustrate this method in a study of protein glycosylation. The method uses in vitro synthesis and selective capture of polyhistidine (His-) tagged proteins from bacterial cell-free lysates onto a self-assembled monolayer and subsequent assessment of glycosylation status by self-assembled monolayers for matrix-assisted laser desorption mass spectrometry (SAMDI-MS).⁶

Two emerging technologies have set the stage for the approach described here. First, cell-free protein synthesis (CFPS) in *Escherichia coli* lysates has undergone substantial development and is now a robust method to obtain g/L titers of a protein of interest from linear or plasmid DNA templates in test tubes in a matter of hours.^{7–9} CFPS reactions can be easily parallelized and interfaced with robotic liquid-handling and can provide known protein concentrations without rate-

Received: September 23, 2019

Accepted: December 19, 2019

Published: December 19, 2019

limiting steps of cell lysis and protein purification, resulting in shorter design-build-test cycles.^{7–9} Although *E. coli* does not contain native glycosylation systems, it can be engineered with heterologous glycosylation enzymes for the study and production of glycoproteins.^{10–13}

Second, SAMDI-MS provides a label-free and high throughput assay for quantifying enzymatic modifications of a broad range of biomolecules. SAMDI-MS uses mass spectrometry to measure the masses of substrates and corresponding products that are immobilized to a self-assembled monolayer and can be performed in plates having an array of 1,536 monolayers in 1 h.¹⁴ Recent examples include the characterization of diverse PTM enzymes including GTs using carbohydrate substrates^{15,16} and the extensive characterization of the peptide acceptor preferences of polypeptide-modifying glycosyltransferases (ppGTs) to inform the design of optimized peptide sequences within target proteins.¹⁰ This previous work used peptides as substrates for glycosylation, with the limitation that peptides might not recapitulate the activities of their corresponding protein substrates.¹⁷ The method reported here allows the use of intact protein substrates and therefore will enable a broad range of studies of protein modification.

■ EXPERIMENTAL SECTION

Reagents and Equipment. All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise specified. NTA-SH for SAM functionalization was prepared as previously described.¹⁸ All SAMDI-MS experiments were performed with an Applied Biosystems SciEx (AB Sciex) MALDI-TOF/TOF 5800 instrument. Protein spectra were acquired in linear positive mode, and peptide spectra were acquired in reflector positive mode. Crude extracts for CFPS were generated from *E. coli* strain BL21 Star (DE3) (NEB). Cell growth, harvest, and lysis were performed as previously described.¹⁹ Generated lysates were flash-frozen on liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$ until use. Excel software was used to analyze and plot the data.

Preparation of SAMDI-MS Plates. Stainless steel plates with 384 gold spots were prepared as explained elsewhere.²⁰ The plates were placed in an ethanolic solution of 0.4 mM symmetric 11-carbon alkyl-disulfide terminated with tri-(ethylene glycol) groups and 0.1 mM asymmetric alkyl-disulfide terminated with a maleimide group and a tri(ethylene glycol) group (ProChimia) for approximately 16 h to form self-assembled monolayers (SAMs). The plates were removed, rinsed with ethanol, and dried with air prior to use. Plates for protein capture were further treated with an aqueous 100 μM NTA-thiol solution in 100 mM Tris-HCl pH 8.0 for 30 min, rinsed with ethanol, and incubated with aqueous 100 mM NiSO_4 for 10 min. Plates were then rinsed with water only before drying with air.

Molecular Cloning and DNA Template Preparation. All strains and DNA constructs used in this study are summarized in Table S1. Full-length coding DNA sequences generated in this study are provided in Note S1. Plasmid DNA encoding NGT (pJL1.NGT);¹⁰ the *Campylobacter jejuni* oligosaccharyltransferase (OST), CjPglB (pJL1.pglB);¹² the *C. jejuni* LLO biosynthesis pathway (pACYCpglB::kan);¹² and the Im7.DQNAT acceptor protein (pJL1.Im7.DQNAT)¹² were prepared as described previously. Plasmid DNA encoding Im7 linear DNA templates containing positional replacements of GGNWTT at each amino acid position in the wildtype

(WT) Im7 sequence were designed and synthesized by Twist Biosciences (Twist). Each template contains a C-terminal 6xHis-tag and a short 5' untranslated region (UTR) sequence containing a portion of the ribosome binding site. To generate the linear DNA templates for linear expression template (LET) CFPS reactions, T7 promoter and terminator regions were added to template genes using PCR amplification with primers 502 and 503 (Integrated DNA Technologies, Table S2). 25- μL PCR reactions for amplification of Im7 variants were performed using 2 ng of DNA template, 0.02 U/ μL Q5 high-fidelity polymerase, 500 nM of each primer, 200 μM dNTPs, in 1X Q5 polymerase buffer (NEB). The PCR was completed using a 30 s initial denaturation at 98 $^{\circ}\text{C}$ followed by 36 rounds of 10 s 98 $^{\circ}\text{C}$ denaturation, and 30 s 59 $^{\circ}\text{C}$ annealing, and 1 min 72 $^{\circ}\text{C}$ extension followed by a final extension for 5 min at 72 $^{\circ}\text{C}$. Sequences for therapeutically relevant proteins Leptin, G-CSF, Interferon-Alfa, and IL1-RA modified with an additional N-terminal leader sequence encoding both a His-tag and an optimized glycosylation sequence (GGNWTT) were synthesized and assembled into the pJL1 vector by Twist Biosciences. The native N-X-S/T glycosylation site on IL1-RA was disabled by an N185Q mutation. These plasmids were then used as templates for PCR amplification by primers 174 and 175 using a similar method to the Im7 library with an annealing temperature of 65 $^{\circ}\text{C}$.

Expression and Purification of Im7-6 and Im7.DQNAT from *E. coli*. Purified Im7-6 and Im7.DQNAT used for the development of SAMDI-MS methods and as a substrate for CjPglB, respectively, were expressed and purified from living *E. coli* cells as described previously¹² with minor modifications. Briefly, proteins were expressed in BL21 Star (DE3) *E. coli* cells that were grown at 37 $^{\circ}\text{C}$ in Miller's Luria Broth (LB) supplemented with appropriate antibiotics; induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) overnight at 20 $^{\circ}\text{C}$; harvested by centrifugation; and then lysed by homogenization in Buffer 1 (50 mM NaH_2PO_4 and 300 mM NaCl, pH 8.0) supplemented with 10 mM imidazole, 1X Halt Protease Inhibitor (Thermo Fisher Scientific), 25 U/mL Benzonase (Millipore), and 0.1 mg/mL lysozyme (Sigma). The proteins were purified using Qiagen Ni-NTA agarose and Buffer 1 by equilibrating with 10 mM imidazole, washing with 10 column volumes of 30 mM imidazole, and eluting with 500 mM imidazole before dialysis into Buffer 1 with 5% glycerol. Proteins were quantified using Nanodrop 2000c UV absorption at 260 nm.

In Vitro Protein Glycosylation Reactions. Glycosylated Im7-6 samples for the development of SAMDI-MS methods were generated by mixing 10 μM Im7-6 purified from *E. coli* cells, 5 mM uracil diphosphate (UDP-) glucose, and 0.05 μM NGT in in vitro glycosylation (IVG) buffer (100 mM HEPES pH 7.4, 0.05% Tween-20) in a final reaction volume of 5 μL and incubated for 4 h at 37 $^{\circ}\text{C}$ before quenching by 1:1 addition of 30% (v/vol) ethanol and buffer. Glycosylation of Im7 library proteins was completed by dilution of completed LET-CFPS reactions to a final concentration of 10 μM in IVG buffer and 16% (v/vol) CFPS reaction (varying CFPS reaction volumes were filled by a completed CFPS reaction that synthesized sfGFP) in a separate 96 well plate. For the positional glycosylation site library, IVG reactions were assembled in a new 96 well plate where each well contained 2.5 μM of an Im7 variant, 5 mM UDP-glucose, and indicated concentrations of NGT in 100 mM HEPES pH 7.4. These 4- μL IVG reactions were then incubated with indicated times

and temperatures. The four therapeutically relevant proteins were glycosylated by assembly of 1 μM of each target protein from completed LET-CFPS reactions, 0.1 μM NGT, and 2.5 mM UDP-glucose. These IVG reactions were incubated overnight at 30 °C. For glycosylation of Im7.DQNAT with CjPglB, all reagents were prepared as previously described.¹² Glycosylation reactions were composed of 4 μL of CjPglB expressed in CFPS, 4 μL of extracted *C. jejuni* lipid-linked oligosaccharides (LLOs), 1 μL of purified Im7-DQNAT, and 1 μL of a master mix (10% (w/v) Ficoll 400, 500 mM HEPES, pH 7.4, and 10 mM MnCl_2). CjPglB IVG reactions were then incubated overnight at 30 °C and analyzed.

Analysis of NGT Preferences for Amino Acids Adjacent to GGNWTT. Peptides were screened with SAMDI-MS according to the method previously described.¹⁰ 50 μM of each peptide was mixed with 2.5 mM UDP-glucose in the presence of NGT, (0.02 μM for X_{-3} peptides or 0.04 μM for X_{+4} peptides) in 100 mM HEPES (pH 8.0) and 500 mM NaCl for 1 h at 30 °C. After this IVG reaction, the peptides were immobilized to a maleimide monolayer and analyzed by SAMDI-MS. A control containing CFPS without NGT was also performed. To predict NGT preferences, corresponding peptide modification values for each Im7 mutant were summed together and then normalized across the Im7 library.

Mass Analysis of Proteins by SAMDI-MS. One μL of sample mixture to be analyzed by SAMDI-MS was dispensed onto each spot of a 384-well SAMDI-MS plate prepared with a Ni-NTA monolayer using a Tecan 96-channel automated liquid handler. This plate was incubated for 30 min at room temperature, rinsed extensively with water, and then dried with compressed air. 0.5 μL of 10 mg/mL Sinapic Acid matrix in 50% acetonitrile/49.9% H_2O /0.1% TFA was applied onto each spot and allowed to dry in ambient conditions prior to SAMDI-MS analysis. Spectra were acquired using a 300 ns extraction delay time. 2000 laser shots (at 400 Hz) were manually collected and averaged over a spot for each spectrum. Data were baseline corrected and smoothed with a Gaussian routine before integration of peak areas using AB Sciex Data Explorer software. External calibration was performed prior to acquisition. All SAMDI-MS mass spectra shown in this paper are normalized such that 100% is equal to the intensity of the most intense peak within the indicated mass ranges (Rel. Int. %). Modification efficiency (Glycosylation %) values shown in Figure 5 were calculated using the following equation.

$$\text{Glycosylation \%} = 100 \times \frac{\text{Product}}{\text{Product} + \text{Substrate}}$$

where Product corresponds to the sum of the area-under-the-curve for protonated $[\text{Glc}_1]$ and sinapic acid $[\text{Glc}_1]_{\text{SA}}$ adducts of the modified protein. Substrate corresponds to sum of protonated $[\text{Glc}_0]$ and sinapic acid $[\text{Glc}_0]_{\text{SA}}$ adducts of the unmodified protein species. Correction calculations for sinapic acid adducts are explained in Method S1. Theoretical and observed mass annotations for SAMDI-MS spectra are shown in Table S3.

Cell-Free Protein Synthesis. CFPS reactions were conducted using a PANOX-SP crude lysate system.²¹ CFPS reactions were prepared as previously described.²² DNA templates encoding proteins to be synthesized were added in the form of either plasmid DNA at a concentration of 13.3 $\mu\text{g}/\text{mL}$ or the addition of 3 μL linear expression template PCR product, as indicated. The CFPS reactions for the Im7 library

of therapeutic proteins were carried out at 30 °C for 6 h. The NGT used for glycosylation was produced in CFPS using pJL1.NGT for 20 h at 23 °C. CFPS reactions used to synthesize CjPglB were carried out as previously described,¹² supplemented with nanodiscs and incubated at 30 °C for 20 h.

Quantification of CFPS Yields. Total and soluble CFPS yields were quantified for the Im7 library and therapeutic protein targets using CFPS reactions identical to those above supplemented with 10 μM ^{14}C -leucine (PerkinElmer). Protein quantification for duplicate CFPS reactions was completed using trichloroacetic acid (TCA) protein precipitation followed by radioactivity quantification using a Microbeta2 scintillation counter (PerkinElmer) as previously described.²¹ Soluble fractions were isolated after centrifugation at 12,000g for 15 min at 4 °C.

LC-TOF Analysis of Glycoprotein and Glycopeptides. Proteins were purified using Dyna-His tag Ni-NTA magnetic beads (Thermo Fisher Scientific) and analyzed by ultra-performance liquid chromatography-time-of-flight mass spectrometry (UPLC-TOF) as previously described with minor modifications.¹⁰ Briefly, samples were diluted 1:4 in Buffer 1, bound to Dyna-His tag beads, washed with Buffer 1 with 5 mM imidazole, eluted with Buffer 1 with 500 mM imidazole, and dialyzed into 1:4 diluted Buffer 1. This dialyzed sample was injected into a Bruker Elute UPLC equipped with a C4 Column (186004495 Waters Corp.) coupled to an Impact-II UHR TOF mass spectrometer (Bruker). Liquid chromatography was performed using a gradient of acetonitrile with 0.1% formic acid (Solvent B) and H_2O with a 0.1% formic acid (Solvent A) system at a flow rate of 0.5 mL/min and a 50 °C column temperature. The LC method used a 1 min hold at 20% B and a 4 min gradient from 20% to 50% B followed by washing and equilibration. For glycopeptide analysis, protein glycosylation targets were digested with 0.0044 $\mu\text{g}/\mu\text{L}$ MS grade Trypsin (Thermo Fisher Scientific) at 37 °C overnight before injection. UPLC-MS of peptides was performed similarly to protein analysis using a C18 Column (186003686 Waters Corp.). The LC method for glycopeptides was completed using a 1 min hold at 0% B and a 4 min gradient from 0% to 50% B followed by washing and equilibration. A scan range of 100–3000 m/z with spectral rates of 2 and 8 Hz was used for proteins and peptides, respectively. External calibration was performed prior to data collection.

LC-TOF Data Analysis. LC-MS data were analyzed using Bruker Compass Data Analysis version 4.1. Glycopeptide MS and intact glycoprotein MS spectra were averaged across the full elution times of the glycosylated and unglycosylated species. Summed LC-MS spectra were selected and annotated manually. Quantification of intact protein modification for verification of SAMDI-MS quantification was completed as previously described by integration of extracted ion chromatograms from dominant charge states.²

Circular Dichroism Spectra of Selected Im7 Mutants. Positional mutants p23, p24, p72, p73, p39, and p40 were expressed and purified as described above. CD spectra were collected from purified proteins at 25 °C in 50 mM sodium phosphate buffer at protein concentrations of 13–15 μM . CD spectra were collected using a Jasco J-810 spectrometer over the wavelength range of 190–280 nm using a 1.0 mm path length quartz cuvette. Scans were acquired in triplicate at a rate of 1 nm/min.

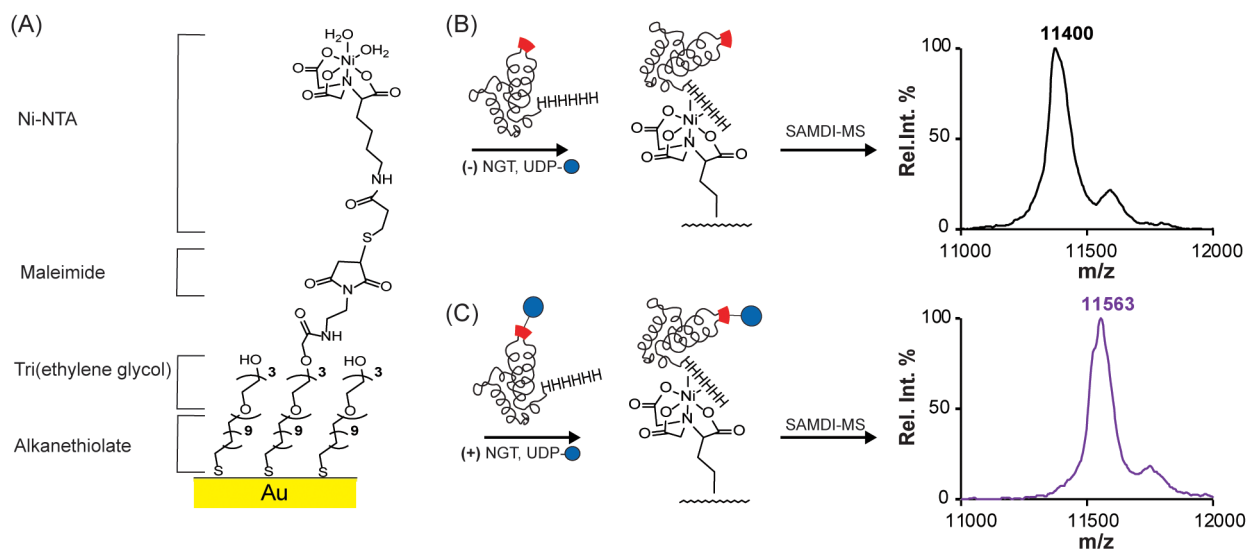


Figure 1. Scheme for assaying glycosylation of Im7-6 by NGT. (A) A monolayer displaying Ni-NTA against a background of tri(ethylene glycol) groups. (B) SAMDI-MS of Im7-6 in the absence of treatment with NGT shows a major G_0 peak at average m/z of 11400 ± 2 and a smaller sinapic acid adduct peak ($m/z = 11593 \pm 5$). (C) Glycosylated Im7-6 (G_1) and its sinapic acid adduct detected with average m/z of 11563 ± 9 and 11759 ± 17 , respectively. The mass difference of 163 Da between G_1 and G_0 corresponds to the addition of glucose and loss of water (expected 162 Da).

Im7 Solvent Accessibility Analysis. The STRIDE Web server (<http://webclu.bio.wzw.tum.de/stride/>) was used to calculate the absolute accessible surface area for each residue of native Im7 from PDB 1AYI.²³ The relative solvent accessibility (RSA) of each residue was calculated by dividing the absolute accessible surface area by the maximum solvent accessibility of the amino acid as previously determined.²⁴

RESULTS

Our assay of glycoprotein modification starts by organizing parallel reaction mixtures containing His-tagged protein substrate, enzyme, and cofactors in a multiwell plate. The reactions are then quenched and transferred to a monolayer array plate, where the protein substrate and product are immobilized onto a self-assembled monolayer presenting nickel-chelated nitrilotriacetic acid (NTA). We chose to use NTA-terminated monolayers because this method is familiar to protein science groups and has been previously validated.^{25–27} Following Ni-NTA immobilization, all other reaction components are removed by rinsing, allowing the protein substrate and product to be quantified by SAMDI-MS mass spectrometry (Figure 1). The amount of modified product is calculated by integration of the peak areas in the resulting SAMDI-MS mass spectrum. Because the signal of a protein is split between the parent protein peak and the adduct species at a constant proportion, its contribution to the product peak can be determined (Method S1).^{28,29}

As a proof-of-concept target for our assay, we chose Im7-6, a recombinant version of Im7 immunity protein from *E. coli* that contains an internal glycosylation sequence GGNWTT and a 6x His-tag at the C-terminus because it can be readily glycosylated by NGT.¹⁰ While any post-translational modification system that results in a mass change could be selected, the NGT glycosylation system was chosen as a model because there is no commercial antibody for this modification and it represents one of the smallest mass shifts reflecting glycosylation, presenting a challenging test case for our method. We confirmed that SAMDI-MS can detect both the unglycosylated and glycosylated forms of Im7-6. Glycosylated

Im7-6 was generated by incubating Im7-6 purified from bacterial cells with unpurified NGT and UDP-glucose (Figure 1).

To further validate the quantitative measurement of glycosylation on Im7-6, we prepared samples with defined glycosylation efficiencies by mixing defined amounts of the substrate and glycosylated product (Figure 2). We observed a linear correlation between SAMDI-MS response and glycosylation efficiency, indicating that glycosylated and unglycosylated analytes have similar ionization efficiency and binding affinity to the monolayer (Figure 2B). Comparison to an established UPLC-TOF method showed a strong linear correlation, validating the use of SAMDI-MS to measure glycosylation (Figure 2C). However, SAMDI-MS requires only 2–3 s of instrument time per sample compared to 10 min for the UPLC-TOF method.

SAMDI-MS and Cell-Free Protein Synthesis. Having shown that protein SAMDI-MS is quantitative, we next tested the strategy of synthesizing His-tagged substrates rapidly from linear DNA transcripts in CFPS to develop the coupled synthesis and analysis scheme shown in Figure 3. We originally considered pursuing a traditional protein expression strategy in living cells; however, the time and cost of cloning and plasmid preparation were a significant bottleneck. The falling cost of commercial gene synthesis enabled us to use a linear expression template (LET) LET-CFPS strategy instead.³⁰ To develop our synthesis method, we tested a variant of Im7, termed “p3” for the position of the modified asparagine residue in the GGNWTT glycosylation sequence replacing native residues 1–6. To avoid the synthesis of redundant promoter and terminator regions, we designed a gene fragment encoding only our intended Im7-p3 protein including a C-terminal His-tag and a leader sequence at the 5' end containing a short portion of the ribosomal binding site. We then designed primers that would install necessary transcriptional and translational elements during PCR amplification of the original gene. We found that the linear template generated in this PCR provided similar CFPS yields compared to a plasmid encoding Im7-p3 after 6 h (Figure S1). With the ability to make protein

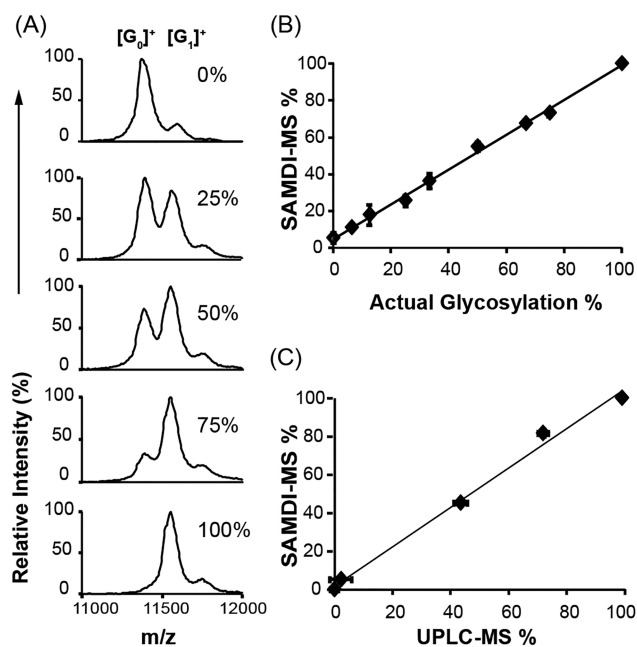


Figure 2. Quantitative assessment of SAMDI-MS protein glycosylation assay. (A) Mixtures of glycosylated (G_1) and unglycosylated (G_0) Im7-6 were analyzed by SAMDI-MS to obtain spectra (0 to 100% glycosylated, increasing downward). (B) Linear correlation of % glycosylation determined by Protein SAMDI-MS to actual values ($R^2 = 0.997$, $n = 3$). (C) Correlation of Protein SAMDI-MS performance and a UPLC-MS method for identical samples ($R^2 = 0.993$, $n = 3$). The sinapic acid adduct of G_0 overlaps with the G_1 peak so both (B) and (C) were calculated with corrections for sinapic acid adducts for both G_0 and G_1 (Method S1). Error bars equal one standard deviation ($n = 3$).

substrates in CFPS, we tested the compatibility of the SAMDI-MS assay with CFPS-derived samples. We found that our assay is robust across a range of CFPS lysate concentrations (Figure 4A). Furthermore, we found an approximate lower boundary

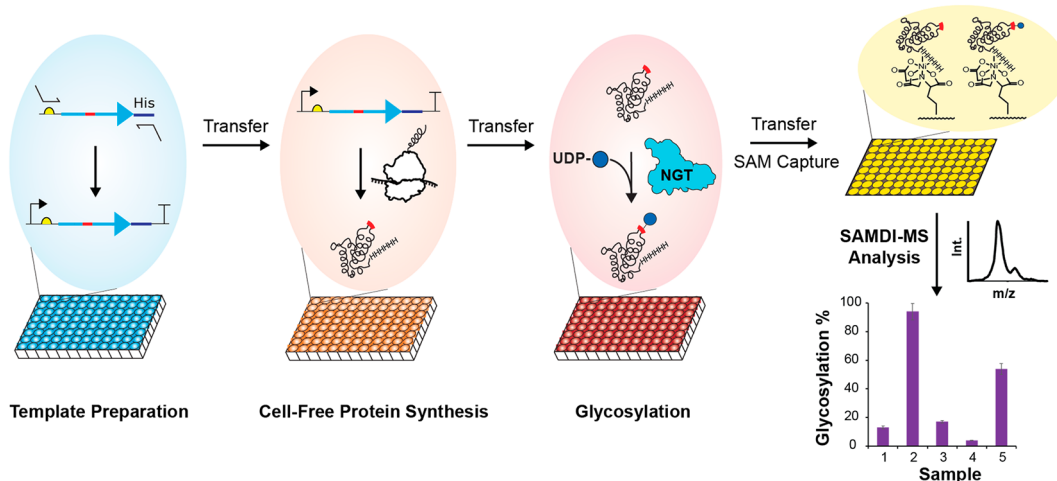


Figure 3. Strategy for high-throughput synthesis and analysis of intact glycoproteins. Linear DNA encoding His-tagged proteins containing an engineered glycosylation sequence are commercially synthesized and then amplified by PCR in a 96 well plate to create linear expression templates (LETs). LETs are then added directly to CFPS reaction mixtures in a new well plate to generate His-tagged substrates. After CFPS is complete, synthesized proteins within lysates undergo glycosylation by NGT and excess UDP-glucose in another well plate. Reactions are quenched and then transferred to a SAMDI-MS plate containing Ni-NTA terminated monolayers (chemical structure shown in Figure 1) for capture of His-tagged protein. After capture, reaction products are analyzed by SAMDI-MS to quantify relative amounts of glycosylation (theoretical bar graph shown).

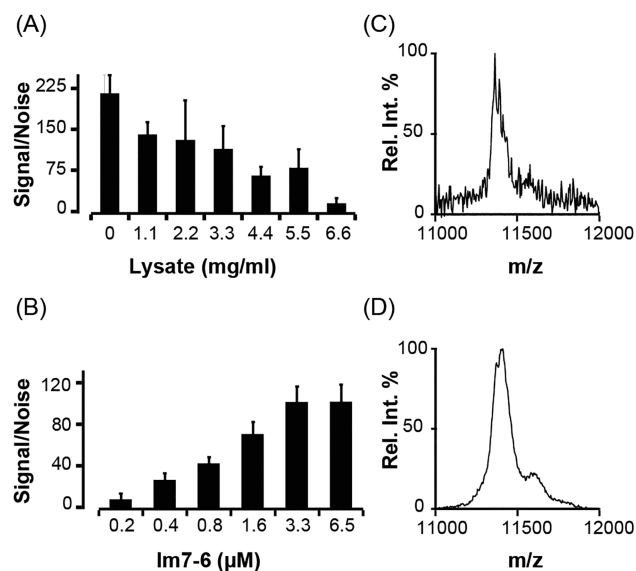


Figure 4. SAMDI-MS compatibility with CFPS lysate samples. (A) SAMDI-MS spectral signal-to-noise of 5μ M Im7-6 in samples containing variable amounts of CFPS lysate (shown as overall protein concentration). (B) SAMDI-MS spectral signal-to-noise of Im7-6 at various concentrations in samples containing 1.1 mg/mL of CFPS lysate. (C) and (D) are representative SAMDI-MS spectra of 5μ M Im7-6 in samples containing 6.6 and 1.1 mg/mL of CFPS lysate, respectively. Error bars equal one standard deviation ($n = 3$).

of detection of His-tagged protein at $\sim 0.5 \mu$ M (Figure 4B). For consistency, we completed each reaction with 10% CFPS lysate (v/v), which corresponds to an approximate concentration of ~ 1.1 mg/mL of total lysate protein and provided an excellent signal-to-noise ratio (Figure 4D).¹⁹ We attribute the decrease in signal spectra quality at the concentrated lysate sample conditions ($>50\%$ CFPS lysate volume or ~ 5.5 mg/mL of protein) to competition of background species that weakly interact with the Ni-NTA ligand (Figure 4C).

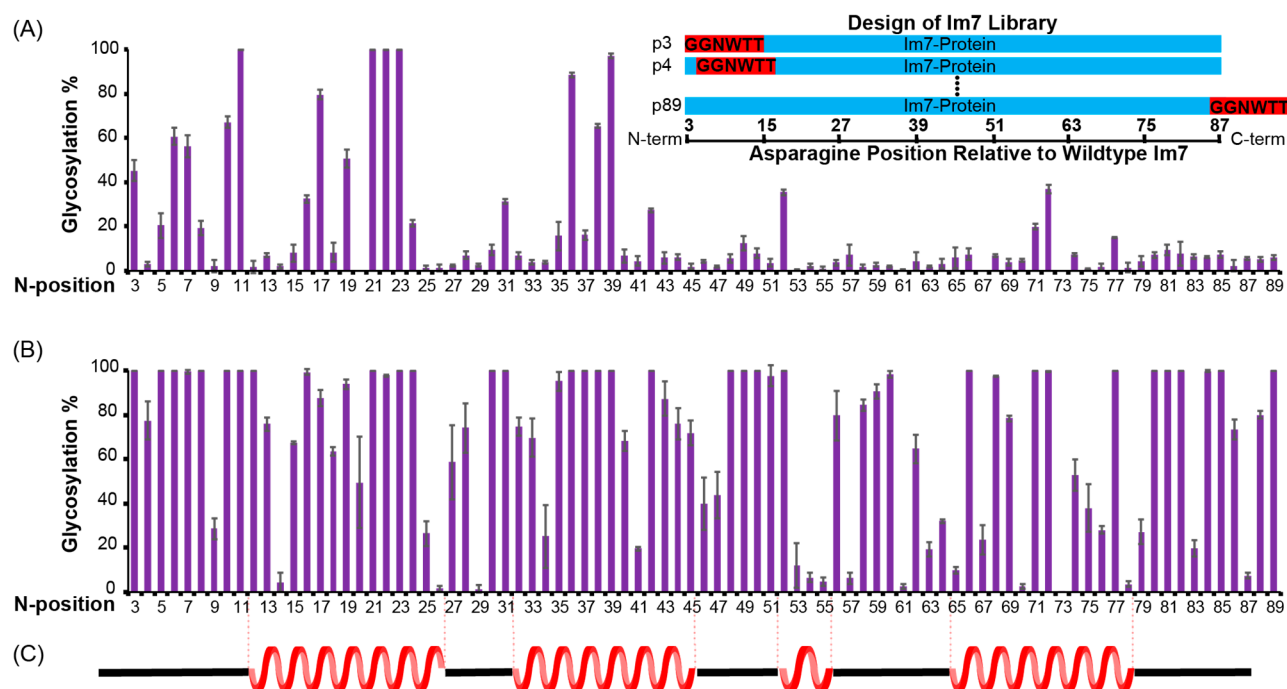


Figure 5. Analysis of Im7 positional glycosylation site library synthesized and glycosylated in vitro. (A) Treatment of 2.5 μM of synthesized Im7-library with “non-saturating” IVG conditions (0.05 μM NGT) ($n = 3$). Design of the Im7 library with GGNWTT (top right inset) substitution of native residues. Representative spectra of select mutants in Figure S3. (B) Identical Im7-library IVG reactions shown in (A) under “saturating” IVG conditions (0.667 μM NGT) ($n = 3$). (C) Location of helical secondary structure in wildtype Im7 based on PDB 1AYI. Im7-library CFPS yields were previously determined by radioisotope incorporation (see Supporting Information). Error bars equal one standard deviation ($n = 3$). Bars showing 100% glycosylation indicate that the aglycosylated protein species was not detected.

Im7 Library Glycosylation Study. Having shown that this assay is compatible with CFPS lysates and can quantitatively determine the glycosylation efficiency of intact protein substrates, we demonstrated the throughput of our CFPS-SAMDI-MS workflow with a library of Im7 mutants having the GGNWTT sequence inserted at each amino acid position in a manner identical to Im7-p3. In this way, each hexameric amino acid sequence in Im7 was replaced with a glycosylation acceptor sequence. We note that a similar approach could be used to optimize the placement of conjugation sites within target proteins. Furthermore, as the first mass spectrometry data set to systematically assess all possible glycosylation positions within a target protein, we hoped to provide a case study to understand if the propensity for glycosylation could be predicted *a priori* or if empirical measurements were required.

We expressed each Im7 variant using LET-CFPS and used [^{14}C]-leucine incorporation to determine the total and soluble protein yields for each construct. We found that all proteins were highly soluble (>90%) and that soluble protein yields for each member of the Im7 library were similar except for the first 20 positions which exhibited lower and more variable yields (Figure S2). We attribute this yield variation to changes in the strength of the ribosome binding site, which is sensitive to mRNA sequence changes near the start codon.^{31,32}

Next, we characterized the in vitro glycosylation of the synthesized Im7 variants following the workflow outlined in Figure 3. This workflow includes PCR amplification of gene templates, cell-free protein expression, in vitro glycosylation, capture of His-tagged proteins to Ni-NTA functionalized monolayers, and SAMDI-MS analysis. Each step in this process is performed using a standard 96 well plate format that is

compatible with automation by robotic liquid handling. All samples consistently yielded spectra with high signal-to-noise with little variation in the relative size of the synaptic acid adduct peak for all protein mutants (Figures S3 and S4, respectively).

To assess the propensity of each variant for glycosylation, we made two measurements of glycosylation efficiency, one after treatment with “non-saturating” IVG conditions that did not fully glycosylate most library members (2.5 μM Im7 mutant with 5 mM UDP-glucose and 0.05 μM NGT incubated for 2 h at 25 $^{\circ}\text{C}$) and another after “saturating” IVG conditions that did fully glycosylate most library members (0.667 μM NGT for 5 h at 37 $^{\circ}\text{C}$), shown in Figure 5A and Figure 5B, respectively. Under nonsaturating conditions, wide variability in glycosylation was observed. These data indicate that positions 11, 21–23, 36, and 39 show the greatest propensity for glycosylation. Data collected after exposure of the library to saturating glycosylation conditions showed that, with enough time and enzyme, most positions can be modified by NGT but that positions 14, 26, 29, 53, 55, 57, 61, 70, 73, 78, and 87 are especially resistant to glycosylation (Figure 5B).

With this data set in hand, we asked if these patterns of glycosylation could have been predicted *a priori*. Because each Im7 protein contained the same optimized glycosylation acceptor sequence (GGNWTT), our initial expectation was that glycosylation efficiencies would correlate with the native secondary structure of the protein (PDB 1AYI) surrounding the glycosylation sites with unstructured and loop regions exhibiting greater modification rates compared to helical regions due to steric hindrance and inflexibility of the substrate. However, visual inspection yielded no clear relationship between the native structure of Im7 and glycosylation

efficiency, with the most heavily modified mutants corresponding to the placement of GGNWTT in helical positions (Figure 5C). Comparing glycosylation efficiencies to the relative solvent accessibility (RSA) values of each amino acid position using the native crystal structure of Im7 also yielded no clear correlation (Figure S5). Circular dichroism data collected from selected Im7-library members clearly showed disruption of the WT Im7 structure (Figure S6). These data suggest that the lack of correlation between glycosylation data and secondary structure may be due to disruption of the native Im7 structure by substitution of WT amino acids with the GGNWTT glycosylation sequence.

Because the native protein structure was unable to predict our observed glycosylation efficiencies, we asked whether the glycosylation pattern reflected differences in NGT selectivity for amino acids adjacent to the core GGNWTT acceptor sequence. To address this possibility, we synthesized a peptide library wherein the identity of amino acids at the X_{-3} and X_{+4} positions relative to the glycosylated asparagine was varied. We screened and analyzed these peptides by SAMDI-MS after treatment with NGT. However, we again found no clear relationship between NGT glycosylation of these peptides with observations at the protein level (Figure S7). These data indicate that adjacent amino acid preferences do not significantly influence the observed glycosylation pattern. This finding is also consistent with previous observations that the X_{-1} to X_{+2} positions are the most important in determining the preference of NGT for acceptor peptides.¹⁰ The fact that observed glycosylation efficiencies do not correlate with *a priori* predictions based on structural information on the native protein sequence or the preference of NGT for adjacent amino acid sequences indicates that a high-throughput method to empirically determine the glycosylation efficiency of each mutant holds great utility in understanding and optimizing protein glycosylation sites.

Generalizability to Other Proteins and Glycosylation Systems. To assess the generalizability of our combined CFPS-SAMDI-MS workflow, we extended the method to four additional proteins: Metreleptin (Leptin), Anakinra (IL1-receptor antagonist, IL1-RA), Filgrastim (granulocyte colony-stimulating factor, G-CSF), and Interferon- α . We chose these approved therapeutic proteins because the installation of glycans or glycan-conjugated PEG structures has been shown to increase the pharmacokinetics (including circulatory half-life) of small protein therapeutics,^{33–36} motivating a method to quickly determine glycosylation efficiency for these proteins. Here, we sought to simply demonstrate cell-free glycoprotein synthesis and SAMDI-MS detection of these proteins, rather than complete a comprehensive positional library scan as we had done with Im7. To achieve this goal, we designed linear DNA templates encoding each protein modified with N-terminal His- and GGNWTT glycosylation acceptor sequence tags and then expressed them by LET-CFPS. To simplify analysis, the known native N-X-S/T glycosylation site on IL1-RA was disabled by an N185Q mutation, leaving all proteins with a single available glycosylation site. After quantifying the CFPS yields of each protein (Figure S8), we performed IVG reactions of each therapeutic protein target with NGT and analyzed reaction products by SAMDI-MS. All protein targets exhibited a characteristic mass shift corresponding to the addition of a glucose by SAMDI-MS as shown in Figure 6. Glycosylation of each protein construct was additionally verified by UPLC-MS of trypsinized peptides (Figure S9).

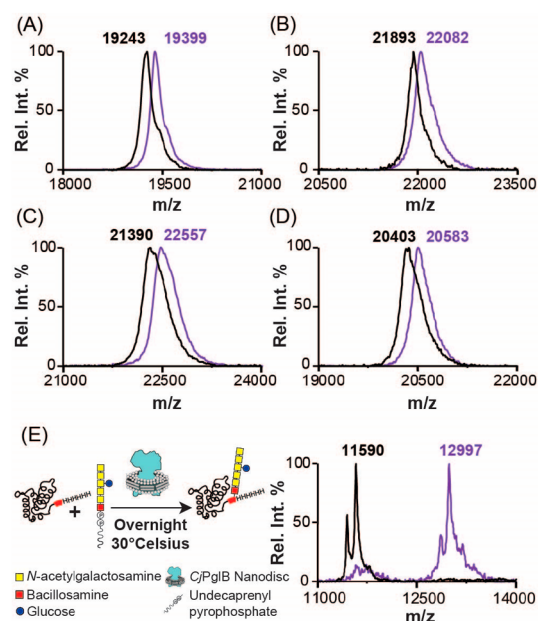


Figure 6. Generalization of SAMDI-MS workflow to analyze NGT glycosylation of therapeutically relevant proteins and oligosaccharyl-transferase glycosylation of Im7. Sequences based on (A) Metreleptin (Leptin), (B) Filgrastim (G-CSF), (C) Interferon- α , and (D) Anakinra (IL1-RA) were modified to include an N-terminal GGNWTT tag, synthesized using LET-CFPS, glycosylated by NGT, and analyzed by SAMDI-MS. (E) Left, glycosylation scheme of acceptor protein by CjPglB with a heptasaccharide from an undecaprenyl-linked oligosaccharide donor. Glycan symbols are according to the Consortium for Functional Glycomics nomenclature. Right, SAMDI-MS of Im7 construct containing a C-terminal DQNAT tag recognized and glycosylated by CjPglB. The difference between G_1 and G_0 in (E) corresponds closely to the addition of the heptasaccharide. ($\Delta m = 1407$ Da, expected is 1404 Da). All peaks are labeled with their average detected m/z values.

Consistent with our SAMDI-MS observations, we detected efficient glycosylation at the N-terminal GGNWTT glycosylation tag.

Finally, we demonstrated the versatility of our approach for detecting modification with larger, more complex glycans by analyzing the glycosylation of Im7 by a bacterial OST from *C. jejuni* (CjPglB) in Figure 6E. Unlike NGT, which transfers only one glucose residue, CjPglB transfers a prebuilt glycan *en bloc* from a lipid-linked oligosaccharide (LLO) donor to an acceptor protein sequence. We tested a version of Im7 containing a C-terminal His-tag and the glycosylation acceptor sequence DQNAT (Im7-DQNAT) which has been optimized for modification by CjPglB.³⁷ We used a recently reported method to synthesize CjPglB *in vitro* using a CFPS reaction supplemented with His-tagged ApoE-lipid nanodiscs.¹² We then combined this completed CFPS reaction with 4 μM of purified Im7-DQNAT and a preparation of extracted *E. coli* lipids containing the *C. jejuni* heptasaccharide LLO. These reactions were incubated overnight before analysis by SAMDI-MS. Spectra acquired from samples before and after treatment with CjPglB show nearly complete glycosylation of Im7 with the heptasaccharide. The difference between G_1 and G_0 corresponds closely to the expected mass of heptasaccharide (measured $\Delta m = 1407$ Da, expected $\Delta m = 1404$ Da), demonstrating the ability to accurately detect more complex glycan structures.

DISCUSSION

This paper demonstrates how SAMDI-MS can be used to perform high-throughput analysis of glycosylation on intact proteins generated using LET-CFPS. This work is significant because it offers an analytical method with significantly higher throughput compared to current mass-based methods for analyzing intact proteins. For example, approximately 1 day is required to analyze 1,000 samples using the strategy outlined in this work. In contrast, a conventional experiment using protein expression in living cells and UPLC-MS would require several weeks of cloning, bacterial transformation, and in vivo expression followed by 2 weeks for purification, dialysis, and preparation of each sample to complete the same experiment. We demonstrate the utility and practicality of this method by quantifying the relative glycosylation of an inserted glycosylation sequence beginning at every amino acid position of a model protein.

The requirement for large numbers of distinct proteins and the use of traditional assays utilizing immunological probes have prohibited the comprehensive study of the relationship between the efficiency of a PTM at a given site and the location of the site within a target protein.⁵ Our approach combining SAMDI-MS and LET-CFPS allowed us to perform such a study for NGT using a library of Im7 mutant proteins modified with the optimized glycosylation acceptor sequence GGNWTT at every amino acid position. We did not find clear rules to explain the propensity of each position for glycosylation. We believe this is due to conformational effects caused by the replacement of stretches of native amino acids in the protein with a glycosylation acceptor sequence. The difficulty in predicting intact protein glycosylation *a priori* using existing structural data or peptide mimics, as we illustrate in this study, reinforces the need for a method to synthesize and analyze large numbers of glycoproteins.

An important caveat to our approach is that the resolving power of SAMDI-MS is essentially identical to that of MALDI-TOF mass spectrometry. Therefore, the ability to detect and quantify a change in mass is dependent on the overall mass of the substrate. With our instrument, we have found that for proteins larger than 20 kDa it becomes increasingly difficult to resolve and quantify the addition of a single hexose. However, most proteins of interest are modified with larger polysaccharides, whose masses are more readily resolved by MALDI-TOF, as in the case of the *C. jejuni* heptasaccharide with a mass addition of ~1.4 kDa vs 162 Da. Therefore, we anticipate that the assay platform described in this work can be generalized to other glycosylation systems, beyond the challenging example of a single-hexose modification presented here. We also anticipate that our assay platform could be further complemented by on-chip proteolytic digestion methods to generate peptide mass fingerprints and assessment of positional glycosylation status as demonstrated by recent work from Flitsch and colleagues.^{38,39}

CONCLUSION

We have developed a combined cell-free protein synthesis and SAMDI-MS approach to rapidly generate and assess the glycosylation efficiency of intact glycoproteins. To our knowledge, this is the first example of seamlessly integrating CFPS of protein substrates from linear DNA templates and SAMDI-MS to study protein glycosylation, greatly increasing the throughput with which defined, synthetic glycoproteins can

be created and analyzed in a versatile fashion without the need for Western blotting or glycosylation-specific antibodies. We expect that this method will substantially increase the study of glycosylation of full-length proteins by providing the means to rapidly evaluate hundreds of protein substrates without the need for molecular cloning, utilization of antibodies, expression in living cells, or substantial quantities of material.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.9b04334>.

CFPS yields, amino acid preference study, methods to account for sinapic acid adduct peak overlap, fraction_(SA) for Im7 library, representative SAMDI-MS spectra from Figure 5A, UPLC-MS data, list of strains and DNA templates used to synthesize proteins, SAMDI-MS of CjPglB-loaded nanodiscs, CD data, annotation of SAMDI-MS masses, and DNA sequences generated in the study (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: m-jewett@northwestern.edu.

*E-mail: milan.mrksich@northwestern.edu.

ORCID

Matthew P. DeLisa: 0000-0003-3226-1566

Michael C. Jewett: 0000-0003-2948-6211

Milan Mrksich: 0000-0002-4964-796X

Author Contributions

#J.M.T. and W.K. contributed equally to this work. J.M.T. and W.K. conceived of the study, performed experiments, analyzed data, and wrote the manuscript. J.M.H. provided materials, assisted with PglB experiments, and performed circular dichroism studies. A.R. assisted with protein synthesis and glycosylation experiments. L.L. synthesized and screened peptide mimic libraries. M.P.D. assisted with conceiving and guiding the study. M.C.J. and M.M. conceived of the study, interpreted the data, and supervised the study. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge J. Stark, A. H. Thames, K. Duncker, and K. Warfel for helpful critiques and the sharing of reagents. We thank X. Zheng and M. Li for helpful discussions. We acknowledge support from the Defense Threat Reduction Agency (HDTRA1-15-10052/P00001), the International Institute for Nanotechnology (Ryan Graduate Fellowship), the Ruth L. Kirschtin National Research Service Award F30 CA196185, National Institutes of Environmental Health Sciences (T32 ES007059), and the National Science Foundation (MCB-1413563 and the Graduate Research Fellowship program under grant No. DGE-1324585). M.C.J. also acknowledges the Packard Foundation and the Dreyfus Teacher-Scholar Program. This work made use of the IMSERC core facility at Northwestern University, the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205).

■ REFERENCES

- (1) Nørregaard Jensen, O. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 33–41.
- (2) Ban, L.; Mrksich, M. *Angew. Chem., Int. Ed.* **2008**, *47* (18), 3396–3399.
- (3) Walsh, G.; Jefferis, R. *Nat. Biotechnol.* **2006**, *24* (10), 1241–1252.
- (4) Clausen, H.; Wandall, H. H.; Steentoft, C.; Stanley, P.; Schnaar, R. L. Glycosylation Engineering. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2015; pp 713–728.
- (5) Sterner, E.; Flanagan, N.; Gildersleeve, J. C. *ACS Chem. Biol.* **2016**, *11* (7), 1773–1783.
- (6) Mrksich, M. *ACS Nano* **2008**, *2* (1), 7–18.
- (7) Carlson, E. D.; Gan, R.; Hodgman, C. E.; Jewett, M. C. *Biotechnol. Adv.* **2012**, *30* (5), 1185–94.
- (8) Silverman, A. D.; Karim, A. S.; Jewett, M. C. *Nat. Rev. Genet.* **2019**, DOI: 10.1038/s41576-019-0186-3.
- (9) Dudley, Q. M.; Karim, A. S.; Jewett, M. C. *Biotechnol. J.* **2015**, *10* (1), 69–82.
- (10) Kightlinger, W.; Lin, L.; Rosztochy, M.; Li, W.; DeLisa, M. P.; Mrksich, M.; Jewett, M. C. *Nat. Chem. Biol.* **2018**, *14* (6), 627–635.
- (11) Jaroentomeechai, T.; Stark, J. C.; Natarajan, A.; Glasscock, C. J.; Yates, L. E.; Hsu, K. J.; Mrksich, M.; Jewett, M. C.; DeLisa, M. P. *Nat. Commun.* **2018**, *9* (1), 2686.
- (12) Schoborg, J. A.; Hershewe, J. M.; Stark, J. C.; Kightlinger, W.; Kath, J. E.; Jaroentomeechai, T.; Natarajan, A.; DeLisa, M. P.; Jewett, M. C. *Biotechnol. Bioeng.* **2018**, *115* (3), 739–750.
- (13) Guarino, C.; DeLisa, M. P. *Glycobiology* **2012**, *22* (5), 596–601.
- (14) O’Kane, P. T.; Dudley, Q. M.; McMillan, A. K.; Jewett, M. C.; Mrksich, M. *Science Advances* **2019**, *5* (6), eaaw9180.
- (15) Ban, L.; Pettit, N.; Li, L.; Stuparu, A. D.; Cai, L.; Chen, W.; Guan, W.; Han, W.; Wang, P. G.; Mrksich, M. *Nat. Chem. Biol.* **2012**, *8* (9), 769–73.
- (16) Gurard-Levin, Z. A.; Kim, J.; Mrksich, M. *ChemBioChem* **2009**, *10* (13), 2159–2161.
- (17) Silverman, J. M.; Imperiali, B. Bacterial N-Glycosylation Efficiency is Dependent on the Structural Context of Target Sequences. *J. Biol. Chem.* **2016**, *291*, 22001
- (18) Zhang, M.; Wu, S. C.; Zhou, W.; Xu, B. J. *Phys. Chem. B* **2012**, *116* (33), 9949–56.
- (19) Kwon, Y.-C.; Jewett, M. C. *Sci. Rep.* **2015**, *5*, 8663.
- (20) Anderson, L. L.; Berns, E. J.; Bugga, P.; George, A. L.; Mrksich, M. *Anal. Chem.* **2016**, *88* (17), 8604–8609.
- (21) Jewett, M. C.; Swartz, J. R. *Biotechnol. Prog.* **2004**, *20* (1), 102–109.
- (22) Jewett, M. C.; Calhoun, K. A.; Voloshin, A.; Wu, J. J.; Swartz, J. R. *Mol. Syst. Biol.* **2008**, *4*, 220.
- (23) Heinig, M.; Frishman, D. *Nucleic Acids Res.* **2004**, *32* (WebServer), W500–W502.
- (24) Tien, M. Z.; Meyer, A. G.; Sydykova, D. K.; Spielman, S. J.; Wilke, C. O. *PLoS One* **2013**, *8* (11), No. e80635.
- (25) Sigal, G. B.; Bamdad, C.; Barberis, A.; Strominger, J.; Whitesides, G. M. *Anal. Chem.* **1996**, *68* (3), 490–7.
- (26) Patrie, S. M.; Mrksich, M. *Anal. Chem.* **2007**, *79* (15), 5878–87.
- (27) Marin, V. L.; Bayburt, T. H.; Sligar, S. G.; Mrksich, M. *Angew. Chem., Int. Ed.* **2007**, *46* (46), 8796–8.
- (28) Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1990**, *62* (17), 1836–1840.
- (29) Beavis, R. C.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1989**, *3* (12), 432–5.
- (30) Schinn, S. M.; Broadbent, A.; Bradley, W. T.; Bundy, B. C. *New Biotechnol.* **2016**, *33* (4), 480–7.
- (31) Espah Borujeni, A.; Channarasappa, A. S.; Salis, H. M. *Nucleic Acids Res.* **2014**, *42* (4), 2646–2659.
- (32) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. *Nat. Biotechnol.* **2009**, *27* (10), 946–950.
- (33) Elliott, S.; Lorenzini, T.; Asher, S.; Aoki, K.; Brankow, D.; Buck, L.; Busse, L.; Chang, D.; Fuller, J.; Grant, J.; Hernday, N.; Hokum, M.; Hu, S.; Knudten, A.; Levin, N.; Komorowski, R.; Martin, F.; Navarro, R.; Osslund, T.; Rogers, G.; Rogers, N.; Trail, G.; Egrie, J. *Nat. Biotechnol.* **2003**, *21* (4), 414–21.
- (34) Keys, T. G.; Wetter, M.; Hang, I.; Rutschmann, C.; Russo, S.; Mally, M.; Steffen, M.; Zuppiger, M.; Müller, F.; Schneider, J.; Faridmoayer, A.; Lin, C.-w.; Aebi, M. *Metab. Eng.* **2017**, *44*, 293–301.
- (35) DeFrees, S.; Wang, Z. G.; Xing, R.; Scott, A. E.; Wang, J.; Zopf, D.; Gouty, D. L.; Sjöberg, E. R.; Panneerselvam, K.; Brinkman-Van der Linden, E. C.; Bayer, R. J.; Tarp, M. A.; Clausen, H. *Glycobiology* **2006**, *16* (9), 833–43.
- (36) Gregoriadis, G.; Jain, S.; Papaioannou, I.; Laing, P. *Int. J. Pharm.* **2005**, *300* (1–2), 125–130.
- (37) Chen, M. M.; Glover, K. J.; Imperiali, B. *Biochemistry* **2007**, *46* (18), 5579–85.
- (38) Gray, C. J.; Sanchez-Ruiz, A.; Sardzikova, I.; Ahmed, Y. A.; Miller, R. L.; Reyes Martinez, J. E.; Pallister, E.; Huang, K.; Both, P.; Hartmann, M.; Roberts, H. N.; Sardzik, R.; Mandal, S.; Turnbull, J. E.; Eysers, C. E.; Flitsch, S. L. *Anal. Chem.* **2017**, *89* (8), 4444–4451.
- (39) Beloqui, A.; Calvo, J.; Serna, S.; Yan, S.; Wilson, I. B.; Martin-Lomas, M.; Reichardt, N. C. *Angew. Chem., Int. Ed.* **2013**, *52* (29), 7477–81.