

NORTHWESTERN UNIVERSITY

High-Throughput Development of Glycosyltransferase Assays Using SAMDI-MS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemistry

By

José-Marc Techner

EVANSTON, ILLINOIS

June 2019

© Copyright by José-Marc Techner 2019

All Rights Reserved

## Abstract

More than half of proteins in humans are modified with carbohydrates in a process called glycosylation, yet this process remains poorly understood even though approximately 1% of the expressed human genome encodes biosynthetic machinery for glycosylation.

Unlike genomics and proteomics where high throughput tools are now routinely used to generate hypothesis-directed research there is a lack of tools that can probe and profile the activity of enzyme mediating protein glycosylation in a similar manner. A promising analytic platform combines self-assembled monolayer surfaces and mass spectrometry to detect and quantitative enzymatic changes on diverse substrates, including proteins and carbohydrates. This approach is called **Self-Assembled Monolayers coupled with Matrix-Assisted Desorption Ionization (SAMDI-MS)** mass spectrometry and can be used to measure glycosyltransferase activity. In this dissertation, I explore the use of SAMDI-MS to measure glycosyltransferases activities on protein and carbohydrate substrates from complex samples such as cellular lysates. Appropriate details and background are discussed in Chapters 1 and 2 regarding glycosyltransferases and SAMDI-MS, respectively. I demonstrate the ability to detect endogenous activity from MDA-MB-231 breast cancer cell lysates. This work is discussed in Chapter 3. I also explore the capability of combining Cell-free Protein Synthesis technology with SAMDI-MS to detect and quantitate relative glycosylation on histidine-tagged protein substrates. I apply this method to study glycosyltransferase activity on a library of recombinant proteins by the activity of a cytoplasmic N-glycosyltransferase. This work is discussed in chapter 4. Additional information regarding this study and its particulars are included in the appendix. This dissertation

demonstrates the potential of SAMDI-MS as an enabling profiling technology and details a rapid assay platform providing the throughput and means to rapidly evaluate hundreds of protein substrates without requiring cloning, utilization of antibodies, expression in living cells, or substantial quantities of material. The work presented addresses a technological gap in the measurement of glycosyltransferase activity relevant to efforts requiring large numbers of protein substrates and complement current efforts to design glycosylation sites on proteins of therapeutic interest.

## Acknowledgements

I aspire to be a physician-scientist because I want to be at the forefront of development of new medical knowledge and technology. I found the needed training to do this with Dr. Milan Mrksich with whom I had the pleasure of learning from these past six years at the forefront of analytical technology. Dr. Mrksich taught me a big picture perspective in addition to helping hone my scientific judgement, especially when it came to judicious selection of problems to address as scientist, since it is incredibly easy to be distracted by the sheer multitude of scientific questions to pursue. Dr. Mrksich taught me a significant amount with respect to proper presentation of data, discussion of research, and how to write competitive fellowship and funding grants. In addition to these ancillary skills I met many prominent visiting scientists to whom Dr. Mrksich introduced me as a graduate student. I fully intend to take the skills and experience I have obtained over these years with me into the clinic as I return to my medical training.

I am very grateful to the National Institutes of Health, specifically the National Cancer Institute which awarded me a Ruth Kirschstein F30 National Research Award Fellowship that has supported my work these past four years and will continue one more year as I return to the clinic. I am also grateful to the International Institute for Nanotechnology here at Northwestern University for the Ryan Fellowship that supported the first two years of my graduate work as well as the first couple of years in the Medical Scientist Training Program. Finally, I also acknowledge the Defense Threat Reduction Agency for their support in my third through sixth year of graduate work.

I want to thank the members of my thesis committee for agreeing to serve. Specifically, I want to thank Dr. Nathan Gianneschi for his agreement to be a last-minute addition to my thesis

committee and graciously allowing me to defend with him one-on-one when he was not available for the set thesis date. I also thank Dr. Hossein Ardehali, director of the Medical Scientist Training Program, for agreeing to serve on my committee despite his busy schedule. A special thank you goes to Dr. Michael Jewett. Dr. Jewett provided significant intellectual assistance and support to my work as I interacted with several of his graduate students for the past several years.

I want to thank and acknowledge the significant effort of my collaborator, Weston Kightlinger, who will be my co-author on an upcoming publication featuring our work combining SAMDI-MS and Cell-Free Protein Synthesis to study glycosyltransferases. I thank him for his patience as I progressed through multiple rounds of troubleshooting and optimization as well as for his assistance in tackling analytical problems together. I had many exciting and stimulating conversations with Weston and I expect we'll see more exciting work from him the future, notwithstanding his already large contributions to engineering glycosylation on recombinant proteins.

I also want to acknowledge Jewett Group members Jasmine Hershewe and the recently minted Dr. Jessica Stark whom I worked with on and off the past several years. I thank them for significant discussions as well as their contribution to my work, especially Jasmine who has significantly contributed to my upcoming publication, in addition to Weston. I also want to thank Dr. Jennifer Schoborg whom I worked with briefly before her graduation several years ago. Finally, I want to thank other members of the Jewett lab since I spent many hours in their lab, performing experiments side-by-side and I found my interactions with this group only ever pleasurable and helpful.

From my own lab I also want to note and acknowledge Dr. Liang Lin who I worked with intimately in my group and who has also contributed significantly to my protein SAMDI-MS work, both experimentally and through fruitful and helpful discussion. He graciously allowed me to use him as a sounding board for crackpot scientific ideas and performed the peptide study that supported Weston and mine's work on protein glycosylation. Thank you again for your help!

Though I did not collaborate with them I must give acknowledgement to members of the Mrksich lab, present and past, who helped shaped my scientific adventure as a PhD student. Specifically, Dr. Courtney Sobers welcomed me to the downtown campus and assisted in my early attempts to perform SAMDI experiments. Dr. Eric Berns also provided considerable intellectual support, especially for my early work exploring glycosyltransferase activity in cellular lysate using click-chemistry. Dr. Justin Modica helped me with conception of the synthetic work for preparing the NTA-ligand used in protein SAMDI and early troubleshooting with MALDI-TOF of protein samples, a small effort on his part with resounding effects on my confidence as a scientist. Newly minted Dr. Pradeep Bugga and Dr. Jennifer Grant also assisted in this capacity, as well as for moral and social support and occasional assistance with the MALDI-TOF mass spectrometer. There were times where no experiment or significant endeavor I was working on was successful for months and I needed a sympathetic ear. In this capacity I want to acknowledge Dr. Hsin-Yu Kuo who I spoke with on many occasions, especially when travelling back and forth on the Intercampus Shuttle. The small and cumulative interactions with other members of the Mrksich group were also important, especially in their feedback to my presentations and proposals as well as day-to-day interactions, to my scientific development. In this manner Alexei Ten, Patrick O'Kane, Lindsey Szymczak, Elamar Mouilly, Sarah Anderson

and Sarah Wood, Che-Fan (Jeffrey) Huang, Adam Pluchinsky, Dan Sykora, Kazi Helal, Dr. James Kornacki, and Dr. Maria Cabezas all assisted me.

Outside of my professional colleagues and friends are my personal friends from the MSTP. Specifically, I must acknowledge Sherry Lee from whom I received significant assistance in writing my F30 NRSA pre-doctoral fellowship. We also used each other as sounding boards and shared many a scientific discussion. I also received significant assistance from MSTP leadership past and present. Dr. Sarah Rice gave me critical feedback on my early NRSA drafts and Dr. Dane Chetkovich provided critical career advice. Dr. Jayms Peterson provided significant administrative support along with the rest of the MSTP staff over the years.

That last individuals I must thank and acknowledge are my mother and father, Dr. Norma Ojeda and Marc Techner. This last decade and half away from home would have been more difficult if it weren't for our weekly skype session where I could vent and discuss the trials and tribulations of the week. You helped me so much that an entire PhD dissertation would be insufficient to describe it.

## List of Abbreviations

SAMDI-MS	Self-assembled Monolayer Ionization Desorption Mass Spectrometry
LET	Linear Expression Template
2eaGlcNAc	2-ethylazido N-acetylglucosamine
GlcNAc	N-acetylglucosamine
CuCC	Copper-based Click Chemistry
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization- Time-of-Flight
UDP-Glc	Uridine-5'-diphospho-glucose
NDP	Nucleotide-Diphosphate
SAM	Self-assembled monolayer
GT	Glycosyltransferase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
PBS	Phosphate-Buffered Saline
TCEP	Tris(2-carboxyethyl) phosphine hydrochloride
THPTA	tris-hydroxypropyltriazolylmethylamine
ELISA	Enzyme-Linked Immunosorbent Assay
BSA	Bovine Serum Albumen
DMEM	Dulbecco's Modified Eagle Media
FBS	Fetal Bovine Serum
LLO	Lipid-Linked Oligosaccharide
OST	Oligosaccharyltransferase

NGT	N-Glycosyltransferase
PCR	Polymerase Chain Reaction
LET	Linear-Expression Template
CFPS	Cell-Free Protein Synthesis
NTA	Nitrilotriacetic Acid
EDTA	Ethylenediaminetetraacetic Acid
DNA	Deoxyribose Nucleic Acid
ESI	Electrospray Ionization
DESI	Desorption Electrospray Ionization
SA	Sinapic Acid
AUC	Area-under-curve
ATP	Adenosine Triphosphate
UTP	Uridine-5'-triphosphate
TFA	Trifluoroacetic Acid

## Table of Contents

<b>Copyright Page</b>	2
<b>Abstract</b>	3
<b>Acknowledgements</b>	5
<b>List of Abbreviations</b>	9
<b>Table of Contents</b>	11
<b>List of Figures</b>	13
<b>Chapter 1 Introduction: Glycosylation and Glycosyltransferases</b>	
1.1 Glycosylation is a Complex Biological Process	16
1.2 Health Relevance of Glycosylation	18
1.3 Glycosyltransferases in N-glycosylation	20
1.4 Glycosyltransferase Assays	26
<b>Chapter 2 Introduction: Mass Spectrometry and Self-Assembled Monolayers</b>	
2.1 Self-Assembled Monolayers	30
2.2 Spatiotemporal Control in Biology and Microfluidic Devices	31
2.3 Mass spectrometry of SAMs	35
2.4 SAMDI-MS and Glycosyltransferases	36
2.5 SAMDI-MS and Intact Proteins	38
<b>Chapter 3 Exploring SAMDI-MS to Profile Endogenous Galactosyltransferase Activity in MDA-MB-231 Breast Cancer Cell Lysates</b>	
3.1 Introduction	43
3.2 Results	46
3.3 Discussion and Conclusion	54

3.4	Experimental	57
<b>Chapter 4</b>	<b>High-Throughput Synthesis and Analysis of Intact Glycoproteins using SAMDI-MS</b>	
4.1	Introduction	61
3.2	Results	63
3.3	Discussion and Conclusion	82
3.4	Experimental	85
<b>Chapter 5</b>	<b>Summary, Final Thoughts and Future Implications</b>	104
<b>References</b>		109
<b>Appendices</b>		
	<b>Appendix A – Pilot Optimization: TM-SPRAYER Matrix Deposition and Protein SAMDI-MS</b>	125
	<b>Appendix B – Gene Sequences for Im7 Library and Engineered Therapeutic Protein Sequences</b>	128

## List of Figures

<b>Figure 1.1</b>	N-linked vs O-linked Glycosylation	16
<b>Figure 1.2</b>	Selected tryptic glycopeptides from rHuEPO	18
<b>Figure 1.1:</b>	2 Major models of N-glycosylation.	21
<b>Figure 2.2:</b>	Structure of Alkanethiolate and common functional headgroups.	31
<b>Figure 2.3:</b>	Screens of putative GTs performed on arrays of sugar acceptors	37
<b>Figure 2.4:</b>	Detection of antigens from human serum using monolayers.	39
<b>Figure 3.5:</b>	SAMDI-MS of glycosyltransferase activity	44
<b>Figure 3.6:</b>	Preparation of alkyne monolayer	46
<b>Figure 3.7:</b>	Immobilization of 2eaGlcNAc to alkyne-terminated monolayer	47
<b>Figure 3.8.</b>	SAMDI detection of GalTase Activity.	48
<b>Figure 3.9:</b>	GalTase vs GlcNAc concentration.	50
<b>Figure 3.10:</b>	GalTase activity by Solution vs Solid Phase Formats.	52
<b>Table 3.1:</b>	Cell line models of breast cancer.	53
<b>Figure 3.11:</b>	High-throughput profiling of Glycosyltransferase activities.	54
<b>Figure 4.12.</b>	Scheme for assaying glycosylation of Im7-6 by NGT.	64
<b>Figure 4.2.</b>	Quantitative Assessment of SAMDI-MS Protein Glycosylation Assay	66
<b>Figure 4.3.</b>	Performance of SAMDI-MS for CFPS lysate samples.	67
<b>Figure 4.1:</b>	Optimization of reaction time and PCR template amount.	69
<b>Figure 4.2:</b>	Measurement of synthesized yields of Im7 library using <sup>14</sup> C-leucine incorporation.	70
<b>Figure 4.6:</b>	Strategy for high-throughput synthesis and analysis of glycosylation on intact proteins.	71

<b>Figure 4.7.</b>	Rapid analysis of Im7 positional glycosylation site library synthesized and glycosylated in vitro	73
<b>Figure 4.8:</b>	Circular Dichroism spectra of selected Im7 mutants from the glycosylation screen.	74
<b>Figure 4.9:</b>	Study of NGT amino acid preferences adjacent to GGNWTT sequence.	76
<b>Figure 4.10.</b>	Examples of NGT glycosylation of proteins of therapeutic interest and oligosaccharide activity on Im7 analyzed by SAMDI-MS	77
<b>Figure 4.11:</b>	Peptide LC-MS verification of therapeutic protein glycosylation.	79
<b>Figure 4.12:</b>	<sup>1</sup> H NMR spectrum of NTA-SH ligand	86
<b>Figure 4.13:</b>	<sup>13</sup> C NMR spectrum of NTA-SH ligand.	87
<b>Figure 4.14:</b>	Sinapic Acid Adduct Peak Area % by SAMDI-MS	94
<b>Figure 4.15:</b>	SDS-PAGE of purified Im7 mutants for CD experiment.	98
<b>Table 4.1:</b>	Plasmid and Strains.	102
<b>Table 4.2:</b>	Primer Sequences for LET-CFPS.	103
<b>Figure A.13:</b>	Performance using TM-SPRAYER vs manual dried-droplet matrix deposition.	126

## Chapter 1

### Introduction: Glycosylation and Glycosyltransferases

Advancements in analytical microarray technology has driven a revolution in biology over the last twenty years. Development of high-throughput microarrays now allows capture of genetic and proteomic information of biological systems previously not attainable. Similar progress, however, has not been paralleled in glycobiology, the study of carbohydrate-driven processes in nature. This gap reflects the lack of tools that can directly measure activity of glycosyltransferases (GT), the enzymes that build complex carbohydrate structures in the body, and subsequently a full understanding of their relation to human disease.

#### **1.1 Glycosylation is a Complex Biological Process.**

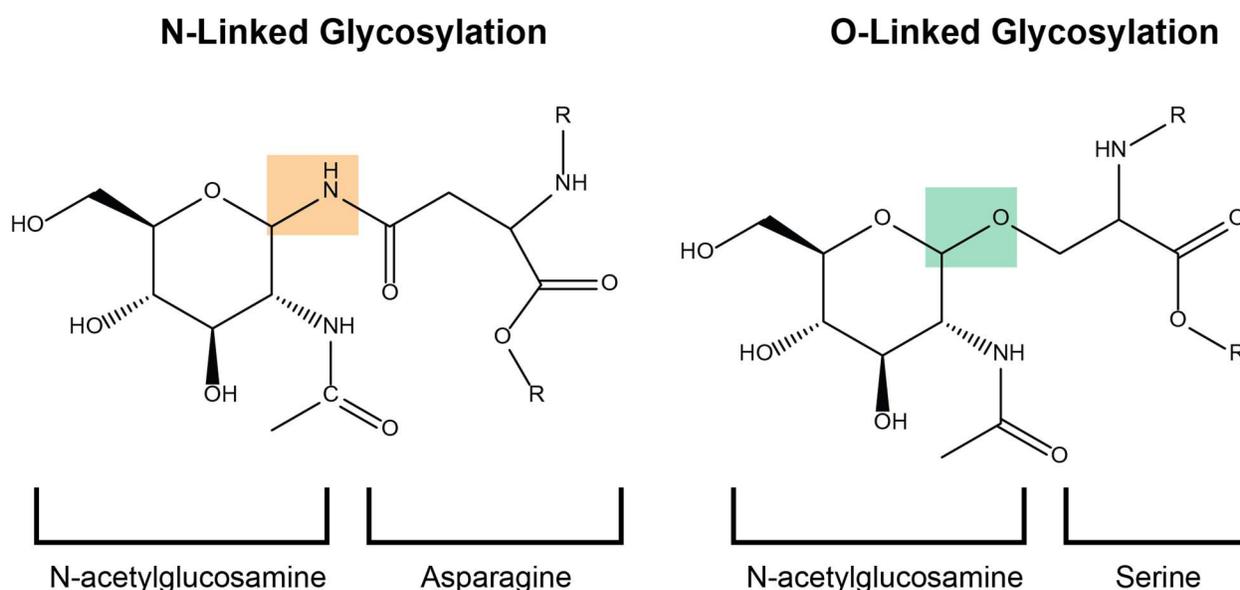
Glycosylation is the transfer of a carbohydrate (glycan) from a donor molecule to an acceptor molecule by a GT to yield complex oligosaccharide structures. The human body has more than 200 such enzymes, which comprises about 1% of the expressed genome.<sup>1</sup> Glycosylation is a modification found in all life. Glycosyltransferases (GT) are the enzymes that catalyze this process and are broadly grouped into two groups: Leloir and non-Leloir GTs. Leloir GTs utilize activated monosaccharide-nucleotides as donors while non-Leloir GTs use non-nucleotide donors such as lipid moieties.<sup>2</sup> Acceptors can in principle be any variety of macromolecule, including proteins, lipids, other carbohydrates, and nucleic acids. Glycosylation can occur with inversion or retention of the anomeric carbon, depending on the glycosyltransferase.<sup>3</sup> This dissertation primarily concerns glycosylation of proteins, although some work on carbohydrate

glycosylation is also presented. Note: the terms glycan, sugar, carbohydrate, and saccharide all used interchangeably to refer to similar class of chemical structures.

This dissertation only focuses on glycosyltransferases although many of the techniques and technologies discussed are also applicable to glycosidases, the enzymes which catalyze the reverse activity. As opposed to glycosylation, the related term glycation is the chemical modification of proteins by sugars driven by non-enzymatic processes, as exemplified by the formation of glycated proteins characteristic of advanced diabetes due to persistently high concentrations of glucose present in the blood stream.<sup>1</sup> Finally, N-glycosylation and not O-glycosylation is the primary focus of work in this work, although again many of the analytical approaches discussed are applicable to the study of O-glycosylation. This dissertation focuses primarily on glycosylation occurring on polypeptide substrates and although some work presented (in chapter 3) has also been performed on monosaccharide arrays, glycosylation focusing on other substrates will not be discussed.

Glycosylation on polypeptides is broadly classified into two categories based on the nature of the covalent linkages: N-glycosylation and O-glycosylation. N-glycosylation (N-linked glycan) is a sugar chain covalently linked to an asparagine via an amide bond to the anomeric carbon on the reducing end of a sugar chain, although recently arginine N-glycosylation has been discovered in enteropathogenic and enterohemorrhagic *Escherichia coli* (*E. coli.*) strains.<sup>2-4</sup> Analogously, O-glycosylation entails a covalent bond between the anomeric carbon of a sugar via the hydroxyl group of a threonine/serine, and tyrosine in some bacterial systems, to form an ether.<sup>5</sup> Examples of these linkages are illustrated in **Figure 1.1**. A glycoprotein/peptide is a protein/peptide which carries one or more glycans of either type. While glycosylation is broadly

classified as a posttranslational modification, N-glycosylation in eukaryotes and some bacterial systems is initiated primarily in a co-translational context in the endoplasmic reticulum as the polypeptide chain continues to be formed and prior to folding. Further elaboration of N-glycans then occurs in a posttranslational manner by trimming/addition of glycans, performed enzymatically by glycoside hydrolases (glycosidases) and glycosyltransferases, respectively.

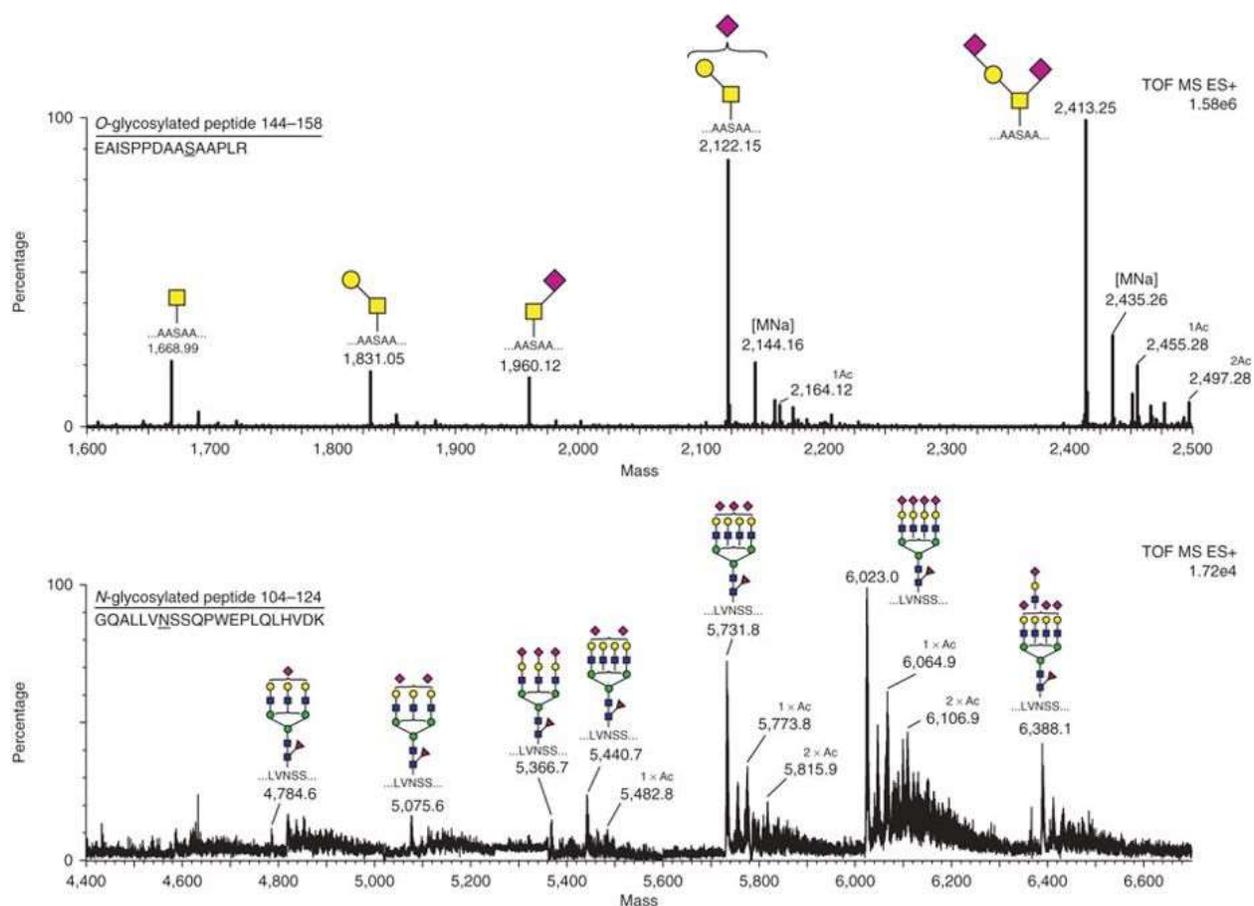


**Figure 1.14: N-linked vs O-linked Glycosylation.** In N-linked glycosylation the glycosyl bond is formed from the amide nitrogen on an Asparagine. In O-linked glycosylation an ether bond is formed between a hydroxyl amino acid threonine or serine, and occasionally tyrosine. Here this is illustrated using Serine. In orange and green shading, the glycosidic bond in N and O-linked glycosylation, respectively.

The roles of glycosylation in proteins are staggering and our collective understanding only scratch the surface. Complicating our understanding is the occurrence of micro- and microheterogeneity of glycosylation. Both describe how a single protein substrate can present in a biological population such as a single cell, with a broad range of distinct glycoproteins. Microheterogeneity describes the diversity of multiple possible glycan structures that can occupy

a single acceptor site in a population. An example of this microheterogeneity can be seen with mass spectra derived from tryptic analysis of a recombinant human erythropoietin (**Figure 1.2**).<sup>6</sup>

Macroheterogeneity describes the diversity in glycosylation occupancy on a protein substrate containing more than one possible glycosylation site. The generation of glycoproteins is a non-template driven process that arises from complex interaction of metabolic and enzymatic pathways influenced by many environmental and genetic factors.<sup>7</sup> The functional underpinning for why this heterogeneity exists in nature is poorly understood although recent work suggests roles in regulating immune recognition and microbiome interactions.<sup>8,9</sup> As a consequence of this heterogeneity most biomanufactured proteins meant for therapeutic applications exist as a complex mixture rather than as a single molecule, greatly complicating their clinical use. Controlling this heterogeneity, particularly with respect to site occupancy, requires significantly greater understanding of the underpinning biophysical processes and requires development of enabling technology to assist this research.



**Figure 1.15: Selected tryptic glycopeptides from rHuEPO.** The glycopeptide approach gives valuable data on the glycan composition on a particular site; structural isomers of glycans with similar composition were identified by analyzing the released glycans. The major N- and O-glycan structures identified after analysis of the released glycans are shown on the glycopeptides. Upper spectrum: The tri- and tetrasaccharide containing one and two sialic acids, respectively, are the major glycoforms found on the O-glycosylation site S153. Lower spectrum: N-glycosylation site N110 contains a mixture of different glycoforms containing up to four sialic acids, providing evidence that sialylated glycopeptides can be detected in ESI-positive mode. Adapted by permission of Springer Nature, *Nature Protocols*. Determination of site-specific glycan heterogeneity on glycoproteins. 7 (9), 1285-1298 Kolarich, D; Jensen, A. P.; Altmann, F.; Packer, N. H., Copyright 2012.

## 1.2 Health Relevance of Glycosylation.

Given its ubiquity in biological processes it is not surprising that glycosylation has many implications to health and disease. Changes in glycosylation often signal the progression of many diseases. For example, glycosylation changes of the HIV viral envelope protein modulate viral infectivity and detection by the host immune system.<sup>4</sup> Glycans are active in many cellular

processes including mediating cell-cell adhesions, determining blood type, and cancer metastasis.<sup>2</sup> GTs in Gram-negative bacteria produce lipo-oligosaccharide, an endotoxin, that can lead to a deadly autoinflammatory reaction known as sepsis.<sup>4</sup> GTs also relate to cancer development. For example, changes in glycan structures of cell membranes occur early during a cell's malignant transformation.<sup>5-8</sup> Genetic studies reveal these changes occur early in carcinogenesis and differ between different subtypes of a cancer.<sup>9</sup> Studies also note that the set of GTs that initiate glycosylation of cell-membrane proteins as well as cellular machinery that encode the synthesis of donors both become upregulated in breast cancer.<sup>10</sup> These attributes hint at potential biomarkers. For example, the Lewis family of glycan structures have been identified as epitopes in several different cancers not present in normal tissue.<sup>11-13</sup> Cancer vaccine and therapeutic antibody development strategies actively target such glycan epitopes.<sup>14,15</sup> That body of research illustrates why development of assays that can measure GT activity directly in cell lysate would greatly benefit current biomedical research.

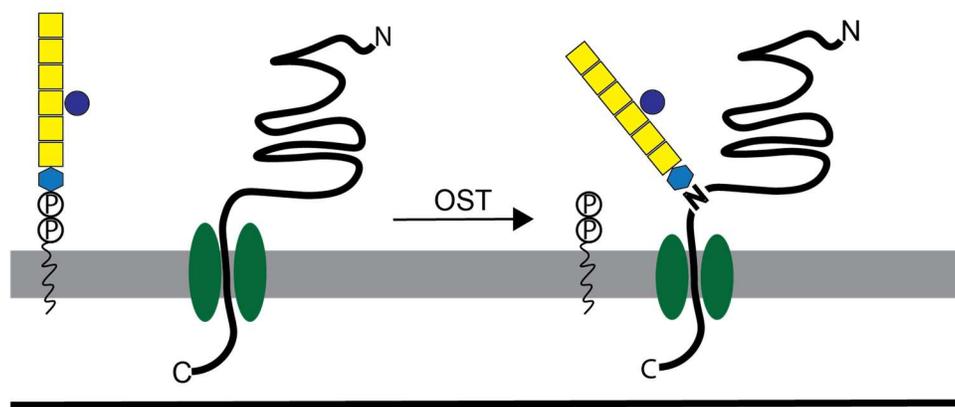
The presence of glycans on proteins has profound effects on their function and is of paramount importance to biopharmaceutical efforts. The composition of glycans on therapeutic proteins can induce deadly immunogenic responses, such as in the case of cetuximab hypersensitivity.<sup>10</sup> Incorporation of glycosylation can be used to engineer enhanced protein stability and *in vivo* circulatory half-life, for example by reducing proteolytic degradation by endogenous proteases.<sup>11</sup> Finally, glycosylation is essential for the function of many antibody therapeutics by modulating which cell-surface receptors they interact with.<sup>12, 13</sup> 40% of clinical products in development are biopharmaceuticals, with the majority of these protein based.<sup>14</sup> In

2013, 11 of the top 20 best-selling drugs were glycoproteins.<sup>15</sup> Consequently methods that enable the study of glycosylation on proteins is of paramount scientific importance.

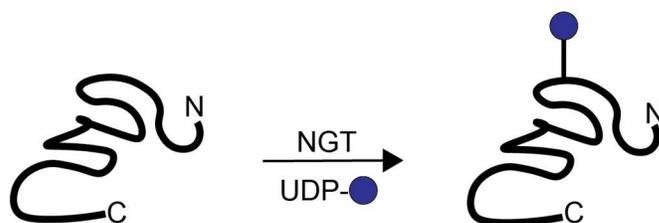
### 1.3 Glycosyltransferases in N-glycosylation

Initially considered a feature of eukaryotic organisms only, it is now widely recognized that N-protein glycosylation is found in all domains of life through a conserved pathway mediated by oligosaccharyltransferase (OST), which initiates the transfer of a block of pre-built glycans from a lipid-linked carrier to an asparagine residue.<sup>16, 17</sup> The OST utilizes oligosaccharides pre-built by glycosyltransferases onto a lipid-carrier (LLO) instead of nucleotide-donors. The oligosaccharide is catalytically transferred by OST en-bloc, the entire pre-built glycan, onto a polypeptide acceptor site minimally comprised of an N-X-S/T motif, where an asparagine is followed by any amino acid except proline followed by a hydroxyl amino acid. This has been the paradigm of initiation of N-glycosylation until 2003 when a novel enzyme class of cytoplasmic N-glycosyltransferases were discovered.<sup>18</sup> These enzymes catalyze the transfer of single monosaccharides from common nucleotide-donor molecules to an asparagine residue without using lipid-linked oligosaccharide donors; in a case of convergent evolution also utilizing the N-X-S/T motif for recognition of potential glycosylation sites.<sup>19</sup> The most well studied bacterial enzyme models for the traditional OST and newly characterized N-glycosylation pathways, respectively, are *Campylobacter jejuni* PglB and ApNGT (NGT) from *Actinobacillus pleuropneumoniae*. These two enzymes are of intense interest by the glycoengineering community to understand and control the N-glycosylation of recombinant proteins and thus are worth discussing in more detail.

## Oligosaccharyltransferase



## Cytoplasmic N-Glycosyltransferase



**Figure 1.16: 2 Major models of N-glycosylation.** Top) N-glycosylation system found in *Campylobacter jejuni* using oligosaccharyltransferase PglB. Pre-built heptasaccharide on a diphosphoundecaprenyl lipid is utilized as the sugar donor by OST on the luminal side of the endoplasmic reticulum membrane. Glycosylation occurs in close proximity to the translocation complex with largely unfolded protein acceptor substrate. Bottom) Cytoplasmic N-glycosylation, as mediated by an N-glycosyltransferase, occurs on acceptor protein substrate posttranslationally and subsequent glycan complexity occurs through sequential glycosylation reactions. Nucleotide-activated donor UDP-GlcNAc is used as the initiating monosaccharide.

The first bacterial N-glycosylation system was discovered in *C. jejuni* 16 years ago, mediated by the single-subunit OST PglB.<sup>20-23</sup> PglB is located primarily in the periplasmic space. In *C. jejuni* a heptasaccharide is built on the cytoplasmic side of the inner membrane on undecaprenyl phosphate (Und-P) via series of glycosylation steps mediated by transmembrane glycosyltransferases.<sup>24, 25</sup> The glycan in the *C. jejuni* pathway is a conserved heptasaccharide with the structure GalNAc- $\alpha$ 1,4-GalNAc- $\alpha$ 1,4-[Glc- $\beta$ 1,3]-GalNAc- $\alpha$ 1,4-GalNAc- $\alpha$ 1,4-GalNAc-

$\alpha$ 1,3-Bac- $\beta$ 1, where Bac is 2,4-diacetamido-2,4,6-trideoxyglucopyranose.<sup>21</sup> The resulting LLO is then translocated across the inner membrane into the periplasmic space by an ATP-dependent flippase, PglK, before subsequent consumption in the initiation of N-glycosylation by OST onto an acceptor protein or peptide.<sup>26, 27</sup> PglB is homologous to the STT3 subunit from the eukaryotic OST complex. Unlike eukaryotic OST which exhibits primarily co-translational glycosylation activity, PglB exhibits additional significant capability transferring sugars post-translationally to flexible structures in folded proteins.<sup>28</sup> Furthermore the glycosylation sequon recognized is more stringent than in eukaryotic OST, with PglB recognizing extended *N*-glycosylation consensus sequence that contains an aspartic or glutamic acid at the -2 position (Asp/Glu-X<sub>1</sub>-Asn-X<sub>2</sub>-Ser/Thr, where X<sub>1</sub> and X<sub>2</sub> represent any amino acid except Proline).<sup>29</sup> PglB also retains hydrolase activity and releases the heptasaccharide as free oligosaccharide into the periplasmic space as a compensatory response to drops of osmotic pressure.<sup>30</sup> The PglB glycosylation system has since been found in all campylobacter species and in considerable numbers of related bacterial families, indicating this is a widespread system among bacteria.<sup>31-33</sup>

Eukaryotic OSTs have broad acceptor-site specificity, which permits glycosylation of asparagine residues in the context of a very short consensus sequon (N-X-S/T; where X is any residue except proline). Indeed, as previously mentioned PglB recognizes more specific sequon with a negatively charged amino acid (aspartic acid or glutamic acid) in the -2 position relative to the asparagine. A possible explanation for this sequence preference comes from the crystal structure of *C. lari* PglB (56% identical to *C. jejuni* PglB) in which a salt bridge between R331, a conserved residue in bacterial OSTs, and the -2 aspartic acid of a bound acceptor peptide is evident.<sup>34</sup> The salt bridge formed with R331 is absent in the eukaryotic complex as R331 is

replaced by D362 in yeast Stt3, which explains the lack of requirement of an aspartic acid in the eukaryotic sequon.<sup>35</sup> Furthermore, eukaryotic OSTs are typically comprised of multiple subunits of which the catalytic STT3 is only one.<sup>36</sup> Thus increased site selectivity is likely conserved in the single subunit bacterial OST whereas in eukaryotes site selectivity is determined by activity and interactions of other subunits.

The importance of sequon placement on resulting glycosylation has been illustrated in important work focusing on N-linked protein glycosylation by oligosaccharyltransferase (OST).<sup>37, 38</sup> Bañó-Polo and others found that proximity of a sequon to the C-terminus and transmembrane regions resulted in substantially lower glycosylation efficiency, attributed to more rapid passage of the sequon past the (OST) active site following chain termination.<sup>39</sup> Other work has largely focused on differences in sequon composition rather than sequon position on glycosylation efficiency.<sup>40-42</sup> However, these studies are limited in that they are typically performed in an *in vivo* context with glycosylation largely occurring co-translationally. One notable exception is recent work by Silverman and Imperiali who assessed glycosylation efficiencies of several positional protein variants by *Campylobacter Jejuni* OST PglB *in vitro* vs *in vivo* and found that positions associated with increased surface accessibility and lower thermostability positively correlated with the degree of glycosylation observed.<sup>43</sup> These findings are consistent with the Aebi group whose work suggest that PglB requires glycosylation sites to be located in regions with high flexibility.<sup>28, 34</sup> However these studies have all relied on western blotting or radiation labelling to monitor glycosylation which are low-throughput and cumbersome strategies and explain the lack of more extensive *in vitro* studies.

Cytoplasmic N-glycosyltransferases (NGTs) were first discovered in *Haemophilus influenzae* to be responsible for the single-monosaccharide glycosylation of endogenous proteins.<sup>18, 19, 44</sup> Homologous NGTs have since been found in other bacterial species including in *Actinobacillus pleuropneumonia*, ApNGT, which is the best characterized NGT currently.<sup>45, 46</sup> This system is characterized by sequential building of glycans beginning with the initial transfer of a monosaccharide to an acceptor asparagine residue. To illustrate, ApNGT in its endogenous system glycosylates significant portions of adhesin proteins on the extracellular membrane that are then further extended by other glycosyltransferases.<sup>42</sup> ApNGT catalyzes a  $\beta$ 1-N-glycosidic bond between the amino group of asparagine residues and glucose or galactose, is an inverting glycosyltransferase, and has had its crystal structure determined.<sup>46, 47</sup> These studies show that ApNGT can also utilize UDP-Galactose and UDP-Xylose but not substituted hexoses (GlcA, GlcNAc, GalNAc and Neu5Ac).<sup>42</sup> This wide intrinsic donor tolerance highlights its potential as starting substrate for protein engineering efforts aimed at expanding acceptor tolerance. In addition to donor specificity, ApNGT like OST, recognizes the N-X<sub>1</sub>-T/S (where X is not proline) motif for glycosylation.<sup>46</sup> Additional substrate specificity studies have revealed that charged residues in proximity to the asparagine residue are disfavored and confirmed that a hydroxyl amino acid is required at the X<sub>+2</sub> downstream from the asparagine for glycosylation to occur.<sup>42, 48, 49</sup> These studies in aggregate reveal that despite the broad selectivity for glycosylation sites, efficiency of glycosylation still varies considerably as a result of the residues immediately surrounding the NXT site.

More recently, ApNGT has been a candidate for engineering glycosylation sites on substrate proteins using a sequential chemoenzymatic format. For example, ApNGT has been

employed for synthesis of polypeptides bearing defined complex oligosaccharides by coupling with an endo-glycosidase (ENGase)-catalyzed transglycosylation.<sup>50</sup> In this case, the initial addition of glucose to the acceptor protein was sufficient for ENGase to transfer the oligosaccharide from a synthetic oxazoline donor. The Glc-Asn linkage demonstrated greater resistance to hydrolysis and thus was proposed to be able to improve the *in vivo* stability of glycopeptides/proteins. Other work has instead focused on using mutagenesis to engineer ApNGT to expand the donor monosaccharide tolerance to include GlcN, with subsequent acetylation to generate GlcNAc on a polypeptide acceptor.<sup>41</sup> Generating this is considered essential as GlcNAc is the linking sugar to asparagine in human glycoforms and the controlled synthesis of human glycoforms is a top priority for pharmaceutical applications.

These previous studies focused on primarily of peptide acceptors with minimal exploration of ApNGT specificity in a protein context, other than validation of activity with a glycosylation site located in a flexible region of a substrate protein. However, there is some information that can be gleaned from its natural substrates, autotransporter adhesins. HMW1 is an adhesin protein used by bacteria to attach to epithelial tissue to aid bacterial colonization. HMW1 adhesin is modified at over 30 asparagine residues by an ApNGT homolog, HMW1C.<sup>44</sup> This extensive glycosylation prevents premature degradation of HMW1 during the process of secretion and promotes tethering of HMW1 to the cell surface.<sup>18</sup> ApNGT shares similar propensity for other autotransporter adhesins including those found in other bacterial species.<sup>42</sup> Autotransporter adhesins typically exhibit slow folding kinetics and thus are present as unfolded substrates for extended periods of time in the cytoplasm, where ApNGT resides.<sup>51, 52</sup> These data thus suggest that ApNGT exhibits general preference for unfolded substrates. In the case of

folded protein substrates, this would be exhibited by higher glycosylation efficiency at sites which are flexible and unstructured versus regions with extensive secondary structure. As with PglB, studies focused on ApNGT activity primarily utilize western blotting or labelled fluorescence peptides as substrates, with the exception of the Mrksich group which used label-free mass spectrometry as their chosen assay technique.<sup>48, 53-55</sup>

#### 1.4 Glycosyltransferase Assays

Radiometric methods are the oldest and classic glycosyltransferase assays which utilize isotopic labelling and detection strategies. They have high sensitivity which makes them compatible with use of lysates which typically contain low quantities of glycosyltransferases.<sup>56</sup> Generally, radiometric assays are based on the detection of transfer of an isotopically labelled sugar donor to an acceptor molecule after a separation step. The most common methods to do this utilize some form of chromatography or solid-phase extraction as the separation method.<sup>57, 58</sup> The detection step is typically scintillation counting but can also take the form of autoradiographs, in this case using isotopically labelled protein acceptors. Scintillation counting is quantitative whereas in the latter mass shifts are measured by densitometry which is semi-quantitative. A constraint with radiometric approaches, however, is limited availability and cost of radiolabeled donor. The use of radiation as the detection format also presents significant challenges for high-throughput analysis applications given the significant cost associated with disposal of radioactive waste in the research laboratory environment. A variation of the radiometric assay, called a scintillation proximity assay, has been developed to circumvent some of these limitations.<sup>59, 60</sup> In this format, acceptors of interest are immobilized onto a microsphere

bead pre-impregnated with scintillation molecule. Following a successful glycosyltransferase reaction, radiolabeled donor sugar is transferred to the acceptor near the scintillation molecule, resulting in the emission of light from the scintillation-sensitive bead itself. In this way unreacted donor contributes little background and no separation steps are required in order to quantify reaction activity.

The need for GT assays without radiation has spawned the development of a whole array of GT assay formats. A popular strategy targets the formation of the nucleotide/nucleoside of sugar donors using coupled enzyme systems. Gosselin et.al developed a spectrophotometric assay which could monitor the formation of nucleoside-diphosphate (NDP) released following a successful glycosylation reaction via coupled lactate dehydrogenase/ pyruvate kinase system.<sup>61</sup> In this example, a phosphate group is transferred from phosphoenolpyruvate to NDP by pyruvate kinase, releasing pyruvate which in turn is consumed to oxidize NADH and detected by a spectrophotometer. Alternatively, pyruvate can be coupled to a different enzymatic system to enable colorimetric detection instead.<sup>62</sup> In a similar approach, Wu and others instead coupled phosphatase digestion of NDP with a malachite green chemisensor to detect levels of inorganic phosphate released upon NDP consumption.<sup>63</sup> More recently, a distinct coupled enzyme system is instead linked to luciferase activity in a commercial variant, UDP-Glo, which utilizes luminescence instead of spectrophotometric detection format.<sup>64</sup>

A similar approach to targeting NDP formation instead focuses on the change of pH associated with the release of NDP, which includes the concomitant release of 1 equivalent of proton. In turn, this pH change can be detected through the use of standard pH indicators phenol

red or bromothymol blue.<sup>65, 66</sup> Both formats were shown to characterize multiple types of glycosyltransferases, however, both assays also require purified reagents and low amounts of buffer to prevent interference with the assay format.

More recent technological variations use fluorescence detection platforms. The Hamachi group and others groups have developed a chemisensors that selectively bind NDP over sugar-nucleotides and subsequently monitoring of fluorescence changes leads to quantitative measurements of glycosylation activity.<sup>67-70</sup> The commercial “Transcreener” assay platform developed by BellBrook Laboratories utilizes FRET to detect the displacement of a fluorescent molecule on an NDP-binding antibody to monitor GT activity.<sup>71</sup> Aside from the indirect detection approach of all these methods, the high possibility of background interference from other metabolic processes prohibits the use of complex samples such as cell lysates from being utilized in these assays without more extensive sample preparation beforehand, limiting throughput. An additional issue is that many GTs exhibit hydrolysis of NDP donors in the absence of acceptors, thereby complicating quantification in assays relying on UDP detection.<sup>72-75</sup> Another distinct disadvantage to these methods is that they apply only to Leloir-type GTs and are not applicable to other glycosylation systems that utilize lipid-linked donors such as oligosaccharyltransferase, the main N-glycosylation system utilized in eukaryotes.

Mass spectrometry offers a potential label-free approach to monitor glycosylation activity. The Davis group routinely utilizes ESI-MS to characterize GT activity profiles with respect to small molecule libraries.<sup>76, 77</sup> Most recently, Song and colleagues used MALDI-TOF to characterize multiple glycosylation sites on proteins after *in vitro* glycosylation.<sup>49</sup> However, the extensive sample preparation required in these examples discourages the use of mass

spectrometry for routine in-vitro glycosylation experiments. Instead mass spectrometry has found a place in proteomic approaches designed to characterize the heterogeneous glycan structures found on endogenous glycosylation sites in proteins and characterize site occupancy, rather than as a workhorse in vitro glycosylation assay.<sup>78</sup> In this capacity mass spectrometry has been exceptionally successful at being used to quantify and characterize a wide array of possible glycoforms. For example, a recent mass spectrometric technique by Christina Woo termed isotope-targeted glycoproteomics (IsoTaG), utilizes isotope recoding to characterize azidosugar-labeled glycopeptides bearing fully intact glycans.<sup>79</sup> Using metabolic labeling and affinity enrichment, 74 glycan structures on over 3500 intact glycopeptides were detected, findings comparable to released N-glycans identified by standard permethylation analysis but without the need of additional laborious chemical derivatization step.

SAMDI-MS is a label-free mass spectrometric platform using self-assembled monolayers that circumvents many of the analytical pitfalls associated with the existing glycosyltransferase assays discussed here and is further discussed in Chapter 2.

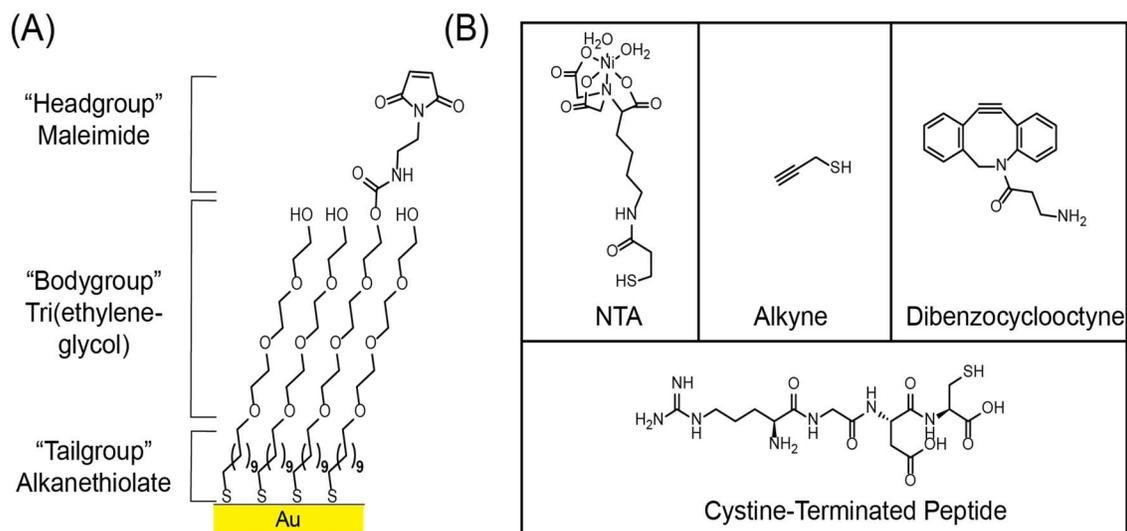
## Chapter 2

### Introduction: Mass Spectrometry and Self-Assembled Monolayers

#### 2.1 Self-Assembled Monolayers:

Self-assembled monolayers (SAMs) are organic assemblies formed by the spontaneous adsorption of molecules from solution or gas phase onto the surface of solids. The molecules that form SAMs typically are comprised of a “headgroup” which presents with affinity for the surface and a tail which can be chemically distinct or identical to the headgroup. The most well-studied SAMs are comprised of alkanethiols on noble and coinage metals, with gold in particular.<sup>80</sup> Thiol groups form covalent bonds with gold resulting in a single-molecule layer on gold surfaces.<sup>81</sup> Alkanethiols with long, saturated, unbranched alkyl chains form well-defined monolayers due to extensive van der Waals interactions resulting tightly-packed structures entirely covering the gold-surface.<sup>82</sup> These alkanethiol tail groups can be further modified with distinct functional groups to confer additional properties to SAMs. For example, incorporation chains of oligo (ethylene glycol) into the tail end of alkanethiols confers strong resistance to nonspecific protein adsorption.<sup>83-86</sup> This is an important factor as proteins have significant propensity to adsorb strongly to a variety of surfaces which complicate the use of generic surfaces for bioanalytical applications. Ethylene glycol prevents this phenomenon due to the tight binding of water molecules that effectively screen the monolayer from non-specific Van der Waals interactions with proteins. Furthermore, alkanethiols with different terminal head groups can be mixed in stoichiometric ratios to generate “mixed” SAMs with defined densities.<sup>80</sup> Incorporation of additional organic functional groups compatible with commonly utilized conjugation chemistry enables SAMs to incorporate biomolecules such as proteins and peptides (**Figure 2.1**).

Collectively, these features make alkanethiolate SAMs on gold ideal to develop novel bioanalytical platforms.



**Figure 2.17: Structure of Alkanethiolate and common functional headgroups.** A) Structure of self-assembled monolayer composed by mixed alkanethiolates. Mixed alkane thiolates with identical tail and body but different terminal headgroups, which confer chemical functionality, are mixed together in an ethanolic solution and exposed to gold overnight, resulting in the formation of a self-assembled monolayer on gold. B.) Examples of thiolated ligands that can be used to modify the maleimide terminated alkanethiolates, displayed on an already formed SAM, to confer functionality for different immobilization chemistries. Nitrilotriacetic acid confers functionality to chelate polyhistidine tagged peptides and proteins, alkyne confers reactivity with copper-based click reactions, dibenzylcyclooctyne allows for strain-promoted "copper-free" click reactions, and peptides containing natural thiols on cysteine residues can react directly with maleimide to conjugate peptide to a surface.

## 2.2 Use of SAMs as platforms for biosensors

Multiple analytical platforms have been developed that take advantage of these properties.

The electronic conductivity of gold underlying the SAMs renders it amenable to surface plasmon resonance (SPR) analysis. SPR measures changes in light reflectivity as a consequence of a molecular binding events have been coupled with SAMs to study interactions between various biomolecules such as proteins as well as whole cells.<sup>86-89</sup> SAMs have further been utilized in the context of biochip assays, wherein biomolecules are arrayed on SAMs and treated with samples

to identify protein-protein interactions and enzyme activities. These have typically been analyzed with use of fluorescent or radioactive probes.<sup>90</sup>

The most common strategy to prepare SAMs for bioanalytical sensor applications requires functionalization of the surface with a reactive intermediate that can then covalently or non-covalently react with the targeted ligand. Early work used carboxylic acid terminated SAMs that could then react with peptides or proteins using well established conjugation chemistry targeted reactive nucleophiles, such as amines, on amino side chains. Another well-established approach utilizes SAMs with maleimide functional groups to react with substrates presenting an accessible-cysteine side chain.<sup>91, 92</sup> In this scenario, maleimide groups primarily conjugate with thiols at physiological pH (6.5–7.5) because amines are typically protonated and not available for nucleophilic addition to maleimide. The possibility of thiol or cysteine insertion displacement of monolayer thiol-disulfides is also not a great concern because the maleimide-thiol conjugation reaction occurs much more rapidly than displacement reaction.<sup>93, 94</sup> However these covalent conjugation approaches often present a problem with the orientation of the bound proteins because partial loss of activity is a common occurrence using these ligation strategies. Furthermore these strategies for protein immobilization rely on purified protein samples to prevent competing reactions of other components in a sample.<sup>92</sup>

An alternative strategy uses well established chemistry to capture proteins modified with a polyhistidine tag with ion-incubated metal chelator functional groups. The two most common such chelators are nitrilotriacetic acid and iminodiacetic acid incubated with divalent cations such as Cu, Ni or Co, with the latter two being the most common utilized and nitrilotriacetic acid (NTA).<sup>95-97</sup> This is a now well established approach to prepare recombinant protein arrays on

chelator-terminated surfaces.<sup>98-100</sup> Because his tags can be easily incorporated into proteins recombinantly at multiple locations of a protein construct, this immobilization strategy allows for the opportunity for controlled display of the folded protein to the supporting SAM.<sup>100, 101</sup> However, a potential limitation of this method is the relatively low binding selectivity since several endogenous proteins have been recognized to be able to bind the metal ion, thus competing with the desired histidine-tagged protein.<sup>102</sup> Furthermore the strength of the binding interaction is also relatively weak at 1-1 $\mu$ M, although this can be effectively lowered by increasing the relative density of the NTA monolayer.<sup>103</sup> To address this several groups have utilized NTA-derivative chelators to overcome this limitation because the binding affinity of the histidine-tag for multivalent chelators increases by several order of magnitude as demonstrated by a systematic study of SAMs formed on gold terminated in mono-, bis-, and tris-NTA moieties by the Tampé group.<sup>103-105</sup>

Early work in the Mrksich group also innovated the use of SAMs for bioanalytical applications, including studying glycosylation, by exploring various immobilization strategies. Early work utilized monolayers presenting carbohydrates to characterize the binding specificities of lectins and glycosyltransferases by imaging monolayers with fluorescently labeled proteins.<sup>106-108</sup> These monolayers were prepared with several different strategies. The earliest strategy utilized mixed SAM solution with alkanethiolates terminated in carbohydrate moieties whereas later strategies utilized thiol and cyclopentadiene oligosaccharides to react with monolayers terminated with maleimide and benzoquinone, respectively. In these studies, the analysis of enzymatic and protein activity was primarily assessed via labelled fluorescence detection and label-free SPR. One example was Concanavalin A, a lectin, bound to mannose

moieties on a monolayer and detected by fluorescence.<sup>109</sup> Later work focused on conjugating fusion-proteins covalently linked using a suicide inhibitor approach for subsequent SPR analysis. In this approach an irreversible inhibitor of an enzyme is presented on the monolayer surface and the target protein is a recombinantly fused to that enzyme. For example a 40 kDa protein comprised of cutinase linked to the cell adhesion domain from fibronectin was conjugated to a monolayer presenting a phosphonate capture ligand, an irreversible inhibitor designed to inhibit cutinase.<sup>110</sup> Similar strategies were later utilized with HALO and SNAP tag ligands.<sup>111</sup> These proteins were then used for subsequent analytical applications.

### **2.3 Mass Spectrometry of SAMs**

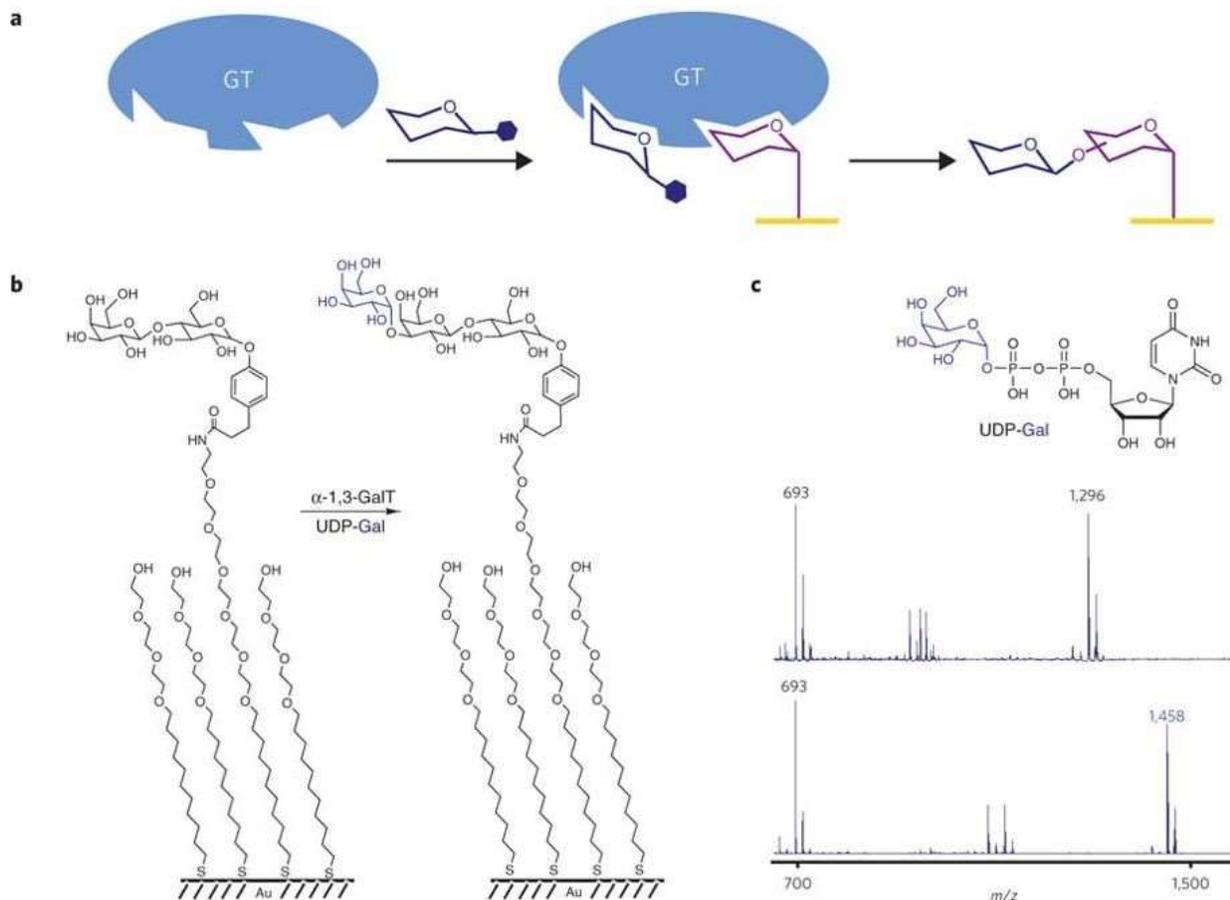
These early examples of SAM biosensors presented several drawbacks, however. The use of indirect detection strategies to assess the quantification products of enzymatic activity does not allow for discovery of unexpected activities. The need for a direct labeling of the reaction is an additional step to optimize and troubleshoot that can become a challenge in implementing high-throughput assays. This requirement is particularly limiting in experiments that seek to identify unanticipated biochemical activities since labelled strategies are often limited in scope of what they detect. Early work in characterizing SAMs with mass spectrometry indicated that this could be a viable label-free alternative. Using direct laser desorption ionization Wilkins and Hanley were able to observe mass species corresponding to intact alkanethiols, including free disulfides and gold-complex forms.<sup>112, 113</sup> The degree of fragmentation of parent ions and nature of adducts observed varied according to the laser fluence. Other work utilized secondary ion mass spectrometry (SIMS) to successfully observe components of SAMs. However, these early

studies utilized custom mass spectrometers and did not characterize the products resulting from interfacial reactions, such as enzymatic changes of substrates on the surface of a SAM. This was overcome with application of matrix-assisted laser desorption/ionization and time of flight mass spectroscopy (MALDI-TOF-MS) using a commercial instrument by Su and Mrksich. They demonstrated that MALDI could be used to observe chemical modification of functional groups as well as exchange of alkanethiolates in a monolayer with alkanethiols in a contacting solution.<sup>114</sup> This approach, named as **Self-Assembled Monolayer matrix-assisted laser Desorption Ionization Mass Spectrometry** or SAMDI-MS, was demonstrated to be useful as an enzymatic assay platform by monitoring the glycosylation of immobilized GlcNAc by a  $\beta$ 1,4-galactosyltransferase.<sup>109</sup>

#### **2.4. SAMDI-MS and Glycosyltransferases**

The SAMDI assay has since been further developed as a platform to study glycosyltransferase activity.<sup>115</sup> In one example, the donor specificity of recombinant  $\beta$ 1,3-N-acetylglucosaminyltransferase was assayed by mixing the enzyme with azido-modified lactose acceptor molecule and UDP-Galactose molecule in solution. In this format, the products of the glycosylation reaction were “pulled-down” using copper-click reaction as opposed to using solid-phase format, where the acceptor is immobilized to a SAM prior to a reaction. In this manner, SAMDI could obtain kinetic data for activity in solution. In another example, 84 different putative glycosyltransferases were screened using a solid-phase format with 60,000 unique combinations of acceptor, donor, glycosyltransferase, and buffer conditions (**Figure 2**). The acceptor library utilized was composed primarily of monosaccharides and several

oligosaccharide substrates. This effort led to the discovery of 4 new glycosyltransferases with novel activity, with additional kinetic characterization using the previously described “pull-down” format.<sup>116</sup> SAMDI has also been utilized to monitor glycosyltransferase activity on peptide substrates. In one particularly impressive example, 3,480 unique peptides and 13,903 unique reaction conditions were used to explore the substrate specificity of ApNGT and several ppGalNAcTransferases.<sup>48</sup> Others, the Fritch group in particular, have also utilized SAMDI to study glycosylation using peptide substrates.<sup>117-121</sup>

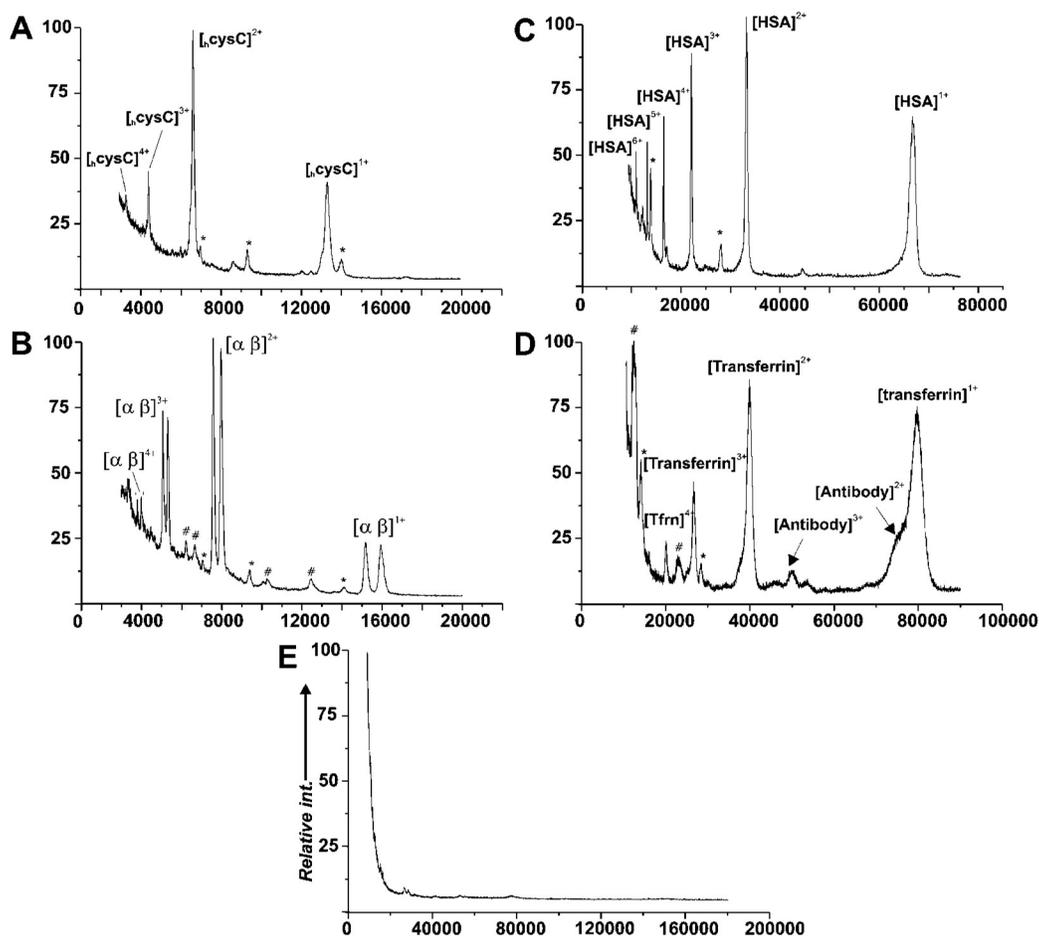


**Figure 2.18: Screens of putative GTs were performed on arrays of sugar acceptors.** (a) Glycosyltransferase (GT) assays were performed by applying solutions containing a GT and a sugar donor (blue) to regions of a self-assembled monolayer presenting carbohydrate acceptors (purple). SAMDI mass spectrometry was then used to analyze the monolayers to identify those combinations of GT, donor and acceptor that give a glycosylation reaction. (b) In one example, GGTA1 and the sugar donor UDP-Gal were applied to a lactose-terminated monolayer. (c) SAMDI spectra revealed that the lactose-substituted alkyl disulfide (at  $m/z$  1296) was glycosylated to give the trisaccharide (at  $m/z$  1458). Reprinted by permission from Springer Nature. Nature Chemical Biology. Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry. 8 (9), 769-73 Ban, L.; Pettit, N.; Li, L.; Stuparu, A. D.; Cai, L.; Chen, W.; Guan, W.; Han, W.; Wang, P. G.; Mrksich, M., Copyright 2012.

## 2.5. SAMDI-MS and Intact Proteins

The predominant use of SAMDI-MS has been monitoring enzymatic activity of small molecule substrates, however SAMDI-MS can also be used to analyze proteins captured or

immobilized to a SAM array, for example the presence of ConA lectin bound to an immobilized mannose.<sup>114</sup> In 2007, Patrie developed a SAMDI-MS based immunosensor to identify the proteolytic fragment of Cystatin C from cerebrospinal fluid samples.<sup>122</sup> In this work, a polyhistidine-tagged Protein A was captured using the triazacyclononane-derived ligand (Actacn), which is functionally similar to nitrilotriacetic acid but exhibits significantly tighter affinity with polyhistidine substrates.<sup>123, 124</sup> Protein A is a surface protein derived from *Staphylococcus aureus* that has strong affinity for the Fc portion of mammalian immunoglobulins to disrupt opsonization and phagocytosis.<sup>125</sup> Protein A is extensively used in preparation of immunosensors and antibody affinity purification methods.<sup>126, 127</sup> Patrie exploited this feature to immobilize antibodies with proper orientation towards contact sample solution and monitored the assembly of the immunosensor by SAMDI-MS as well as the quantitation of proteolytic activity. Patrie has since gone and further refined the technical features of the assay for enhanced precision and reproducibility.<sup>128, 129</sup> In another example, Marin utilized a similar immobilization strategy, this time using the more traditional NTA ligand, to immobilize rhodopsin-impregnated nanodiscs. These nanodiscs were comprised of histidine-tagged membrane scaffold protein (MBP) and the lipid 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC).<sup>130</sup> In this case, SAMDI-MS was used to monitor the light-dependent binding of transducing to rhodopsin.



**Figure 2.19: Detection of antigens from human serum using monolayers.** Monolayers present antibodies for cysC (A), hemoglobin (B), HSA (C), and transferrin (D). Each spectrum reveals prominent peaks for the intended analyte and lower intensity peaks for protein G (\*) and unidentified components derived from the stock antibody (#). A control experiment using an identical monolayer that was not treated with Ni<sup>2+</sup>, protein G, or antibody prior to incubation with serum revealed little binding of serum components (E). Adapted with permission from Patrie, S. M.; Mrksich, M., Self-assembled monolayers for MALDI-TOF mass spectrometry for immunoassays of human protein antigens. *Anal Chem* 2007, 79 (15), 5878-87. Copyright 2007. American Chemical Society.

Beyond the Mrksich group other groups have utilized SAMs as platforms to incubate proteins for later mass spectrometric analysis.<sup>131, 132</sup> A common form is termed Immuno-MALDI-TOF, which similarly to the previously described immunosensor uses surface chemistry to incubate antibodies to capture proteins of interest. In one variant using SAMs, SPR analysis is conducted to monitor the capture of antigenic protein followed by MALDI-TOF analysis.<sup>131</sup>

Other non-chip variants have been utilized such as using microbeads or nanoparticles mixed with sample followed by a wash step before an elution step for MALDI-TOF analysis.<sup>133-135</sup> This has been further explored in a multiplexed format wherein inflammatory markers C-reactive protein, serum amyloid A, and calprotectin and the kidney function marker cystatin C were quantitated from common samples.<sup>132</sup> A distinctive drawback with these approaches is the lower throughput versus the SAMDI-MS approach. Secondly, in several of these studies wash buffers required use of detergents to prevent non-specific binding, which is obviated in the case of SAMs which are intrinsically resistant to non-specific adsorption via integration of poly (ethylene glycol) groups.

The Mrksich group and others have extensively demonstrated the versatility of combining the use of self-assembled monolayers with mass spectrometry as a versatile biosensor platform generalizable to a large variety of assays. It is label free and the detection of mass changes allows for discovery of unanticipated activity. Furthermore, the use of alkanethiolate monolayers terminated in ethylene glycol allows for sample compatibility with complex lysate samples that otherwise would require extensive washing steps to prevent non-specific interactions with the surface. Despite its extensive use in characterizing and screening for novel glycosyltransferase activities, however, there remains a need for its application in intact glycoprotein analysis. This would enable addressing questions such as the significance of inserting glycosylation sites throughout the structure of a known protein. Existing immunolabeling approaches such as western blotting and ELISA are susceptible to variable antibody quality as well as sensitivity to changes in epitope, which are often comprised of specific secondary structure or sensitive to changes in posttranslational modification.<sup>136</sup> SAMDI-MS is not affected by any of these limitations, thus highlighting the need to extend SAMDI-MS assays

to folded protein substrates. Furthermore, the compatibility with complex lysates and recent work demonstrating its viability as an adequate platform for activity-based profiling of deacetylase activities opens the door to investigating a similar use for glycosyltransferase activity profiling.<sup>137, 138</sup> This is a potentially useful application as many disease processes, cancer foremost among them, is characterized by changes in glycosylation patterns of cell-surface glycoproteins.<sup>139-141</sup> In this dissertation I explore the use of SAMDI-MS in both of these directions.

## Chapter 3

### Exploring SAMDI-MS to Profile Endogenous Galactosyltransferase Activity in MDA-MB-231 Breast Cancer Cell Lysates.

#### 3.1 Introduction

Molecular profiling has demonstrated its usefulness by enabling the development of personalized medical strategies that improve clinical outcomes breast cancer, the most common cancer in women.<sup>142</sup> For example, targeted treatment is now given based on receptor expression status which has improved patient outcomes.<sup>143, 144</sup> Tumors that lack these markers, however, have no targeted therapy and typically have lower patient outcomes. Increased availability of genomic and proteomic technology continues to enable discovery of more subtypes of therapeutic interest.<sup>145</sup> Contrarily, the profiling of glycans, which consist of complex structures of oligosaccharides, is less prevalent. This is so even though evidence demonstrating that glycosylation is important in breast cancer and plays a functional role in determining cancer invasiveness. For example, ST6GalNAc I is  $\alpha$ 2,6-sialyltransferase that adds a sialic acid (NeuAc) to O-GlcNAc.<sup>5</sup> ST6GalNAc I was recently shown to be key to generation of the sTn epitope which is upregulated in breast cancer.<sup>146, 147</sup> Increased sTn is associated with a poor prognosis and aggressive phenotype.<sup>148-151</sup> Another glycosyltransferase, Gnt-V, is also upregulated in breast cancer and recently shown to promote an invasive phenotype.<sup>152, 153</sup>

These insights motivated glycoproteomic analysis of several established breast cancer cell lines that reveal increases in sialylation and high-mannose structures in invasive breast cancer cells.<sup>154-158</sup> Also increased were the presence of a branched triantennary glycan attributed to be produced by Gnt-V.<sup>157</sup> Another study linked CD44v4, an alternative spliced variant of

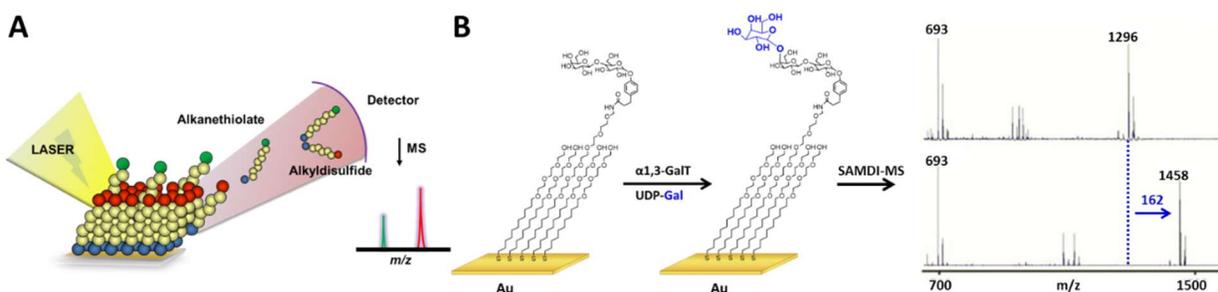
putative breast cancer stem cell marker CD44 adorned with sialyl-Lewis<sup>X</sup> moieties, to promote cell invasion by acting as a ligand for E-selectin.<sup>159</sup> In contrast to membrane bound proteins, cytosolic proteins from invasive breast cancer cells were found to have dramatic decreases in sialylation and fucosylation.<sup>154, 158</sup> Despite these studies it remains unclear **A)** Which glycosyltransferases are involved; **B)** How these enzymes are regulated; and **C)** Whether a specific combination of activity of these enzymes can effect an invasive phenotype.

Classic methods to study glycosyltransferase activity rely on either radiometric or colorimetric methods whose utility are generally limited to *in vitro* studies.<sup>160</sup> The current standard practice to study products of glycosylation relies on mass spectrometric analysis of glycan structures enzymatically released from conjugation to parent biomolecules in parallel with genetic analysis and lectin or antibody profiling.<sup>161</sup> These studies are limited by specificity of enzyme digestion and the antibodies and lectins used for profiling are limited in the range of specificities they exhibit. These methods are also low-throughput and require tedious protein purification. The limitations preclude the ability to easily screen for unanticipated novel activity relevant of mechanistic investigation or information that could then be utilized for chemoenzymatic synthesis of carbohydrates, a priority for the NIH.

Carbohydrate arrays are solid surfaces to which functionalized sugars are immobilized. The Mrksich and several other research groups first introduced this technology in 2002 utilizing a Diels-Alder reaction to link a cyclopentadienyl containing sugar to surface monolayers displaying benzoquinone groups.<sup>106, 107</sup> Microarrays are now utilized to characterize the specificity of carbohydrate-binding proteins which in turn can be used to profile the activities of GTs.<sup>162, 163</sup> However, arrays present limited applicability to high-throughput screening of

complex lysates because they fail to prevent non-specific protein adsorption or use fluorescently-labeled proteins to detect glycans. Furthermore, these profiling studies are limited by the range of specificities of said labeled proteins.

The Mrksich group has developed a powerful method that does not carry these limitations. This method uses self-assembled monolayers (SAMs) of alkanethiolates on gold.<sup>107, 116, 164, 165</sup> Alkanethiolate monolayers on gold are prepared where 10% of the chains present a modifiable functional group against a background of tri(ethylene glycol) chains. The former immobilize the substrate molecules and the latter prevent non-specific adsorption of proteins.<sup>91, 166</sup> SAMs present immobilized ligands in a homogeneous environment and allow for controlled uniform density.<sup>167</sup> SAMs are then analyzed using Matrix-Assisted Laser Desorption-Ionization mass spectrometry (SAMDI).<sup>109, 114, 168</sup> The Mrksich group has demonstrated the quantitative utility of this assay system in over twenty (20) papers with a range of biochemical activities and recent work demonstrates compatibility with cell lysates.

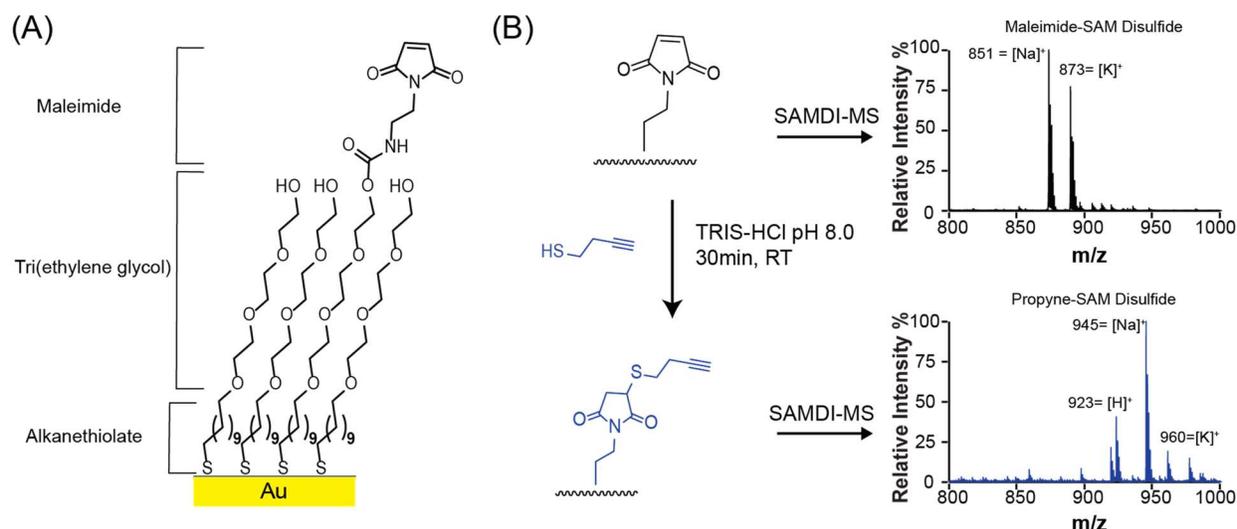


**Figure 3.20: SAMDI-MS of glycosyltransferase activity** (A) A laser desorbs and ionizes alkanethiolates that are then detected by mass spectrometry. (B) A lactose-substituted alkyl disulfide (at  $m/z$  1296) is reacted with a  $\alpha 1,3$ -galactosyltransferase and sugar donor UDP-Gal to generate trisaccharide (at  $m/z$  1458). Glycosylation can be identified by mass shifts of the mass spectrum before, and after, treatment with the enzyme. In this example the addition of galactose results in a mass shift of 162.

The SAMDI assay has been previously demonstrated to measure glycosyltransferase activities in high-throughput.<sup>115</sup> In one example, the donor specificity of recombinant  $\beta$ 1,3-N-acetylglucosaminyltransferase was assayed by mixing the enzyme with azido-modified lactose acceptor molecule and UDP-Galactose molecule in solution (Figure 1). In another example, 84 different putative glycosyltransferases were screened with 60,000 unique combinations of acceptor, donor, glycosyltransferase, and buffer conditions. This led to the discovery of 4 new glycosyltransferases.<sup>116</sup> Also demonstrated was a ‘pull-down’ assay where the enzyme, donor and acceptor were mixed in a buffer and allowed to react in the homogeneous phase – as opposed to the solid-phase, where the acceptor is immobilized to the surface prior to reaction. At the end of the reaction, a quenching reagent was added, and the reaction mixture applied to alkyne-terminated SAM, where the acceptor sugar immobilized to the monolayer. I further explore both formats for applications in profiling glycosyltransferase activities from a model breast cancer cell line, MDA-MD-231, using copper-click chemistry.

### 3.2 Results

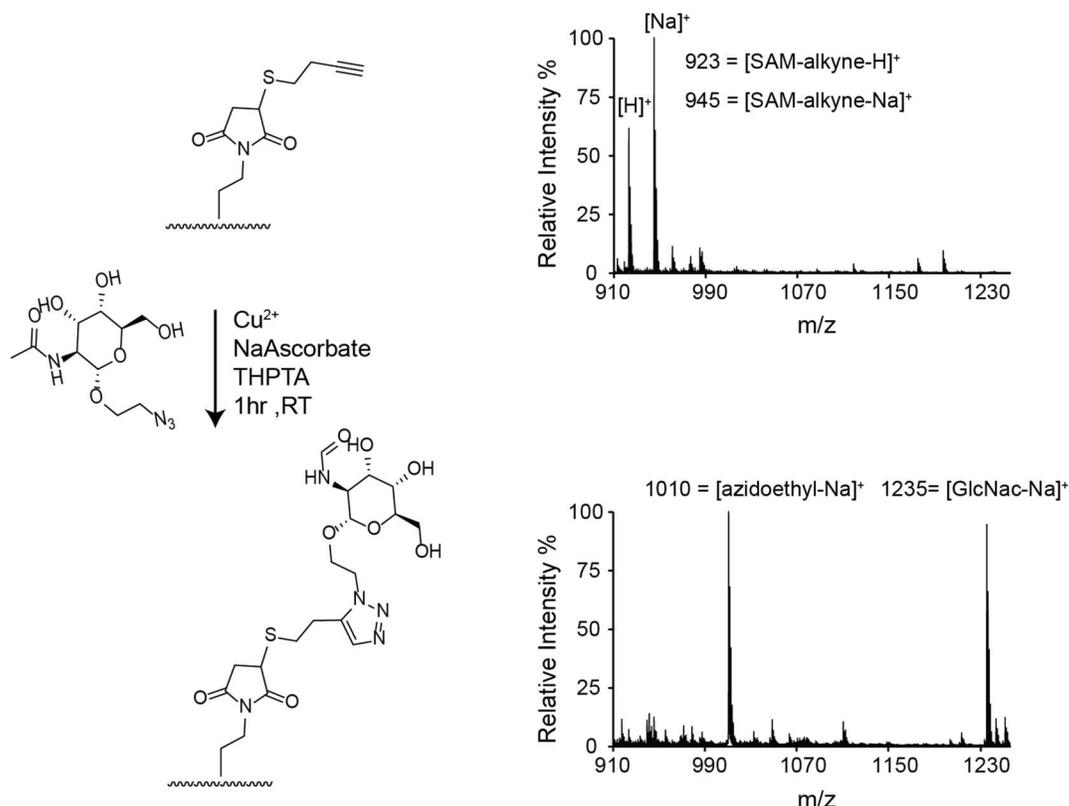
**Preparation of alkyne monolayers:** Copper-based click chemistry (CuCC) involves the reaction of azide molecule to an alkyne moiety catalyzed by a copper ion resulting in a cyclic product. This is also known as an azide-alkyne Huisgen cycloaddition. For SAMDI-MS this requires preparing an alkyne terminated SAM. To prepare the alkyne-SAM I first synthesized 2-propyne-1-thiol from propargyl-bromide and potassium thioacetate using a protocol adapted from Schuster et. al.<sup>169</sup> After synthesis I covalently immobilized the alkyne-thiol to a maleimide-terminated SAM via a Michael addition. This resulted in a SAM that displayed only the alkyne as all the maleimide was reacted, which was confirmed by mass spectrometry (**Figure 2**).



**Figure 3.21: Preparation of alkyne monolayer** A.) Schematic of alkanethiolate self-assembled monolayer prior to modification with 2-propyne-1-thiol. 10% of maleimide alkanethiolate stoichiometry is reflected in surface SAM, resulting in a 10% density of maleimide-terminated alkanethiolates on the gold surface B.) Treatment of maleimide-terminated SAM with 2-propyne-1-thiol at slightly alkaline pH allows for addition of thiol component in a Michael addition, resulting in full conversion of all maleimide terminated alkanethiolates to propyne. SAMDI-MS spectra detect the alkanethiolates in their mixed-disulfide forms (with ethylene-glycol alkanethiolates) in proton, sodium or potassium adducts, depending on species propensity to form said adducts. The mass difference corresponding to addition of 2-propyne-1-thiol (72 Da) is confirmed in the mass spectra (851 to 923).

**CuCC of 2-ethylazide N-acetylglucosamine onto alkyne monolayers:** I screened the suitability of the monolayer for click-chemistry for ideal conditions to perform the CuCC of 1-2-ethylazideN-acetylglucosamine (2eaGlcNAc) and the surface alkyne. The reaction cocktail consisted of copper sulfate, sodium ascorbate, and accelerant ligand tris-(benzyltriazolylmethyl)amine (THPTA). Sodium ascorbate reduces Cu(II) to generate the catalytic Cu(I) while THPTA prevents the generation of oxidative side-reactions that can oxidize the Cu(I) back to Cu(II), increasing reaction efficiency. After screening different ratios, I obtained full conversion of the alkyne peak to the 2eaGlcNAc conjugated peak using 1 mM

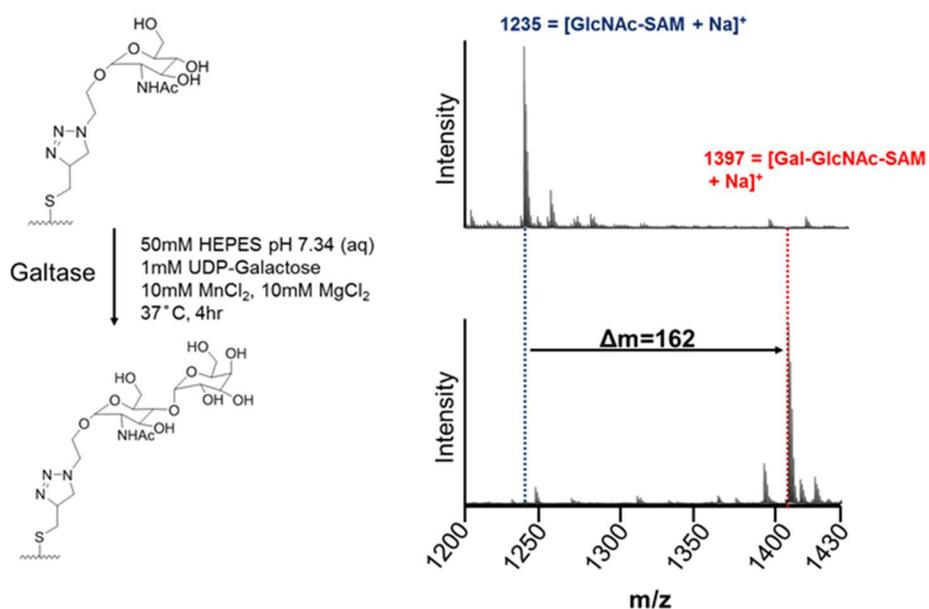
CuSO<sub>4</sub>, 10 mM NaAsc, and 4 mM THPTA after one hour of incubation. Longer incubations resulted in degradation of the surface with loss of the recognizable tri(ethylene glycol) disulfide peak ( $m/z = 693$ ) that forms the backbone of the monolayer. I attribute this to the reactivity of the copper moiety for alkanethiols.<sup>80</sup>



**Figure 3.22: Immobilization of 2eaGlcNAc to alkyne terminated monolayer.** Azide terminated GlcNAc allowed to react with alkyne terminated monolayer in copper-catalyzed reaction. Full conversion occurs after 1 hour. Mass shift corresponding to 290Da (945 to 1235) is consistent with the addition of the 2eaGlcNAc molecule via click reaction and detected as a sodium adduct. Carbohydrates are predominantly detected as Na adduct via MALDI-TOF MS.<sup>170</sup> Presence of azidoethyl contaminant in original GlcNAc stock solution also detected as a peak at 1010  $m/z$ .

**Measuring  $\beta$  1,4-galactosyltransferase with SAMDI-MS and CuCC:** Bovine  $\beta$ 1,4-galactosyltransferase I (GalTase) was chosen as a model enzyme to test the new click-enabled monolayer because it is well characterized in the literature and commercially available. This

enzyme also known as N-acetyllactosamine synthetase and displays acceptor specificity for N-acetylglucosamine (GlcNAc) to which it transfers galactose (Gal) in a 1-4 configuration of the glycosyl bond to form N-acetyllactosamine (LacNAc).<sup>171</sup> To confirm that our SAMDI-MS based assay could detect transfer of a galactose to the acceptor GlcNAc displayed on the surface, I treated GlcNAc terminated monolayers with GalTase for an extended period of time with excess UDP-Galactose. Upon SAMDI-MS, I confirmed to be able to fully convert the GlcNAc to its LacNAc form (**Figure 4**).



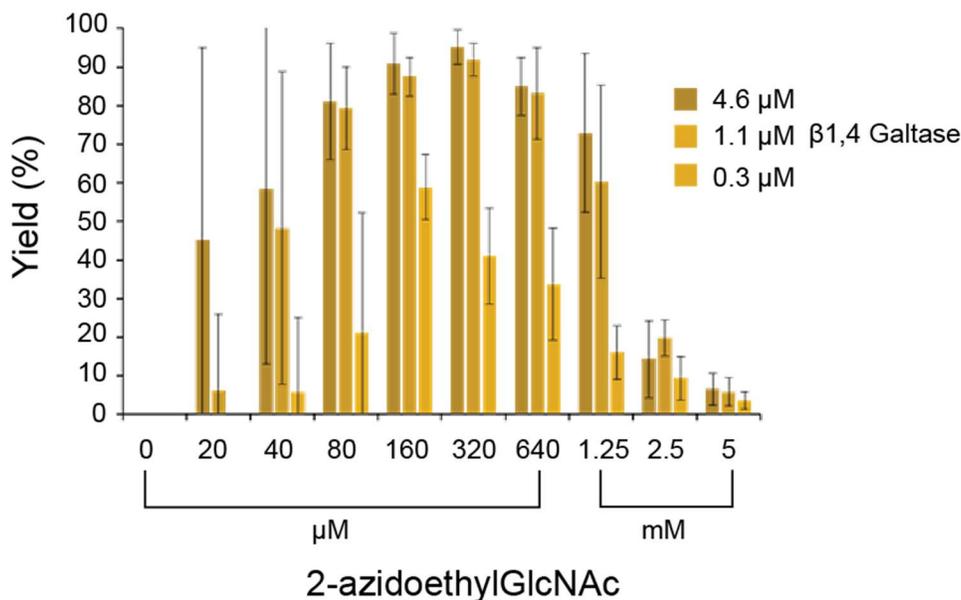
**Figure 3.23. SAMDI detection of GalTase Activity.** GalTase transfers a galactose from UDP-Gal to GlcNAc displayed on a SAM. This results in a mass difference of +162 which is subsequently detected on the mass spectrum (1235 to 1397 m/z). All peaks are detected as sodium adducts only.

To test the limits of the assay to different amounts of GalTase and substrate 2-aeGlcNAc was screened in the presence of 1mM UDP-galactose. These experiments were performed using a pull-down format: enzyme, donor and acceptor were allowed to react separately in a 96-well plate for 4 hours at 37 °C. I chose to perform glycosylation reactions in solution before

immobilization for SAMDI because endogenous galactosylation reactions involving MDA-MB-231 were previously performed in this manner based upon literature.<sup>172</sup> After glycosylation reactions in-vitro were completed and quenched with ethanol, I then added a CuCC cocktail to before transfer of 2 ul of the mixture to a surface displaying alkyne. After 1-hour incubation, surfaces were washed and dried and analyzed by mass spectrometry (**Figure 5**). Calculation of the product conversion used the following relationship by integrating the area under the curve (AUC) of the substrate and product peaks:

$$\% \text{ Galactosylation} = 100 * \frac{\text{AUC}_{1397}}{\text{AUC}_{1235} + \text{AUC}_{1397}} \quad (3.1)$$

This experiment showed several observations. At lower 2-aeGlcNAc levels there is an increase in the yield of the reaction with increasing 2-aeGlcNAc concentration. This trend is reversed at higher concentrations of 2-aeGlcNAc. This trend can be explained that at lower concentrations the rate of the enzymatic reaction depends on availability of the substrate. At higher concentration, GalTase becomes saturated and reaches maximal velocity.

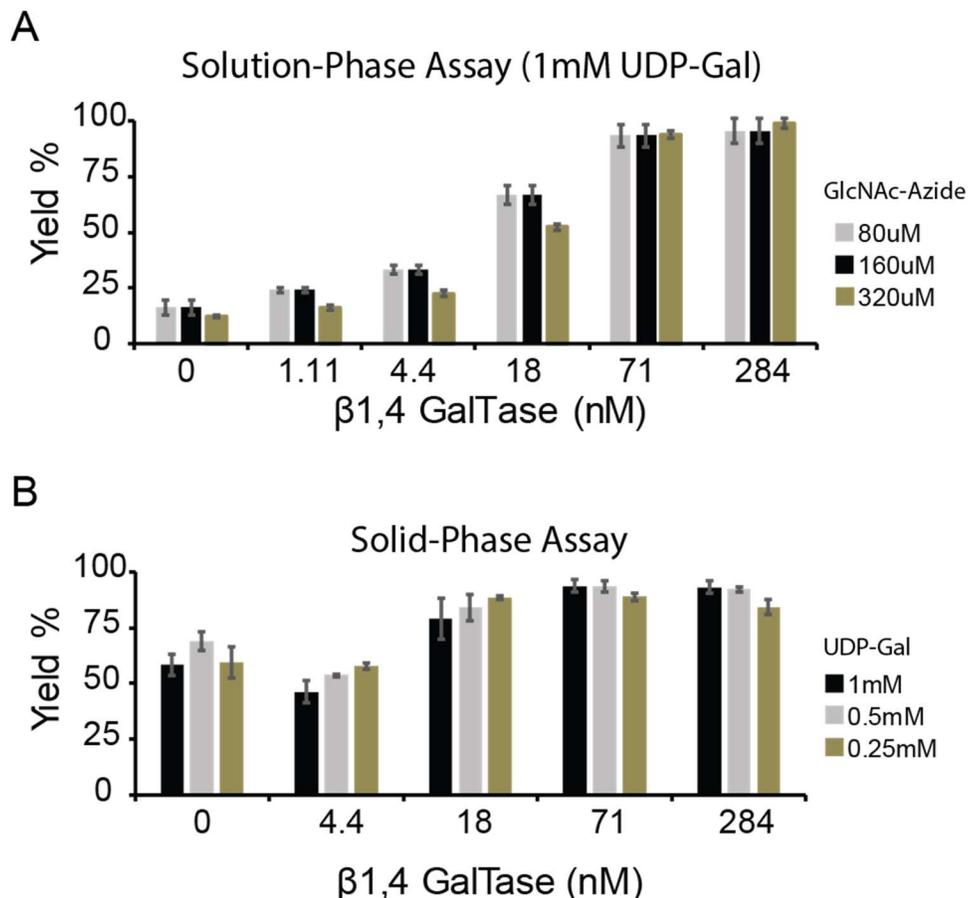


**Figure 3.24: GalTase vs GlcNAc concentration.** Optimal yield is attained between 0.156mM and 0.313mM depending on enzyme concentration. Smaller enzyme concentrations shift the substrate-yield dependence to lower concentrations, as lower levels are needed to achieve the maximal reaction rate. Reaction was performed for 4 hours at 37°C. (Error bars) standard deviation, n=3 averages for data.

As enzyme concentration increases, this trend shifts towards the right to higher 2-eaGlcNAc concentrations but repeats in a similar fashion, as expected for enzymes following standard Michaelis-Menten kinetics. At low 2-eaGlcNAc concentrations immobilization efficiency becomes a concern as the peak corresponding to the unmodified 2-eaGlcNAc occurs in significantly smaller intensities than at higher concentrations. Accordingly, it became analytically difficult to distinguish product peaks from background because of a steep loss in signal-to-noise ratio (S/N). S/N lower than 15 were disregarded in the data analysis.

The next set of experiments compared solid-phase and solution-phase experiments on observed GalTase activity. Solid-phase entails immobilization the 2-eaGlcNAc to the surface prior to exposure of a buffered mixture containing GalTase and UDP-Gal to allow for in-vitro glycosylation directly onto the surface whereas solution-phase reactions were performed in

similar manner to the previous experiment. Both lysate buffer and lysate cell lysate derived from MDA-MB-231 were tested in these two formats. Lysate was prepared from pellets of approximately 1-2 million cells and 100-200uL of lysis buffer, which consisted of the same content as reaction buffer in addition to 2% Triton X-100 detergent. After 30 minutes of chemical lysis, cell homogenate was centrifuged and then supernatant separated for use as lysate in subsequent experiments. The cell lysate condition contained approximately 0.2 mg/mL of total protein independent of GalTase spiked in. Surface reactions were incubated for 4 hours at 37°C in a humidified chamber before undergoing SAMDI-MS analysis (**Figure 6**).



**Figure 3.25: GalTase activity by Solution vs Solid Phase Formats.** A.) Different concentrations of GalTase were reacted with varying 2eaGlcNAc concentrations for 2 hours at 37°C in both buffer and in lysate containing 0.2mg/mL total protein, not including GalTase supplemented, as in-vitro solution reactions. All reactions utilized 1mM UDP-Gal as the donor molecule. After 4 hours, reactions were quenched with ethanol, mixed with CuCC reaction cocktail, and reaction products immobilized to a surface and analyzed by SAMDI-MS. B.) Different concentrations of GalTase was allowed to pre-immobilized 2eaGlcNAc with varying amounts of UDP-Gal for 4 hours at 37°C in humidified chambers before SAMDI-MS analysis. In both experiments higher rates of activity are detected in the lysate activity, with endogenous galactosylation activity detected. (Error bars) standard deviation, n=3 for buffer condition, n=2 for lysate condition, averages for data.

The first notable observation is the difference in activity exhibited between the buffer and lysate conditions. A higher apparent % conversion across all enzyme and acceptor concentrations was seen in the presence of lysate. This could be due either to the accumulated effect of endogenous enzyme activity in conjunction with supplemented GalTase or because of the

presence of endogenous cofactors that improve enzyme efficiency. More importantly, the presence of endogenous activity can be detected in the absence of any supplemented GalTase. This observation is supported by literature as MDA-MB-231 cells are documented to exhibit increased expression of GalTase, among other galactosyltransferases.<sup>172</sup> At this time, the specific galactosyltransferase enzyme(s) has not been specifically confirmed. Furthermore, in additional experiments a small amount glucosylation activity was also detected on 2eaGlcNAc at higher lysate concentration (1.2mg) implying either lower intrinsic enzyme activity under these conditions for this specific activity or lower abundance of the responsible.

### **3.3 Discussion and Conclusion**

The observation of endogenous activity is important because it supports the feasibility of the SAMDI platform to measure other GT activities. GTs are often present as type II transmembrane proteins which traditionally present challenges to solubilize. That native activity was detected using the both solid-phase platform indicates that with adequate optimization of lysis conditions and choice of detergent, endogenous GTs can access substrate displayed on the surface of the array as confirmed in this experiment. This is advantageous because less enzyme is needed to attain high levels of detection with solid-phase experiments as indicated by previous experiments.

The observed degradation of the monolayer points to potential limitation with use of copper-based click immobilization chemistry. While the concentrations of azido substrate used in this assay were enough under these conditions (10 $\mu$ M-1mM) it is conceivable that lower concentrations could be utilized absent the 1-hour incubation time limitation. The advent of

copper-free click approaches utilizing dibenzylcyclooctyne where bond-strain promotes the reaction without need of a copper catalyst, could overcome this limitation.<sup>173</sup> Alternatively, other immobilization strategies can be explored such as more traditional thiol-maleimide chemistry which is rapid but can cross-react with other substances endogenous to the lysate.

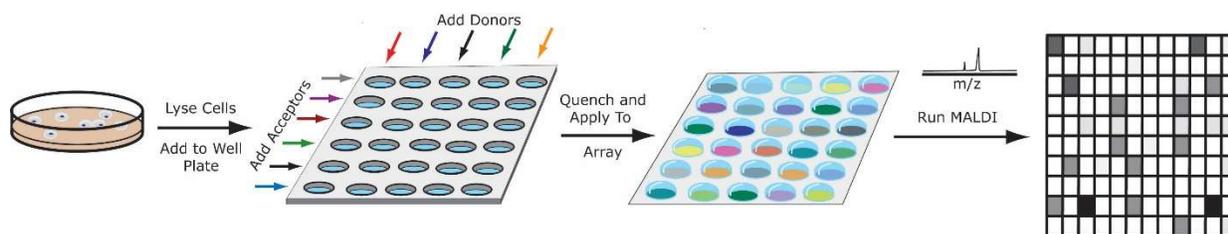
<b>Cell Line</b>	<b>Invasive <i>in-vitro</i></b>
MCF10A	No
MCF-7	No (need estrogen)
T47D	No (need estrogen)
MDA-MB-231	Yes
MDA-MB-468	Yes

**Table 3.2: Cell line models of breast cancer.** Adapted from literature<sup>174, 175</sup>

The next step enabled by this study would be to profile glycosyltransferase patterns in following breast cancer cell lines (See table 1) because there are documented changes that correlate with invasiveness vs benign phenotype and because these cell lines are heavily utilized in the breast cancer field. SAMDI-MS can be used to explore the hypothesis that the malignant phenotype of breast cancer cells has a distinct glycosyltransferase activity pattern unique from benign tissue with three key questions. Several breast tumors exhibit distinct phenotypes in cell condition where a hormone or other signaling molecule is supplemented, such as MCF-7 in the presence of estrogens.<sup>175</sup>

Lysate from cell cultures can be assessed against one of 11 azide-oligosaccharides with the 9 common donor molecules. Acceptors shown in Table 2 have been chosen primarily for their commercial availability. As more complex acceptors become available, they too can be integrated in the SAMDI screening to increase the range of activities profiled. The copper-click SAMDI

assay can then be used generate a profile of observed activities. This process is illustrated below in Fig 5.



**Figure 3.26: High-throughput profiling of Glycosyltransferase activities.** Profiling will use combinations of donors and acceptors to discover glycosyltransferase activities in breast cancer lysate. Lysates will be prepared and transferred to a 96-well plate. Unique combinations of each of 11 acceptors and each of 8 sugar donors will be added to the wells. Then the reaction mixtures will be transferred to a SAM array. The arrays will be analyzed by MALDI mass spectrometry to determine the extent of glycosylation and generate activity profiles

Peaks obtained from these spectra will then be integrated to measure the ratio of modified to unmodified modified acceptor to generate heatmaps. This process can be repeated using a panel of nucleotide-donor analogs that suppress activity of classes of glycosyltransferases. More selective inhibitors, if available, will also be used. This combination of reaction screening coupled with enzyme inhibition will be used to identify characteristic GT activity profiles for each cell line. Lysate will be fractionated so that enrichment of specific cellular compartments may be assayed, as the activity of glycosyltransferases is likely distinct as previously indicated.<sup>154</sup>

There are two prime concerns overall with respect to profiling GT activities as currently suggested in this exploratory study. First, that profiles will not be easily discernable via standard statistical measures available at this time and may require a more sophisticated statistical informatics approach. The second limitation would be if no activity profile is discernable as might be suggested by the lack of diversity of activities presented. This may be because the given acceptor library is insufficient to detect a large range of activities. However, given that several

glycosyltransferases catalyze similar reactions, diversity of activities can be explored using specific inhibitor strategies selective for glycosyltransferase isoform or via a molecular biology approach utilizing siRNA and genetically modified cell lines that knock down or increase expression of endogenous glycosyltransferases to be explored. An alternative assay building on existing peptide array techniques would be to limit the activity profiling to O-glycosylation via immobilization of peptide substrate. The Mrksich lab has established assays using immobilized peptides and has extensive experience in peptide-synthesis. One promising peptide motif would be the 20-amino acid long tandem repeats found on the Mucin I protein, which is documented to be aberrantly glycosylated in breast cancer cells and has been used to generate peptide arrays.<sup>48</sup>  
<sup>119</sup> However, such an array has not been used to profile endogenous cell glycosylation activity.

The work presented here is significant because the study of protein glycosylation is currently hindered by both technological and knowledge gaps. In terms of technological gap, no high-throughput assay exists capable of studying activity of glycosyltransferases, particularly in complex mixtures such as cell lysates. It is also the first example of using SAMDI-MS for detection of endogenous glycosyltransferase activity from human-derived cell lysate.

### **3.4 Experimental**

#### **Reagents**

Unless specifically stated all reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise specified. 2-azidoethylGlcNAc was purchased from TCI chemical. Disulfides used to form self-assembled monolayers were purchased from ProChimia Surfaces

(Sopot, Poland). Tris(3-hydroxypropyltriazolylmethyl) amine (THPTA) was purchased from Click Chemistry Tools (Scottsdale, AZ). 2-Propyne-1-thiol was prepared according to the procedure in Schuster et al.<sup>169</sup>

### **Preparation of SAMDI Plates**

Stainless steel plates with dimensions matching standard MALDI plates (8 cm × 12.3 cm) were cleaned with hexanes, ethanol, and water. A mask with an array of 2.8 mm circles laid out in a standard 384-well format was placed over the plates and an electron beam evaporator (Thermionics Laboratory Inc., Hayward, CA) was used to deposit a layer of titanium (5 nm at 0.02 nm/s) onto the exposed circles, followed by a layer of gold (35 nm at 0.05 nm/s). Plates were stored under vacuum until use. The plates were placed in a solution of two alkyl-disulfides in ethanol, 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 (both from ProChimia Surfaces, Poland), for approximately 16 h. This is expected to produce monolayers with 10% of the alkanethiolates terminated with maleimide groups. The plates were rinsed with ethanol and then immersed in a 10 mM ethanolic solution of hexadecyl phosphonic acid for 10 min. The plates were removed, rinsed with ethanol dried with air prior to use. An aqueous 100 μM 2-propyne-1-thiol solution in 100mM TRIS-HCL pH 8.0 ethanol was applied to each spot for 30 minutes, then rinsed with ethanol and dried with air. The plates were then used for copper-based click immobilization of azido-tagged carbohydrate substrates prior to SAMDI-MS analysis.

## **In vitro Glycosylation Reactions**

Unless otherwise specified in the text the following conditions were used for glycosylation reactions. Solution-phase reactions: Glycosylation reactions were typically performed in a reaction buffer composed of comprised of 50mM HEPES pH 7.4 (sigma Aldrich), 10mM MnCl<sub>2</sub>, 10mM MgCl<sub>2</sub>, 2% Triton-X100 in a total volume of 5uL. Reactions were quenched with addition of 5uL of 30% ethanol (v/vol) before transferred to alkyne monolayers for immobilization. Solid-phase: glycosylation reactions were performed using identical buffer conditions as stated above on SAMs pre-immobilized with 2eaGlcNAc in total reaction volumes of 2uL per spot in a pre-warmed humidifying chamber to prevent drying over time. After reactions were completed, surfaces were washed with water, waste collected for disposal, and dried by air before SAMDI-MS analysis.

## **Copper-Click Immobilization and SAMDI-MS**

Glycosylation reaction samples were mixed 4:1 with a solution of “click” reagents -10 mM CuSO<sub>4</sub>, 40 mM sodium ascorbate, and 20 mM THPTA. A volume of 1 μL of the sample mixture was dispensed onto each spot on the MALDI plate and incubated for 60 min at room temperature, then rinsed extensively with water and ethanol. A 10 mg/mL solution of 2,4,6-trihydroxyacetophenone in acetone was delivered to each spot on the array and the surfaces were analyzed using an AB Sciex 5800 MALDI TOF/TOF instrument in positive reflector mode. The area under the curves for the  $[M + H]^+$  peaks of disulfides was measured with the Data Explorer software (AB Sciex). Presented data represent the means and standard deviations of all spots.

## Cell culture and lysate preparation

MDA-MB-231 cells were obtained from ATCC and cultured high-glucose DMEM medium supplemented with 10% fetal bovine serum, glutamax, penicillin and streptomycin. All cells were cultured in a humidified incubator at 37 °C and CO<sub>2</sub>. Cells were trypsinized and suspended in media, and the average number of cells per  $\mu\text{L}$  was counted using a hemocytometer and Countess automated cell counter (Life Technologies), and cell concentrations were adjusted to seed the desired number of cells per spot in 3  $\mu\text{L}$  media. After culture, media was removed and lysis buffer composed of 20 mM Tris, 136 mM NaCl, 1 mM EDTA, 0.5% Triton-X 100, pH 7.4. A protease inhibitor tablet obtained from Roche was added to the lysis buffer. A Pierce™ BCA Assay Kit (Thermo Fisher Scientific) was utilized to measure the approximate protein concentration of lysate samples.

## Chapter 4

### High-Throughput Synthesis and Analysis of Intact Glycoproteins using SAMDI-MS

This chapter is adapted from the following unpublished work:

José-Marc Techner<sup>#</sup>, Weston Kightlinger<sup>#</sup>, Liang Lin, Jasmine Hershewe, Ashvita Ramesh, Matthew P. DeLisa, Michael C. Jewett, Milan Mrksich. “*High-Throughput Synthesis and Analysis of Intact Glycoproteins using SAMDI-MS.*” With permission from Analytical Chemistry, in preparation for publication. Unpublished work copyright 2018 American Chemical Society." #Authors contributed equally to this work.

#### INTRODUCTION

Recombinant therapeutic proteins are a common class of approved drugs and 70% of protein therapeutics are glycosylated. Despite their increasing presence in the clinic many these recombinant proteins suffer from suboptimal stability and half-life *in vivo*. Glycoengineering, which entails the design of therapeutic protein sequences and expression systems to control and alter glycosylation, is an approach to tune these pharmaceutical properties.<sup>11, 176-178</sup> Plaguing these engineering efforts, however, is a lack of high-throughput methods to quickly prepare and analyze the glycosylation states of these engineered proteins.

Existing approaches for designing and testing protein glycosylation suffer from two main technical challenges. First, the synthesis of a wide variety of glycoprotein designs requires labor-intensive effort to construct, transform, express, and purify large numbers of proteins from live cell systems. The time and dollar costs to prepare such protein libraries are too prohibitive for most laboratories to pursue and new cell lines are typically generated in order to access the needed glycoforms to test these substrates against.<sup>179</sup> Second, current assays to determine glycosylation states of proteins suffer from either low-throughput or lack generalizability. For example, the use of immunological probes (such as a lectin or antibody conjugate) to detect

glycosylation in western blotting or in enzyme-linked immunosorbent assays (ELISA) often lack specificity and are not available for all types of glycoforms possibly present. Alternatively, the combination of electrospray-mass spectrometry and chromatographic separation platforms (LCMS) have been developed that circumvent these disadvantages.<sup>180</sup> However, the chromatographic separation steps required entail lengthy run-times that make this approach impractical for the assessment of hundreds of samples. While emerging cell-free protein (CFPS) *in vitro* glycosylation platforms open access to complex glycosylation pathways without the need of extensive cell line engineering, there is still a need for analytical methods that allow for quick assessment of glycosylation state for hundreds of protein samples.<sup>181-183</sup>

We have developed a technique based on a combination of self-assembled monolayers and matrix-assisted laser desorption mass spectrometry (MALDI-MS) which appears well poised to fill this need. SAMDI uses alkanethiolate self-assembled monolayers (SAMs) on gold chips to capture enzyme reaction substrates and is compatible with complex lysates.<sup>168</sup> SAMDI-MS has been utilized to aid characterization of glycosyltransferases using carbohydrate substrates and in recent work, we have applied SAMDI to extensively characterize the acceptor peptide specificities of N-linked glycosyltransferases to design optimized peptide sequences.<sup>48, 116, 165</sup> However, neither of these studies utilized proteins as substrates and in the latter effort the designed sequences required further cloning into a protein template in order to assess how glycosylation is affected in a protein context.

Here, we report a high-throughput and generalizable strategy to quantify glycosylation on intact proteins using SAMDI-MS. This method uses CFPS to generate polyhistidine(His)-tagged protein variants containing engineered glycosylation sequences directly from linear expression

templates (LET) which are produced by gene synthesis and amplified by polymerase chain reaction (PCR).<sup>184</sup> Lysate containing these protein substrates were then subjected to *in-vitro* glycosylation reactions by *Actinobacillus pleuropneumoniae* N-glycosyltransferase (NGT) before transfer to a self-assembled monolayer comprised of nickel-incubated nitrilotriacetic acid (Ni-NTA) against a background of tri(ethylene glycol) groups followed by SAMDI-MS. Each step in the process is performed using a standard 96 well plate format that is compatible with robotic liquid handling. We determined optimum assay conditions and validated the ability of the assay to quantitatively determine glycosylation efficiency. To demonstrate the high-throughput nature of the assay, we synthesized and determined the glycosylation efficiency of a library of 87 variants of the *E. coli* Immunity Protein 7 (Im7), where each variant contained a single glycosylation sequence at a unique position. We also demonstrate generalizability of this assay against different protein substrates with glycan modifications by producing and determining the glycosylation state of several proteins of therapeutic interest and detecting modification of Im7 with a heptasaccharide by the oligosaccharyltransferase (OST) pglB from *Campylobacter jejuni*.

## RESULTS

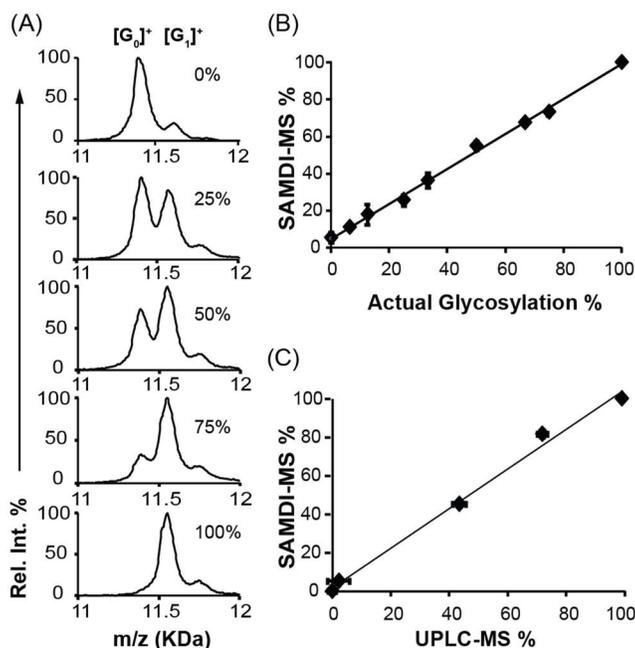
We chose the NGT glycosylation system as a test case because the addition of a single glucose (+162 Da) cannot be readily detected by western blotting as there is no commercial antibody for this modification to our knowledge and it reasonably represents the smallest mass shift required for quantification of glycosylation status. Furthermore, there has been widespread interest in the use of NGT for biocatalysis because it can be easily expressed functionally in bacterial systems.<sup>41, 49, 185</sup> As a proof-of-concept for our approach, we chose Im7-6, a

recombinant version of Im7 immunity protein from *E. coli* that contains an internal glycosylation sequence GGNWTT and a hexahistidine tag at the C-terminus. Im7-6 has been demonstrated to be readily glycosylated by NGT and thus makes an ideal candidate to test our assay strategy. We used nitrilotriacetic acid-terminated (NTA) monolayers as our immobilization functionality because of the widespread use of hexahistidine tags in recombinant protein engineering laboratories and its prior success in selective capture of proteins onto SAMs.<sup>88, 122, 130</sup>

To evaluate the ability to characterize *in vitro* glycosylation (IVG) of Im7-6 by NGT previously prepared by CFPS using SAMDI, reactions containing purified Im7-6 (5 $\mu$ M), excess UDP-Glucose and NGT (0.05 $\mu$ M) were carried out in aqueous conditions buffered with 100mM HEPES pH 7.5 for approximately 4 hours. After reactions were quenched with ethanol, the Im7-6 substrate and glycosylated product were captured onto monolayers presenting terminal Ni-NTA groups against a background of tri(ethylene glycol) groups. The monolayers were prepared on plates having an array of 384 gold islands in standard plate geometry. Following immobilization and application of sinapic acid matrix, SAMDI mass spectrometry was performed which confirmed both capture of Im7-6 and presence of glycosylation (Figure 1).



accounted for to adjust the final integration value.<sup>186, 187</sup> In this way we determined, as illustrated in Figure **2B**, that there is a linear correlation of SAMDI-MS response to increasing amounts of glycosylated product ( $R^2 = 0.997$ ). This, in addition to the accuracy of the SAMDI-MS readings, indicate that glycosylated and unglycosylated analytes have similar binding and ionization efficiency to the monolayer. To further validate SAMDI's ability to quantify Im7-6 glycosylation, we compared its quantitative performance against a previously established UPLC-TOF method.<sup>48</sup> As shown in Figure **2C**, we observed a strong linear correlation between SAMDI and this method ( $R^2 = 0.993$ ), indicating that SAMDI can quantitatively measure intact protein glycosylation with similar accuracy to state-of-the-art methods. However, SAMDI only required 30 seconds of instrument time per sample compared with UPLC-TOF method which required at least 10 min of instrument time per run.

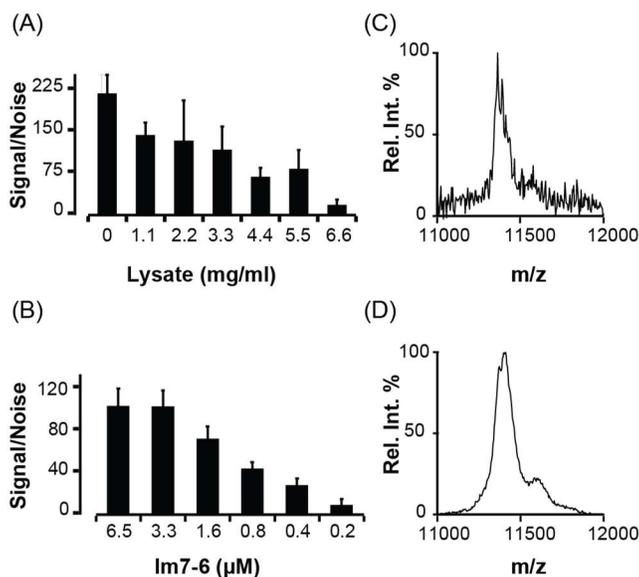


**Figure 4.2. Quantitative Assessment of SAMDI-MS Protein Glycosylation Assay (A)**

Separate mixtures of glycosylated ( $G_1$ ) and unglycosylated ( $G_0$ ) Im7-6 were subjected to SAMDI-MS to obtain spectra (0 to 100% glycosylated, increasing downwards). Sinapic acid adduct of  $G_0$  overlaps with presence of  $G_1$  peak, as noted by apparent increase in  $G_1$  peak area for 50% glycosylated condition. (B) Linear correlation of Protein SAMDI-MS determined glycosylation % to actual values, ( $R^2=0.997$ ,  $n=3$ ). (C) Correlation of Protein SAMDI-MS performance vs LCMS method for identical set of samples, ( $R^2=0.993$ ,  $n=3$ ). Both (B) and (C) are calculated after applying a correction for the presence of sinapic acid adducts for both  $G_0$  and  $G_1$ . (Error bars) One standard deviation.

An important aspect of the assay design was its performance over a range of complex sample conditions. The assay is applicable over a range of protein concentrations with an approximate lower boundary of detection around  $\sim 0.5\mu\text{M}$ . Furthermore, our assay is robust under a range of lysate concentrations making this technique suitable for a wide range of assay conditions as illustrated in **Figure 3**. For experimental consistency we chose for each reaction to be 10% CFPS lysate (v/v), which corresponds to an approximate concentration of  $\sim 1.1\text{mg/mL}$  of lysate protein since that provided excellent signal-to-noise ratio.<sup>188</sup> We attribute the decrease in signal spectra quality at very concentrated lysate sample conditions ( $>50\%$  CFPS lysate volume

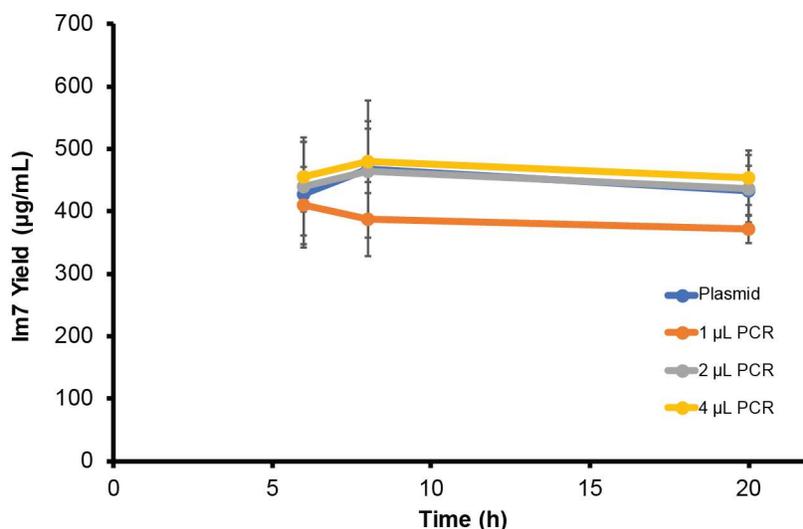
or ~5.5mg/mL of protein) to competition of weakly-interacting background species with the Ni-NTA moiety.



**Figure 4.3. Performance of SAMDI-MS for CFPS lysate samples.** (A) SAMDI-MS spectral signal-to-noise of 5µM Im7-6 in samples containing variable amounts of CFPS lysate as quantified by protein concentration. B.) SAMDI-MS spectral signal-to-noise of serial dilutions of Im7-6 concentration in samples containing a constant 4mg/mL of CFPS lysate. (C) and (D) are representative SAMDI-MS spectrum of 5µM Im7-6 in samples containing 24 and 4 mg/mL of CFPS lysate, respectively. (Error bars) One standard deviation.

Our next step was to test the strategy of synthesizing His-tagged glycosylation substrates rapidly from linear DNA transcripts. We originally considered pursuing a traditional strategy using plasmids templates however the time and cost of cloning and plasmid preparation was a significant bottleneck. However, the falling cost of commercial gene synthesis enabled us to use a linear expression template (LET) CFPS strategy instead.<sup>184</sup> We chose to test a variant of Im7, termed “p3” with respect to the position of the inserted asparagine residue in the GGNWTT glycosylation sequence replacing native residues 1-6. Performing cell-free protein synthesis requires DNA templates with promoter and terminator sequences in addition to coding regions.

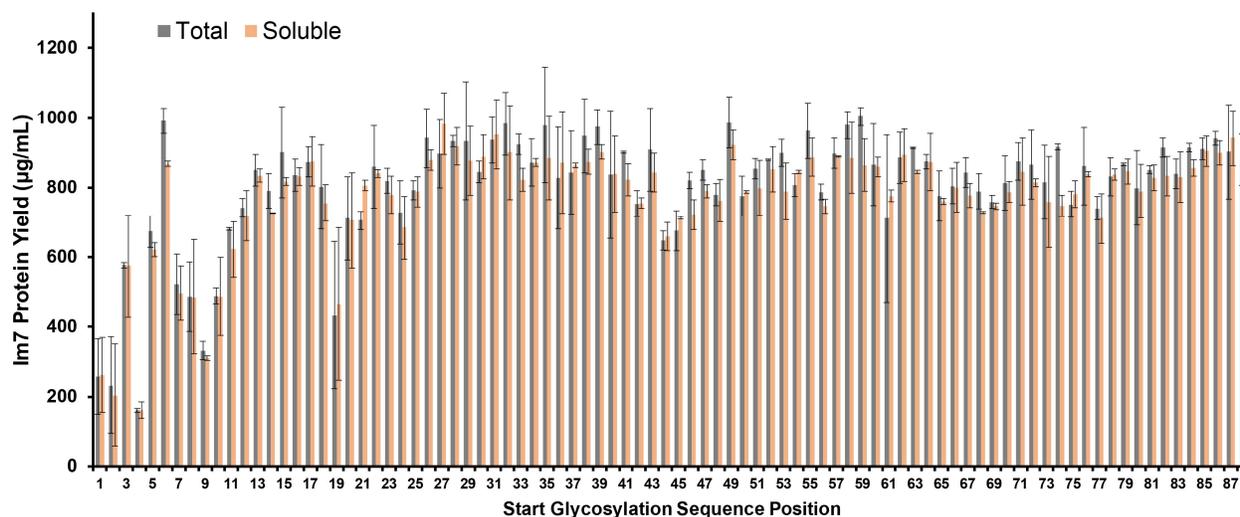
To avoid the cost of synthesizing redundant promoter and terminator regions, we designed our initial gene to encode only our intended Im7-p3 protein including a C-terminal hexahistidine-tag and a short portion of the ribosomal binding site comprising a leader sequence at the 5' end. We then designed primers that upon PCR amplification of the original incomplete gene would result in a linear DNA product including the necessary transcriptional and translational elements at the 3' leader sequence needed for cell-free protein synthesis. We tested this strategy against a plasmid-encoded version of Im7-p3 and found that both linear and plasmid expression templates achieved similar yields within 6 hours (Figure 4). Yields were determined by incorporation of radioactive [ $^{14}\text{C}$ ]-leucine as previously described.<sup>189</sup>



**Figure 4.3: Optimization of reaction time and PCR template amount.** CFPS reactions supplemented with  $^{14}\text{C}$ -leucine with indicated amounts and types of DNA templates encoding Im7-6 were incubated at  $30^\circ\text{C}$  for indicated time periods before being flash frozen on liquid nitrogen. CFPS reactions, PCR reactions, and quantification of protein synthesis by incorporation of  $^{14}\text{C}$ -leucine were performed as described in Materials and Methods. Background radioactivity from CFPS samples without any DNA template was subtracted from each sample. Soluble protein fractions were isolated by centrifugation of CFPS reaction products at  $12,000\times g$  for 15 min at  $4^\circ\text{C}$ . Data are averages at each time-point. (Error bars) One standard deviation.

With confirmation of our synthetic strategy we next sought to demonstrate the ability of our approach to enable high-throughput measurement of protein glycosylation efficiencies by designing a library of Im7 mutants where the GGNWTT sequence is placed at each amino acid position in a manner identical to Im7-p3, beginning with the first amino acid after methionine and ending with the last amino acid position in wildtype Im7. Using the previously described PCR amplification strategy, yields for each PCR product were assessed by incorporation of radioactive [ $^{14}\text{C}$ ]-leucine. Soluble protein yields for each member of the library were similar at 600-1000  $\mu\text{g}/\text{mL}$ , except for the first 20 positions which exhibited lower and more variable

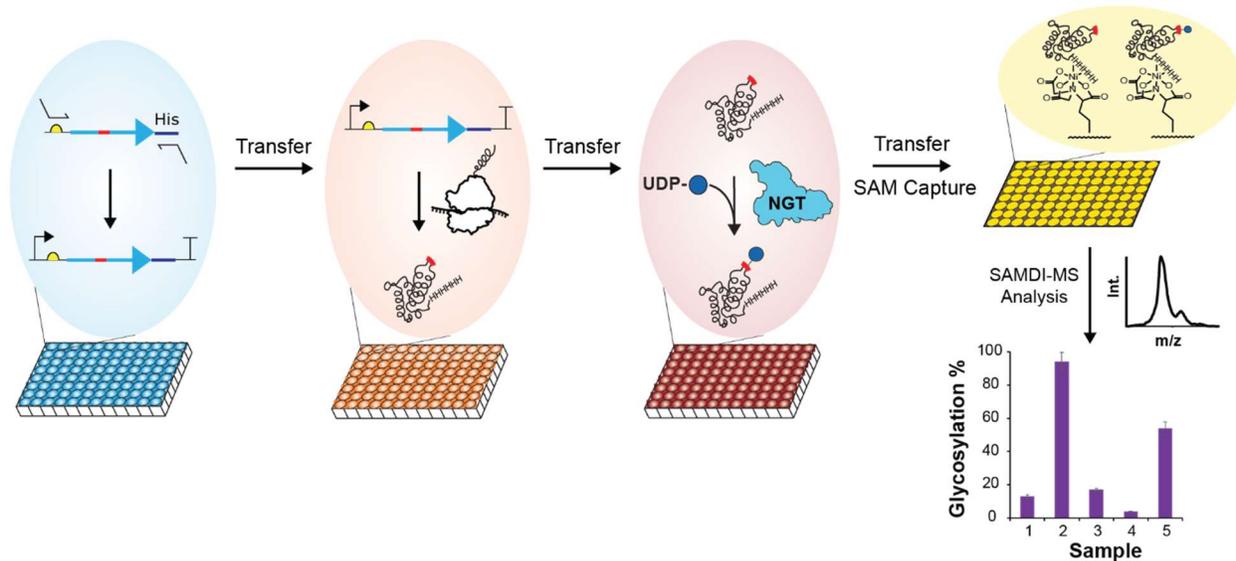
yields. We interpret this yield variation to be caused by changes in the strength of the ribosome binding site, which is sensitive to changes in mRNA sequence near the start codon.<sup>190, 191</sup>



**Figure 4.4: Measurement of synthesized yields of Im7 library using <sup>14</sup>C-leucine incorporation.** Total and soluble yields for each positional Im7 mutant were measured. (Error bars) One standard deviation from n=2 at each time-point. CFPS reactions supplemented with <sup>14</sup>C-leucine and 2 µl of DNA template from PCR reactions encoding each Im7 glycosylation sequence position variants were incubated at 30°C for 6 h. CFPS reactions, PCR reactions, and quantification of protein synthesis by incorporation of <sup>14</sup>C-leucine were performed as described in Materials and Methods. Background radioactivity from CFPS samples without any DNA template was subtracted from each sample. Soluble protein fractions were isolated by centrifugation of CFPS reaction products at 12,000xg for 15 min at 4°C. Data are averages of n=2 at each time-point. (Error bars) One standard deviation.

After quantification of the Im7-library we then subjected the Im7 variants to an *in vitro* glycosylation study in a manner outlined by Figure 2. This utilized a library synthesized in parallel to the quantification library without radioisotope leucine and using the same stock of linear DNA templates. Briefly, 2 µl of the original PCR reaction performed was transferred to another plate and subjected to a CFPS reaction for 6 hours. After CFPS was complete, we then prepared *in vitro* glycosylation reactions where each protein concentration was normalized to

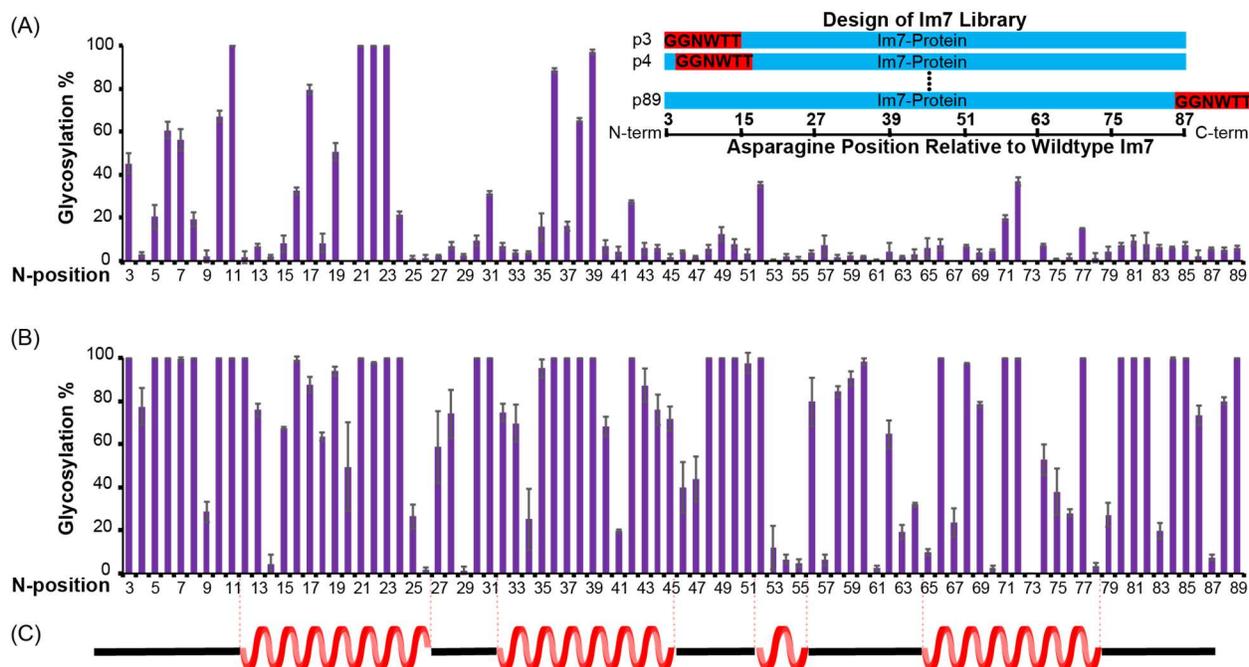
2.5 $\mu$ M based on the previous radioactive quantification measurements. These were then subjected to IVG reactions with NGT and UDP-glucose at a final reaction volume of 4  $\mu$ l. A typical library synthesis provided enough protein substrate for several IVG reactions. After reactions were diluted 1:1 with 30% ethanol to quench the IVG reaction, 1  $\mu$ l of each reaction products were then transferred to plates previously prepared with Ni-NTA SAMs. Following incubation and capture steps, plates were washed and dried before quantification of glycosylation efficiency by SAMDI-MS. This entire process was performed in a 96 well standard plate format with robotic liquid handling of the transfer steps performed to minimize variation in sample handling at each step. Each plate can accommodate up to 384 unique experiments allowing an experimenter to obtain triplicate experimental replicates of a single library condition in a minimal amount of time.



**Figure 4.6: Strategy for high-throughput synthesis and analysis of glycosylation on intact proteins.** Linear transcripts of His-tagged proteins containing an engineered glycosylation site (GGNWTT) substituting native residues beginning at each amino acid position are commercially synthesized and then amplified by PCR with promoter and terminator regions to create linear expression templates (LET). LETs are then added directly to cell-free reaction mixtures to initiate protein synthesis of His-tagged substrates. After CFPS is complete, lysates containing synthesized protein then undergo glycosylation by NGT and excess UDP-Glucose in well plates. Reactions are quenched with addition of ethanol and then products transferred to a MALDI-compatible assay plate containing 384 gold islands covered in Ni-NTA terminated monolayers 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). Each step of the workflow is compatible with standard well formats and robotic liquid handling.

Because each Im7 protein contained the same optimized glycosylation acceptor sequence (GGNWTT), our initial expectation was that glycosylation efficiencies would correlate roughly with the native structure of the protein. We expected that glycosylation sites within unstructured and loop regions in the native context would exhibit greater modification rates compared to those within a more rigid helical context. Our data show this is not the case with our library exhibiting no clear relationship between the native structure of Im7 and glycosylation efficiency. Under relatively minor IVG conditions a wide variability in glycosylation is already observed with the most heavily modified mutants corresponding to the placement of GGNWTT in helical positions.

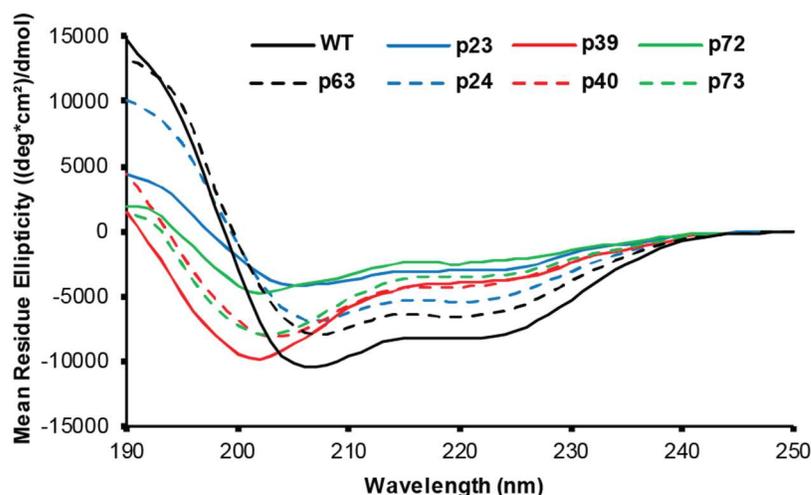
Furthermore, upon further testing with saturating IVG conditions, it becomes clear that most positions can be eventually modified by NGT, given enough time. These results are illustrated in Figure 7.



**Figure 4.7. Rapid analysis of Im7 positional glycosylation site library synthesized and glycosylated in vitro.** (A) Treatment of 2.5 $\mu$ M of synthesized Im7-library with 5mM UDP-glucose and 0.05 $\mu$ M NGT with 100mM HEPES pH 7.4 for 2hr @ 25°C, (n=3). Design of the Im7 library with GGNWTT (top right inset) substitution of native residues. (B) Identical reaction conditions to above except treatment with 0.667 $\mu$ M NGT for 5hr @ 37 °C, (n=3). (C) Location of helical secondary structure in wildtype Im7 based on PDB 1AY1. Im7-library CFPS yields were previously determined by radioisotope incorporation (see supplemental). (Error bars) One standard deviation.

It is likely that the replacement of native residues with GGNWTT, a sequence rich in helically unfavorable amino acids glycine and threonine due to their respective conformational flexibility and to bulky natures, likely cause disruption of expected native secondary structure.<sup>192</sup> This possibility is supported by numerous prior studies on Im7 folding which indicate that single

point mutations throughout the protein are sufficient for the Im7 to lose its native conformation. To explore this possibility, we performed circular dichroism experiments of several positional mutants after purification. We chose several pairs of mutants located in positions expected to be helical in the wildtype structure and compared the obtained CD spectra against wildtype Im7 (Figure 4.8). These data reveal significant disruption of native structure in our synthesized mutants.

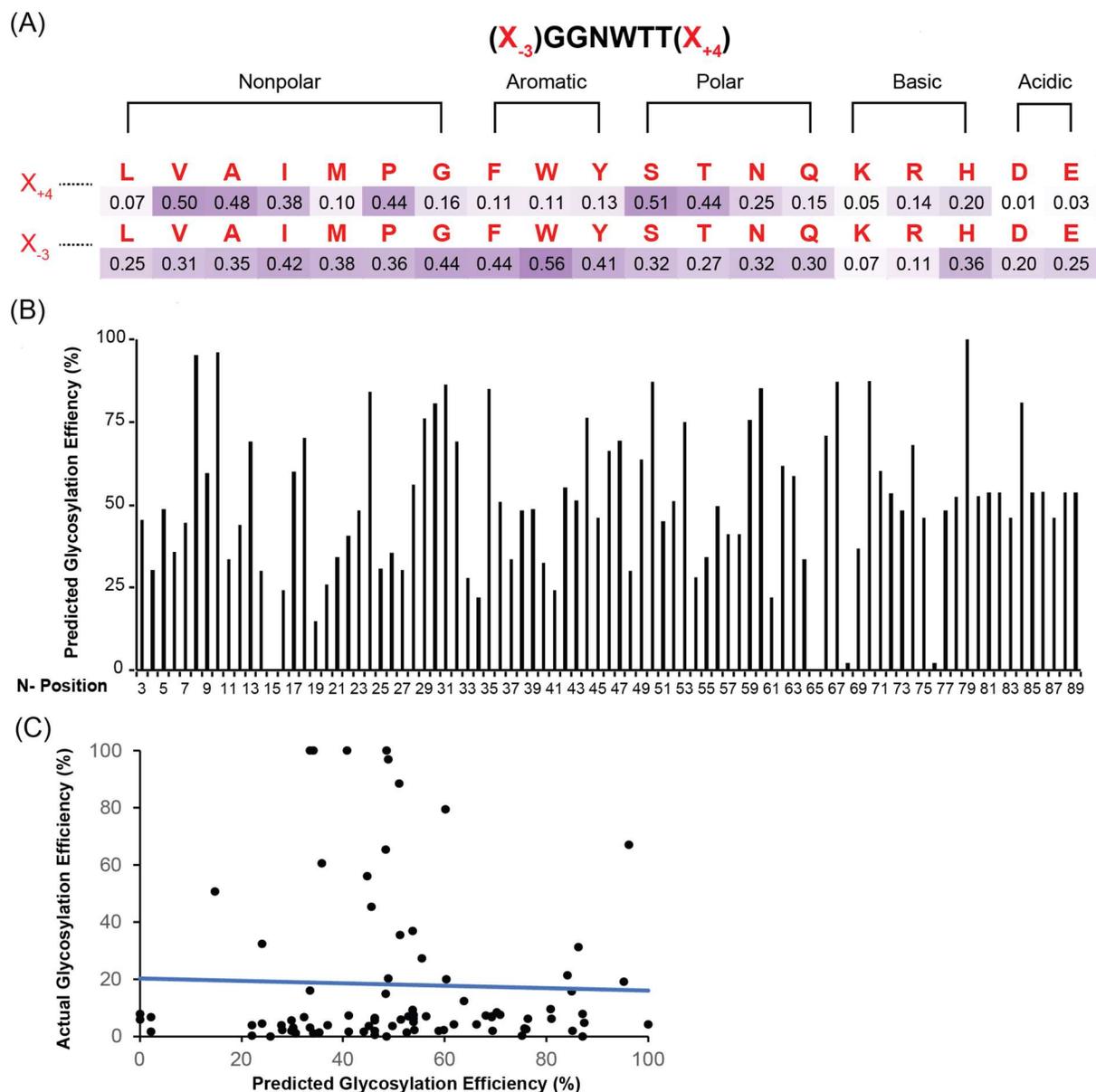


**Figure 4.8: Circular Dichroism spectra of selected Im7 mutants from the glycosylation screen.** Positional mutants p23 and p24, p72 and p73, and p39 and p40 were selected and compared against WT and p61, where GGNWTT is located in a loop region. CD spectra were collected from purified proteins at 25°C in 50mM sodium phosphate buffer at protein concentrations 13-15uM. All mutants show significant disruption from WT with p63 showing the least.

Numerous non-native contacts, amino acid residue interactions not found in the final native structure, are present in intermediate folding conformations of Im7 which are required for formation of helical secondary structure before Im7 adopts the final native state configuration.<sup>193,</sup>  
<sup>194</sup> Native structure-disrupted mutants of Im7 have been shown to form organized, but distinct structures, from the wildtype.<sup>195-197</sup> Collectively, these observations might then explain the

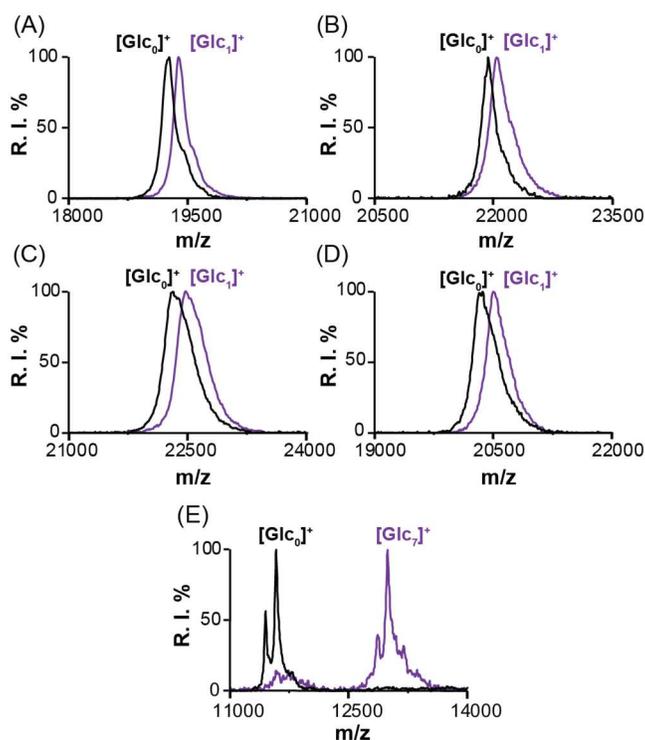
variability in glycosylation efficiencies observed here, where even adjacent positional mutants exhibit drastically different amounts of modification.

To explore the possibility that differences in NGT selectivity for peptide positions outside of the core GGNWTT sequon were causing the observed differences in protein glycosylation, we created a customized peptide library containing peptides with all possible amino acids in the  $X_{-3}$  and  $X_{+4}$  positions relative to the glycosylated asparagine, as shown in **Figure 9**. However, we again found that modification patterns of these peptide substrates did not closely correlate with observations at the intact protein level. The difficulty in predicting intact protein glycosylation *a priori* using structural data or customized peptide mimics, reinforces the need for a high-throughput method to synthesize and analyze glycoproteins such as what we illustrate here.



**Figure 4.9: Study of NGT amino acid preferences adjacent to the GGNWTT sequence.**

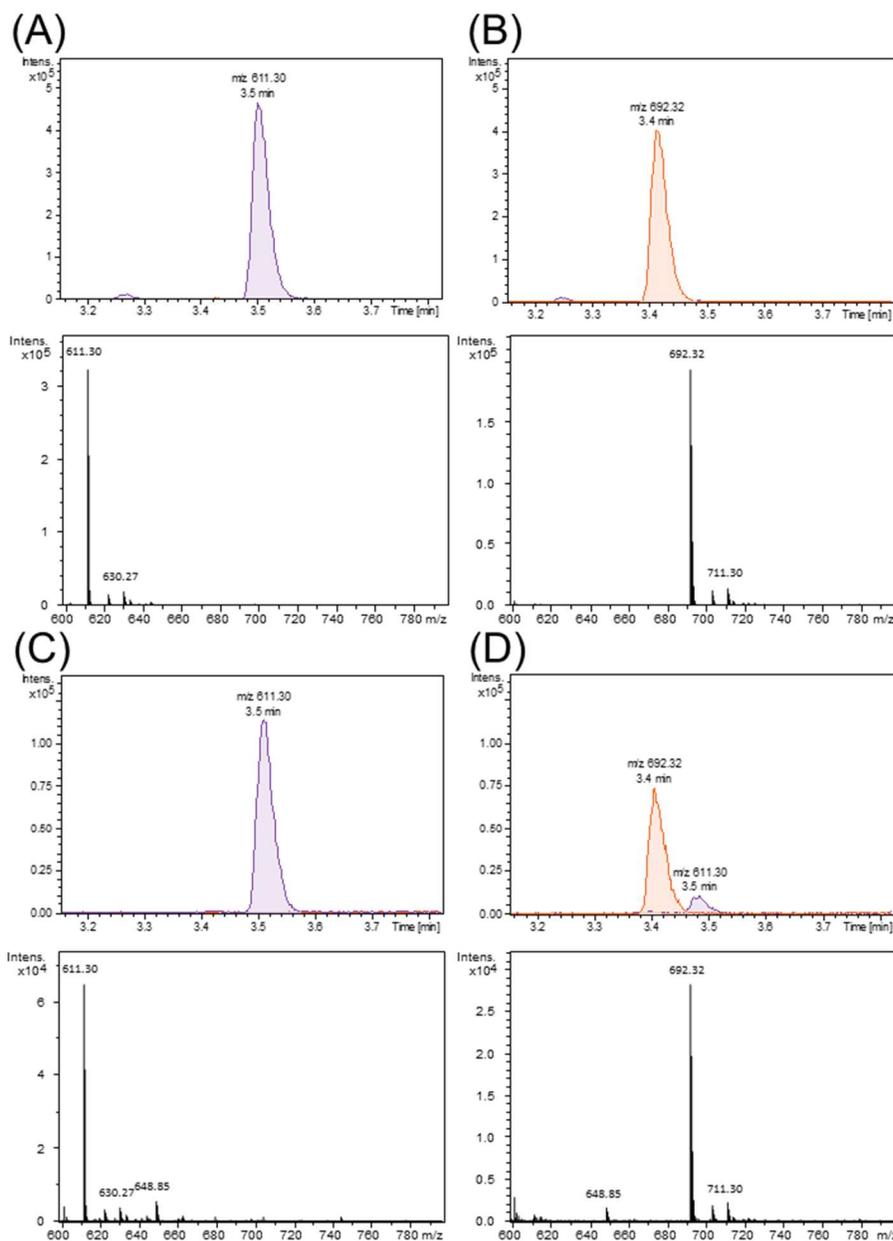
Library of peptides in form of either  $X_{-3}\text{GGNWTRC}$  or  $\text{GGNWTTX}_{+4}\text{RC}$  were synthesized by solid-phase peptide synthesis.  $\sim 50\mu\text{M}$  of each peptide was allowed to react with  $2.5\text{mM}$  UDP-Glucose in the presence of  $0.02\mu\text{M}$  of NGT (for  $X_{-3}$  peptides) or  $0.04\mu\text{M}$  (for  $X_{+4}$  peptides) for 30 minutes at  $30^\circ\text{C}$ , followed by immobilization and SAMDI-MS analysis. A) Relative glycosylation efficiencies observed for each peptide categorized by amino acid type. Each is average of  $n=3$  experiments. B.) Predicted glycosylation efficiency for each Im7 GGNWTT position calculated by multiplying the efficiency of the relevant peptides and normalizing across the entire library set. C.) Linear correlation analysis of the predicted glycosylation efficiency based on the peptide library and the protein data in **Figure 5A**. Lack of linear correlation between peptide library results and Im7 protein samples suggest that amino acid preferences at the  $X_{-3}$  and  $X_{+4}$  position play little role dictating NGT efficiency.



**Figure 4.10: Examples of NGT glycosylation of proteins of therapeutic interest and oligosaccharide activity on Im7 analyzed by SAMDI-MS.** Sequences based on based on (A) Metreleptin (B) Filgrastim (C) Interferon-alfa and (D) Anakinra were modified to include an N-terminal GGNWTT tag, synthesized from LET-CFPS, glycosylated by NGT and analyzed by SAMDI-MS. (E) SAMDI-MS of Im7 construct containing a C-terminal DQNAT tag recognized and glycosylated by *C. jejuni* oligosaccharyltransferase PglB with a heptasaccharide from a lipid-linked undecaprenyl donor. The difference between  $G_7$  and  $G_0$  in (E) corresponds closely to the addition of the heptasaccharide. ( $\Delta m = 1413$  Da, expected is 1404 Da).

Next, to demonstrate the generalized nature of SAMDI-MS for detection of glycosylation of diverse protein substrates, we selected proteins of therapeutic interest from THPdb, a curated repository of FDA-approved peptides and proteins.<sup>198</sup> We chose four proteins, Leptin (Metreleptin), IL1-Receptor Antagonist (Anakinra), Granulocyte Colony-Stimulating Factor (G-CSF) and Interferon-Alpha (Interferon- $\alpha$ ) to test compatibility of our SAMDI method to detect glycosylation. To simplify analysis the known native N-X-S/T glycosylation site on Anakinra

was also disabled by an N185Q mutation. We designed linear DNA transcripts of each protein modified with N-terminal hexahistidine and GGNWTT glycosylation sequon tags and then synthesized them by LET-CFPS. Next, we performed IVG reactions of each therapeutic protein using 0.1  $\mu$ M NGT and 2.5mM UDP-Glucose at 37°C for 4 hours to ensure full glycosylation and performed SAMDI-MS for each protein. All protein targets exhibited a characteristic mass shift corresponding to the addition of a glucose by SAMDI-MS as shown in **Figure 10**. Because of the possibility that several of the protein constructs might contain natural other natural N-glycosylation sites, glycosylation of each protein construct was additionally verified by UPLC-MS of trypsinized peptides (**Figure 11**). For each protein, we detected efficient glycosylation at the N-terminal GGNWTT tag only.



**Figure 4.11: Peptide LC-MS verification of therapeutic protein glycosylation.** Extracted ion chromatograms and MS spectra from peptide LC-MS analysis of trypsinized and engineered forms of (A) Metreleptin, (B) Metreleptin treated with NGT, (C) Filgrastim, (D) Filgrastim treated with NGT, (E) Anakinra, (F) Anakinra treated with NGT, (G) Interferon-alpha, and (H) Interferon-alpha treated with NGT. Extracted ion chromatograms show the elution of doubly-charged, aglycosylated (purple) and glycosylated (orange) forms of the tryptic peptide ATTGGNWTTAGGK engineered into the N-terminus of each protein. MS spectra are summed over the elution times of all detected glycosylated and aglycosylated forms of the tryptic peptide.

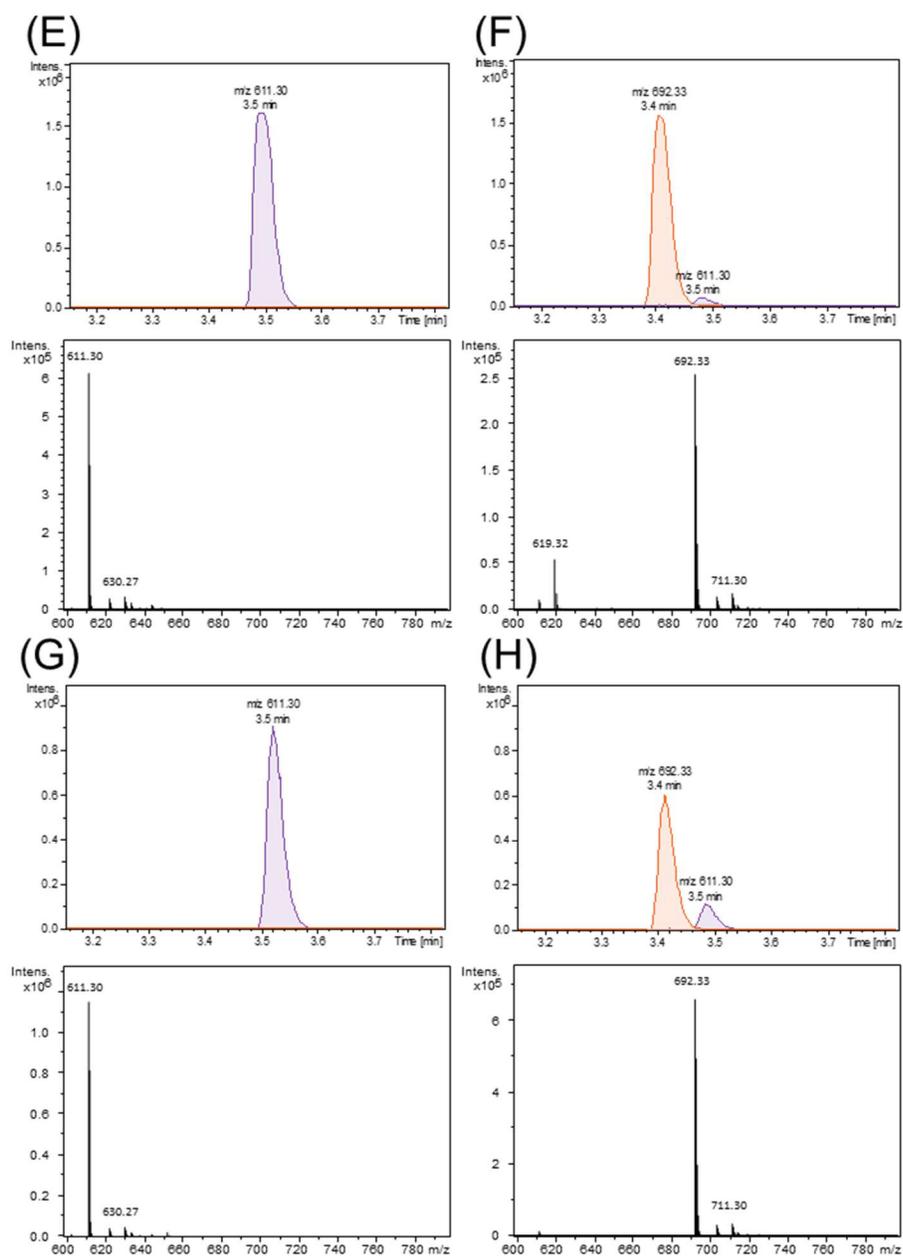


Figure 4.11 (continued).

Finally, we demonstrated the versatility of our approach for detecting modification with diverse glycans by observing the glycosylation of Im7 by a bacterial oligosaccharyltransferase (OST) in **Figure 10E**. Unlike NGT which transfers only one glucose residue, OSTs transfer 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 the PglB OST from *C. jejuni*.<sup>199</sup> We used a recently reported method to synthesize PglB *in vitro* using a CFPS reaction supplemented with protein-lipid nanodiscs.<sup>181</sup> We then combined this completed CFPS reaction with 4  $\mu$ M of purified Im7-DQNAT and a membrane extraction of *E. coli* producing the *C. jejuni* heptasaccharide LLO and incubated the reaction overnight before analysis by SAMDI-MS. SAMDI-MS spectra acquired from samples before and after treatment with OST clearly show the nearly complete glycosylation of Im7 with this heptasaccharide. The difference between  $G_7$  and  $G_0$  corresponds closely to the expected mass of heptasaccharide ( $\Delta m = 1413$  Da, expected is 1404 Da), demonstrating the ability to accurately detect more complex glycan structures.

## DISCUSSION AND CONCLUSION

This chapter demonstrates how SAMDI-MS can be used to perform high-throughput analysis of glycosylation on intact proteins generated using CFPS methods. This work is significant because it utilizes mass spectrometry to quantitate the relative glycosylation of an inserted glycosylation sequence beginning at every amino acid position of a model protein, the

first such example to our knowledge. This work is also important because compared to current mass spectrometric methods for analyzing intact protein analysis it is significantly higher-throughput. To generate the data presented in Figure 7, a total of 88 protein variants X 2 glycosylation conditions X 3 triplicate X 2 technical duplicates = 1,056 independent data points, approximately 12-14 hours of experimentation time (excluding preparation steps) and required 7.5 hours of MALDI-TOF instrument time were needed. In contrast, a conventional UPLC-MS experiment would require purification, dialysis, and preparation of each sample by hand and 176 hours of instrument time to obtain the same number of measurements, excluding sample prep. While in our study we additionally analyzed tryptic digests of the engineered therapeutic proteins, we anticipate that our approach of using SAMDI-MS on intact proteins can be complemented with additional on-chip proteolytic digestion to generate peptide mass fingerprints of captured proteins on SAMs.<sup>121, 200</sup>

We expect the approach outlined in this chapter to be particularly useful in investigating questions regarding positional glycosylation efficiency. Methods for the site-specific glycosylation of proteins enable fundamental studies of the functional roles played by glycans and are important in the development of biologic therapeutics. One strategy uses a short peptide ‘acceptor’ that can be glycosylated by an oligosaccharyl transferase (OST) and engineers the target protein with this ‘sequon’ at the desired glycosylation site.<sup>37, 38</sup> The rules that describe what positions the sequon can be installed, while still preserving protein structure and allowing for efficient glycosylation are not fully elucidated. Bañó-Polo and others, for example, found that insertion of a sequon near the C-terminus or in transmembrane regions generally resulted in substantially lower glycosylation efficiency.<sup>39</sup> The technical challenges in these studies—due to the expression of large numbers of

proteins and use of immunological probes, which often recognize epitopes incorporating peptide backbone,—make it difficult to comprehensively correlate efficiency of glycosylation with the site of the sequon. One example is recent work by Silverman and Imperiali measured glycosylation efficiencies of seven positional mutants of *Campylobacter jejuni* PEB3 with the finding that positions associated with increased surface accessibility and lower thermostability resulted in higher yields for glycosylation.<sup>43</sup> They utilized a radiometric assay strategy, which suffers from significant throughput drawbacks as discussed in chapter 1, in addition to traditional protein expression approach.

An important caveat to our approach is that resolving power of our method is essentially identical to MALDI-TOF. Therefore, the ability to detect and quantify a change in mass relies on that change in mass being of large enough size in relation to the parent peak in order to be resolved by mass spectrometry. 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 glycoforms whose masses are more readily resolved by MALDI-TOF, as in the case of the PglB heptasaccharide with a mass addition of ~1.4 KDa vs 162 Da. Additionally, recent work in Desorption Ionization Electron Spray Mass spectrometry (DESI-MS) shows great promise as a potential alternative to MALDI-MS with greater resolution and overcoming traditional limitations to protein detection in that technology.<sup>201,</sup>  
<sup>202</sup> Therefore, we anticipate that the assay platform described here can be generalized to other glycosylation systems, beyond the challenging example of a single-hexose modification presented here.

In conclusion, we have developed a combined cell-free protein synthesis and SAMDI approach to rapidly generate and assess the glycosylation efficiency of intact glycoproteins. To our knowledge, this is the first example of combining LET-CFPS and SAMDI mass spectrometry to study glycosylation, resulting in a paradigm shift in the throughput with which defined, synthetic glycoproteins can be created and analyzed in a versatile fashion without the need for Western blotting or specific antibodies. We expect that this method will enable fast characterization of positional glycosylation efficiency and aid efforts at designing optimal glycosylation sites in therapeutic proteins. This rapid approach also offers the potential to substantially increase our ability to study structure-activity relationships by providing the throughput and means to rapidly evaluate hundreds of protein substrates without requiring cloning, utilization of antibodies, expression in living cells, or substantial quantities of material.

## **EXPERIMENTAL**

### **Synthesis of NTA-SH ligand**

Preparation NTA-SH ligand for SAM functionalization adapted from Zhang et.al.<sup>203</sup> Reactions were performed in flame-dried glassware under a nitrogen atmosphere using dry, deoxygenated solvents passed over a column of activated alumina. All reagents were purchased from Sigma-Aldrich. Purification was performed on a Waters 600 HPLC purification system. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on a Varian Inova (500MHz) and Bruker Avance III (125MHz), respectively. Data for <sup>1</sup>H spectra is reported in terms of chemical shift relative to Me<sub>4</sub>Si ( $\delta$  0.0 ppm) while that for <sup>13</sup>C NMR is reported relative to CDCl<sub>3</sub> ( $\delta$  77.16 ppm). HRMS

was acquired on an Agilent 6210A LC–TOF mass spectrometer using electrospray ionization (ESI).

Briefly, to a flame-dried flask (10mL, round-bottom) 3,3-Dithiodipropionic acid (30.2mg, 0.14mmol), N-hydroxysuccinimide (41.2mg, 0.36 mmol) and N,N-diisopropylcarbodiimide (48.8uL, 0.31 mmol) in tetrahydrofuran (2.5mL). The resulting mixture was stirred continuously under room temperature for 2 hours resulting in a cloudy white precipitate. Solid precipitate was removed and discarded via vacuum filtration and the solution dried under vacuum until solid white-yellow crude powder remained. To this crude mixture was added triethylamine (46.2uL, 0.629 mmol) and **N<sup>2</sup>,N<sup>2</sup>-Bis(carboxymethyl)-N<sup>6</sup>-(12-mercapto-1-oxododecyl)-L-lysine** (43.5mg, 0.166mmol) **in dimethyl sulfoxide**. The reaction was stirred continuously and allowed to proceed overnight at room temperature. The next day, (25.6mg, 0.15mmol) of Tris(2-carboxyethyl)phosphine hydrochloride was added and the mixture stirred for another 2 hours at room temperature. Purification via HPLC purification (water: acetonitrile:0.1% TFA) following fraction collection and lyophilization resulted in a white-flaky product. Yield was 56.5%. <sup>1</sup>H NMR (500 MHz, Acetone-*d*<sub>6</sub>) δ 7.18 (s, 1H), 3.69 (q, *J* = 18.2 Hz, 5H), 3.58 – 3.46 (m, 1H), 3.24 (q, *J* = 5.7 Hz, 2H), 2.75 (q, *J* = 7.2 Hz, 2H), 2.54 – 2.42 (m, 2H), 1.86 (t, *J* = 8.2 Hz, 2H), 1.73 (q, *J* = 11.4, 9.5 Hz, 1H), 1.65 – 1.47 (m, 5H); <sup>13</sup>C NMR (125 MHz, Acetone) δ 174.37, 171.20, 158.86, 66.33, 55.64, 40.72, 39.41, 24.56, 21.07 (Figures 4.13 and 4.14, respectively).

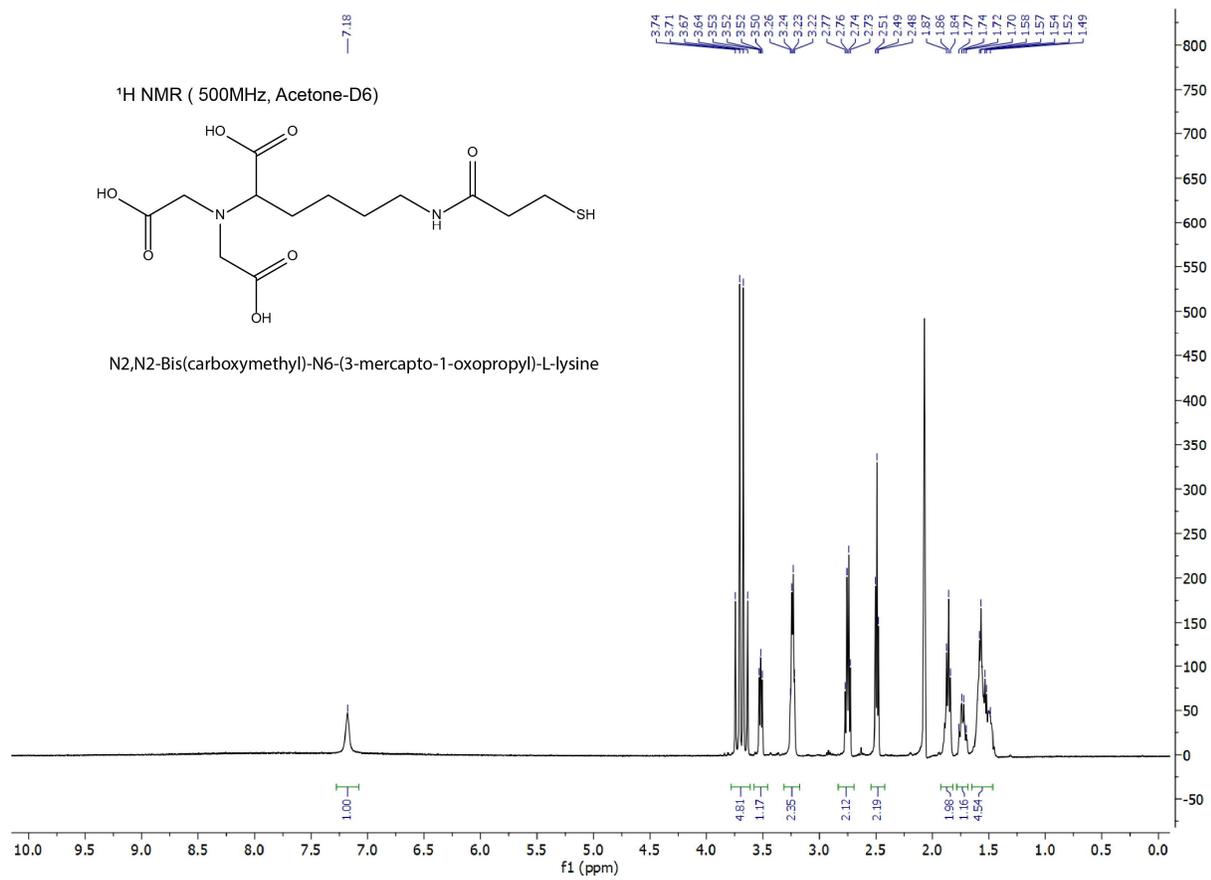
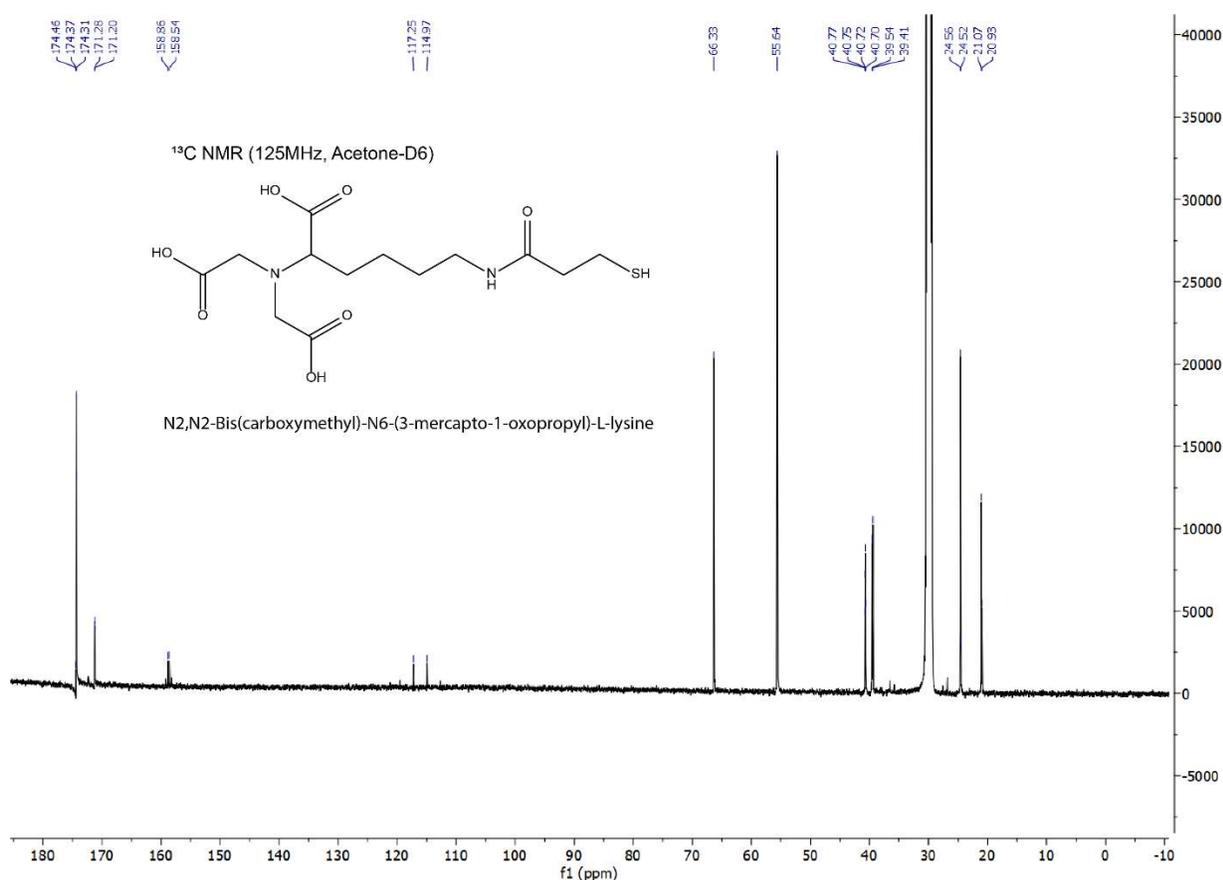


Figure 4.12: <sup>1</sup>H NMR spectrum of NTA-SH ligand



**Figure 4.13:** <sup>13</sup>C NMR spectrum of NTA-SH ligand.

### Preparation of SAMDI Plates

Stainless steel plates with dimensions matching standard MALDI plates (8 cm × 12.3 cm) were cleaned with hexanes, ethanol, and water. A mask with an array of 2.8 mm circles laid out in a standard 384-well format was placed over the plates and an electron beam evaporator (Thermionics Laboratory Inc., Hayward, CA) was used to deposit a layer of titanium (5 nm at 0.02 nm/s) onto the exposed circles, followed by a layer of gold (35 nm at 0.05 nm/s). Plates were stored under vacuum until use. The plates were placed in a solution of two alkyl-disulfides

in ethanol, 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 (both from ProChimia Surfaces, Poland), for approximately 16 h. This is expected to produce monolayers with 10% of the alkanethiolates terminated with maleimide groups. The plates were removed, rinsed with ethanol dried with air prior to use. Plates intended for peptide analysis were used as prepared; plates for protein capture were further treated with an aqueous 100  $\mu$ M NTA-thiol solution in 100mM TRIS-HCL pH 8.0. 2 $\mu$ L was applied to each spot for 30 minutes at ambient conditions, then rinsed with ethanol and dried with air. Next, an aqueous solution of 100mM NiSO<sub>4</sub> was applied to each spot for 10 minutes then rinsed with water only before drying with air. The plates were then used for immobilization of histidine-tagged protein prior to SAMDI-MS analysis.

### **In Vitro Glycosylation Reactions of Proteins**

Reactions with Im7-6: Im7-6 was prepared and purified as previously described.<sup>48</sup> Unpurified NGT was produced in E. Coli CFPS as previously described.<sup>48</sup> For glycosylation reactions 10 $\mu$ M Im7-6, 5 $\mu$ M UDP-Glucose (Sigma Aldrich), and 0.05 $\mu$ M NGT were allowed to react in 1X PBS (Gibco) without Mg or Ca according to indicated amounts and times in a final reaction volume of 5 $\mu$ L. Reactions were quenched with addition of 1:1 of 15% (v/vol) ethanol (Decon Laboratories) and 1X PBS. Concentrations of Im7-6 were determined using Nanodrop 2000c UV absorption at 260nm against blank 1X PBS buffer and NGT concentration were previously determined using [<sup>14</sup>C] leucine incorporation.

Im7 library glycosylation reactions: Im7 proteins were prepared from LET-CFPS with original concentrations determined by [<sup>14</sup>C]leucine incorporation and diluted into final concentration of 10 $\mu$ M in an aqueous buffer of 100mM HEPES pH 7.4 and 16% (v/vol) CFPS reaction (the varying CFPS reaction volume was filled by a completed CFPS reaction that synthesized sfGFP) in a separate 96 well plate (Dot-Bioscientific). In vitro glycosylation reactions were then assembled in a new 96 well plate, each well containing one 2.5 $\mu$ M of Im7 variant, 0.05 $\mu$ M (for non-saturated) or 0.667 $\mu$ M NGT (for saturated), 5mM UDP-Glucose with 100mM HEPES pH 7.4. Each reaction was performed in a final 4 $\mu$ L volume, 2.5 $\mu$ L of diluted Im7 variant and 1.5 $\mu$ L of a reagent mix containing NGT, UDP-Glucose, and HEPES pH 7.4 at 2.67X stock. To quench glycosylation reactions 4 $\mu$ L of 15% (v/vol) ethanol was added for a final volume of 8 $\mu$ L.

PglB glycosylation reaction: All reagents utilized were prepared as previously described.<sup>181</sup> PglB expressed in CFPS, crude LLOs in *E. coli* membrane fraction, and purified acceptor protein were combined. Glycosylation reactions were composed of 4 $\mu$ L of PglB expressed in CFPS, 4 $\mu$ L of crude LLOs in *E. coli* membrane fraction, 1 $\mu$ L of purified Im7-DQNAT or Im7-AQNAT, and 1 $\mu$ L of a master mix (10% (w/v) Ficoll 400, 500mM HEPES, pH 7.4, and 10mM manganese chloride) was added just prior to incubation to reduce precipitation of reaction components. Reactions were then incubated overnight at 30°C.

Therapeutic protein glycosylation reactions: 2.5 $\mu$ L of original CFPS reactions were added to 7.5 $\mu$ L of IVG buffer (100mM HEPES pH 7.4, 0.05% Tween-20). In vitro glycosylation reactions were then assembled in a new 96 well plate, each well containing 1 $\mu$ L of diluted therapeutic protein variant, 0.1 $\mu$ M NGT, and 2.5mM UDP-Glucose with 100mM HEPES pH

7.4, 0.05% Tween-20. Each reaction was performed in a final 4 $\mu$ L volume, 2.5 $\mu$ L of diluted Im7 variant and 1.5 $\mu$ L of a reagent mix containing NGT, UDP-Glucose, and IVG buffer at 2.67X concentrated stock. To quench glycosylation reactions 4 $\mu$ L of 15% (v/vol) ethanol was added for a final volume of 8 $\mu$ L.

### **Analysis of GGNWTT Adjacent Amino Acid Preferences of NGT**

Peptides were screened with SAMDI-MS according to the method previously described.<sup>48</sup> After reduction with tris(2-carboxyethyl)phosphine (TCEP) reducing gel (Thermo Fisher Scientific), 50  $\mu$ M of each peptide was allowed to react with 2.5mM UDP-Glucose in the presence of unpurified NGT produced in *E. coli* CFPS, (0.02 $\mu$ M for X<sub>3</sub> peptides or 0.04 $\mu$ M for X<sub>4</sub> peptides) in 100 mM HEPES pH 8.0 for 30 minutes at 30°C. As a control, the same volume of CFPS after 20 h of sfGFP synthesis was used instead of NGT from CFPS. The reaction was not quenched. 2  $\mu$ L TCEP reducing gel were added to each 10 $\mu$ L of reaction solution and incubated at 37 °C for 1 h. 2  $\mu$  L of the reduced solutions were then transferred to a 384-well SAMDI plate using Tecan 96-channel arm and incubated at room temperature (~25°C) for 0.5 h. SAMDI plates were washed with two rounds of water and ethanol, and dried with flowing nitrogen. After application of 10 mg/mL of 2',4',6' -trihydroxyacetophenone monohydrate (THAP) matrix (Sigma-Aldrich) in acetone onto the entire SAMDI plate, an Applied Biosystems SciEx MALDI-TOF/TOF 5800 instrument was used to perform mass spectrometry on each spot. Applied Biosystems SciEx Time of Flight Series Explorer Software version 4.1.0 was used to analyze MS spectra.

## Mass Analysis of Proteins by SAMDI-MS

MALDI plates displaying Ni-NTA monolayers are prepared as described above. A volume of 1  $\mu$ L of sample mixture to be analyzed by SAMDI was dispensed onto each spot and incubated for 30 minutes at room temperature, then rinsed extensively with water and dried with compressed air. 0.5  $\mu$ L of 10 mg/mL of Sinapic Acid (SA) matrix (Sigma-Aldrich) in 49% Acetonitrile/49.9% H<sub>2</sub>O/0.1% TFA onto each monolayer spot and allowed to dry in ambient conditions. Mass analysis was performed on an Applied Biosystems SciEx (AB SciEx) MALDI-TOF/TOF 5800 (Framingham, MA, USA) instrument. A 349 nm Nd:YAG laser was used as a desorption/ionization source. The accelerating voltage was 15 kV and the extraction delay time was 300 ns. All spectra were acquired in positive linear mode. A total of 2000 laser shots (at a pulse rate of 400 Hz) were manually collected and averaged over a spot for the acquisition of 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 spectral acquisition using Opti-TOF Cal Mix 3 Plus High Mass Calibration Insert (AB SciEx). The modification efficiency of proteins was calculated using the following equation.

$$\text{Modification \%} = 100 \times \frac{\text{Product}_{\text{AUC}}}{\text{Product}_{\text{AUC}} + \text{Substrate}_{\text{AUC}}}$$

Where Product<sub>(AUC)</sub> corresponds to sum of protonated(Glc<sub>0</sub>) and sinapic acid (Glc<sub>0</sub>-SA) adducts of the modified protein species. Substrate<sub>(AUC)</sub> corresponds to sum of protonated(Glc<sub>1</sub>) and sinapic acid (Glc<sub>1</sub>-SA) adducts of the unmodified protein species. The protonated species of

the modified protein and the sinapic acid adduct of the unmodified protein overlap, which was taken into account using the following equations to generate  $\text{Product}_{\text{AUC}}$  and  $\text{Substrate}_{\text{AUC}}$  values.

$$\begin{aligned}\mathbf{Product}_{\text{AUC}} &= [\mathbf{Glc}_1]_{\text{AUC}}^+ + [\mathbf{Glc}_1 - \mathbf{SA}]_{\text{AUC}}^+ - \mathbf{p}[\mathbf{Glc}_0 - \mathbf{SA}]_{\text{AUC}}^+ \\ \mathbf{Substrate}_{\text{AUC}} &= [\mathbf{Glc}_1]_{\text{AUC}}^+ + [\mathbf{Glc}_1 - \mathbf{SA}]_{\text{AUC}}^+ + \mathbf{p}[\mathbf{Glc}_0 - \mathbf{SA}]_{\text{AUC}}^+ \\ \mathbf{p}[\mathbf{Glc}_0 - \mathbf{SA}]_{\text{AUC}}^+ &= [\mathbf{Glc}_0]_{\text{AUC}}^+ \times \mathbf{Fraction}_{(\text{SA})}\end{aligned}$$

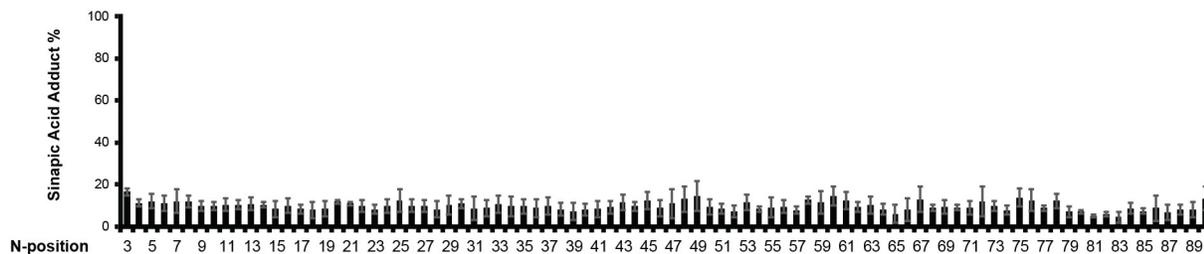
Where  $\mathbf{p}[\mathbf{Glc}_0 - \mathbf{SA}]$  is the predicted amount of unmodified protein species present in its sinapic acid adduct overlapping with the protonated species of the modified protein.  $\mathbf{p}[\mathbf{Glc}_0 - \mathbf{SA}]$  is calculated from the relative fraction of sinapic acid adduct detected for each protein species.

### Measuring Fraction of Sinapic Acid Adduct for Im7 proteins

Each Im7 protein was analyzed by SAMDI without supplementation of UDP-Glucose or NGT to determine the relative fraction of sinapic acid adduct present for each sample (**Figure 4.15**). Sinapic acid fractions  $\text{Fraction}_{(\text{SA})}$  were calculated using the equation below.

$$\mathbf{Fraction}_{(\text{SA})} = \frac{[\mathbf{Protein} - \mathbf{SA}]_{\text{AUC}}^+}{[\mathbf{Protein}]_{\text{AUC}}^+ + [\mathbf{Protein} - \mathbf{SA}]_{\text{AUC}}^+}$$

Where (Protein-SA) and (Protein) correspond to the sinapic acid and singly protonated adduct species, respectively.  $\text{Fraction}_{(\text{SA})}$  for the Im7 library are provided in **Figure 4.15**.



**Figure 4.14: Sinapic Acid Adduct Peak Area % by SAMDI-MS.** SAMDI-MS of 2.5 $\mu$ M of synthesized Im7-library with 10mg/mL Sinapic acid in 50:49.9:0.1% mixture of acetonitrile/water/trifluoroacetic acid. Im7-library CFPS yields were previously determined by radioisotope incorporation. (Error bars) One standard deviation from n=3.

## DNA Template preparation

All DNA templates and plasmid used in this study are summarized in supplemental table. The coding sequences of plasmids and DNA templates generated during this study are also provided in an additional supporting information file. Plasmid DNA encoding *Actinobacillus pleuropneumonia* (NGT) and Im7-6 were prepared using standard molecular biology techniques and then purified from transformed DH5-alpha *E. coli* strains and Zymo Midiprep kits as previously elsewhere.<sup>204</sup> DNA templates containing positional substitutions of GGNWTT at each amino acid position in wildtype Im7 sequence as well as engineered forms of Leptin, IL1-RA, interferon alfa, and G-CSF were designed and synthesized by Twist Biosciences. For the Im7 library, each template contained a C-terminal 6xHis-tag and a short 5' untranslated region (UTR) sequence containing a portion of the ribosome binding site. Leptin, IL1-RA, IFN-alpha, and G-CSF were designed with an N-terminal 6x-Histag followed by a glycosylation tag previously optimized for NGT modification. The native N-X-S/T glycosylation site on IL1-RA was also disabled by an N185Q mutation. To generate the linear DNA templates for LET-CFPS

reactions, T7 promoter and terminator regions were added to template genes using PCR amplification with primers 502 and 503 (sequences provided in Table S1).

### **Preparation of cell extracts for CFPS**

Crude extracts for CFPS were generated from *E. coli* strain BL21 (DE3)-Star (NEB). Cell growth, harvest, and lysis were performed as previously described.<sup>188</sup> Briefly, BL21(DE3) *E. coli* cells were grown in 1 L of 2xYTPG (yeast extract 10 g/L, tryptone 16 g/L, NaCl 5 g/L, K<sub>2</sub>HPO<sub>4</sub> 7 g/L, KH<sub>2</sub>PO<sub>4</sub> 3 g/L, and glucose 18 g/L, pH 7.2) in a 2.5 L Tunair flask at 37°C and 250 r.p.m. with initial inoculation to OD<sub>600</sub> = 0.08. At OD<sub>600</sub> = 0.6-0.8, production of the T7 RNA polymerase was induced by the addition of 1 mM IPTG. At OD<sub>600</sub> = 3.0, cells were pelleted by centrifugation at 5,000 × g at 4 °C for 15 min. The pellets were then washed three times with cold S30 buffer (10 mM Tris–acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, and 2 mM dithiothreitol (DTT)) and flash frozen on liquid nitrogen and stored at –80 °C. Cells were thawed, resuspended in 1.0 mL of S30 buffer per gram wet weight, and lysed in 1 mL aliquots on ice using a Q125 Sonicator (Qsonica) at 50% amplitude, 10 seconds on and 10 seconds off until 700 J had been delivered. After sonication, 3 μL of DTT (1 M) was added and the cell lysate was centrifuged at 12,000 × g and 4 °C for 10 min. The supernatant was then clarified again by an additional centrifugation at 10,000 × g and 4 °C for 10 min. The supernatant was flash-frozen on liquid nitrogen and stored at –80 °C until use.

### Cell-free protein synthesis

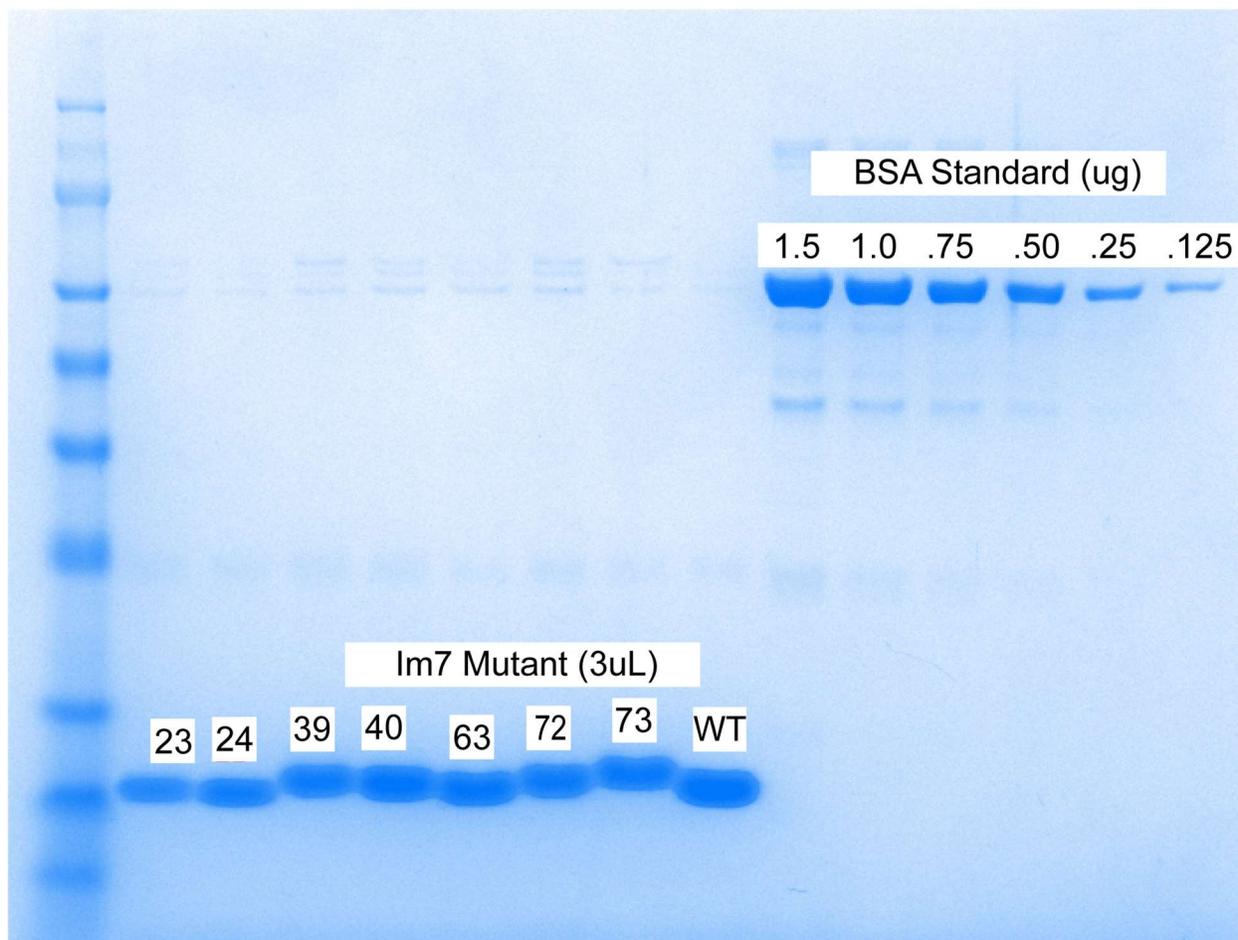
CFPS reactions were conducted using a PANOX-SP crude lysate system.<sup>189</sup> A standard reaction contained 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 34 µg/mL folinic acid; 170 µg/mL of *E. coli* tRNA mixture; 2 mM for each of the 20 standard amino acids; 0.33 mM nicotinamide adenine dinucleotide (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; 12 mM magnesium glutamate; 57 mM HEPES, pH 7.2; 33 mM phosphoenolpyruvate (PEP); and 27% v/v of *E. coli* cell extract. DNA templates encoding proteins to be synthesized were added either in the form of plasmid DNA at a concentration of 13.3 µg/mL or the addition of 3 µL linear expression template PCR product, as indicated. *E. coli* total tRNA mixture (from strain MRE600) and phosphoenolpyruvate were purchased from Roche Applied Science.<sup>205</sup> ATP, GTP, CTP, UTP, the 20 amino acids, and other materials were purchased from Sigma-Aldrich. NGT and AGT were synthesized as previously described.<sup>204</sup> The CFPS reactions were carried out at 30 °C for 6 hours.

### 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 [<sup>14</sup>C]leucine (PerkinElmer). Protein quantification for duplicate CFPS reactions was completed using trichloroacetic acid (TCA) protein precipitation followed by radioactivity quantification using a Microbeta2 liquid scintillation counter (PerkinElmer) according to established protocols.<sup>189</sup> Soluble fractions were taken after centrifugation at 12,000 × g for 15 min at 4 °C.

### **Protein Purification from CFPS for Circular Dichroism Experiments**

Proteins selected for circular dichroism analysis were prepared from multiple individual CFPS reactions and then pooled together for each respective mutant. Pooled protein reactions were then purified using Dyna-His tag beads (Thermo Fisher Scientific) according to manufacturer's protocol. After elution with 40 $\mu$ L with 500mM imidazole, the samples were dialyzed against 50 mM NaH<sub>2</sub>PO<sub>4</sub> and 150 mM NaCl, pH 7.5 in 3.5 kDa MWCO 96-well plate dialysis cassettes (Thermo Fisher Scientific). After dialysis, 3  $\mu$ L was analyzed by standard SDS-PAGE gel analysis confirmed to confirm purity. Concentrations were quantified for purity using Image Lab software version 6.0.0 densitometry with BSA standard curve after separation by SDS-PAGE, staining with InstantBlue Coomassie stain, and destaining in water (**Figure 4.15**).



**Figure 4.15: SDS-PAGE of purified Im7 mutants for circular dichroism experiment.** 3uL of each purified sample were separated by SDS-PAGE, stained with InstantBlue Coomassie stain, and destained in water. A standard curve of BSA was also analyzed and used to determine purity (>95%) by densitometry.

### Circular Dichroism

All protein samples were prepared in 50mM sodium phosphate buffer. Concentrations were determined by UV absorption at 260nm using a Nanodrop 2000 spectrophotometer and samples prepared at a final concentration of 13 $\mu$ M. CD spectra were collected using a Jasco J-810 spectrometer over the wavelength range of 190–280 nanometers using a 1.0-mm path length quartz cuvette. Scans were acquired at a rate of 1 nm/min; each spectrum represents the

accumulation of three scans. The temperature of the cuvette was maintained at 25°C throughout analysis to minimize solvent evaporation and temperature fluctuations. All CD spectra were baseline-corrected of the protein samples by subtracting the 50mM sodium phosphate buffer spectrum. Baseline-corrected CD spectra were converted from raw ellipticity ( $\theta$ , mdeg) to mean residue ellipticity using the following equation.

$$MRE = \frac{\theta \times MW}{10 \times N \times C \times L}$$

Where MW is the molecular weight of the protein (g/mol), C is the protein concentration ( $\mu\text{mol/L}$ ), L is the optical pathlength through buffer (cm), and N is the number of amino acid residues.

### **LC–TOF analysis of glycoprotein modification**

Proteins expressed by CFPS and then incubated in IVG reactions were purified using Dyna-His tag Ni-NTA magnetic beads (Thermo Fisher Scientific) and analyzed by LC-TOF in a manner previously described<sup>48</sup>. Briefly, samples were diluted 1 to 4 in Buffer 1 (50 mM NaH<sub>2</sub>PO<sub>4</sub> and 300 mM NaCl, pH 8.0) for protein binding followed by washing with Buffer 1 and elution using 35  $\mu\text{l}$  of Buffer 1 + 500 mM imidazole. After elution, proteins were dialyzed into Buffer 1 diluted 1:4 in H<sub>2</sub>O overnight and then 5  $\mu\text{l}$  was injected Bruker Elute UPLC equipped with an ACQUITY UPLC Peptide BEH C4 Column, 300Å, 1.7  $\mu\text{m}$ , 2.1 mm X 50 mm (186004495 Waters Corp.) with a 10 mm guard column of identical packing (186004495 Waters Corp.) coupled to an Impact-II UHR TOF Mass Spectrometer (Bruker Daltonics, Inc.). The chromatographic separation method was based on the manufacturer's instructions for XBridge column with minor modifications. Liquid chromatography was performed using 100% H<sub>2</sub>O and

0.1% formic acid as Solvent A and 100% acetonitrile and 0.1% formic acid as Solvent B at a flow rate of 0.5 mL/min and a 50°C column temperature. An initial condition of 20% B was held for 1 min before elution of the proteins of interest during a 4 min gradient from 20% to 50% B. The column was washed and equilibrated by 0.5 min at 71.4% B, 0.1 min gradient to 100% B, 2 min wash at 100% B, 0.1 min gradient to 20% B, and then a 2.2 min hold at 20% B, giving a total 10 min run time. An MS scan range of 100-3000 m/z with a spectral rate of 2 Hz was used. External calibration was performed prior to data collection.

### **LC-MS analysis of glycopeptide modification**

The method used for the analysis of glycopeptides by LC-MS was like the method used for glycoproteins above. Briefly, His-tag purified, dialyzed glycosylation targets were digested with 0.0044 µg/µl MS Grade Trypsin (Thermo Fisher Scientific) at 37°C overnight. LC-MS was performed by injection of 5 µl of digested glycopeptides into a Bruker Elute UPLC equipped with an ACQUITY UPLC Peptide BEH C18 Column, 300Å, 1.7 µm, 2.1 mm X 100 mm (186003686 Waters Corp.) with a 10 mm guard column of identical packing (186004629 Waters Corp.) coupled to an Impact-II UHR TOF Mass Spectrometer. Liquid chromatography was performed using 100% H<sub>2</sub>O and 0.1% formic acid as Solvent A and 100% acetonitrile and 0.1% formic acid as Solvent B at a flow rate of 0.5 mL/min and a 40°C column temperature. An initial condition of 0% B was held for 1 min before elution of the peptides of interest during a 4 min gradient to 50% B. The column was washed and equilibrated by a 0.1 min gradient to 100% B, a 2 min wash at 100% B, a 0.1 min gradient to 0% B, and then a 1.8 min hold at 0% B, giving a

total 9 min run time. For LC-MS of glycopeptides, a scan range of 100-3000 m/z with a spectral rate of 8 Hz was used. External calibration was performed prior to data collection.

### **LC-TOF data analysis**

LC-MS data was 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 glycoforms (as determined by extracted ion chromatograms of theoretical glycopeptide and glycoprotein charge states with tolerances of  $\pm 0.05$  m/z). 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<sup>2</sup>. Briefly, the masses of the three most abundant charge states of both the glycosylated and unglycosylated forms of the protein were used to generate extracted ion chromatograms which were then integrated and used to quantify relative peak areas to measure glycosylation efficiency. LC-MS spectra and chromatograms for figures was obtained by export from Compass Data Analysis.

### **DNA Sequences.**

Plasmids and DNA sequences used in this study with sources and details are shown in **Table 4.1** or in Appendix C. Primer sequences used for preparation of Im7 library for LET-CFPS are provided in **Table 4.2**. Codon optimized sequences (using GeneArt) for Metreleptin, Filgrastim, Interferon-alfa and Anakinra with customized additional N-terminal leader sequence encoding both the polyhistidine (6x) histidine tag and the glycosylation GGNWTT sequence

were synthesized by Twist Bioscience and assembled into PJL1 vector. Im7-library genes were codon optimized and synthesized by Twist Bioscience. Primer sequences were synthesized from Integrated DNA Technologies.

**Table 4.1: Plasmid and Strains.** Plasmid and strains used in this study.

Plasmid and Strain	Relevant Characteristics	Source
<b>Strains</b>		
DH5- $\alpha$	<i>fhuA2</i> $\Delta$ ( <i>argF-lacZ</i> )U169 <i>phoA</i> <i>glnV44</i> $\Phi$ 80 $\Delta$ ( <i>lacZ</i> )M15 <i>gyrA96</i> <i>recA1</i> <i>relA1</i> <i>endA1</i> <i>thi-1</i> <i>hsdR17</i>	New England Biolabs
BL21 Star (DE3)	<i>F-ompT</i> <i>hsdSB</i> ( <i>rB-</i> , <i>mB-</i> ) <i>galdcmrne131</i> (DE3)	Thermo Fisher Scientific
CLM24	W3110 $\Delta$ <i>wecA</i> $\Delta$ <i>waaL::Kan</i>	Thapakorn and Stark, et. al. <sup>182</sup>
<b>Plasmids</b>		
pJL1.sfGFP	Kan <sup>R</sup> , <i>P</i> <sub>T7</sub> ,super folder green fluorescent protein (sfGFP), C-term strep-tag	Bundy and Swartz <sup>206</sup>
pJL1.NGT	<i>A. pleuropneumoniae</i> NGT, (NGT_ACTP2)	Kightlinger and Lin, et. al. <sup>48</sup>
pJL1.Im7-6	<i>E. coli</i> Im7, (IMM7_ECOLX), 26_31=ATTGGNWTTAGG, C-term 6xHis-tag	Kightlinger and Lin, et. al. <sup>48</sup>
pSF-pglB	PglB Plasmid for <i>E. coli</i> expression	Schoborg et. al. <sup>181</sup>
pACYCpglB::kan	LLO Plasmid for <i>E. coli</i> expression	Schoborg et. al. <sup>181</sup>
pJL1.Im7.DQNAT	Target Plasmid OST glycosylation in vitro	Schoborg et. al. <sup>181</sup>
pJL1.His.NGTtag.Lep	Metreleptin, 1InsMEKKIHHHHHHGSKATTGGNWTTAGGKGS	This Study (Appendix C.2)
pJL1.His.NGTtag.G-CSF	Filgrastim, 1InsMEKKIHHHHHHGSKATTGGNWTTAGGKGS	This Study (Appendix C.2)
pJL1.His.NGTtag.IFN-alpha	Interferon-alfa, 1InsMEKKIHHHHHHGSKATTGGNWTTAGGKGS	This Study (Appendix C.2)

pJL1.His.NGTtag.IL1-RA	Anakinra, 1InsMEKKIHHHHHHGSKATTGGNWTTAGGKG S, N185Q	This Study (Appendix C)
<b>Linear Im7 DNA Sequences</b>		
Im7 P-X (where X=1-87)	<i>E. coli</i> Im7, (IMM7_ECOLX), X_X+12=GGNWTT, C-term 6xHis-tag	This Study (Appendix C)

<b>Primer Name</b>	<b>Primer Sequence (5'→3')</b>
503F_LAmp_Long	gcgaaattaatacgactcactatagggagaccacaacggttccctctagaataatttGTTTAA CTTTAAGAAGGAGATATACATATG
502R_LAmp_Long_Histag	ggctttcagcaaaaaaccctcaagaccgtttagaggccccaaggggttatgctaggtcgacTTA ATGGTGATGATGGTGATG

**Table 4.2: Primer Sequences for LET-CFPS.** Capital letters indicate area binding to 5' (untranslated region and start codon) or 3' (His-tag and stop codon) of synthesized DNA template.

## Chapter 5

### Summary, Final Thoughts and Future Implications

In this dissertation I explored the application of SAMDI-MS to measuring glycosyltransferase activities on protein substrates as well as extending it to profiling eukaryotic cellular activities. These are both natural extensions of the SAMDI-MS platform as well as essential analytical problems relevant to multiple fields of glycobiology. In chapter 3, a pilot project successfully demonstrated SAMDI-MS ability to detect endogenous activity from breast cancer cell line MDA-MB-231, an important model of metastasis, that correlates with known data on galactosyltransferase upregulation in breast cancer. This study utilized copper-free click chemistry that is orthogonal to known functional groups as the immobilization strategy and both solid-phase “chip” and more standard solution-phase demonstrated similar activity correlations. In chapter 4, SAMDI-MS was paired with cell-free protein synthesis to generate and screen a large his-tagged protein library for glycosylation efficiency by ApNGT, a cytoplasmic N-glycosyltransferase. This work was able to monitor modification of intact proteins with a single monosaccharide, for which no antibody label currently exists, with higher resolution and throughput than current technology currently provides. Despite the steps forward in these studies there is still significant progress to be made in exploring questions of glycosylation positional efficiency as well as studying glycosyltransferase profiles using SAMDI-MS.

Activity-based profiling (ABP) is a growing approach to characterizing complex biological systems. ABP aims to only characterize the active forms of particular enzymes using chemical probes in contrast to standard proteomic and genomic profiling, which indicate levels of a protein present regardless of its functional state. This approach was popularized by

Benjamin Cravatt and has largely been utilized in the realm of proteomics.<sup>207, 208</sup> However, SAMDI-MS coupled with large-scale peptide libraries was previously demonstrated to be useful in developing activity profiles for lysine deacetylases.<sup>137, 138</sup> Subsequently the idea of using SAMDI-MS to explore glycosyltransferase activities in relevant biological context, such as breast cancer epithelial-mesenchymal transitions, could be feasible. Chapter 3 was a pilot study to determine if SAMDI-MS could be feasible to detect endogenous levels of glycosyltransferase activity from lysates as opposed to overexpressed lysates or in-vitro purified enzymes, as previously done.<sup>116, 165</sup> I was successful at demonstrating this feasibility although the lack of a diverse and large enough acceptor library hindered further research efforts. Future work centering on O-glycosylation on polypeptide substrates should be feasible and doable, especially given the success that recent work on characterizing ppGalNAc transferases using SAMDI-MS.<sup>48</sup>

Intact glycoprotein analysis is a challenging and complex analytical problem for any glycobiology lab. Standard practice utilizes immunological probes in western blot or enzyme-linked immunosorbent assays (ELISA) to determine if glycosylation occurred, as detected through a mass shift. However, these approaches typically lack resolution to detect single monosaccharide shifts on intact proteins whereas mass spectrometry does not share this same limitation (Figure 5.1). This analytical challenge is important when investigating questions regarding positional differences of glycosylation in protein substrates. Important early work using intact protein substrates has been focused almost exclusively with oligosaccharyltransferase because the transfer of a large glycan is more readily detected by western blot. For single-monosaccharide glycosyltransferases, small molecule substrates such as peptides are used instead although these often lack significant secondary structure that can alter

glycosyltransferase preferences. The introduction a SAMDI-MS based method that can allow ready detection of monosaccharide posttranslational modification of intact proteins is an enabling technology that should these questions to be readily approached for single-monosaccharide glycosyltransferases. One example of such an approach would be to select and prepare libraries of small proteins with different secondary structures and introduced minimal N-X-T glycosylation sites. To account for intrinsic glycosylation efficiency differences between sequons, a tiling peptide approach can be taken in parallel to mimic each sequon position. In this way, the structural preferences of the glycosyltransferase in question can be readily characterized. The SAMDI-MS technique introduced in chapter 4 can be readily used for such a study, particularly in combination with powerful Linear Expression Template- Cell free protein synthesis strategy.

Aside from addressing pressing questions of acceptor glycosylation, ongoing work in the Mrksich lab have provided some technical advancements relevant to these analytical challenges that can be essential in work going forward. First, development of higher density SAM plates and MALDI-TOF imaging present an order of magnitude higher in sample processing than currently existed just a year or two ago. During my tenure in the lab, 384 and 1,536 gold-spot plates were standard for SAMDI-MS based projects. However, there now exists 6,144 gold-spot plates and SAMD-MS imaging technology.<sup>209, 210</sup> Both technological advancements present the possibility of extremely high throughput analysis of in-vitro glycosylation as well as other posttranslational modifications. These high-density chips and imaging technology are currently limited to analysis of small-molecules covalently bound to the surface SAM and not larger protein substrates. Pairing with matrix-deposition optimization technology originally designed

for protein imaging applications, such as that demonstrated with the TM-SPRAYER (see Appendix 1), presents the possibility to extend these technological developments to analysis of non-covalently bound protein substrates such as that demonstrated in chapter 4 of this dissertation. Additionally, parallel developments of surface-based mass spectrometry techniques outside of the Mrksich group offer future opportunities for analysis of SAM-bound substrates. Desorption ESI (DESI) uses the stream of charged solvent from an electrospray source to solubilize and desorb molecules from the MS surface.<sup>211</sup> Like MALDI-TOF, sample preparation is minimal, and the solvent composition can be varied to enrich for specific molecules, an essential concern with glycosylated substrates. Already a variant of DESI has been used to analyze substrates from an captured by antibody-coated surface plasmon resonance (SPR) chip.<sup>212</sup> Given the non-covalent nature of antibody capture as well as the SAMDI study in chapter 4, one would suppose that DESI could likewise be used to analyze glycosylation on intact proteins.

A chief hurdle to wider adoption of SAMDI-MS as an in-house technique in many laboratories is the cost and maintenance of MALDI-TOF mass spectrometers. Top of the line mass spectrometers cost as much or more than a standard National Institutes of Health RO1 grant provides to an independent investigator, and thus the primary academic consumers are core-facilities, limiting their use for less frequent experimental studies. However, new developments have been made in the development of benchtop MALDI-TOF mass spectrometers, such as the microflex™ LRF by Bruker, driven in part by the demands of clinical biomarker assay development.<sup>213</sup> While these mass spectrometers have overall lower resolution than standalone MALDI-TOFs, they herald an ongoing cost decrease and wider adoption of MALDI-TOF

technology. Thus SAMDI-MS as a standard workhorse technique may only be a matter of time and ride the wave of decreasing MALDI-TOF technology costs. This will be particularly useful for the glycobiology field where analytical challenges have hampered its development compared to other biological fields.

In this dissertation, I have successfully applied the SAMDI-MS assay to report endogenous activity of glycosyltransferase activities from human cell lines as well as demonstrated its quantitative ability to characterize what reasonably represents the smallest mass shift required for glycosylation status on intact protein substrates. This has established a platform that can be used for further structure-function studies with respect to a wide array of possible glycosylation activities in what I hope will be an enabling technique for glycobiology laboratories.

## References

- Wolff, S. P.; Jiang, Z. Y.; Hunt, J. V., Protein glycation and oxidative stress in diabetes mellitus and ageing. *Free Radic Biol Med* **1991**, *10* (5), 339-52.
2. Pearson, J. S.; Giogha, C.; Ong, S. Y.; Kennedy, C. L.; Kelly, M.; Robinson, K. S.; Lung, T. W.; Mansell, A.; Riedmaier, P.; Oates, C. V.; Zaid, A.; Muhlen, S.; Crepin, V. F.; Marches, O.; Ang, C. S.; Williamson, N. A.; O'Reilly, L. A.; Bankovacki, A.; Nachbur, U.; Infusini, G.; Webb, A. I.; Silke, J.; Strasser, A.; Frankel, G.; Hartland, E. L., A type III effector antagonizes death receptor signalling during bacterial gut infection. *Nature* **2013**, *501* (7466), 247-51.
3. Li, S.; Zhang, L.; Yao, Q.; Li, L.; Dong, N.; Rong, J.; Gao, W.; Ding, X.; Sun, L.; Chen, X.; Chen, S.; Shao, F., Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. *Nature* **2013**, *501* (7466), 242-6.
4. Gao, X.; Wang, X.; Pham, T. H.; Feuerbacher, L. A.; Lubos, M. L.; Huang, M.; Olsen, R.; Mushegian, A.; Slawson, C.; Hardwidge, P. R., NleB, a bacterial effector with glycosyltransferase activity, targets GAPDH function to inhibit NF-kappaB activation. *Cell Host Microbe* **2013**, *13* (1), 87-99.
5. Zarschler, K.; Janesch, B.; Pabst, M.; Altmann, F.; Messner, P.; Schaffer, C., Protein tyrosine O-glycosylation--a rather unexplored prokaryotic glycosylation system. *Glycobiology* **2010**, *20* (6), 787-98.
6. Kolarich, D.; Jensen, P. H.; Altmann, F.; Packer, N. H., Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc* **2012**, *7* (7), 1285-98.
7. Zhang, P.; Woen, S.; Wang, T.; Liao, B.; Zhao, S.; Chen, C.; Yang, Y.; Song, Z.; Wormald, M. R.; Yu, C.; Rudd, P. M., Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today* **2016**, *21* (5), 740-765.
8. Smilowitz, J. T.; Lebrilla, C. B.; Mills, D. A.; German, J. B.; Freeman, S. L., Breast milk oligosaccharides: structure-function relationships in the neonate. *Annu Rev Nutr* **2014**, *34*, 143-69.
9. Varki, A., Biological roles of glycans. *Glycobiology* **2017**, *27* (1), 3-49.
10. Chung, C. H.; Mirakhor, B.; Chan, E.; Le, Q. T.; Berlin, J.; Morse, M.; Murphy, B. A.; Satinover, S. M.; Hosen, J.; Mauro, D.; Slebos, R. J.; Zhou, Q.; Gold, D.; Hatley, T.; Hicklin, D. J.; Platts-Mills, T. A., Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *N Engl J Med* **2008**, *358* (11), 1109-17.
11. 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., Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat Biotechnol* **2003**, *21* (4), 414-21.
12. Jennewein, M. F.; Alter, G., The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol* **2017**, *38* (5), 358-372.

13. Shinkawa, T.; Nakamura, K.; Yamane, N.; Shoji-Hosaka, E.; Kanda, Y.; Sakurada, M.; Uchida, K.; Anazawa, H.; Satoh, M.; Yamasaki, M.; Hanai, N.; Shitara, K., The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J Biol Chem* **2003**, *278* (5), 3466-73.
14. Walsh, G., Biopharmaceutical benchmarks 2018. *Nat Biotechnol* **2018**, *36* (12), 1136-1145.
15. Walsh, G., Biopharmaceutical benchmarks 2014. *Nat Biotechnol* **2014**, *32* (10), 992-1000.
16. Nothaft, H.; Szymanski, C. M., Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem* **2013**, *288* (10), 6912-20.
17. Schwarz, F.; Aebi, M., Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* **2011**, *21* (5), 576-82.
18. Grass, S.; Buscher, A. Z.; Swords, W. E.; Apicella, M. A.; Barenkamp, S. J.; Ozchlewski, N.; St Geme, J. W., The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Molecular Microbiology* **2003**, *48* (3), 737-751.
19. Grass, S.; Lichti, C. F.; Townsend, R. R.; Gross, J.; St Geme, J. W., 3rd, The Haemophilus influenzae HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog* **2010**, *6* (5), e1000919.
20. Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M., N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* **2002**, *298* (5599), 1790-3.
21. Young, N. M.; Brisson, J. R.; Kelly, J.; Watson, D. C.; Tessier, L.; Lanthier, P. H.; Jarrell, H. C.; Cadotte, N.; St Michael, F.; Aberg, E.; Szymanski, C. M., Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *J Biol Chem* **2002**, *277* (45), 42530-9.
22. Szymanski, C. M.; Yao, R.; Ewing, C. P.; Trust, T. J.; Guerry, P., Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol* **1999**, *32* (5), 1022-30.
23. Linton, D.; Allan, E.; Karlyshev, A. V.; Cronshaw, A. D.; Wren, B. W., Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in Campylobacter jejuni. *Mol Microbiol* **2002**, *43* (2), 497-508.
24. Reid, C. W.; Stupak, J.; Chen, M. M.; Imperiali, B.; Li, J.; Szymanski, C. M., Affinity-capture tandem mass spectrometric characterization of polyprenyl-linked oligosaccharides: tool to study protein N-glycosylation pathways. *Anal Chem* **2008**, *80* (14), 5468-75.
25. Reid, C. W.; Stupak, J.; Szymanski, C. M.; Li, J., Analysis of bacterial lipid-linked oligosaccharide intermediates using porous graphitic carbon liquid chromatography-electrospray ionization mass spectrometry: heterogeneity in the polyisoprenyl carrier revealed. *Anal Chem* **2009**, *81* (20), 8472-8.

26. Kelly, J.; Jarrell, H.; Millar, L.; Tessier, L.; Fiori, L. M.; Lau, P. C.; Allan, B.; Szymanski, C. M., Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J Bacteriol* **2006**, *188* (7), 2427-34.
27. Alaimo, C.; Catrein, I.; Morf, L.; Marolda, C. L.; Callewaert, N.; Valvano, M. A.; Feldman, M. F.; Aebi, M., Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J* **2006**, *25* (5), 967-76.
28. Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M., N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **2006**, *314* (5802), 1148-50.
29. Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M., Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* **2006**, *25* (9), 1957-66.
30. Nothaft, H.; Liu, X.; McNally, D. J.; Li, J.; Szymanski, C. M., Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proc Natl Acad Sci U S A* **2009**, *106* (35), 15019-24.
31. Nothaft, H.; Szymanski, C. M., Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **2010**, *8* (11), 765-78.
32. Larkin, A.; Imperiali, B., The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **2011**, *50* (21), 4411-26.
33. Ielmini, M. V.; Feldman, M. F., *Desulfovibrio desulfuricans* PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **2011**, *21* (6), 734-42.
34. Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P., X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011**, *474* (7351), 350-5.
35. Bai, L.; Wang, T.; Zhao, G.; Kovach, A.; Li, H., The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature* **2018**, *555* (7696), 328-333.
36. Yan, A.; Lennarz, W. J., Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology* **2005**, *15* (12), 1407-15.
37. Cheung, J. C.; Li, J.; Reithmeier, R. A., Topology of transmembrane segments 1-4 in the human chloride/bicarbonate anion exchanger 1 (AE1) by scanning N-glycosylation mutagenesis. *Biochem J* **2005**, *390* (Pt 1), 137-44.
38. Cheung, J. C.; Reithmeier, R. A., Scanning N-glycosylation mutagenesis of membrane proteins. *Methods* **2007**, *41* (4), 451-9.
39. Bano-Polo, M.; Baldin, F.; Tamborero, S.; Marti-Renom, M. A.; Mingarro, I., N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci* **2011**, *20* (1), 179-86.

40. Huang, Y. W.; Yang, H. I.; Wu, Y. T.; Hsu, T. L.; Lin, T. W.; Kelly, J. W.; Wong, C. H., Residues Comprising the Enhanced Aromatic Sequon Influence Protein N-Glycosylation Efficiency. *J Am Chem Soc* **2017**, *139* (37), 12947-12955.
41. Xu, Y.; Wu, Z.; Zhang, P.; Zhu, H.; Song, Q.; Wang, L.; Wang, F.; Wang, P. G.; Cheng, J., A novel enzymatic method for synthesis of glycopeptides carrying natural eukaryotic N-glycans. *Chem Commun (Camb)* **2017**, *53* (65), 9075-9077.
42. Naegeli, A.; Neupert, C.; Fan, Y. Y.; Lin, C. W.; Poljak, K.; Papini, A. M.; Schwarz, F.; Aebi, M., Molecular analysis of an alternative N-glycosylation machinery by functional transfer from *Actinobacillus pleuropneumoniae* to *Escherichia coli*. *J Biol Chem* **2014**, *289* (4), 2170-9.
43. Silverman, J. M.; Imperiali, B., Bacterial N-Glycosylation Efficiency Is Dependent on the Structural Context of Target Sequons. *J Biol Chem* **2016**, *291* (42), 22001-22010.
44. Gross, J.; Grass, S.; Davis, A. E.; Gilmore-Erdmann, P.; Townsend, R. R.; St Geme, J. W., 3rd, The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *J Biol Chem* **2008**, *283* (38), 26010-5.
45. Choi, K. J.; Grass, S.; Paek, S.; St Geme, J. W., 3rd; Yeo, H. J., The *Actinobacillus pleuropneumoniae* HMW1C-like glycosyltransferase mediates N-linked glycosylation of the *Haemophilus influenzae* HMW1 adhesin. *PLoS One* **2010**, *5* (12), e15888.
46. Schwarz, F.; Fan, Y. Y.; Schubert, M.; Aebi, M., Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J Biol Chem* **2011**, *286* (40), 35267-74.
47. Kawai, F.; Grass, S.; Kim, Y.; Choi, K. J.; St Geme, J. W., 3rd; Yeo, H. J., Structural insights into the glycosyltransferase activity of the *Actinobacillus pleuropneumoniae* HMW1C-like protein. *J Biol Chem* **2011**, *286* (44), 38546-57.
48. Kightlinger, W.; Lin, L.; Rosztoczy, M.; Li, W.; DeLisa, M. P.; Mrksich, M.; Jewett, M. C., Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat Chem Biol* **2018**, *14* (6), 627-635.
49. Song, Q.; Wu, Z.; Fan, Y.; Song, W.; Zhang, P.; Wang, L.; Wang, F.; Xu, Y.; Wang, P. G.; Cheng, J., Production of homogeneous glycoprotein with multisite modifications by an engineered N-glycosyltransferase mutant. *J Biol Chem* **2017**, *292* (21), 8856-8863.
50. Lomino, J. V.; Naegeli, A.; Orwenyo, J.; Amin, M. N.; Aebi, M.; Wang, L. X., A two-step enzymatic glycosylation of polypeptides with complex N-glycans. *Bioorg Med Chem* **2013**, *21* (8), 2262-2270.
51. Junker, M.; Schuster, C. C.; McDonnell, A. V.; Sorg, K. A.; Finn, M. C.; Berger, B.; Clark, P. L., Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci U S A* **2006**, *103* (13), 4918-23.
52. Charbonneau, M. E.; Janvore, J.; Mourez, M., Autoprocessing of the *Escherichia coli* AIDA-I autotransporter: a new mechanism involving acidic residues in the junction region. *J Biol Chem* **2009**, *284* (25), 17340-51.

53. Lolli, F.; Mulinacci, B.; Carotenuto, A.; Bonetti, B.; Sabatino, G.; Mazzanti, B.; D'Ursi, A. M.; Novellino, E.; Pazzagli, M.; Lovato, L.; Alcaro, M. C.; Peroni, E.; Pozo-Carrero, M. C.; Nuti, F.; Battistini, L.; Borsellino, G.; Chelli, M.; Rovero, P.; Papini, A. M., An N-glycosylated peptide detecting disease-specific autoantibodies, biomarkers of multiple sclerosis. *Proc Natl Acad Sci U S A* **2005**, *102* (29), 10273-8.
54. Kong, Y.; Li, J.; Hu, X.; Wang, Y.; Meng, Q.; Gu, G.; Wang, P. G.; Chen, M., N-Glycosyltransferase from *Aggregatibacter aphrophilus* synthesizes glycopeptides with relaxed nucleotide-activated sugar donor selectivity. *Carbohydr Res* **2018**, *462*, 7-12.
55. Lolli, F.; Mazzanti, B.; Pazzagli, M.; Peroni, E.; Alcaro, M. C.; Sabatino, G.; Lanzillo, R.; Brescia Morra, V.; Santoro, L.; Gasperini, C.; Galgani, S.; D'Elia, M. M.; Zipoli, V.; Sotgiu, S.; Pugliatti, M.; Rovero, P.; Chelli, M.; Papini, A. M., The glycopeptide CSF114(Glc) detects serum antibodies in multiple sclerosis. *J Neuroimmunol* **2005**, *167* (1-2), 131-7.
56. Palcic, M. M.; Seto, N. O.; Hindsgaul, O., Natural and recombinant A and B gene encoded glycosyltransferases. *Transfus Med* **2001**, *11* (4), 315-23.
57. Yeoh, K. K.; Butters, T. D.; Wilkinson, B. L.; Fairbanks, A. J., Probing replacement of pyrophosphate via click chemistry; synthesis of UDP-sugar analogues as potential glycosyl transferase inhibitors. *Carbohydr Res* **2009**, *344* (5), 586-91.
58. Palcic, M. M.; Heerze, L. D.; Pierce, M.; Hindsgaul, O., The Use of Hydrophobic Synthetic Glycosides as Acceptors in Glycosyltransferase Assays. *Glycoconjugate Journal* **1988**, *5* (1), 49-63.
59. Ahsen, O.; Voigtmann, U.; Klotz, M.; Nifantiev, N.; Schottelius, A.; Ernst, A.; Muller-Tiemann, B.; Parczyk, K., A miniaturized high-throughput screening assay for fucosyltransferase VII. *Anal Biochem* **2008**, *372* (1), 96-105.
60. Baker, C. A.; Poorman, R. A.; Kezdy, F. J.; Staples, D. J.; Smith, C. W.; Elhammer, A. P., A scintillation proximity assay for UDP-GalNAc:polypeptide, N-acetylgalactosaminyltransferase. *Anal Biochem* **1996**, *239* (1), 20-4.
61. Gosselin, S.; Alhussaini, M.; Streiff, M. B.; Takabayashi, K.; Palcic, M. M., A continuous spectrophotometric assay for glycosyltransferases. *Anal Biochem* **1994**, *220* (1), 92-7.
62. Zhang, L.; Ren, F.; Li, J.; Ma, X.; Wang, P., A Modified Coupled Enzyme Method for O-linked GlcNAc Transferase Activity Assay. *Biol Proced Online* **2009**, *11* (1), 170-83.
63. Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W., Universal phosphatase-coupled glycosyltransferase assay. *Glycobiology* **2011**, *21* (6), 727-33.
64. Das, D.; Walvoort, M. T.; Lukose, V.; Imperiali, B., A Rapid and Efficient Luminescence-based Method for Assaying Phosphoglycosyltransferase Enzymes. *Sci Rep* **2016**, *6*, 33412.
65. Deng, C.; Chen, R. R., A pH-sensitive assay for galactosyltransferase. *Anal Biochem* **2004**, *330* (2), 219-26.

66. Persson, M.; Palcic, M. M., A high-throughput pH indicator assay for screening glycosyltransferase saturation mutagenesis libraries. *Anal Biochem* **2008**, *378* (1), 1-7.
67. Lee, H. S.; Thorson, J. S., Development of a universal glycosyltransferase assay amenable to high-throughput formats. *Anal Biochem* **2011**, *418* (1), 85-8.
68. Wongkongkatep, J.; Miyahara, Y.; Ojida, A.; Hamachi, I., Label-free, real-time glycosyltransferase assay based on a fluorescent artificial chemosensor. *Angew Chem Int Ed Engl* **2006**, *45* (4), 665-8.
69. Ryu, J.; Eom, M. S.; Ko, W.; Han, M. S.; Lee, H. S., A fluorescence-based glycosyltransferase assay for high-throughput screening. *Bioorg Med Chem* **2014**, *22* (8), 2571-5.
70. Chen, X.; Jou, M. J.; Yoon, J., An "Off-On" type UTP/UDP selective fluorescent probe and its application to monitor glycosylation process. *Org Lett* **2009**, *11* (10), 2181-4.
71. Zielinski, T.; Reichman, M.; Donover, P. S.; Lowery, R. G., Development and Validation of a Universal High-Throughput UDP-Glycosyltransferase Assay with a Time-Resolved FRET Signal. *Assay Drug Dev Technol* **2016**, *14* (4), 240-51.
72. Soya, N.; Shoemaker, G. K.; Palcic, M. M.; Klassen, J. S., Comparative study of substrate and product binding to the human ABO(H) blood group glycosyltransferases. *Glycobiology* **2009**, *19* (11), 1224-34.
73. Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H., Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J Mol Biol* **2006**, *359* (3), 708-27.
74. Lazarus, M. B.; Nam, Y.; Jiang, J.; Sliz, P.; Walker, S., Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **2011**, *469* (7331), 564-7.
75. Sun, H. Y.; Lin, S. W.; Ko, T. P.; Pan, J. F.; Liu, C. L.; Lin, C. N.; Wang, A. H.; Lin, C. H., Structure and mechanism of Helicobacter pylori fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J Biol Chem* **2007**, *282* (13), 9973-82.
76. Yang, M.; Brazier, M.; Edwards, R.; Davis, B. G., High-throughput mass-spectrometry monitoring for multisubstrate enzymes: determining the kinetic parameters and catalytic activities of glycosyltransferases. *Chembiochem* **2005**, *6* (2), 346-57.
77. Yang, M.; Fehl, C.; Lees, K. V.; Lim, E. K.; Offen, W. A.; Davies, G. J.; Bowles, D. J.; Davidson, M. G.; Roberts, S. J.; Davis, B. G., Functional and informatics analysis enables glycosyltransferase activity prediction. *Nat Chem Biol* **2018**, *14* (12), 1109-1117.
78. Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H., Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* **2012**, *7* (7), 1299-310.
79. Woo, C. M.; Felix, A.; Byrd, W. E.; Zuegel, D. K.; Ishihara, M.; Azadi, P.; Iavarone, A. T.; Pitteri, S. J.; Bertozzi, C. R., Development of IsoTaG, a Chemical Glycoproteomics Technique for Profiling Intact N- and O-Glycopeptides from Whole Cell Proteomes. *J Proteome Res* **2017**, *16* (4), 1706-1718.

80. Love, J. C.; Estroff, L. A.; Kriebel, J. K.; Nuzzo, R. G.; Whitesides, G. M., Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem Rev* **2005**, *105* (4), 1103-69.
81. Hakkinen, H., The gold-sulfur interface at the nanoscale. *Nat Chem* **2012**, *4* (6), 443-55.
82. Ulman, A., Formation and Structure of Self-Assembled Monolayers. *Chem Rev* **1996**, *96* (4), 1533-1554.
83. Prime, K. L.; Whitesides, G. M., Adsorption of Proteins onto Surfaces Containing End-Attached Oligo(Ethylene Oxide) - a Model System Using Self-Assembled Monolayers. *Journal of the American Chemical Society* **1993**, *115* (23), 10714-10721.
84. Deng, L.; Mrksich, M.; Whitesides, G. M., Self-assembled monolayers of alkanethiolates presenting tri(propylene sulfoxide) groups resist the adsorption of protein. *Journal of the American Chemical Society* **1996**, *118* (21), 5136-5137.
85. Prime, K. L.; Whitesides, G. M., Self-Assembled Organic Monolayers - Model Systems for Studying Adsorption of Proteins at Surfaces. *Science* **1991**, *252* (5009), 1164-1167.
86. Mrksich, M.; Sigal, G. B.; Whitesides, G. M., Surface-Plasmon Resonance Permits in-Situ Measurement of Protein Adsorption on Self-Assembled Monolayers of Alkanethiolates on Gold. *Langmuir* **1995**, *11* (11), 4383-4385.
87. Hodneland, C. D.; Lee, Y. S.; Min, D. H.; Mrksich, M., Selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. *Proc Natl Acad Sci U S A* **2002**, *99* (8), 5048-52.
88. Sigal, G. B.; Bamdad, C.; Barberis, A.; Strominger, J.; Whitesides, G. M., A self-assembled monolayer for the binding and study of histidine-tagged proteins by surface plasmon resonance. *Anal Chem* **1996**, *68* (3), 490-7.
89. Subramanian, A.; Irudayaraj, J.; Ryan, T., A mixed self-assembled monolayer-based surface plasmon immunosensor for detection of E. coli O157:H7. *Biosens Bioelectron* **2006**, *21* (7), 998-1006.
90. Herne, T. M.; Tarlov, M. J., Characterization of DNA probes immobilized on gold surfaces. *Journal of the American Chemical Society* **1997**, *119* (38), 8916-8920.
91. Houseman, B. T.; Gawalt, E. S.; Mrksich, M., Maleimide-functionalized self-assembled monolayers for the preparation of peptide and carbohydrate biochips. *Langmuir* **2003**, *19* (5), 1522-1531.
92. Hong, H. G.; Jiang, M.; Sligar, S. G.; Bohn, P. W., Cysteine-Specific Surface Tethering of Genetically-Engineered Cytochromes for Fabrication of Metalloprotein Nanostructures. *Langmuir* **1994**, *10* (1), 153-158.
93. Smith, E. A.; Thomas, W. D.; Kiessling, L. L.; Corn, R. M., Surface plasmon resonance imaging studies of protein-carbohydrate interactions. *J Am Chem Soc* **2003**, *125* (20), 6140-8.

94. Wegner, G. J.; Lee, H. J.; Corn, R. M., Characterization and Optimization of Peptide Arrays for the Study of Epitope–Antibody Interactions Using Surface Plasmon Resonance Imaging. *Analytical Chemistry* **2002**, *74* (20), 5161-5168.
95. Hochuli, E.; Bannwarth, W.; Dobeli, H.; Gentz, R.; Stuber, D., Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio-Technology* **1988**, *6* (11), 1321-1325.
96. Gaberc-Porekar, V.; Menart, V., Potential for using histidine tags in purification of proteins at large scale. *Chemical Engineering & Technology* **2005**, *28* (11), 1306-1314.
97. Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G., Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **1975**, *258* (5536), 598-9.
98. Ludden, M. J.; Mulder, A.; Schulze, K.; Subramaniam, V.; Tampe, R.; Huskens, J., Anchoring of histidine-tagged proteins to molecular printboards: self-assembly, thermodynamic modeling, and patterning. *Chemistry* **2008**, *14* (7), 2044-51.
99. Zhen, G. L.; Falconnet, D.; Kuennemann, E.; Voros, J.; Spencer, N. D.; Textor, M.; Zurcher, S., Nitrilotriacetic acid functionalized graft copolymers: A polymeric interface for selective and reversible binding of histidine-tagged proteins. *Advanced Functional Materials* **2006**, *16* (2), 243-251.
100. Gizeli, E.; Glad, J., Single-step formation of a biorecognition layer for assaying histidine-tagged proteins. *Anal Chem* **2004**, *76* (14), 3995-4001.
101. Gamsjaeger, R.; Wimmer, B.; Kahr, H.; Tinazli, A.; Picuric, S.; Lata, S.; Tampé, R.; Maulet, Y.; Gruber, H. J.; Hinterdorfer, P.; Romanin, C., Oriented Binding of the His6-Tagged Carboxyl-Tail of the L-type Ca<sup>2+</sup>Channel  $\alpha$ 1-Subunit to a New NTA-Functionalized Self-Assembled Monolayer. *Langmuir* **2004**, *20* (14), 5885-5890.
102. Hengen, P., Purification of His-Tag fusion proteins from Escherichia coli. *Trends Biochem Sci* **1995**, *20* (7), 285-6.
103. Valiokas, R.; Klenkar, G.; Tinazli, A.; Reichel, A.; Tampe, R.; Piehler, J.; Liedberg, B., Self-assembled monolayers containing terminal mono-, bis-, and tris-nitrilotriacetic acid groups: characterization and application. *Langmuir* **2008**, *24* (9), 4959-67.
104. Valiokas, R.; Klenkar, G.; Tinazli, A.; Tampe, R.; Liedberg, B.; Piehler, J., Differential protein assembly on micropatterned surfaces with tailored molecular and surface multivalency. *Chembiochem* **2006**, *7* (9), 1325-9.
105. Lata, S.; Reichel, A.; Brock, R.; Tampe, R.; Piehler, J., High-affinity adaptors for switchable recognition of histidine-tagged proteins. *J Am Chem Soc* **2005**, *127* (29), 10205-15.
106. Houseman, B. T.; Mrksich, M., Model Systems for Studying Polyvalent Carbohydrate Binding Interactions. In *Host-Guest Chemistry: Mimetic Approaches to Study Carbohydrate Recognition*, Penadés, S., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2002; pp 1-44.
107. Houseman, B. T.; Mrksich, M., Carbohydrate arrays for the evaluation of protein binding and enzymatic modification. *Chem Biol* **2002**, *9* (4), 443-54.

108. Houseman, B. T.; Mrksich, M., The Role of Ligand Density in the Enzymatic Glycosylation of Carbohydrates Presented on Self-Assembled Monolayers of Alkanethiolates on Gold. *Angew Chem Int Ed Engl* **1999**, *38* (6), 782-785.
109. Su, J.; Mrksich, M., Using mass spectrometry to characterize self-assembled monolayers presenting peptides, proteins, and carbohydrates. *Angew Chem Int Ed Engl* **2002**, *41* (24), 4715-8.
110. Murphy, W. L.; Mercurius, K. O.; Koide, S.; Mrksich, M., Substrates for cell adhesion prepared via active site-directed immobilization of a protein domain. *Langmuir* **2004**, *20* (4), 1026-30.
111. Grant, J.; Modica, J. A.; Roll, J.; Perkovich, P.; Mrksich, M., An Immobilized Enzyme Reactor for Spatiotemporal Control over Reaction Products. *Small* **2018**, *14* (31), e1800923.
112. Kornienko, O.; Ada, E. T.; Hanley, L., Organic surface analysis by two laser-ion trap mass spectrometry. *Analytical Chemistry* **1997**, *69* (8), 1536-1542.
113. Trevor, J. L.; Lykke, K. R.; Pellin, M. J.; Hanley, L., Two-laser mass spectrometry of thiolate, disulfide, and sulfide self-assembled monolayers. *Langmuir* **1998**, *14* (7), 1664-1673.
114. Su, J.; Mrksich, M., Using MALDI-TOF mass spectrometry to characterize interfacial reactions on self-assembled monolayers. *Langmuir* **2003**, *19* (12), 4867-4870.
115. Guan, W.; Ban, L.; Cai, L.; Li, L.; Chen, W.; Liu, X.; Mrksich, M.; Wang, P. G., Combining carbochips and mass spectrometry to study the donor specificity for the *Neisseria meningitidis* beta1,3-N-acetylglucosaminyltransferase LgtA. *Bioorg Med Chem Lett* **2011**, *21* (17), 5025-8.
116. Ban, L.; Pettit, N.; Li, L.; Stuparu, A. D.; Cai, L.; Chen, W.; Guan, W.; Han, W.; Wang, P. G.; Mrksich, M., Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry. *Nat Chem Biol* **2012**, *8* (9), 769-73.
117. Laurent, N.; Haddoub, R.; Voglmeir, J.; Wong, S. C.; Gaskell, S. J.; Flitsch, S. L., SPOT synthesis of peptide arrays on self-assembled monolayers and their evaluation as enzyme substrates. *Chembiochem* **2008**, *9* (16), 2592-6.
118. Laurent, N.; Voglmeir, J.; Flitsch, S. L., Glycoarrays--tools for determining protein-carbohydrate interactions and glycoenzyme specificity. *Chem Commun (Camb)* **2008**, (37), 4400-12.
119. Laurent, N.; Voglmeir, J.; Wright, A.; Blackburn, J.; Pham, N. T.; Wong, S. C.; Gaskell, S. J.; Flitsch, S. L., Enzymatic glycosylation of peptide arrays on gold surfaces. *Chembiochem* **2008**, *9* (6), 883-7.
120. Sanchez-Ruiz, A.; Serna, S.; Ruiz, N.; Martin-Lomas, M.; Reichardt, N. C., MALDI-TOF mass spectrometric analysis of enzyme activity and lectin trapping on an array of N-glycans. *Angew Chem Int Ed Engl* **2011**, *50* (8), 1801-4.
121. Beloqui, A.; Calvo, J.; Serna, S.; Yan, S.; Wilson, I. B.; Martin-Lomas, M.; Reichardt, N. C., Analysis of microarrays by MALDI-TOF MS. *Angew Chem Int Ed Engl* **2013**, *52* (29), 7477-81.

122. Patrie, S. M.; Mrksich, M., Self-assembled monolayers for MALDI-TOF mass spectrometry for immunoassays of human protein antigens. *Anal Chem* **2007**, *79* (15), 5878-87.
123. Warden, A.; Graham, B.; Hearn, M. T.; Spiccia, L., Synthesis of novel derivatives of 1,4,7-triazacyclononane. *Org Lett* **2001**, *3* (18), 2855-8.
124. Johnson, D. L.; Martin, L. L., Controlling protein orientation at interfaces using histidine tags: an alternative to Ni/NTA. *J Am Chem Soc* **2005**, *127* (7), 2018-9.
125. Deisenhofer, J., Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry* **2002**, *20* (9), 2361-2370.
126. Nakanishi, K.; Muguruma, H.; Karube, I., A novel method of immobilizing antibodies on a quartz crystal microbalance using plasma-polymerized films for immunosensors. *Analytical Chemistry* **1996**, *68* (10), 1695-1700.
127. Anderson, G. P.; Jacoby, M. A.; Ligler, F. S.; King, K. D., Effectiveness of protein A for antibody immobilization for a fiber optic biosensor. *Biosens Bioelectron* **1997**, *12* (4), 329-36.
128. Roth, M. J.; Kim, J.; Maresh, E. M.; Plymire, D. A.; Corbett, J. R.; Zhang, J.; Patrie, S. M., Thin-layer matrix sublimation with vapor-sorption induced co-crystallization for sensitive and reproducible SAMDI-TOF MS analysis of protein biosensors. *J Am Soc Mass Spectrom* **2012**, *23* (10), 1661-9.
129. Roth, M. J.; Maresh, E. M.; Plymire, D. A.; Zhang, J.; Corbett, J. R.; Robbins, R.; Patrie, S. M., Surface preparation strategies for improved parallelization and reproducible MALDI-TOF MS ligand binding assays. *ACS Appl Mater Interfaces* **2013**, *5* (1), 6-10.
130. Marin, V. L.; Bayburt, T. H.; Sligar, S. G.; Mrksich, M., Functional assays of membrane-bound proteins with SAMDI-TOF mass spectrometry. *Angew Chem Int Ed Engl* **2007**, *46* (46), 8796-8.
131. Kim, Y. E.; Yi, S. Y.; Lee, C. S.; Jung, Y.; Chung, B. H., Gold patterned biochips for on-chip immuno-MALDI-TOF MS: SPR imaging coupled multi-protein MS analysis. *Analyst* **2012**, *137* (2), 386-92.
132. Gao, J.; Meyer, K.; Borucki, K.; Ueland, P. M., Multiplex Immuno-MALDI-TOF MS for Targeted Quantification of Protein Biomarkers and Their Proteoforms Related to Inflammation and Renal Dysfunction. *Anal Chem* **2018**, *90* (5), 3366-3373.
133. Chou, P. H.; Chen, S. H.; Liao, H. K.; Lin, P. C.; Her, G. R.; Lai, A. C.; Chen, J. H.; Lin, C. C.; Chen, Y. J., Nanoprobe-based affinity mass spectrometry for selected protein profiling in human plasma. *Anal Chem* **2005**, *77* (18), 5990-7.
134. Sparbier, K.; Wenzel, T.; Dihazi, H.; Blaschke, S.; Muller, G. A.; Deelder, A.; Flad, T.; Kostrzewa, M., Immuno-MALDI-TOF MS: new perspectives for clinical applications of mass spectrometry. *Proteomics* **2009**, *9* (6), 1442-50.
135. Suzuki, T.; Israr, M. Z.; Heaney, L. M.; Takaoka, M.; Squire, I. B.; Ng, L. L., Prognostic Role of Molecular Forms of B-Type Natriuretic Peptide in Acute Heart Failure. *Clin Chem* **2017**, *63* (4), 880-886.

136. Lombana, T. N.; Matsumoto, M. L.; Berkley, A. M.; Toy, E.; Cook, R.; Gan, Y.; Du, C.; Schnier, P.; Sandoval, W.; Ye, Z.; Schartner, J. M.; Kim, J.; Spiess, C., High-resolution glycosylation site-engineering method identifies MICA epitope critical for shedding inhibition activity of anti-MICA antibodies. *MAbs* **2019**, *11* (1), 75-93.
137. 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* (22), 10635-10642.
138. Peek, C. B.; Affinati, A. H.; Ramsey, K. M.; Kuo, H. Y.; Yu, W.; Sena, L. A.; Ilkayeva, O.; Marcheva, B.; Kobayashi, Y.; Omura, C.; Levine, D. C.; Bacsik, D. J.; Gius, D.; Newgard, C. B.; Goetzman, E.; Chandel, N. S.; Denu, J. M.; Mrksich, M.; Bass, J., Circadian clock NAD<sup>+</sup> cycle drives mitochondrial oxidative metabolism in mice. *Science* **2013**, *342* (6158), 1243417.
139. Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **2005**, *4* (6), 477-88.
140. Meany, D. L.; Chan, D. W., Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clin Proteomics* **2011**, *8* (1), 7.
141. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **2015**, *15* (9), 540-55.
142. Jemal, A.; Siegel, R.; Ward, E.; Hao, Y.; Xu, J.; Murray, T.; Thun, M. J., Cancer statistics, 2008. *CA Cancer J Clin* **2008**, *58* (2), 71-96.
143. Dawood, S.; Broglio, K.; Buzdar, A. U.; Hortobagyi, G. N.; Giordano, S. H., Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. *J Clin Oncol* **2010**, *28* (1), 92-8.
144. Mauri, D.; Polyzos, N. P.; Salanti, G.; Pavlidis, N.; Ioannidis, J. P., Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *J Natl Cancer Inst* **2008**, *100* (24), 1780-91.
145. Curtis, C.; Shah, S. P.; Chin, S. F.; Turashvili, G.; Rueda, O. M.; Dunning, M. J.; Speed, D.; Lynch, A. G.; Samarajiwa, S.; Yuan, Y.; Graf, S.; Ha, G.; Haffari, G.; Bashashati, A.; Russell, R.; McKinney, S.; Group, M.; Langerod, A.; Green, A.; Provenzano, E.; Wishart, G.; Pinder, S.; Watson, P.; Markowitz, F.; Murphy, L.; Ellis, I.; Purushotham, A.; Borresen-Dale, A. L.; Brenton, J. D.; Tavaré, S.; Caldas, C.; Aparicio, S., The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **2012**, *486* (7403), 346-52.
146. Ikehara, Y.; Kojima, N.; Kurosawa, N.; Kudo, T.; Kono, M.; Nishihara, S.; Issiki, S.; Morozumi, K.; Itzkowitz, S.; Tsuda, T.; Nishimura, S. I.; Tsuji, S.; Narimatsu, H., Cloning and expression of a human gene encoding an N-acetylgalactosamine- 2,6-sialyltransferase (ST6GalNAc I): a candidate for synthesis of cancer-associated sialyl-Tn antigens. *Glycobiology* **1999**, *9* (11), 1213-1224.
147. Sewell, R.; Backstrom, M.; Dalziel, M.; Gschmeissner, S.; Karlsson, H.; Noll, T.; Gatgens, J.; Clausen, H.; Hansson, G. C.; Burchell, J.; Taylor-Papadimitriou, J., The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *J Biol Chem* **2006**, *281* (6), 3586-94.

148. Julien, S.; Adriaenssens, E.; Ottenberg, K.; Furlan, A.; Courtand, G.; Vercoutter-Edouart, A. S.; Hanisch, F. G.; Delannoy, P.; Le Bourhis, X., ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumourigenicity. *Glycobiology* **2006**, *16* (1), 54-64.
149. Julien, S.; Krzewinski-Recchi, M.-A.; Harduin-Lepers, A.; Gouyer, V.; Huet, G.; Le Bourhis, X.; Delannoy, P., Expression of Sialyl-Tn antigen in breast cancer cells transfected with the human CMP-Neu5Ac: GalNAc  $\alpha$ 2,6-sialyltransferase (ST6GalNAc I) cDNA. *Glycoconjugate Journal* **2001**, *18* (11/12), 883-893.
150. Julien, S.; Lagadec, C.; Krzewinski-Recchi, M. A.; Courtand, G.; Le Bourhis, X.; Delannoy, P., Stable expression of sialyl-Tn antigen in T47-D cells induces a decrease of cell adhesion and an increase of cell migration. *Breast Cancer Res Treat* **2005**, *90* (1), 77-84.
151. Miles, D. W.; Happerfield, L. C.; Smith, P.; Gillibrand, R.; Bobrow, L. G.; Gregory, W. M.; Rubens, R. D., Expression of sialyl-Tn predicts the effect of adjuvant chemotherapy in node-positive breast cancer. *Br J Cancer* **1994**, *70* (6), 1272-5.
152. Handerson, T.; Camp, R.; Harigopal, M.; Rimm, D.; Pawelek, J., Beta1,6-branched oligosaccharides are increased in lymph node metastases and predict poor outcome in breast carcinoma. *Clin Cancer Res* **2005**, *11* (8), 2969-73.
153. Guo, H. B.; Randolph, M.; Pierce, M., Inhibition of a specific N-glycosylation activity results in attenuation of breast carcinoma cell invasiveness-related phenotypes: inhibition of epidermal growth factor-induced dephosphorylation of focal adhesion kinase. *J Biol Chem* **2007**, *282* (30), 22150-62.
154. Goetz, J. A.; Mechref, Y.; Kang, P.; Jeng, M. H.; Novotny, M. V., Glycomic profiling of invasive and non-invasive breast cancer cells. *Glycoconjugate J* **2009**, *26* (2), 117-31.
155. Kang, P.; Mechref, Y.; Kyselova, Z.; Goetz, J. A.; Novotny, M. V., Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal Chem* **2007**, *79* (16), 6064-73.
156. Kyselova, Z.; Mechref, Y.; Kang, P.; Goetz, J. A.; Dobrolecki, L. E.; Sledge, G. W.; Schnaper, L.; Hickey, R. J.; Malkas, L. H.; Novotny, M. V., Breast cancer diagnosis and prognosis through quantitative measurements of serum glycan profiles. *Clin Chem* **2008**, *54* (7), 1166-75.
157. Saldova, R.; Asadi Shehni, A.; Haakensen, V. D.; Steinfeld, I.; Hilliard, M.; Kifer, I.; Helland, A.; Yakhini, Z.; Borresen-Dale, A. L.; Rudd, P. M., Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J Proteome Res* **2014**, *13* (5), 2314-27.
158. Liu, X.; Nie, H.; Zhang, Y.; Yao, Y.; Maitikabili, A.; Qu, Y.; Shi, S.; Chen, C.; Li, Y., Cell surface-specific N-glycan profiling in breast cancer. *PLoS One* **2013**, *8* (8), e72704.
159. Zen, K.; Liu, D. Q.; Guo, Y. L.; Wang, C.; Shan, J.; Fang, M.; Zhang, C. Y.; Liu, Y., CD44v4 is a major E-selectin ligand that mediates breast cancer cell transendothelial migration. *PLoS One* **2008**, *3* (3), e1826.

160. Patil, S. A.; Chandrasekaran, E. V.; Matta, K. L.; Parikh, A.; Tzanakakis, E. S.; Neelamegham, S., Scaling down the size and increasing the throughput of glycosyltransferase assays: activity changes on stem cell differentiation. *Anal Biochem* **2012**, *425* (2), 135-44.
161. Feizi, T., Carbohydrate microarrays — a new set of technologies at the frontiers of glycomics. *Current Opinion in Structural Biology* **2003**, *13* (5), 637-645.
162. Park, S.; Shin, I., Carbohydrate microarrays for assaying galactosyltransferase activity. *Org Lett* **2007**, *9* (9), 1675-8.
163. Park, S.; Gildersleeve, J. C.; Blixt, O.; Shin, I., Carbohydrate microarrays. *Chem Soc Rev* **2013**, *42* (10), 4310-26.
164. Cai, L.; Ban, L.; Guan, W.; Mrksich, M.; Wang, P. G., Enzymatic synthesis and properties of uridine-5'-O-(2-thiodiphospho)-N-acetylglucosamine. *Carbohydr Res* **2011**, *346* (12), 1576-80.
165. Ban, L.; Mrksich, M., On-chip synthesis and label-free assays of oligosaccharide arrays. *Angew Chem Int Ed Engl* **2008**, *47* (18), 3396-9.
166. Mrksich, M.; Whitesides, G. M., Using Self-Assembled Monolayers That Present Oligo(ethylene glycol) Groups To Control the Interactions of Proteins with Surfaces. In *Poly(ethylene glycol)*, American Chemical Society: 1997; Vol. 680, pp 361-373.
167. Lee, Y. S.; Mrksich, M., Protein chips: from concept to practice. *Trends Biotechnol* **2002**, *20* (12 Suppl), S14-18.
168. Mrksich, M., Mass spectrometry of self-assembled monolayers: a new tool for molecular surface science. *ACS Nano* **2008**, *2* (1), 7-18.
169. Schuster, T.; Schellenberger, S.; Friedrich, R.; Klapper, M.; Mullen, K., Branched fluorinated amphiphiles based on carbohydrates. *Journal of Fluorine Chemistry* **2013**, *154*, 30-36.
170. MilliporeSigma Mass Spectrometry of Glycans. <https://www.sigmaaldrich.com/technical-documents/articles/biology/glycobiology/mass-spectrometry-of-glycans.html> (accessed 04/01/2019).
171. Schanbacher, F. L.; Ebner, K. E., Galactosyltransferase acceptor specificity of the lactose synthetase A protein. *J Biol Chem* **1970**, *245* (19), 5057-61.
172. Villegas-Comonfort, S.; Serna-Marquez, N.; Galindo-Hernandez, O.; Navarro-Tito, N.; Salazar, E. P., Arachidonic acid induces an increase of beta-1,4-galactosyltransferase I expression in MDA-MB-231 breast cancer cells. *J Cell Biochem* **2012**, *113* (11), 3330-41.
173. Orski, S. V.; Poloukhine, A. A.; Arumugam, S.; Mao, L.; Popik, V. V.; Locklin, J., High density orthogonal surface immobilization via photoactivated copper-free click chemistry. *J Am Chem Soc* **2010**, *132* (32), 11024-6.
174. Holliday, D. L.; Speirs, V., Choosing the right cell line for breast cancer research. *Breast Cancer Res* **2011**, *13* (4), 215.

175. Katzenellenbogen, B. S.; Kendra, K. L.; Norman, M. J.; Berthois, Y., Proliferation, hormonal responsiveness, and estrogen receptor content of MCF-7 human breast cancer cells grown in the short-term and long-term absence of estrogens. *Cancer Res* **1987**, *47* (16), 4355-60.
176. Mansell, T. J.; Guarino, C.; DeLisa, M. P., Engineered genetic selection links in vivo protein folding and stability with asparagine-linked glycosylation. *Biotechnol J* **2013**, *8* (12), 1445-51.
177. Fonseca-Maldonado, R.; Vieira, D. S.; Alponi, J. S.; Bonneil, E.; Thibault, P.; Ward, R. J., Engineering the pattern of protein glycosylation modulates the thermostability of a GH11 xylanase. *J Biol Chem* **2013**, *288* (35), 25522-34.
178. Egrie, J. C.; Dwyer, E.; Browne, J. K.; Hitz, A.; Lykos, M. A., Darbepoetin alfa has a longer circulating half-life and greater in vivo potency than recombinant human erythropoietin. *Experimental Hematology* **2003**, *31* (4), 290-299.
179. Clausen, H.; Wandall, H. H.; Steentoft, C.; Stanley, P.; Schnaar, R. L., Glycosylation Engineering. In *Essentials of Glycobiology*, rd; 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 (NY), 2015; pp 713-728.
180. Staub, A.; Guillarme, D.; Schappler, J.; Veuthey, J. L.; Rudaz, S., Intact protein analysis in the biopharmaceutical field. *J Pharm Biomed Anal* **2011**, *55* (4), 810-22.
181. Schoborg, J. A.; Hershewe, J. M.; Stark, J. C.; Kightlinger, W.; Kath, J. E.; Jaroentomeechai, T.; Natarajan, A.; DeLisa, M. P.; Jewett, M. C., A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnol Bioeng* **2018**, *115* (3), 739-750.
182. Jaroentomeechai, T.; Stark, J. C.; Natarajan, A.; Glasscock, C. J.; Yates, L. E.; Hsu, K. J.; Mrksich, M.; Jewett, M. C.; DeLisa, M. P., Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat Commun* **2018**, *9* (1), 2686.
183. Ollis, A. A.; Zhang, S.; Fisher, A. C.; DeLisa, M. P., Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat Chem Biol* **2014**, *10* (10), 816-22.
184. 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. *N Biotechnol* **2016**, *33* (4), 480-7.
185. Keys, T. G.; Aebi, M., Engineering protein glycosylation in prokaryotes. *Current Opinion in Systems Biology* **2017**, *5*, 23-31.
186. Beavis, R. C.; Chait, B. T., High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry. **1990**, *62* (17), 1836-1840.
187. Beavis, R. C.; Chait, B. T., Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun Mass Spectrom* **1989**, *3* (12), 432-5.
188. Kwon, Y. C.; Jewett, M. C., High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci Rep* **2015**, *5*, 8663.

189. Jewett, M. C.; Swartz, J. R., Rapid expression and purification of 100 nmol quantities of active protein using cell-free protein synthesis. *Biotechnol Prog* **2004**, *20* (1), 102-9.
190. Espah Borujeni, A.; Channarasappa, A. S.; Salis, H. M., Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* **2014**, *42* (4), 2646-59.
191. Salis, H. M.; Mirsky, E. A.; Voigt, C. A., Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **2009**, *27* (10), 946-50.
192. Nick Pace, C.; Martin Scholtz, J., A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal* **1998**, *75* (1), 422-427.
193. Capaldi, A. P.; Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E., Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol* **2001**, *8* (1), 68-72.
194. Capaldi, A. P.; Kleanthous, C.; Radford, S. E., Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* **2002**, *9* (3), 209-16.
195. Spence, G. R.; Capaldi, A. P.; Radford, S. E., Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **2004**, *341* (1), 215-26.
196. Whittaker, S. B.; Spence, G. R.; Gunter Grossmann, J.; Radford, S. E.; Moore, G. R., NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* **2007**, *366* (3), 1001-15.
197. Gsponer, J.; Hopearuoho, H.; Whittaker, S. B.; Spence, G. R.; Moore, G. R.; Paci, E.; Radford, S. E.; Vendruscolo, M., Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci U S A* **2006**, *103* (1), 99-104.
198. Usmani, S. S.; Bedi, G.; Samuel, J. S.; Singh, S.; Kalra, S.; Kumar, P.; Ahuja, A. A.; Sharma, M.; Gautam, A.; Raghava, G. P. S., THPdb: Database of FDA-approved peptide and protein therapeutics. *PLoS One* **2017**, *12* (7), e0181748.
199. Chen, M. M.; Glover, K. J.; Imperiali, B., From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*. *Biochemistry* **2007**, *46* (18), 5579-85.
200. 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.; Eyers, C. E.; Flitsch, S. L., Label-Free Discovery Array Platform for the Characterization of Glycan Binding Proteins and Glycoproteins. *Anal Chem* **2017**, *89* (8), 4444-4451.
201. Garza, K. Y.; Feider, C. L.; Klein, D. R.; Rosenberg, J. A.; Brodbelt, J. S.; Eberlin, L. S., Desorption Electrospray Ionization Mass Spectrometry Imaging of Proteins Directly from Biological Tissue Sections. *Anal Chem* **2018**, *90* (13), 7785-7789.
202. Javanshad, R.; Honarvar, E.; Venter, A. R., Addition of Serine Enhances Protein Analysis by DESI-MS. *J Am Soc Mass Spectrom* **2019**, *30* (4), 694-703.

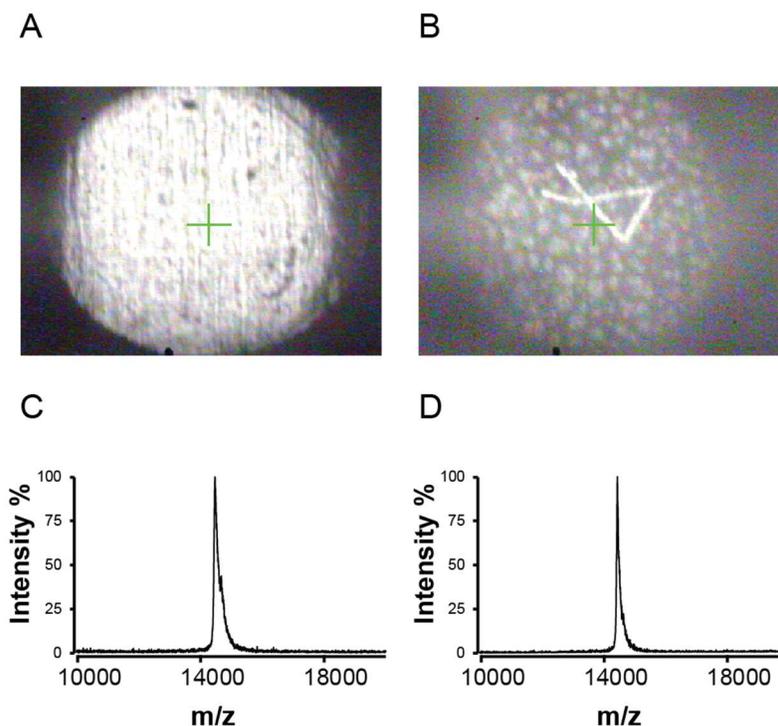
203. Zhang, M.; Wu, S. C.; Zhou, W.; Xu, B., Imaging and measuring single-molecule interaction between a carbohydrate-binding module and natural plant cell wall cellulose. *J Phys Chem B* **2012**, *116* (33), 9949-56.
204. Bondarenko, P. V.; Second, T. P.; Zabrouskov, V.; Makarov, A. A.; Zhang, Z., Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer. *J Am Soc Mass Spectrom* **2009**, *20* (8), 1415-24.
205. Jewett, M. C.; Calhoun, K. A.; Voloshin, A.; Wu, J. J.; Swartz, J. R., An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol* **2008**, *4*, 220.
206. Bundy, B. C.; Swartz, J. R., Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. *Bioconjug Chem* **2010**, *21* (2), 255-63.
207. Liu, Y.; Patricelli, M. P.; Cravatt, B. F., Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* **1999**, *96* (26), 14694-9.
208. Cravatt, B. F.; Wright, A. T.; Kozarich, J. W., Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* **2008**, *77* (1), 383-414.
209. Grant, J.; Goudarzi, S. H.; Mrksich, M., High-Throughput Enzyme Kinetics with 3D Microfluidics and Imaging SAMDI Mass Spectrometry. *Anal Chem* **2018**, *90* (21), 13096-13103.
210. Grant, J.; O'Kane, P. T.; Kimmel, B. R.; Mrksich, M., Using Microfluidics and Imaging SAMDI-MS To Characterize Reaction Kinetics. *ACS Cent Sci* **2019**, *5* (3), 486-493.
211. Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G., Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **2004**, *306* (5695), 471-3.
212. Joshi, S.; Zuilhof, H.; van Beek, T. A.; Nielen, M. W., Biochip Spray: Simplified Coupling of Surface Plasmon Resonance Biosensing and Mass Spectrometry. *Anal Chem* **2017**, *89* (3), 1427-1432.
213. Bruker microflex™ LRF. <https://www.bruker.com/products/mass-spectrometry-and-separations/maldi-toftof/microflex/microflex-lrf.html> (accessed March 27, 2019).

## Appendix A

### Pilot Optimization: TM-SPRAYER Matrix Deposition and Protein SAMDI-MS

A significant technical challenge for protein SAMDI-MS applications is moving towards fully-automated spectral collection. Current condition requires manual collection of MALDI-MS spectra by manual sampling of regions along the periphery of SAM spots, where matrix typically concentrates as it dries. A challenge for automated collection is programming spectrometer to trace around a spot which is currently beyond the means of the AB-SCIEX 5800 with great precision. Instead I proposed to use existing matrix spray technologies originally developed for imaging applications, as these might lead to adequate matrix crystallization with minimal delocalization of signal to the periphery as currently is the case using simple “dried-droplet” methods. To do this I ran several optimization studies using the HTX TM-Sprayer ranging from “drier” to “wet” depositions. In general, higher signal was obtained with wetter matrix applications, likely due to the require of the matrix to facilitate protein elution from the Ni-NTA ligand before crystallization. Dryer applications led to little or no signal which due to the rapid crystallization process probably contained very little protein integrated.

Overall signal/noise was higher for manual deposited samples along with higher resolution as seen in **Figure A.1**. However, I found that I could collect adequate protein spectra from center of the spot, opening the possibility of automated spectral collection as opposed to collecting spectra at the periphery of matrix crystallization, as needed for dried-droplet samples.



**Figure A.28: Performance using TM-SPRAYER vs manual dried-droplet matrix deposition.** A) image of dried-droplet spot, most of the matrix concentrated to the periphery of the spot. B.) matrix evenly spread throughout the spot, with noticeably thick layer of crystals. C.) Spectra obtained from manual collection. D.) Spectra obtained from TM-SPRAYER prepared spots.

*Materials:* A polyhistidine-tagged protein, VopE, was generously donated from the Sarah Rice lab (Feinberg School of Medicine) for these exploratory experiments. Sinapic acid and all other reagents were purchased from Sigma Aldrich. Ni-NTA SAM plates were prepared as previously discussed in this dissertation.

*Methods:* Matrix was applied to VopE treated SAM spots either manually using standard a dried-droplet technique or via the HTX TM-Sprayer (HTX Technologies, Chapel Hill, NC, USA). For dried-droplet treated spots, 0.1mL of 10 mg/mL sinapic acid in 50:49:0.1% acetonitrile/water/Trifluoroacetic acid (TFA) solution was applied with a pipette and allowed to

dry under ambient conditions. For TM-SPRAYER treated spots, 5mg/mL was instead used in an identical solvent mixture to prevent excessive clogging during matrix deposition. Matrix was sprayed with six passes at 100  $\mu$ L/min and at 40°C, with a spray spacing of 3 mm. Spray pressure was 10 psi (N<sub>2</sub>), a spray velocity of 600 mm/min, and a 46 mm sprayer nozzle distance from the sample. SAMDI-MS performed as previously described.

*Core Facility:* This work used the Northwestern University Integrated Molecular Structure Education and Research Center (IMSERC).

## Appendix B

### Gene Sequences for Im7 Library and Engineered Therapeutic Protein Sequences (Chapter 4 Study)

#### B.1 Im7-GGNWTT Positional Variant Library Sequences (5'→3') before PCR Amplification into Linear Expression Template.

Im7\_P-1

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGGAGGTA  
ACTGGACAACAGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-2

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGGAGGTA  
ACTGGACAACAGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-3

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGGAGGTA  
ACTGGACAACAGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-4

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGGAGGTA  
ACTGGACAACAGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-5

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGGAGGTA  
ACTGGACAACAGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-6

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGGAGGTA  
ACTGGACAACAGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT

GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-7

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTGGAGGTAAGTGGACAACAGAGTTTGTCAACTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-8

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGGAGGTAAGTGGACAACATTTGTCAACTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-9

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGGAGGTAAGTGGACAACAGTTCAACTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-10

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACGGAGGTAAGTGGACAACACAACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-11

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGGAGGTAAGTGGACAACACTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-12

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGGAGGTAAGTGGACAACACTTAAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-13

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGGAGGTAAGTGGACAACAAAGGAAATTGAAAAAGAGAATGT  
TGCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-14

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGGGAGGTAAGTGGACAACAGAAATTGAAAAAGAGAATGT  
TGCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-15

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGGAGGTAAGTGGACAACAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-16

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTGGAGGTAAGTGGACAACAGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-17

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTCAAGGAGGTAAGTGGACAACAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-18

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTCAACTTGGAGGTAAGTGGACAACAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-19

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTCAACTTCTTGGAGGTAAGTGGACAACAAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG

AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-20

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGGAGGTAAGTGGACAACAGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-21

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAGGAGGTAAGTGGACAACA  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-22

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGGAGGTAAGTGGACA  
ACAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-2"

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAGGAGGTAAGTGG  
ACAACAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-24

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGGAGGTAAC  
TGGACAACAGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-25

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGGGAGGT  
AACTGGACAACAGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-26

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGGA  
GGTAACTGGACAACAGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-27

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GGAGGTAAGTGGACAACATTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-28

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGGAGGTAAGTGGACAACAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-29

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAGGAGGTAAGTGGACAACAGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-30

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGGAGGTAAGTGGACAACACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-31

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGGAGGTAAGTGGACAACACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-32

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGGAGGTAAGTGGACAACAGAACACTTTGTAAAAATTACTGAGCATCCAGATGG

GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-33

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGGGAGGTAAGTGGACAACACACTTTGTAAAAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-34

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGGAGGTAAGTGGACAACACTTTGTAAAAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-35

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGGAGGTAAGTGGACAACAGTAAAAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-36

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGGGAGGTAAGTGGACAACAAAAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-37

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTAGGAGGTAAGTGGACAACAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-38

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGGAGGTAAGTGGACAACAATTACTGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-39

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAAGGAGGTAAGTGGACAACAGAGCATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-40

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACGGAGGTAAGTGGACAACACATCCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-41

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGGAGGTAAGTGGACAACACCAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-42

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAGGAGGTAAGTGGACAACAGATG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-43

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAGGAGGTAAGTGGACAACAG  
GAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-44

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAATTGGAGGTAAGTGGACAA  
CAACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-45

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAATTACTGGAGGTAAGTGGAC

AACAGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
 AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
 ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-46

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGGGAGGTA  
 GGACAACACTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-47

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATGGAGGTA  
 CTGGACAACAATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
 AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
 ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-48

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGGAG  
 GTAAGTGGACAACATATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-49

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATG  
 AGGTAAGTGGACAACATATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATT  
 AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-50

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATG  
 AGGAGGTAAGTGGACAACACCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATT  
 AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-51

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATG  
 AACGGGAGGTAAGTGGACAACAAGTATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-52

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
 AACGGATGGAGGTAAGTGGACAACAGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-53

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
 AACGGATCTGGGAGGTAAGTGGACAACAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-54

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
 AACGGATCTGATCGGAGGTAAGTGGACAACAAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 TAAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-55

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
 AACGGATCTGATCTATGGAGGTAAGTGGACAACAGACGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-56

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
 AACGGATCTGATCTATTATGGAGGTAAGTGGACAACAGATAGCCCCGAAGGGATTGTCAAGGAAAT  
 AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-57

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
 AACGGATCTGATCTATTATCCTGGAGGTAAGTGGACAACAAGCCCCGAAGGGATTGTCAAGGAAAT  
 AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
 CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-58

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
 ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
 GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG

AACGGATCTGATCTATTATCCTAGTGAGGTAAGTGGACAACACCCGAAGGGATTGTCAAGGAAATT  
AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-59

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATGGAGGTAAGTGGACAACAGAAGGGATTGTCAAGGAAATT  
AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-60

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATGGAGGTAAGTGGACAACAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-61

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGGAGGTAAGTGGACAACAATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-62

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGGAGGTAAGTGGACAACAGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-63

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATGGAGGTAAGTGGACAACAAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-64

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCGGAGGTAAGTGGACAACAGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-65

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGGAGGTAAGTGGACAACAATT  
AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-66

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGAGGTAAGTGGACAACA  
AAAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-67

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGGGAGGTAAGTGGACA  
ACAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-68

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGGAGGTAAGTGG  
ACAACATGGCGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-69

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCCGAGGTAAC  
TGGACAACACGAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-70

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGGAGGT  
AACTGGACAACAGCTGCTAACGGTAAGCCAGGATTTAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-71

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG

AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAGGA  
GGTAACTGGACAACAGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-72

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATT  
GGAGGTAAGTGGACAACAAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCAC  
CATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-73

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGGAGGTAAGTGGACAACAGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-74

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAAGGAGGTAAGTGGACAACAAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACCA  
TTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-75

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGGGAGGTAAGTGGACAACACCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-76

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGGAGGTAAGTGGACAACAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-77

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGGAGGTAAGTGGACAACATTTAAACAGGGCGGATCCCATCACCATCATCACCA  
TTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-78

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTGGAGGTAAGTGGACAACAAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-79

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGAGGTAAGTGGACAACACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-80

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTGGAGGTAAGTGGACAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-81

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGGGAGGTAAGTGGACAACAGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-82

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGAGGTAAGTGGACAACAGGATCCCATCACCATCATC  
ACCATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-83

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGAGGTAAGTGGACAACAGGATCCCATCACCATC  
ATCACCATTAAGGCTAGGTGGAGGCTCAGTG

Im7 P-84

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGCTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG

AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTGGAGGTAAGTGGACAACAGGATCCCATCACC  
ATCATCACCATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-85

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAAGGAGGTAAGTGGACAACAGGATCCCATC  
ACCATCATCACCATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-86

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGAGGTAAGTGGACAACAGGATCCC  
ATCACCATCATCACCATTAAAGGCTAGGTGGAGGCTCAGTG

Im7 P-87

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGAGGTAAGTGGACAACAGGAT  
CCCATCACCATCATCACCATTAAAGGCTAGGTGGAGGCTCAGTG

Im7\_WT

GAAGTGCCATTCCGCCTGACCTGTTTAACTTTAAGAAGGAGATATACATATGGAAGTGGAAA  
ATAGTATTAGTGATTACACAGAGGCTGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATGTT  
GCAGCAACCGATGATGTGTTAGATGTGTTACTCGAACACTTTGTAAAAATTACTGAGCATCCAGATGG  
AACGGATCTGATCTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGATTGTCAAGGAAATTA  
AAGAATGGCGAGCTGCTAACGGTAAGCCAGGATTTAAACAGGGCGGATCCCATCACCATCATCACC  
ATTAAGGCTAGGTGGAGGCTCAGTG

## B.2 GGNWTT Engineered Therapeutic Proteins Sequences (5'→3')

pJL1.His.NGTtag.Lep Vector: pJL1

ATGGAGAAAAAATCcatcaccatcatcaccatGGTAGCAAAGcgactaccGGAGGTAAGTGGACAACAg  
cgggaggaAAAGGATCCATGGTGCCGATTCAGAAAGTTCAGGATGATACCAAACGCTGATTTAAACCA  
TTGTGACCCGCATCAACGATATTTACATACCCAGAGCGTTAGCAGCAAACAGAAAGTTACCGGTCT  
GGATTTTATTCCGGGTCTGCATCCGATTCTGACCCTGAGCAAAATGGATCAGACCCTGGCAGTTTAT  
CAGCAGATTCTGACAAGCATGCCGAGCCGTAATGTTATTAGATTAGCAATGATCTGGAAAACCTGC  
GTGATCTGCTGCATGTTCTGGCATTAGCAAAAGCTGTCATCTGCCGTGGGCAAGCGGTCTGGAAAC  
CCTGGATAGCTTAGGTGGTGTCTGGAAGCAAGCGGTTATAGCACCGAAGTTGTTGCACTGAGCCG  
TCTGCAGGGTAGTCTGCAGGATATGCTGTGGCAGCTGGATCTGAGTCCGGGTTGTTAA

pJL1.His.NGTtag.G-CSF Vector: pJL1

ATGGAGAAAAAATCcatcaccatcatcaccatGGTAGCAAAGcgactaccGGAGGTAAGTGGACAACAg  
cgggaggaAAAGGATCCATGACACCGCTGGGTCTGCAAGCAGCCTGCCGCAGAGCTTTCTGCTGAAA  
TGCTGGAACAGGTTTCGTAATCAAGGTGATGGTGCAGCACTGCAAGAAAACTGTGTGCAACCT

ATAAACTGTGTCATCCGGAAGAACTGGTTCTGCTGGGTCATAGCCTGGGTATTCCGTGGGCACCGCT  
GAGCAGCTGTCCGAGCCAGGCACTGCAGCTGGCAGTTGTCTGAGTCAGCTGCATAGCGGTCTGTT  
TCTGTATCAGGGTCTGCTGCAGGCACTGGAAGGTATTAGTCCGGAAGTGGTCCGACACTGGATAC  
CCTGCAACTGGATGTTGCAGATTTTGCAACCACCATTTGGCAGCAGATGGAAGAATTAGGTATGGCA  
CCAGCGCTGCAGCCGACACAGGGTGCAATGCCTGCATTTGCAAGCGCATTTACAGCGTCGTGCCGGT  
GGTGTCTGGTTGCAAGCCATCTGCAGAGTTTTCTGGAAGTTAGCTATCGTGTTCTGCGTCATCTGG  
CACAGCCGTAA

pJL1.His.NGTtag.IFN-alpha Vector: pJL1

ATGGAGAAAAAATCcatcaccatcatcaccatGGTAGCAAAGcgactaccGGAGGTAACGGACAACAg  
cgggaggaAAAGGATCCTGTGATCTGCCGCAGACACATAGCCTGGGTAGCCGTCGTACCCTGATGCTG  
CTGGCACAGATGCGTCGTATTAGCCTGTTTAGCTGTCTGAAAGATCGTCACGATTTTGGTTTTCCGCA  
AGAAGAATTTGGCAACCAGTTTCAGAAAGCAGAAACCATTCCGTTCTGCATGAAATGATTCAGCAG  
ATTTTTAACCTGTTTCAGCACCAAAGATAGCAGCGCAGCATGGGATGAAACCCTGCTGGATAAATTCTA  
TACCGAACTGTATCAGCAGCTGAATGATCTGGAAGCATGTGTTATTCAAGGTGTTGGTGTACCGAA  
ACACCGCTGATGAATGAAGATAGCATTCTGGCAGTTCGCAAATATTTCCAGCGTATTACCCTGTATCT  
GAAAGAGAAAAAATACAGCCCGTGTGCATGGGAAGTTGTTTCGTGCAGAAATTATGCGTAGCTTTAGC  
CTGAGCACCAATCTGCAAGAAAGCCTGCGTAGCAAAGAATAA

pJL1.His.NGTtag.IL1-RA Vector: pJL1

ATGGAGAAAAAATCcatcaccatcatcaccatGGTAGCAAAGcgactaccGGAGGTAACGGACAACAg  
cgggaggaAAAGGATCCAGCTATAATCTGCTGGTTTTCTGCAGCGTAGCAGCAATTTTCAGAGCCAGA  
AACTGCTGTGGCAGCTGAATGGTCTGCTGGAATATTGTCTGAAAGATCGCATGAACTTTGATATCCC  
GGAAGAAATTAACAGCTGCAGCAGTTCCAGAAAGAAGATGCAGCACTGACCATTTATGAAATGCTG  
CAGAACATCTTTGCCATCTTTTCGTGAGGATAGCAGCAGCACCGGTTGGCAGGAAACCATTGTTGAAA  
ACCTGCTGGCCAATGTGTATCATCAGATTAATCATCTGAAAACGGTGCTGGAAGAGAAGCTGGAAAA  
AGAAGATTTTACACGCGGTAAACTGATGAGCAGCCTGCATCTGAAACGTTATTATGGTCGTATTCTGC  
ACTACCTGAAAGCCAAAGAATATAGCCATTGTGCATGGACCATTGTGCGTGTTGAAATTCTGCGTAAC  
TTCTATTTCAATTAATCGCCTGACCGGTTATCTGCGCAATTA