

**Genetically Encoded Fragment-Based Discovery of Inhibitors for Galectin-3**

by

Daniel Ferrer Vinals

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry  
University of Alberta

© Daniel Ferrer Vinals, 2019

## Abstract

Genetically-encoded fragment-based discovery (GE-FBD) holds the promise for accelerate ligand discovery for protein targets. By chemical bridging of readily available genetically encoded (GE) libraries of polypeptides with constant fragments, this methodology opens a fast route to discover superior lead compounds. In GE-FBD the constant fragment guides the interactions of peptide and target to the close vicinity of the binding site on the target protein. Then, biochemical screening can sample millions of peptide sequences and filter those that provide synergistic binding contributions and enhance the overall binding affinity towards the target. Glycan binding proteins (GBP) represent an important target, that is challenging to address for ligand development with conventional high-throughput screenings (HTS). Conveniently, glycans are in fact, excellent candidates for GE-FBD, however, to tackle challenging targets such as GBP, the need for optimization remains unmet on GE-FBD.

This thesis presents an optimization study for GE-FBD to enable the discovery of glycopeptides that inhibit human galectin-3(Gal3). First, I interrogated the limits of GE-FBD with respect to the initial affinity of the constant fragment (Chapter 2). To this end, GE-libraries of glycopeptides displayed on phage and expressing combinations of redundant Ser and Gly codons, enabled to track libraries individually by DNA sequencing. Next, aldehyde-based oxime ligation with hydroxylamine-containing monosaccharides, decorated each barcoded library with glycans of variable affinity towards Gal3. GE-FBD campaigns of those libraries, against plate-immobilized Gal3 discovered

glycopeptide ligands that bind to Gal3 in a sequence dependent manner. The sequence Gal-WHVP exhibited a 40-fold increase in binding potency for Gal3 compared to random Gal-peptide but was inactive as monovalent ligand. GE-FBD of fragments with no detectable initial affinity discovered ligands that bind Gal3 equally well with or without the fragment. Further implementations of GE-FBD employed a glycan fragment of higher affinity towards Gal3(Chapter 3). The glycan fragment was chemically attached to a library of  $10^8$  diverse phage displayed heptapeptides. Solution phase panning allowed to screen library-target interactions in 1:1 correspondence and in using selective elution of the bound phages I discovered monovalent ligands for Gal3. Next, the use of “silent barcodes” technology allowed to develop a genetically-encoded glycan array technology from which I performed the screening of 80+ glycan fragments (Chapter 4). This study permitted to identify new potential starting point glycan that could maximize the success of future GE-FBD campaigns on Gal3. This approach could be extended to other protein targets for which the initial fragments selection might not be obvious.

## Preface

**Chapter 1** was written as an early stage draft for an invited review on “Enabling synthetic chemistry for Genetically-encoded Fragment-Based Discovery” for the *Trends in Chemistry* journal. R. Derda wrote the initial outline and drafted the initial sections and figures. I reviewed the literature, wrote the manuscript and created the figures with advice from R. Derda.

**Chapter 2** is based on published work and reproduced with permission of D.F. Vinals, P.I. Kitov, Z. Tu, C.X. Zou, C.W. Cairo, H.C. Lin and R. Derda, “Selection of galectin-3 ligands derived from genetically encoded glycopeptide libraries”, *Pept. Sci.* 2018, *111*(1), e24097. R. Derda conceived the project and I was responsible for all the phage selection experiments. P.I. Kitov synthesised all hydroxylamine monosaccharides, Z.J.Tu and H.C. Lin synthesized the fluorescent lactose probes. I synthesised selected glycopeptides with help from A. Cruz and K. Akkerman in the synthesis of peptide precursors. I performed the binding assays by ELISA and fluorescence polarization techniques. N. Bennet designed the primers and performed the PCR amplification of clonal M13 plasmids for transformation. I was responsible for bacterial transformation, propagation and isolation and identification of monoclonal phage. I performed phage clones ELISA binding assay. The barcoded phage libraries used in this study were kindly donated by K.F. Tjhung. I analyzed sequencing data with help of R. Derda using MATLAB script developed by him. Flow cytometry experiments were designed by R. Derda and were performed by C.X. Zou. C.W. Cairo, H.C. Lin and R. Derda reviewed the manuscript and provided useful discussions and edits.

**Chapter 3** is based on non-published work initiated by R. Derda. I was responsible for phage library functionalization, optimization of screening conditions, optimization of basic elution conditions, selection screenings, data analysis and synthesis and validation of selected hits. R. Derda developed the MATLAB scripts used for Illumina data analysis. Z.J. Tu and P.I. Kitov synthesized hydroxylamine glycans derivatives. N. Bennet cloned and amplified fluorescent phage clones mNeonGreen and mCherry. E. Kitova executed the high definition MS binding measurements on biotinylated proteins, under supervision of J. Klassen.

**Chapter 4** is based on non-published work initiated by R. Derda. I was responsible for collecting and organizing the experimental data as well as modification density studies on phage, screening on carbohydrate binding proteins and titring samples from mouse studies. R. Derda performed the experiments on cultured cells. S. Sarkar, J. Maghera and M. Macauley performed mouse injections and animal manipulations. N. Bennet constructed and transformed the SDB library. S. Sarkar, J. Maghera and M. Sojitra isolated, amplified and glycosylated the phage clones. S. Sarkar, and J. Maghera characterized the glycosylation reactions on phage and performed regioselectivity of modification on phage.

## **Acknowledgments**

First and foremost, I want to thank my wife Betsy and my sons Gabriel and Alyssa for her unconditional love and care, for being always there to encourage my pace in life. I thank to my mother Nancy, my father Daniel and my sister Denisse for raising in me curiosity and discovery passions, for educating me under the principles of hard work and commitment. To them I owe everything in my life and I am endlessly grateful.

I want to thank my advisor Prof. Ratmir Derda for accepting me in his group, for his support and patience teaching and guiding me in this academic exercise. From Ratmir guidance I learnt to think proactively, to push my boundaries, and to solve problems in a real-world time scale.

I am thankful to Prof. Christopher Cairo and Prof. Robert Campbell for being my committee members and for providing opportune and constructive advice during the entire program. I would also like to thank Prof. Rylan Lundgren and Prof Florence Williams for accepting being arm's length examiners and thanks to Prof Matthew Macauley accepting being the exam chair. I appreciate Prof Matthew Pratt for coming to Edmonton to be my thesis jury.

I am thankful to lab members Simon Ng, Wadim Matochko, and Reza Jafari for their training and advices at the beginnings of my research. I thank Gareth Lambking for training me, in several instrument of Biological services facility. I thank Bela Reiz and Randy Wittal for his training and advices on analytical HPLC and MS. I thank Chunxia (Cecilia) Zou for his training and help with protein expression.

Next, I want to thank Pavel Kitov his friendship and endless advices in any lab-related questions. I am especially grateful to Alieski Cruz and Kristoffer Akkerman who actively participated on the early stages of the galectin-3 project. I am also deeply grateful to my father Daniel, Isis and Mery for their physical support to my Edmonton journey by helping me with the kids.

Finally, I want to acknowledge the financial support from University of Alberta and from my advisor Ratmir Derda for the completion of this program.

## TABLE OF CONTENTS

<b>Abstract</b> .....	<b>ii</b>
<b>Preface</b> .....	<b>iv</b>
<b>Acknowledgments</b> .....	<b>vi</b>
<b>List of tables</b> .....	<b>xii</b>
<b>List of figures</b> .....	<b>xiii</b>
<b>List of schemes</b> .....	<b>xvi</b>
<b>List of abbreviations</b> .....	<b>xvii</b>
<b>Chapter 1: Enabling synthetic chemistry for Genetically-encoded Fragment-Based Discovery</b> .....	<b>1</b>
<b>1.1. Introduction</b> .....	<b>1</b>
1.1.1. Overview and premises of fragment-based drug discovery .....	1
1.1.2. High throughput considerations on GE-FBD .....	2
1.1.3. Biophysical analysis of cPTM as a matrix/array chemistry .....	3
<b>1.2. Modification Handles and Scars</b> .....	<b>4</b>
1.2.1. Aldehyde modification handles and scars. ....	8
1.2.2. Cysteine modification handles and scars.....	12
<b>1.3. Non-canonical amino acids and scars</b> .....	<b>17</b>
<b>1.4. Enzymatic modifications and scars</b> .....	<b>19</b>
<b>1.5. Thesis overview</b> .....	<b>23</b>
<b>Chapter 2: Selection of Galectin-3 ligands derived from Genetically Encoded Glycopeptide Libraries</b> .....	<b>26</b>
<b>2.1. Introduction</b> .....	<b>26</b>
<b>2.2. Results and discussion</b> .....	<b>30</b>
2.2.1. Selection from phage libraries with genetically traceable chemical post-translational modifications (cPTM) .....	30
2.2.2. Validation of selected ligands derived from non-Galactose scaffolds	38
2.2.3. Validation of selected leads as monovalent G3C ligands.....	40
<b>2.3. Conclusions</b> .....	<b>43</b>
<b>2.4. Experimental procedures</b> .....	<b>45</b>
2.4.1. Materials and general information.....	45
2.4.2. Phage Libraries construction .....	46
2.4.3. Production of glycopeptide displayed libraries .....	47



2.4.4.	Panning of modified phage libraries.....	48
2.4.5.	Deep-sequencing analysis.....	49
2.4.6.	Production of Monoclonal phages .....	50
2.4.7.	Binding assay of monoclonal phage to immobilized G3C .....	51
2.4.8.	Fluorescence polarization assays to test the binding of glycopeptides to soluble G3C .....	52
2.4.9.	Inhibition of G3C-FL binding to Jurkat cells .....	52
2.4.10.	Synthesis of anomeric hydroxylamine conjugates of monosaccharides	53
	<i>N</i> -(2,3,4,6-Tetra- <i>O</i> -acetyl- $\beta$ -D-glucopyranosyloxy)phthalimide .....	55
	<i>N</i> -(2,3,4-Tri- <i>O</i> -acetyl- $\beta$ -L-rhamnopyranosyloxy)phthalimide.....	55
	<i>N</i> -(2,3,4-Tri- <i>O</i> -acetyl- $\beta$ -D-xylopyranosyloxy)phthalimide .....	56
	<b>Deprotection of per-acetylated glycosides of <i>N</i>-hydroxyphthalimide .....</b>	<b>56</b>
	<i>O</i> -( $\beta$ -D-galactopyranosyl)hydroxylamine .....	57
	<i>O</i> -( $\beta$ -D-glucopyranosyl)hydroxylamine.....	57
	<i>O</i> -( $\beta$ -L-rhamnopyranosyl)hydroxylamine.....	58
	<i>O</i> -( $\beta$ -D-xylopyranosyl)hydroxylamine .....	58
2.4.11.	Solid phase peptide synthesis .....	58
2.4.12.	Purification of peptides.....	60
2.4.13.	Representative example of synthesis of Gal-peptide conjugates.....	60
2.4.14.	Synthesis of FPP-1.....	61
2.4.15.	Synthesis of labelling reagent Lac-ABAO .....	62
2.4.16.	Conjugation of horse radish peroxidase with Lac-ABAO.....	66
2.4.17.	Conjugation of G3C to Fluorescein isothiocyanate.....	67
	<b>Chapter 3: Selection of monovalent Galectin-3 ligands from phage-displayed glycopeptide libraries. ....</b>	<b>68</b>
	<b>3.1. Introduction .....</b>	<b>68</b>
	<b>3.2. Previous work on the development of galectin ligands .....</b>	<b>69</b>
	<b>3.3. Results and Discussion.....</b>	<b>71</b>
3.3.1.	Design of TAZTDG-derived phage displayed libraries .....	71
3.3.2.	Solution-phase panning with affinity bead capture .....	75
3.3.3.	Probing reproducibility of discovery with color-coded phage clones	78
3.3.4.	Solution-phase screenings of glycopeptide libraries .....	81

3.3.5.	Validation of selected glycopeptides .....	85
<b>3.4.</b>	<b>Conclusions .....</b>	<b>87</b>
<b>3.5.</b>	<b>Experimental procedures .....</b>	<b>87</b>
3.5.1.	Materials and general information.....	87
3.5.2.	Synthesis of ZJ glycan.....	89
3.5.3.	Study of ZJ ester hydrolysis. ....	93
3.5.4.	Chemical conjugation of biotin to glycan binding proteins .....	93
3.5.5.	ESI-MS binding studies.....	94
3.5.6.	Construction of blocking phage.....	96
3.5.7.	Construction of phage clone SWYD that express mCherry fluorescent protein. 97	
3.5.8.	Serine-X <sub>7</sub> phage library amplification.....	98
3.5.9.	Production of phage displayed glycopeptide conjugates.....	99
3.5.10.	Panning with fluorescent selection controls .....	99
3.5.11.	Solution-phase panning with ZJ-X <sub>7</sub> and Man-X <sub>7</sub> glycopeptide phage libraries 100	
3.5.12.	DNA amplification and sequencing .....	101
3.5.13.	Differential enrichment analysis.....	101
3.5.14.	Synthesis of glycopeptides .....	102
3.5.15.	Fluorescence polarization assays to test the binding of glycopeptides to soluble G3C .....	102
<b>Chapter 4:</b>	<b>Unsupervised study of the interactions between glycan and glycan binding proteins in Genetically Encoded Fragment Based Discovery .....</b>	<b>104</b>
<b>4.1.</b>	<b>Introduction .....</b>	<b>104</b>
<b>4.2.</b>	<b>Liquid Glycan Array (LiGA) platform.....</b>	<b>104</b>
<b>4.3.</b>	<b>Results and discussion.....</b>	<b>107</b>
4.3.1.	Construction of the LiGA building block: glycan-phage conjugate. 107	
4.3.2.	Screening LiGA against glycan binding proteins in vitro .....	114
4.3.3.	LiGA platform on cell-based and <i>in vivo</i> studies .....	119
<b>4.4.</b>	<b>Conclusions .....</b>	<b>123</b>
<b>4.5.</b>	<b>Experimental procedures .....</b>	<b>124</b>
4.5.1.	Materials and general information.....	124
4.5.2.	Construction of silent distal barcoded (SDB) M13 phage library ....	125

4.5.3.	SDB clone isolation and amplification .....	127
4.5.4.	Construction of phage that transduce mCherry and mNeonGreen fluorescent proteins .....	128
4.5.5.	Monitoring phage viability under CuAAC reaction .....	129
4.5.6.	Monitoring phage viability under SPAAC reaction .....	130
4.5.7.	Analysis of glycosylation of phage samples by MALDI-TOF MS..	130
4.5.8.	Quantification of phage-surface chemical modifications. ....	131
4.5.9.	Regioselectivity experiments on pVIII.....	131
4.5.10.	Building a phage displayed Liquid Glycan Array .....	132
4.5.11.	Preparation of biotinylated lectins BioConA, BioG3C and BioGal-1 133	
4.5.12.	Screening LiGA against glycan binding proteins.....	133
4.5.13.	Screening LiGA against CHO cells expressing CD22 .....	134
4.5.14.	Panning LiGA in wild-type and CD22 <sup>hu</sup> mice.....	135
4.5.15.	Cell sorting experiments of LiGA in wild-type and CD22 <sup>hu</sup> mice...	136
4.5.16.	Illumina sequencing.....	137
4.5.17.	Data analysis.....	137
<b>Chapter 5: Conclusions and outlook.....</b>		<b>139</b>
<b>5.1. Conclusions .....</b>		<b>139</b>
<b>5.2. Future directions .....</b>		<b>140</b>
<b>References.....</b>		<b>142</b>
<b>Appendix A: Supporting information for Chapter 2 .....</b>		<b>157</b>
<b>Appendix B: Supporting information for Chapter 3 .....</b>		<b>185</b>
<b>Appendix C: Supporting information for Chapter 4.....</b>		<b>194</b>
<b>Appendix D: Selection of lactose-modified phage libraries against galectin family proteins .....</b>		<b>283</b>
<b>Appendix E: Selection of galactose-modified macrocyclic phage libraries against galectin-3.....</b>		<b>287</b>

## **List of tables**

<b>Table 2- 1.</b> Raw copy numbers from deep sequencing analysis. ....	33
<b>Table 3- 1</b> $K_D$ values of natural and synthetic ligands with galectin-3. ....	70

## List of figures

<b>Figure 1-1.</b> Modification handles and scars.....	6
<b>Figure 1-2.</b> Examples of scars introduced on several fragment-peptide conjugates identified on GE-FBD studies.....	7
<b>Figure 1-3.</b> Modification scars on phage displayed peptides.....	9
<b>Figure 1-4.</b> Synthetic chemistries that enable incorporation of diverse fragments into a library of aldehyde-decorated peptides.....	10
<b>Figure 1-5.</b> Cysteine based point-modifications and scars on GE-FBD campaigns..	13
<b>Figure 1-6.</b> Cysteine based linchpin-type modifications and its chemical scars.	15
<b>Figure 1-7.</b> Chemical scars introduced during bioconjugations on peptide display technologies, using non-canonical amino acids.....	17
<b>Figure 1-8.</b> Proposed enzymatic modifications on phage displayed peptides and scars.....	20
<b>Figure 1-9.</b> Enzymatic modifications on phage displayed peptides.....	22
<b>Figure 2-1</b> Selection of mixed libraries with genetically encoded cPTM.....	31
<b>Figure 2-2.</b> Production of Genetically encoded cPTM phage libraries.....	32
<b>Figure 2- 3.</b> Sequence composition of peptides displayed in serine-terminated phage libraries SX4.....	34
<b>Figure 2-4.</b> Validation of binding of selected glycopeptides in ELISA-like inhibition assay against immobilized G3C. ....	35
<b>Figure 2- 5.</b> Production single-clone phage-carbohydrate conjugates.....	36
<b>Figure 2-6.</b> Binding of glycopeptide-phage constructs to immobilized G3C.....	37
<b>Figure 2-7.</b> Binding of clonal phage-monosaccharide conjugates to immobilized target analyzed by phage-ELISA assay..	38
<b>Figure 2-8.</b> Discovered combinations Xylose Glucose and Rhamnose with tetramer peptide exhibited no binding to G3C.....	39
<b>Figure 2-9.</b> Fluorescence polarization inhibition assay for the analysis of the binding potency of selected leads in monovalent presentation.....	40

<b>Figure 2-10.</b> Inhibition of the binding of fluorescent LacNAc to G3C by simple carbohydrates fragments. ....	41
<b>Figure 2-11.</b> Cell membrane binding of fluorescent Galectin-3(G3C-FL) on Jurkat cells.. ....	42
<b>Figure 3-1</b> Study of the binding hot-spots near the glycan binding site of galectin-3 using FTMap.....	72
<b>Figure 3-2.</b> Design of a TAZTDG cleavable linker. ....	74
<b>Figure 3-3.</b> Stability study of ester group on ZJ glycan.....	75
<b>Figure 3-4.</b> Chemical ligation of biotin tag and mass-spectrometry characterization of the protein G3C.....	77
<b>Figure 3-5.</b> Development of colorcoded phage clones introduced as screening positive-controls.....	79
<b>Figure 3-6.</b> Validation of the reproducibility of discovery in pull-down screenings using selection control phage clone. ....	80
<b>Figure 3-7.</b> Screenings of chemically-modified phage libraries against G3C. ....	82
<b>Figure 3-8.</b> Selection of phage libraries .....	84
<b>Figure 3-9.</b> Fluorescence polarization assays to validate the binding affinity of selected glycopeptides.. ....	86
<b>Figure 4-1.</b> Cloning of SDB and Silent SVEK regions. ....	108
<b>Figure 4-2.</b> Chemical glycosylation of the phage surface using bi-functional tethers.....	110
<b>Figure 4-3.</b> Chemical ligation of azido-glycans onto pVIII protein of M13 phage. ....	111
<b>Figure 4-4.</b> MALDI mass spectrometry characterization of chemical modifications on coat protein pVIII.....	112
<b>Figure 4-5.</b> Site-specific modification of pVIII at N-terminus. ....	113
<b>Figure 4-6..</b> The loading density of glycans on the phage surface monitored by MALDI. ....	114

<b>Figure 4-7.</b> Screening LiGA against carbohydrate binding proteins. ....	116
<b>Figure 4-8.</b> Glycan binding profiles of different carbohydrate binding proteins. .....	118
<b>Figure 4-9.</b> Analysis of glycan-binding preferences of CD22 expressed on the surface of CHO cells using LiGA.....	120
<b>Figure 4- 10.</b> Injection of model LiGA into mouse. ....	122
<b>Figure 4-11.</b> PCR amplification of phage DNA extracted from organs CD22 <sup>ko</sup> /CD22 <sup>hu</sup> chimera mouse.....	123

## List of schemes

<b>Scheme 3- 2.</b> Structure of selected glycopeptides tested for G3C binding affinity in fluorescence polarization assays.....	85
--	----



## List of abbreviations

AE	Azido-ethanol
AOB	Aminoxy-biotin
BSA	Bovine serum albumin
ConA	Concanavalin A
CTB	Cholera toxin subunit B
cPTM	Chemical post-translational modification
DIPEA	<i>N,N</i> -Diisopropylaminoethylamine
DMF	Dimethyl formamide
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
eq	Equivalent
ESI	Electrospray ionization
FBS	Fetal bovine serum
FBDD	Fragment-based drug design
Fmoc	Fluorenylmethyloxycarbonyl
Gal	Galactose
Glu	Glucose
G3C	Carbohydrate recognition domain of human Galectin-3
GE-FBD	Genetically encoded fragment based drug discovery
h	Hour
HBTU	2-(1 <i>H</i> -Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	High-performance liquid chromatography
HRMS-ESI	High-resolution mass spectrometry, electrospray ionisation
LacNAc	<i>N</i> -acetyllactosamine
LNT	Lacto- <i>N</i> -tetraose

Man	Mannose
ManL	3 carbon linker- Mannose
MeCN	Acetonitrile
min	Minute
MOPS	3-Morpholinopropane-1-sulfonic acid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFU	Plaque-forming unit
Rha	Rhamnose
RP	Reversed phase
rpm	Revolutions per minute
RT	Room temperature
s	Second
ssDNA	Single-stranded DNA
tBy	<i>tert</i> -Butyl
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
UPLC-MS	Ultra performance liquid chromatography coupled to mass spectrometer
Xyl	Xylose

# **Chapter 1: Enabling synthetic chemistry for Genetically-encoded Fragment-Based Discovery**

## **1.1. Introduction**

Genetically encoded (GE) libraries are a major source of discovery of peptide-based drugs. Screening from libraries built on different display platforms such as bacteriophages, mRNA, and yeast, have discovered many therapeutic leads that are now approved drugs or are in clinical trials for treatment of hemorrhages<sup>1</sup>, cancer<sup>2-4</sup> and immune disorders.<sup>5-7</sup> Genetically-encoded fragment based discovery (GE-FBD) holds the promise for bridging readily available GE-libraries of polypeptides with foreign fragments incorporated by synthesis. The chemistry for this conjugation, is like a matrix computation that converts all elements of a large array of starting materials to an array of products in one step. The conversion incorporates new chemical features into the constituents of the array.

In this chapter we review the available synthetic strategies that can be used for chemical post-translational modification (cPTM) of GE-libraries, incorporate non-natural recognition moieties into GE-libraries, and enable GE-FBD. We also review major consideration for future development of chemistry with the emphasis on the scar introduced by modification.

### **1.1.1. Overview and premises of fragment-based drug discovery**

Fragment-based drug discovery (FBDD) has become one of the standard strategies for the development of new drug candidates. The theoretical foundations of FBDD were established ~40 years ago<sup>8</sup> and enabling technology

for the practical implementations of FBDD emerged ~25 years ago when Fesik and co-workers<sup>9</sup> performed NMR-based screenings of binding of small molecule “fragments” to a protein. Linking two fragments of micromolar affinities gave rise to a nanomolar ligand for the FK506 binding protein. Today, more than 30 drug candidates derived from FBDD, have entered clinical trials.<sup>10</sup>

The major advantage of FBDD when compared with classical high-throughput screenings (HTS) is more efficient sampling of the chemical space. Typical HTS libraries contain around  $10^6$  compounds, which represent a  $\sim 10^{-60}$  sample from the calculated number of possible drug-like molecules with MW  $\sim 500$  Da estimated at  $10^{63}$ .<sup>11</sup> The chemical space for a fragment is smaller. For a fragment of MW  $\sim 100$  Da the number of possible molecules is  $\sim 10^{11}$  therefore even a relatively small library of 1000 fragments covers a much larger fraction of the available fragment space.<sup>12</sup>

### **1.1.2. High throughput considerations on GE-FBD**

The use of DNA to encode fragments was described by Liu,<sup>13</sup> Harbury,<sup>14</sup> Nery,<sup>15</sup> and others.<sup>16</sup> DNA-encoding methods are used to build DNA encoded libraries (DEL). Libraries of complexity larger than  $10^8$  are now regularly synthesized and screened. DEL makes small footprint molecules from small fragments. Building large surface area molecules by DEL is hard because it involves either many steps on DNA or ligation of large preformed fragments.<sup>17-20</sup> GE-FBD starts from a large fragment that comes with no additional DNA ligation.

The GE-FBD technology is complimentary to DEL. Instead of linking 2~3 fragments, GE-FBD covalently incorporates a fragment (F) into a library of ~billion diverse peptides (P) and yields a new library (F-P). The peptide “fragment” can have deleterious, neutral or synergistic effect on the binding of fragment F. The structure-activity relationships (SAR) of peptides that emerged from selections can be used to predict consensus motifs within the “P” fragment.<sup>20-24</sup> Several GE-FBD campaigns successfully established SAR in discovered ligands and identified fragment-peptide combinations that exhibited improved affinity and specificity towards the target when compared with the parent F fragment.<sup>20-26</sup>

### 1.1.3. Biophysical analysis of cPTM as a matrix/array chemistry

In GE-FBD the binding affinity of a discovered F-P ligand is related to that of F and P alone. W.P. Jencks in his seminal work<sup>8</sup> postulated that contributions to the observed binding of two covalently attached fragments defined as  $\Delta G^{FP}$  can be divided into three parts:

$$\Delta G^{FP} = \Delta G^F + \Delta G^P + \Delta G^S \quad (1)$$

Where  $\Delta G^F$  and  $\Delta G^P$  are the intrinsic binding energies of F and P fragments, and  $\Delta G^S$  represents the change in Gibbs energy that results from the connection of F and P on F-P. The term  $\Delta G^S$  contains an entropic factor that describes the connection of the fragments. For example, the linkage F-P may decrease the freedom of movement of the P moiety when F binds, and this process might be less favourable than the binding of free F. Binding of F, also results in loss of internal entropy on F-P that is not required for the binding of free F. This

term is small for a small linker within the fragments *e.g.* single-bond rotation on a typical small ligand but may become large when F and P are large molecules or are connected with a structurally complex linker.

To quantitatively review the  $\Delta G^S$  in GE-FBD we noted that the cPTM that connects F and P fragments always introduces additional bonds (scars) in the structure of F-P to permit covalent linking of F and P moieties. These “scars” range from simple amide bonds, to large condensed multicyclic structures. Scars of large size add significantly to the structure of F-P and make it difficult to decouple the role of scars from the role of the fragments.

In this chapter we review the development and applications of technologies enabling GE-FBD, with emphasis on cPTM, or enzymatic post-translational modifications (PTM), and scars that these cPTM and PTM strategies produce.

## 1.2. Modification Handles and Scars

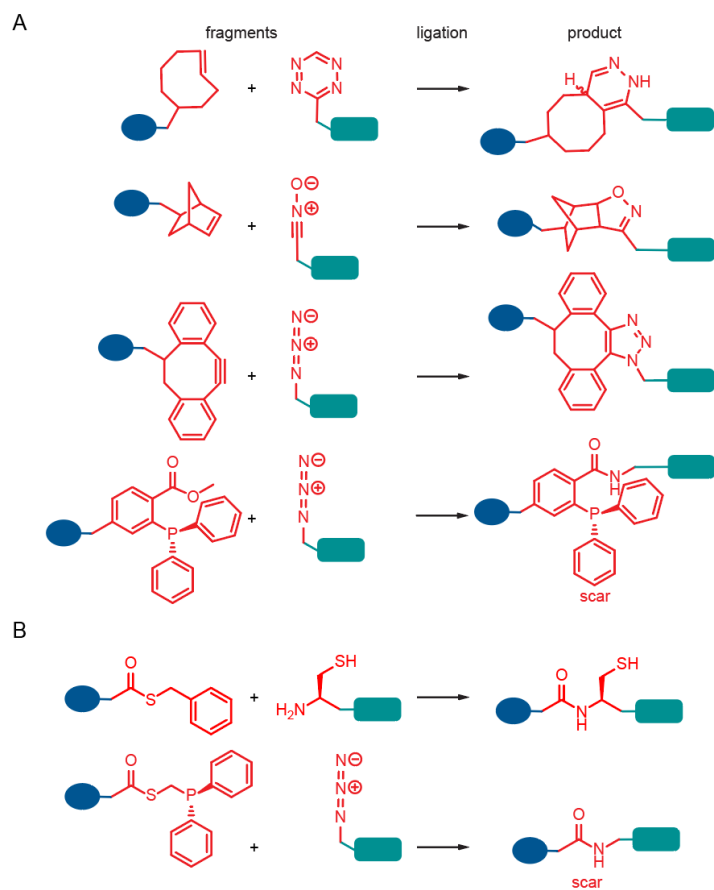
The chemical ligation of a fragment into a library of peptides displayed on phage, RNA or other carrier in GE-FBD, occurs in the presence of functional groups present on the library and the carrier. Thus, synthesis of F-P libraries requires site specific ligation chemistry<sup>27-29</sup> that produce stable products in high yields, in mild reaction conditions and that have minimal impact on the encoding carrier (DNA, RNA or phage)

A related field of bioorthogonal chemistry<sup>30</sup> focus on development of highly selective, fast, and modular reactions and place no specific emphasis on the size of scars derived from the chemical modification handles in ligated products.

Examples of bioorthogonal modifications forming prominent scars include *trans*-cyclooctene-tetrazine, norbornene-nitrile oxide, dibenzocyclooctyne-azide and Staudinger ligation reactions<sup>31-35</sup> (Figure1-1A).

Another related field of total synthesis of proteins places 100% emphasis on the exact structure of the final product and strives to develop “scar-free” (aka traceless) strategies for assembly of glycoproteins and introduction of natural PTM into intact proteins.

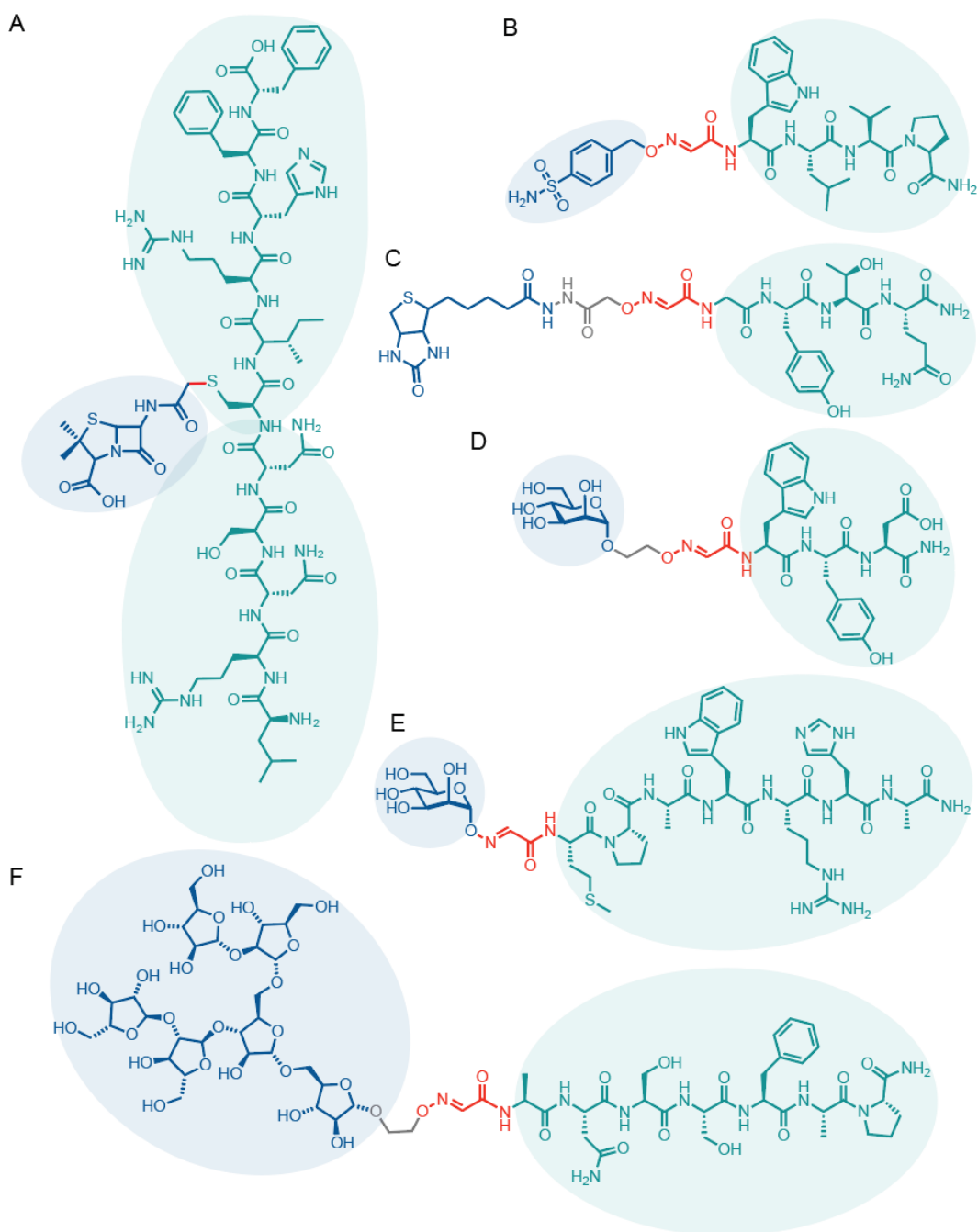
Demands for scar size in GE-FBD are somewhere between those of total protein synthesis and biorthogonal ligations. GE-FBD campaigns could aim to build a library with a known natural linkage and demand synthesis of GE-libraries with native N/O-glycosylation,<sup>36</sup> native lysine methylation<sup>37</sup> or other known PTM. In most GE-FBD campaigns, there is not a notion of a “native” linkage, making it possible to use any linkage between F and P that do not obstruct the binding.



**Figure 1-1.** Modification handles and scars. (A) Examples of bioorthogonal reactions that introduce bulky scars into ligated fragments. (B) Chemoselective ligations of fragments that leaves minor scars on conjugated fragments.

Examples chemical ligations that yield native linkage and no scars (Figure1-1B), are native chemical ligation,<sup>38</sup> traceless Staudinger ligation<sup>39</sup> and recent example of chemical ligation to cysteine-derived dehydroalanine.<sup>40</sup> The later yields mixture of two enantiomers.





**Figure 1-2.** Examples of scars introduced on several fragment-peptide conjugates identified on GE-FBD studies. (A) Roberts 2011. (B) Derda 2016. (C) Derda 2015. (E) Derda 2018. (F) Derda 2018. Fragments incorporated into the library, are in blue. Peptide is highlighted in green. Chemical scars, in red.

Some chemical conjugation strategies can leave as little as one-bond-long scar between fragment and peptide portions (Figure1-2A). Most frequent

reported examples of GE-FBD introduced five atom scars between fragment and peptide portions (Figure1-2B:E).

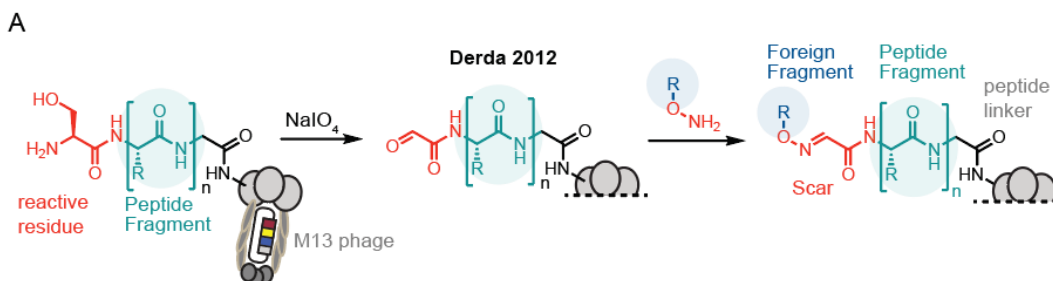
In classical FBDD, major challenge resides in finding the optimal linker that connects prospectively discovered weak fragments.<sup>41</sup> The role of scar in GE-FBD, in success and reproducibility of discovery, is less clear. It is tempting to propose that, scars generated by cPTM introduce a major selection factor but the evidence for this hypothesis is currently insufficient. Only one example of GE-FBD to date tested two linkers in similar selection campaigns and observed a larger number of differentially enriched ligands associated with “short” but not “long” linker between a mannose fragment and heptapeptide.

All other examples of GE-FBD employed only one linker geometry and, naturally, yielded a set of ligands in which linker/scar productively connected the F and P portions. It is not clear whether re-running the same campaign with the same fragment and P library connected via a different linker might yield more or less successful discovery campaign. At least two factors are needed: (i) diverse method for ligation of fragments to P libraries; (ii) robust definition and measure of the “success” of the screen.

### **1.2.1. Aldehyde modification handles and scars.**

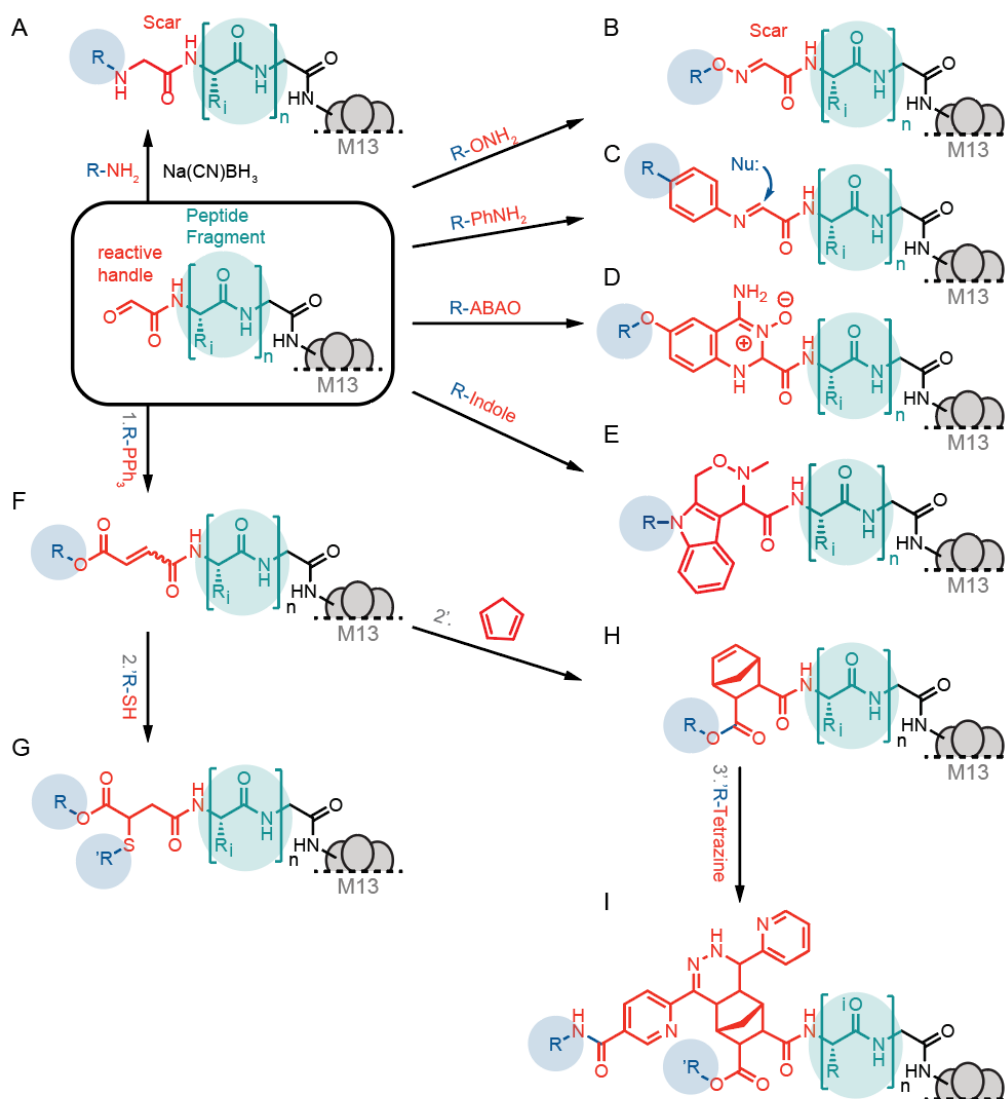
Serine residues at the N-terminus of phage displayed peptides can be selectively oxidized with sodium periodate, yielding aldehyde handles.<sup>42</sup> Aldehyde handles can be conjugated to fragments containing hydroxylamine groups (Figure1-3). This approach had been exploited in multiple GE-FBD

campaigns to enable ligation of foreign fragments to linear and cyclic peptide libraries.<sup>20, 22, 25-26</sup>



**Figure 1-3.** Modification scars on phage displayed peptides. Aldehydes are formed from oxidative cleavage of N-terminus serine and subsequently coupled with aldehyde reactive fragments.

In addition to oxime ligation, many other bioconjugation reactions could in principle, be implemented on aldehyde displayed peptide libraries (Figure 1-4). Few considerations are important for the development of new ligations: first, high rate of reaction allow to perform bioconjugations on a practical time scale using lower concentration of reagents. Reductive amination of aldehydes is a plausible strategy that introduces minimal chemical scar but reported examples of reductive amination, required 6 to 48 h at pH = 6 for 30-70% conversion<sup>43</sup> (Figure 1-4A). Hydroxylamine or substituted anilines in the same conditions (pH 4~5) yield the corresponding oximes and imines<sup>44</sup> at higher rate (Figure 1-4B,C). In buffer containing 100 mM aniline, the speed of oxime bond formation *via* attack of intermediate aniline-imine can be increased by a factor of 400 when compared to oxime bond formation from nascent aldehyde.<sup>45</sup>



**Figure 1-4.** Synthetic chemistries that enable incorporation of diverse fragments into a library of aldehyde-decorated peptides. (A) Reductive amination. (B) reaction with hydroxylamine to form of oximes. (C) Condensation with substituted anilines. Generated imines can be further functionalized with nucleophilic reagents. (D) Reaction with aminobenzamidoxime derivatives. (E) Pictet-spengler reaction with indoles. (F) Wittig reaction on peptides generates a reactive olefin that can undergo Michael addition (H) or Diels-Alder reaction (I) with cyclopentadiene to generate a norbornene handle that can be further reacted with tetrazines.

Other important consideration is the stability of the ligated product, oximes can undergo hydrolysis in aqueous media depending on the pH and conjugates' stability. Unlike classical oxime ligation, reaction of aldehydes with 2-aminobenzamidoxime (ABAO) fragments (Figure1-4D) produced conjugation product that proved to be hydrolytically stable. Ligation of ABAO fragments gave rise to a new class of condensed-ring structures based on dihydroquinazoline, as chemical scar.

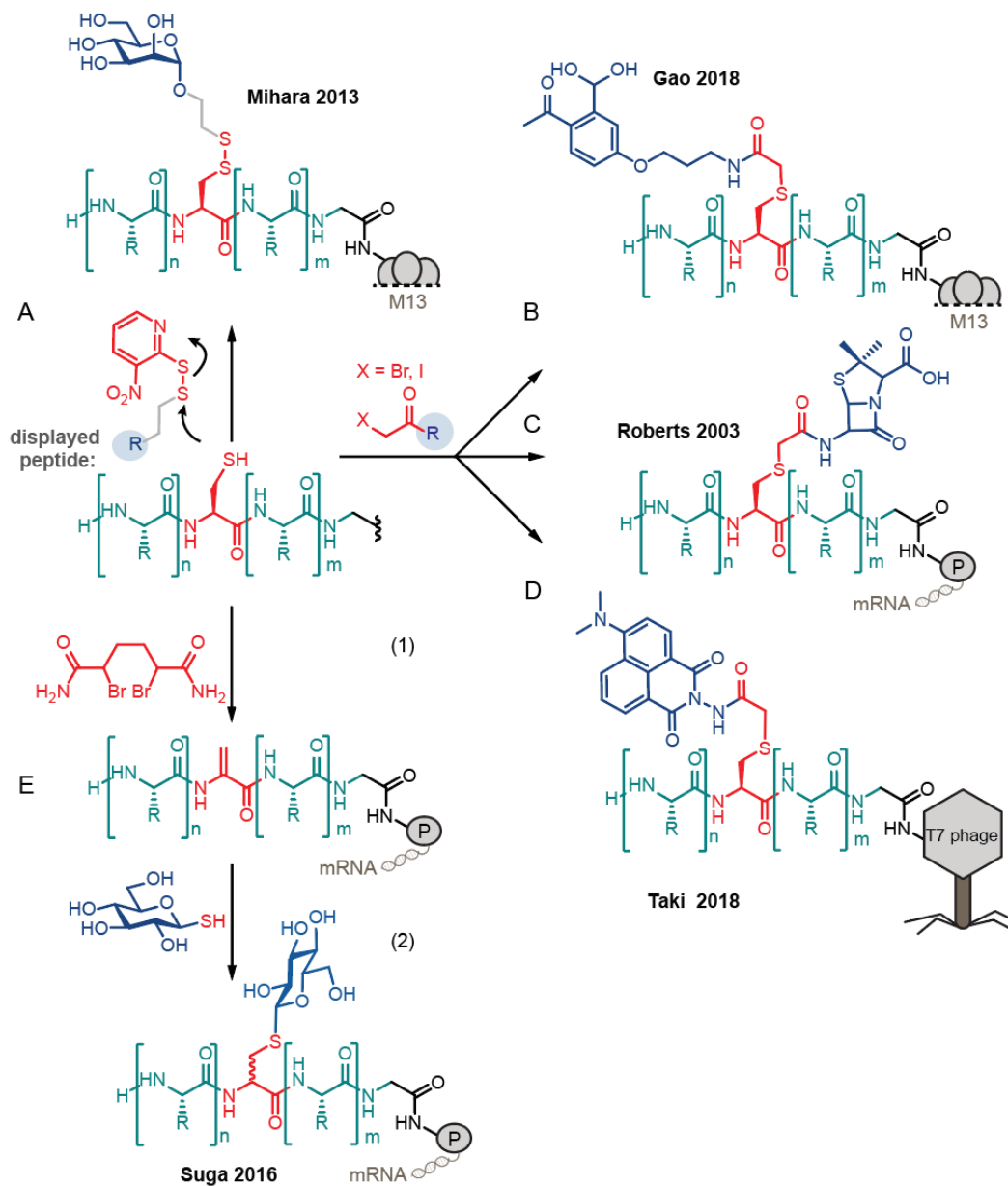
A variation of the Pictet-Spengler reaction introduces another hydrolytically stable conjugation of similar size/scar using indolyl-substituted nucleophiles.<sup>46</sup> This bioconjugation achieved rate constants of about  $k \sim 10 \text{ M}^{-1} \text{ s}^{-1}$  in aqueous conditions at pH 4.5.

Wittig reaction of aldehyde terminated peptide libraries with stabilized ylides, successfully diversified genetically encoded phage libraries<sup>47</sup> (Figure1-4F). The conjugation exhibited rates of  $0.7$  to  $5 \text{ M}^{-1} \text{ s}^{-1}$  on model peptides containing N-terminal glyoxal handles. Wittig product contains a Michael acceptor that can be further functionalized with thiols yielding potential for ligation of two fragments to phage displayed peptides (Figure1-4G). Additionally, these Wittig-generated olefins can undergo Diels-Alder reaction with cyclopentadiene under biocompatible conditions, to produce norbornene scar (Figure1-4H). The norbornene handle could, in principle, be used for further bioorthogonal conjugation of tetrazine derivatives via inverse electron-demand Diels-Alder cycloaddition (Figure1-4I). These conjugations are often praised for rate constants exceeding  $1000 \text{ M}^{-1} \text{ s}^{-1}$ <sup>33,48</sup>, but such tetrazine ligations introduce

bulky chemical scars in ligated products, and the use of such reaction for GE-FBD might not be optimal.

### **1.2.2. Cysteine modification handles and scars.**

Cysteine residues had been frequently employed as chemical modification handles for regioselective bioconjugation strategies in protein and peptides.<sup>49</sup> High nucleophilicity of thiolate groups under physiological conditions and its relatively low abundance in native proteins make them amenable for the development of chemical modification strategies in GE-FBD.<sup>50</sup> If needed, cysteine residues can be removed from phage virions prior to expression of libraries, allowing control over the site of modification. Such cysteine-free phages are much easier to produce than phage that uniformly lack other amino acids.<sup>4, 51</sup>



**Figure 1-5.** Cysteine based point-modifications and scars on GE-FBD campaigns. (A) Disulfide bond exchange allowed the generation of glycopeptide libraries. (B-D) Modification by substitution with haloalkyl electrophiles. (E) RNA-displayed glycopeptides generated after bis-alkylation-elimination of cysteine to dehydroalanine, followed by conjugate addition of thioglycosides.

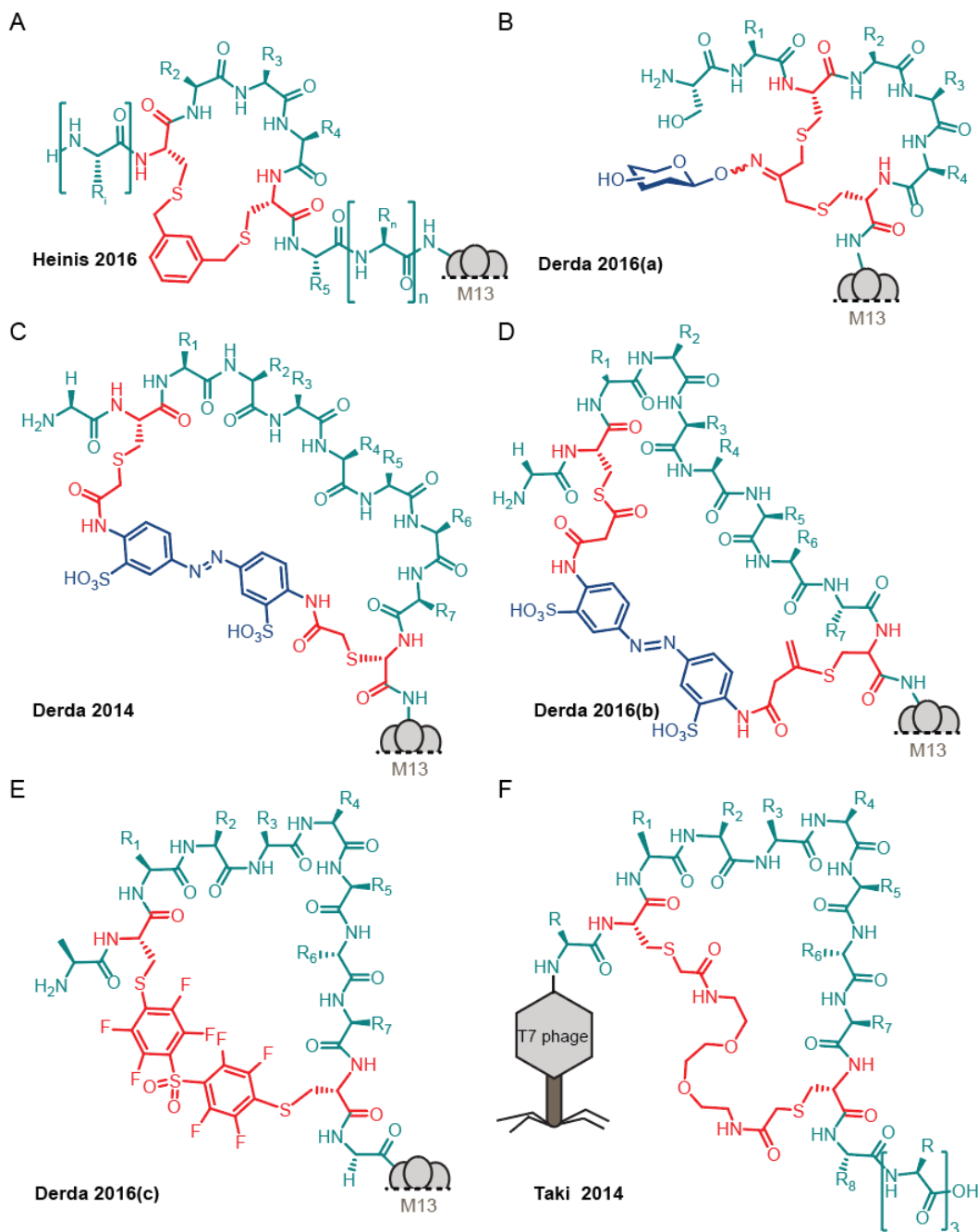
Modifications may target individual cysteine side chains, or they may cross-link at least two cysteine residues forming macrocyclic structures on peptides.

Asymmetric disulfide bridging had been used to construct F-P libraries by Mihara and coworkers<sup>21, 23</sup> to form glycopeptide libraries with a constant mannose fragment. The use of Man-dithio(5-nitro-pyridine) allowed disulfide formation between the fragment and the library (Figure 1-5A).

Roberts and co-workers<sup>24</sup> used S<sub>N</sub>2 reaction between cysteine and 6-bromoacetyl penicillanate to construct mRNA display libraries containing a penicillin constant fragment (Figure 1-5B). Gao and co-workers<sup>52</sup> decorated M13 displayed peptides with 2-acetylphenylboronic acid moieties and successfully screened those libraries against bacterial cells (Figure 1-5C). Similarly, Taki and colleagues<sup>53</sup> created chemically conjugated libraries on bacteriophage T7 display system, bearing solvatochromic bait fragments for the identification of covalent binders in model protein targets (Figure 1-5D). Halo-electrophile substrates introduce conveniently small chemical scars on ligated products.

Even smaller scars can be achieved by first converting cysteine to dehydroalanine (DHA), followed by conjugate addition of nucleophiles. Specifically, Suga and co-workers<sup>40</sup> modified mRNA libraries by first converting the cysteine to (DHA) with  $\alpha,\alpha'$ -dibromoadipic-bis-amide. The conjugate addition of 1-thio- $\beta$ -D-glucose yielded glycan-peptide libraries with short thioglycoside bond, the only drawback of this approach is formation of two diastereomers (Figure 1-5D).



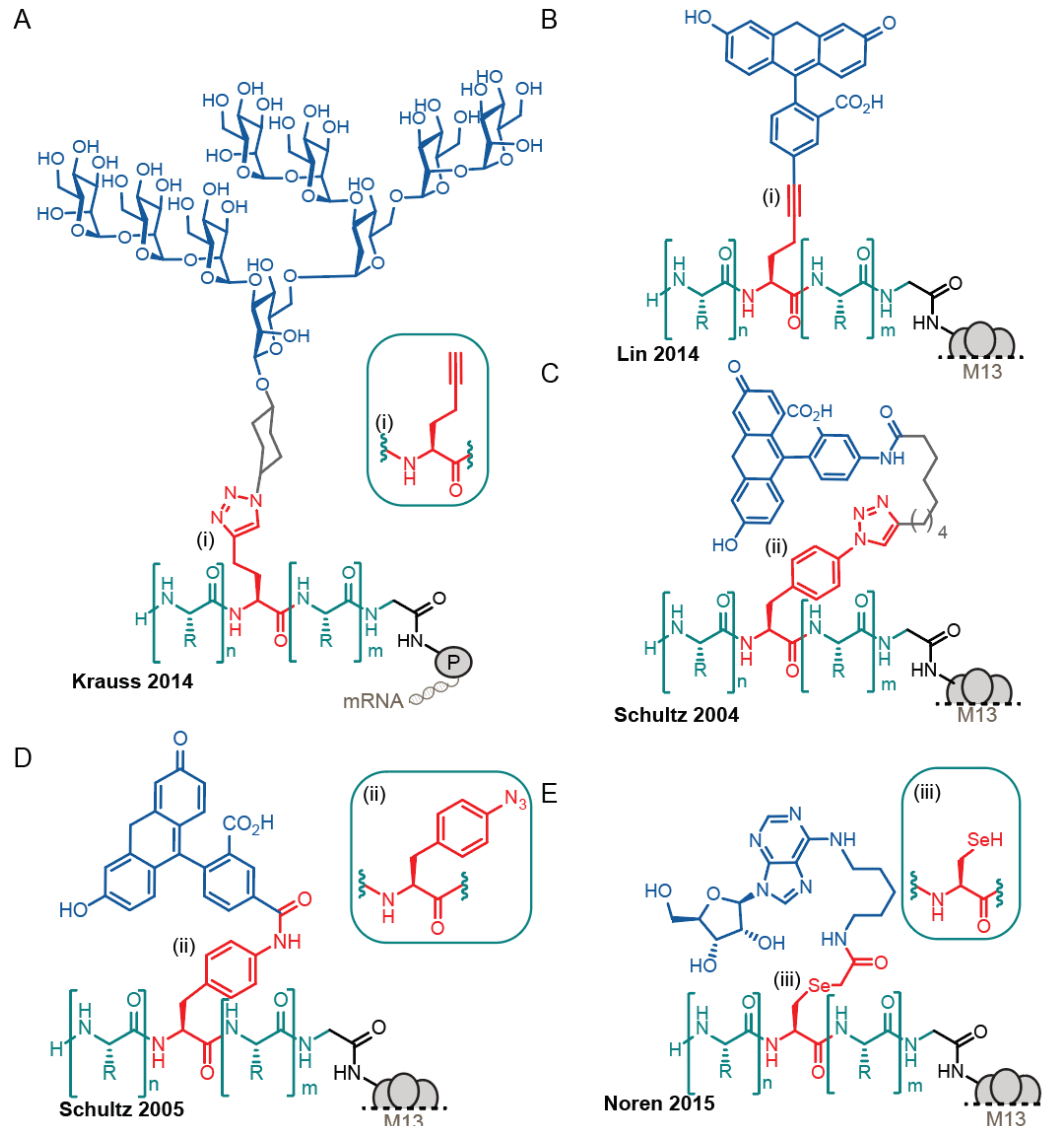


**Figure 1-6.** Cysteine based linchpin-type modifications and its chemical scars. (A) General structure of a library of phage displayed peptides after cyclization with bisbromoxylene (B) Cyclization and further glycosylation of displayed peptides.(C-D) Structures of light-responsive libraries of phage displayed peptides constructed by bis-alkylhalide (C) and bis-allenamide (D) diazobenzene core. (E) Phage displayed library after modification with decafluoro-diphenylsulfone. (F) Genetically encoded library of crown ether analogs displayed on bacteriophage T7.

The cyclization of genetically encoded libraries displayed on phage or mRNA systems by molecular “linchpins” like bisbromoxylene or trisbromoxylene are common<sup>4, 54</sup> but these strategies are not directly applicable to GE-FBD (Figure1-6A). In contrast, alkylation of cysteines with dichloroacetone linchpins followed by addition of carbohydrate hydroxylamine generated libraries of macrocyclic glycopeptides<sup>55</sup>, that can be screened on GE-FBD campaigns (Figure1-6B).

Linchpins can add a functional fragment that controls the conformation of peptide backbone in response to light. Diazobenzene cores had been successfully installed into cysteine containing phage displayed peptides via  $S_N2$  and conjugate addition reactions, with bis-alkyl halide and bis-allenamide substrates respectively<sup>56-57</sup> (Figure1-6C,D) both methods leave behind significant scar. Perfluoroaromatic linker incorporated into phage displayed libraries *via*  $S_NAr$  (Figure 1-6E) could in principle be quoted as “fragment”.<sup>58</sup> Such fragment library was successfully screened against albumin that binds to a variety of lipophilic structures. Similarly, genetically encoded libraries of crown ether analogues had been constructed by cysteine alkylation with oligoethyleneglycol units containing bis-bromoacetamide handles (Figure 1-6E).<sup>59</sup>

### 1.3. Non-canonical amino acids and scars



**Figure 1-7.** Chemical scars introduced during bioconjugations on peptide display technologies, using non-canonical amino acids. (A-B) Diversification of displayed libraries expressing propargylglycine: (A) glycopeptides generated, by azide-alkyne cycloaddition on mRNA display, (B) Phage-displayed peptides modified by Sonogashira cross-coupling reaction. (C-D) Phage display systems expressing *para*-azidophenylalanine: (C) structure of azide-alkyne cycloadducts and (D) Staudinger ligation, on displayed peptides. (E) Modification by nucleophilic substitution using a selenocysteine residue incorporated in phage displayed peptides.

The incorporation of unnatural amino acids (UAA) building blocks into genetically encoded display technologies allowed the expression of peptides with unique reaction handles that can be chemically modified.<sup>60</sup>

Krauss and co-workers<sup>61</sup> repurposed a methionine codon to express homopropargylglycine in methionine-depleted cells, enabling bioorthogonal conjugation of Man9-azide glycan into displayed peptides (Figure 1-7A). Similar UAA was employed by Lin and co-workers<sup>62</sup> in M13 libraries, to achieve a carbon-carbon bond formation using palladium assisted cross-coupling reaction (Figure 1-7B).

Orthogonal aminoacyl-tRNA synthetase/tRNA pairs and suppression of nonsense codons had been implemented by Schultz and co-workers to generate phage displayed peptide libraries containing *para*-azidophenylalanine.<sup>63-64</sup> Copper-assisted azide-alkyne cycloaddition (CuAAC) reaction, introduced a triazole scar between the peptide and ligated fragment (Figure 1-7C). The convenience of CuAAC ligation that introduce only a minor triazole scar, should be taken with caution as several studies had reported reduction of phage viability due to copper-induced toxicity.<sup>65-66</sup> With the use of phosphine substrate, Staudinger reaction introduced a fluorescein fragment with a smaller amide scar (Figure 1-7D).

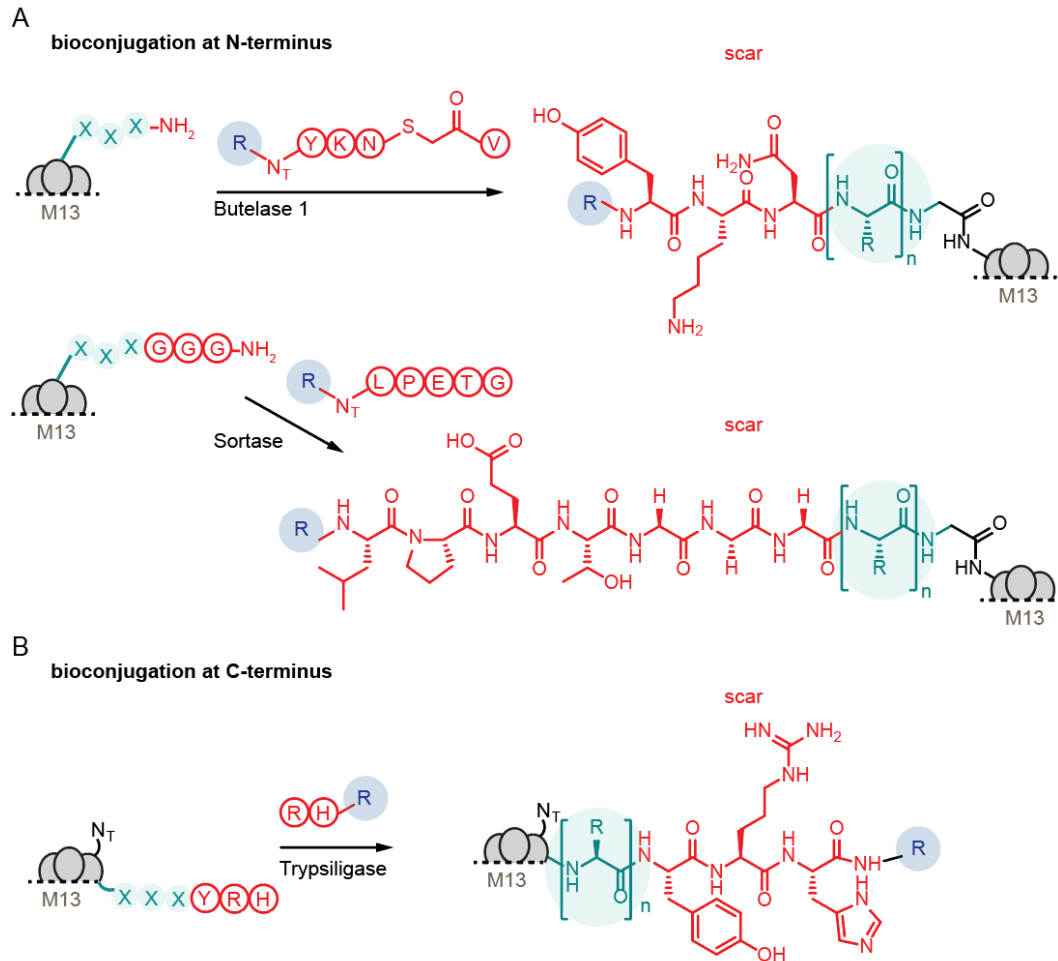
Using a selenocysteine (Sec) operon that responds to UGA nonsense codon by means of a distinctive seryl-tRNA and a selenocysteine insertion sequence, Noren and colleagues introduced selenopeptides into M13 phage displayed libraries.<sup>67</sup> Nucleophilic substitution with alkyl halide containing fragments reaction allowed the attachment of adenosine A1 receptor binding

fragment through a minor acetamide scar (Figure 1-7D). Quoted advantages are low pKa of selenol side chain of ~5.5 compared to 8.1 for the chemically related thiol side chain on cysteine, and higher nucleophilicity of selenol, but direct exploitation of this property such as selective formation of Se-R bond in the presence of SH groups in libraries has yet to be demonstrated.

Although the scars introduced by chemistries based on UAA are generally subtle adducts on ligated products, the implementation of UAA technology on GE-FBD could be challenging in terms of fidelity of incorporation, efficiency and stability of UAA containing libraries.<sup>60, 68</sup>

#### **1.4. Enzymatic modifications and scars**

The substrate specificity and homogeneity of the product on enzyme-catalyzed reactions, makes them attractive as an alternative tool to achieve the ligation of constant fragment into displayed peptides in GE-FBD.



**Figure 1-8.** Proposed enzymatic modifications on phage displayed peptides and scars. (A) Ligases Butelase 1 and Sortase A can selectively attach fragments into N-terminal position (B) Tripsyligase achieve C-terminal modification leaving a three amino acids scar.

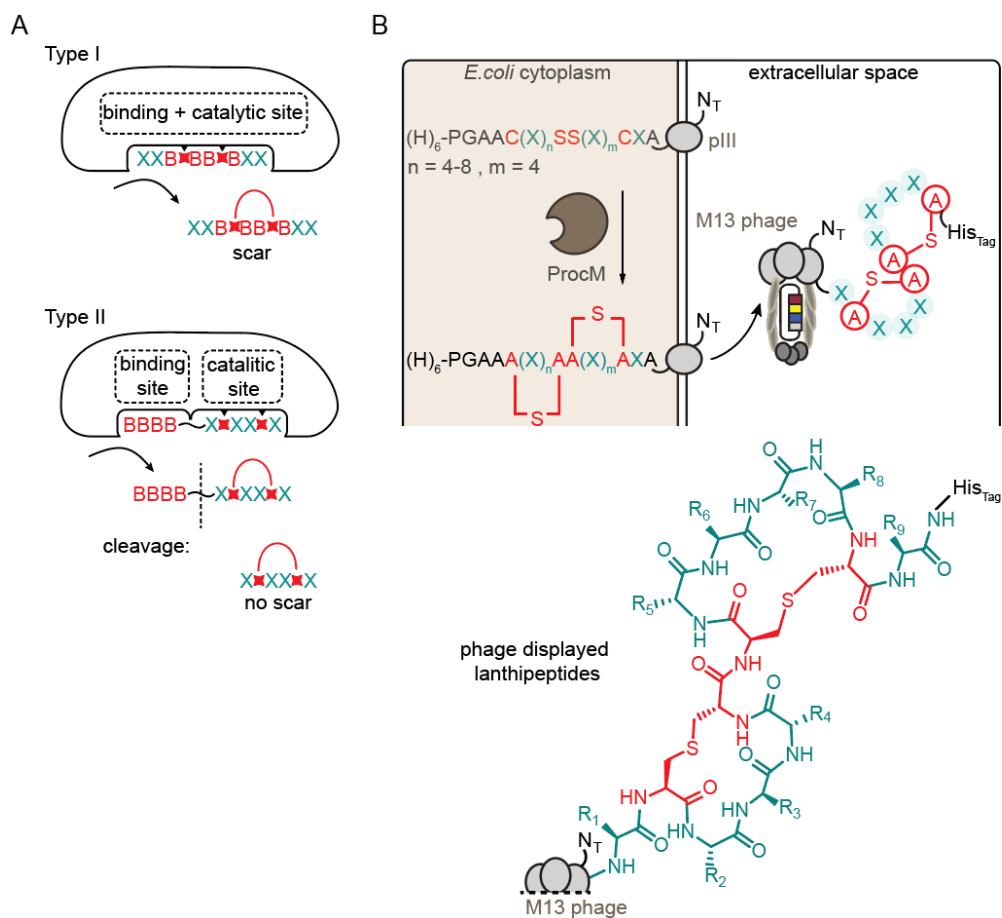
Enzymatic methods are frequently employed to label proteins and antibodies<sup>69-70</sup>. The implementation of classical ligase –BirA biotin ligase– for modification of phage libraries highlighted several challenges at once:

- (1) The recognition motif for BirA is a 15 amino acid long and modification of library by BirA or later evolved variants leave behind a ~150 atoms scar between the fragment and a putative library.
- (2) Kay and co-workers<sup>71</sup> reported that modification of phage by BirA was efficient in vitro, by treating expressed phage library by BirA. In

contrast, expression of BirA in the *E.coli* that produced the phage yielded only marginal (< 5% incorporation). The rate of modification must be faster than the rate of expression/trafficking/packaging of phage.

Butelase 1, has been used as an asparagine and aspartic acid specific ligase to conjugate thiopeptide derivative substrates to N-terminal residues.<sup>72</sup> Studies on libraries of short peptides demonstrated that this enzyme catalyzes the conjugation of any N-terminal residue with the exception of proline.<sup>73</sup> The reaction introduces three amino acids trace on ligated products. Sortase A is another frequently used enzyme to catalyze transpeptidation reactions<sup>74</sup> including phage displayed peptides. It binds the pentapeptide LPXTG where X could be glutamic acid, aspartic acid, glycine or lysine.<sup>75-76</sup> Sortase A tolerates a broad range of modifications at the N-terminal lysine residue. The enzyme cleaves the backbone between threonine and glycine, to generate an acyl-enzyme intermediate that is attacked by N-terminal amine of the oligoglycine tag in the peptide being modified. The amide bond formation with the C-terminal threonine of the substrate releases the ligated product. The resulting ligation contains the heptapeptide LPETGGG as a modification scar (Figure1-9A).

Enzyme mediated bioconjugations are feasible at the C-terminus and in principle on C-terminal displayed libraries on M13<sup>77</sup> or T7. Trypsinligase can perform C-terminal protein bioconjugations<sup>78-80</sup> to three amino acids sequence YRH at C-terminus (Figure1-9A).



**Figure 1-9.** Enzymatic modifications on phage displayed peptides. (A) General type of enzymatic substrate recognition and catalysis. Type I the substrate binding site and the catalytic site overlap in a single active site. Type II the substrate binding site is spatially separated from the catalytic site (B) Production of phage displayed lanthipeptides displayed at C-terminus of pIII.

The modification scars introduced into the ligated fragments are substantial when compared to the size of fragments in GE-FBD studies (MW~1000). Bioconjugation strategies involving Type I chemo-enzymatic labeling in which the substrate recognition site and the catalytic site overlaps, result in the introduction of significant scars. After enzymatic activity the residues required for binding of the substrate contribute to the modification scar. Type II enzymes in which binding site is spatially separated from the catalytic site lead to ligated product in which the recognition tag could be trimmed from the product



leaving efficient scar-less modifications (Figure1-9A). A recent report from Urban and co-workers<sup>81</sup> describes the successful implementation of Type II chemo-enzymatic modifications on M13 phage libraries of lanthipeptides. Using *E.coli* heterologous expression of ProcM enzyme from *Prochlorococcus* MIT9313<sup>82</sup> the authors accomplished the construction of lanthipeptides displayed on the C-terminus of pIII. ProcM recognizes peptide precursors of lanthipeptides with conserved leader sequence and catalyze thioether bridge formation on spatially separated core of random peptide sequences (Figure1-9B). The system is optimized for C-terminal display screenings, and it constitutes an important demonstration of the flexibility and feasibility of enzymatic approaches towards scar-less peptide-fragment conjugations on GE-FBD.

## **1.5. Thesis overview**

In this thesis, I optimize a novel technology for the discovery of ligands for inhibiting the carbohydrate binding protein galectin-3. This technology operates with genetically encoded libraries of high diversity random peptides, covalently attached to constant glycan fragments. In this way, the power of genetically encoded libraries enables fragment-guided biological selection of optimal combinations of glycan and short peptides that synergize in binding to the target protein. The overall success rate in glycopeptide ligand discovery is governed by three basic parameters: (i) the initial binding affinity of the constant glycan fragment, (ii) the properties of the covalent linkage between glycan and peptide fragments, and (iii) the screening technique.

In Chapter 1, I reviewed the current synthetic chemistry toolbox enabling GE-FBD with emphasis on chemical scars introduced by modifications. I discussed the biophysics basis of general fragment-based drug discovery and analyzed the chemical modification as a matrix/array chemistry. I described the chemical modification handles and scars based on canonical and non-canonical amino acids side chains and finally reviewed the possibilities of enzymatic modifications towards scar-less modifications.

Chapter 2 describes the fragment-based discovery screenings of peptide library modified with six single monosaccharide glycan fragments with different initial affinity towards galectin-3. In this study both, glycan modifier and random tetramer peptide were genetically encoded allowing to explore the role of the affinity of the initial fragment, in the outcome of the selection. Starting from low affinity fragments I discovered selective glycopeptide ligands with enhanced binding affinity towards galectin-3, but these ligands were active only when presented in a multivalent fashion.

Chapter 3 describes the implementation of the GE-FBD campaign that started from a monovalent galectin-3 binding fragment. This application built on the lessons about the affinity of the initial fragment and the outcome of the selection learnt from simple monosaccharide screenings and explored selection screenings that started from high diversity library of glycopeptides and used soluble target as a bait. I discovered peptide sequences and conserved motifs that showed *in vitro* binding affinity towards galectin-3, as monovalent ligands.

Chapter 4 describes the development of a novel glycan array technology that allowed screening of 80+ glycan fragments to identify new potential starting points for GE-FBD campaigns on galectin-3. The characterization and validation of the presentation methodology demonstrated the potentialities of expanding this approach to interrogate the interactions between carbohydrate and carbohydrate binding proteins beyond *in vitro* studies.

## **Chapter 2: Selection of Galectin-3 ligands derived from Genetically Encoded Glycopeptide Libraries**

### **2.1. Introduction**

Glycan binding proteins (GBP) represent important, underdeveloped class of therapeutic targets. Shallow topology of ligand-binding sites in GBPs make them a challenging target for high-throughput screens (HTS) and virtual screening/docking approaches.<sup>83</sup> The presence of multiple binding sites, many of which do not coincide with a native Carbohydrate Recognition Domain (CRD) poses an additional challenge for application of general selection strategies to identify ligands for GBP.<sup>84</sup> The ligands emanating from such screens often do not bind to CRD sites.<sup>84-85</sup> Fragment-based discovery could overcome the aforementioned problem by employing ligand building blocks that are known to bind to CRD. The obstacle that every fragment-based approach faces is identification of the optimal linking strategy that brings multiple residues together into a single ligand. We previously demonstrated that genetically encoded fragment-based discovery (GE-FBD) in which the glycan fragment is covalently linked to a library of  $10^8$  peptide fragments can minimize challenges of classical genetically encoded peptide libraries, high-throughput screens, and fragment-based discovery alone. In several reports, GE-FBD applied to glycan-binding proteins yielded peptide-glycan combinations 10-50 times more potent than the original glycan fragment<sup>20, 22-23, 26, 86</sup> In GE-FBD the carbohydrate fragment directs the interactions of peptide and target to the CRD and the screening process could identify peptides that provide synergistic binding contributions that enhance overall binding affinity. Few questions however remain open about GE-FBD: (i)

What is the upper and lower limit of affinity for the initial fragment? Previous GE-FBD-like approaches employed fragments with affinities of 10-100 fM,<sup>22</sup> 1.2 nM,<sup>87</sup> 1  $\mu$ M,<sup>22</sup> 60  $\mu$ M,<sup>26</sup> 150  $\mu$ M<sup>20, 22</sup> and 1 mM<sup>86</sup> (ii) Can GE-FBD work effectively when it employs a mixture of glycan fragments? With one exception<sup>22</sup> all GE-FBD approaches applied to date, used only one glycan,<sup>20, 22, 26</sup> or one type of fragment.<sup>23, 27, 87-89</sup> To address these fundamental questions, we employed GE-FBD to conduct discovery of glycopeptide ligands for GBPs called galectins starting from a mixture of fragments that have low ( $IC_{50} >10$  mM) or no detectable affinity for a galectin protein.

Galectins are a family of mammalian  $\beta$ -D-galactopyranoside binding lectins that facilitate the expression of the biological information encoded by the glycosylation system at the cellular level.<sup>90</sup> Among galectins, the chimera type galectin-3 (Gal3) with a single CRD is well known for its roles in the development, malignancy, and resistance of cancer cells to therapeutic drugs.<sup>91</sup> Overexpression of Gal3 contribute to tumor escape by the regulation of multiple immune editing responses.<sup>92</sup> The carbohydrate-protein interactions of Gal3 can be attributed to several highly conserved aminoacids at the CRD, that recognize preferentially beta-galactosides.<sup>93-94</sup> Galactose is the major binding fragment in naturally occurring ligands,<sup>95</sup> such as lactose, lacto-*N*-tetraose (LNT) and other glycan<sup>96-97</sup> that bind to the CRD of Gal3 (referred to as G3C below). Structural and crystallographic studies of Gal3 demonstrated that hexoses like glucose can be accommodated in different subsites over an extended cavity on the Gal3 binding site.<sup>98-99</sup> The CRD is permissive for saccharides with the “same chair”

conformation like xylose and glucose<sup>100-101</sup> and it has been observed that oligosaccharides containing units of mannose, xylose, and rhamnose interact with Gal3.<sup>100,102</sup> The plasticity of Gal3 in glycan recognition, makes it an interesting platform for fragment-based discovery that start from several weak glycan fragments. GE-FBD that contains Gal, Man, Xyl, Rha, and Glu fragments can test fundamental questions raised above; it could also give rise to new class of glycopeptidic ligands for Gal3.

The development of small-molecule galectin-specific inhibitors had been actively pursued by many groups and pharmaceutical companies.<sup>96,103</sup> For example, fluorinated thiodigalactoside derivatives were recently reported as selective Gal3 inhibitors.<sup>104</sup> These studies indicated that the target, indeed, can be considered as druggable with ligands composed of glycan and small molecule fragment and it is thus, a suitable model system for pilot tests of new discovery and screening strategies. Gal3 has also been subject of classical fragment-based ligand design which employed oxime, derived by linking small molecule aldehyde fragments to hydroxylamine  $\beta$ -galactoside. Such approach produced hybrid galactose-indole ligands with affinity 24 times better than the parental galactose.<sup>105</sup> Although small-molecule inhibitors exist, the discovery of glycopeptide inhibitors could be interesting for exploring ligands with alternative modes of interaction with Gal3. The peptidic nature of ligand also facilitates development of combined therapies involving simultaneous inhibition of galectins and other proteins because the peptidic portion could be expressed as part of an existing antibody or biological drug. Gal3 inhibition combined with existing

biological drugs and antibodies already has been shown to exhibit synergistic benefits on cancer treatment.<sup>106-107</sup> A recent study unveiled that tumors refractory to anti-VEGFR therapy, evade direct blocking of anti-VEGFR antibody, by engaging a secondary mechanism for activation of VEGFR *via* galectin-mediated dimerization of the glycosylated domains on these receptors.<sup>108</sup> These findings open new avenues to explore the benefits of complement the antibody therapy with peptide-based galectin inhibitors.

To address the above questions and goals, we used genetically encoded fragment-based discovery (GE-FBD) approach to identify glycopeptides that bind to the carbohydrate recognition domain of the human Galectin-3 (G3C). We generated six phenotypically identical phage libraries Ser-[X]<sub>4</sub>-Gly-Gly-Gly where X represent a random amino acid, built on variable combinations of redundant Ser and Gly codons. Oxime ligation of hydroxylamine derivatives of galactose (Gal), glucose (Glu), mannose (Man), rhamnose (Rha), and xylose (Xyl), produced a glycopeptide library in which both peptide, and glycan, can be decoded *via* DNA sequencing. Screening of this library, against G3C identified 1062 combinations of monosaccharides and peptides that exhibited a significant ( $p < 0.05$ ) enrichment on G3C and not on control selections. Glycopeptides Gal-WKPE, Gal-WHVP and Gal-LSMA displayed on phage exhibited up to 63-fold increase in binding potency to G3C when compared to phage displaying random glycopeptide or non-glycosylated SWKPE, SWHVP, and SLSMA. This work mapped the boundary conditions of the GE-FBD approach with respect to the affinity of individual fragments. We observed that fragments with no detectable affinity (glucose,

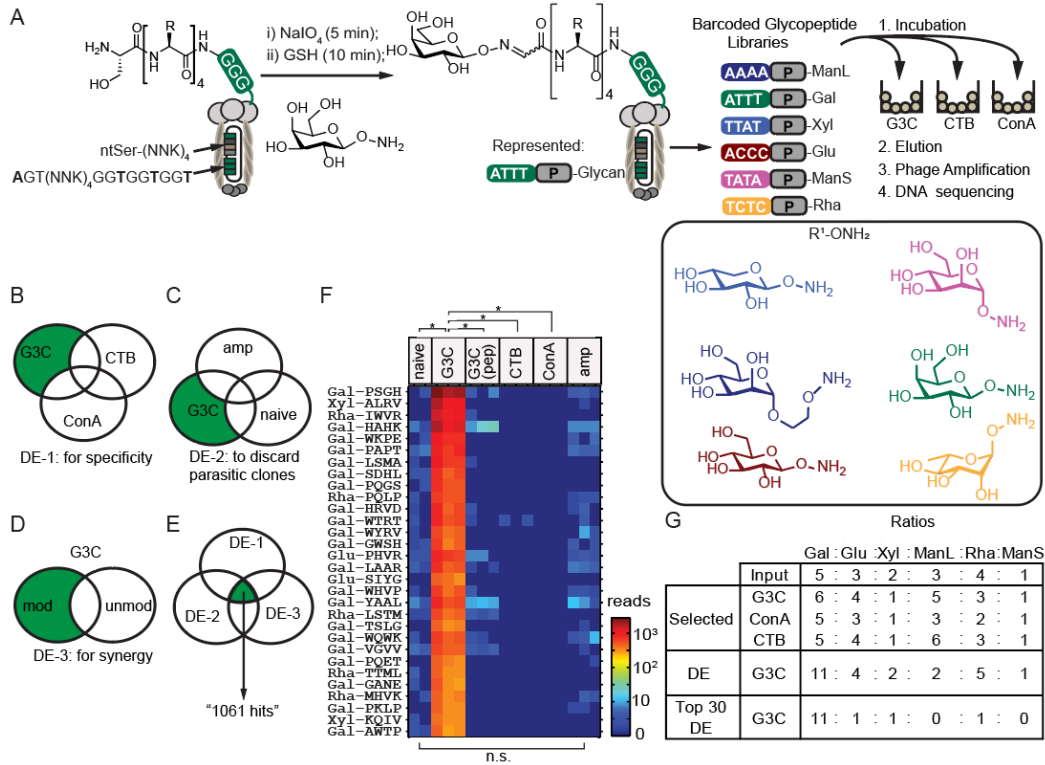
xylose, and rhamnose) diverted the selection towards ligands that bind to G3C equally well with, or without the glycan. Weak fragments (galactose,  $IC_{50} \sim 10$  mM) could effectively steer the selection towards G3C ligands in which glycan and peptide bind synergistically.

## **2.2. Results and discussion**

### **2.2.1. Selection from phage libraries with genetically traceable chemical post-translational modifications (cPTM)**

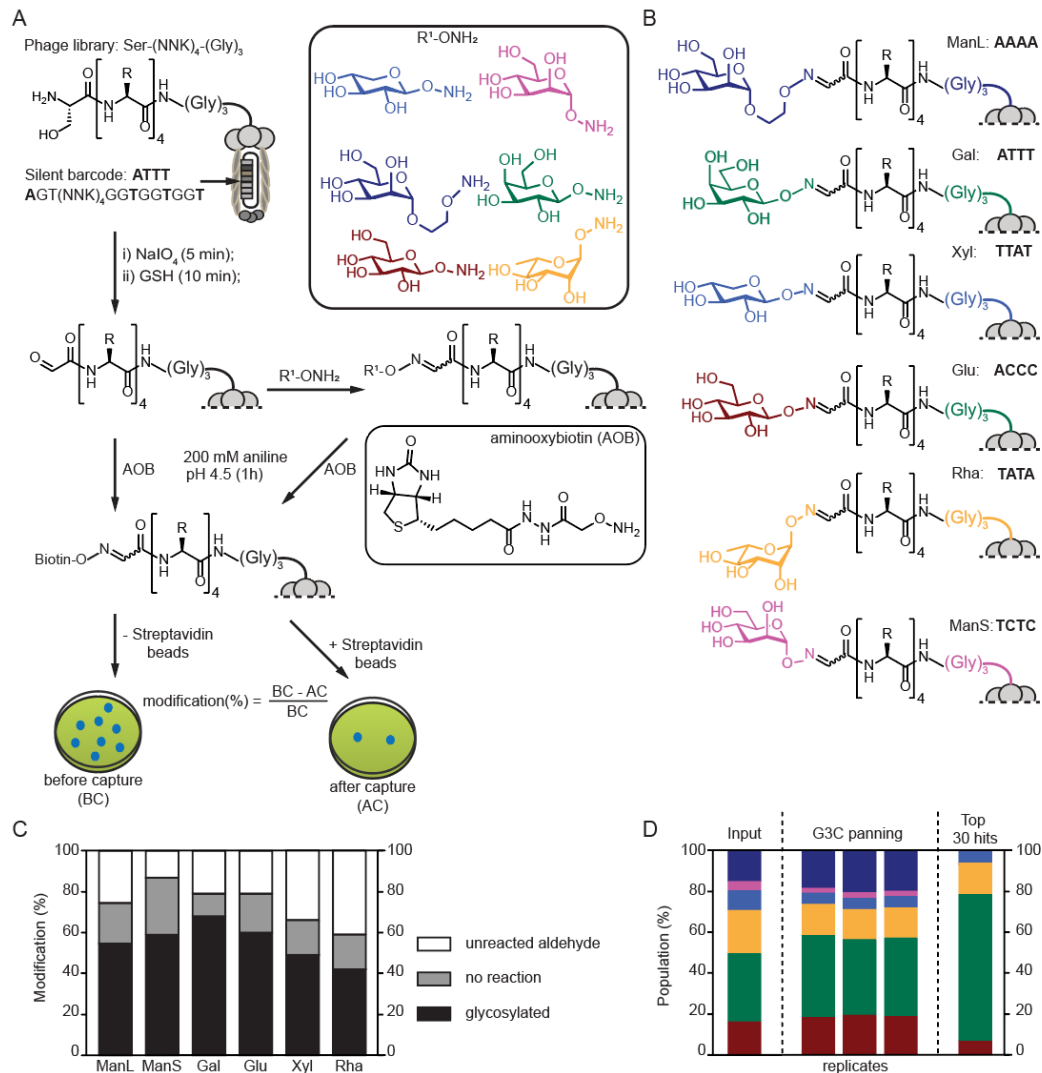
To minimize the loss of rotational entropy in ligands that combine peptide and glycan fragment, we synthesized each glycan with anomeric hydroxylamines which resulted in “one-bond-long” linker between the phage libraries and Gal, Glu, Rha, Man and Xyl fragments (Figure 2-1A). We ligated each glycan<sup>42</sup> to six SXXXXGGG libraries of identical chemical composition. In each library S and GGG portions were encoded by different combination of codons. The resulting glycopeptide libraries allowed tracking and decoding both peptide and carbohydrate portion of the displayed ligand by sequencing of the phage DNA (Figure2-2).





**Figure 2-1** Selection of mixed libraries with genetically encoded cPTM. (A) Libraries of tetrapeptides displayed on M13 phage decorated with carbohydrates through oxime ligation mixed and incubated with immobilized G3C, ConA or CTB. (B-E) Venn diagrams of the subtractive selection of controls leading to the discovered hits. (F) Top 30 from  $10^3$  total hits ( $p < 0.05$ ,  $R > 10$ ). (G) The ratios of each individual barcodes observed before and after selection on each target.

We panned this glycopeptide library against G3C, Concanavalin-A (ConA) and cholera toxin subunit B (CTB) in three replicates and sequenced all selections by Illumina deep-sequencing. In the resulting data, we discarded combinations of monosaccharide and tetramer peptide that appeared in screen against G3C and CTB or G3C and ConA, as these sequences could be prone to target-unrelated enrichment (Figure 2-1B, Appendix A-2). Fast growing clones could often dominate in phage display screenings.<sup>109-110</sup>



**Figure 2-2.** Production of Genetically encoded cPTM phage libraries. (A) Chemical modification, and biotin-capture based quantification of genetically traceable phage libraries. (B) Structure of each genetically encoded glycosylation of displayed as glycopeptides on phage libraries. (C) The yield of glycosylated libraries estimated from biotin capture experiments. (D) The fraction of each library modification before and after selection against G3C.

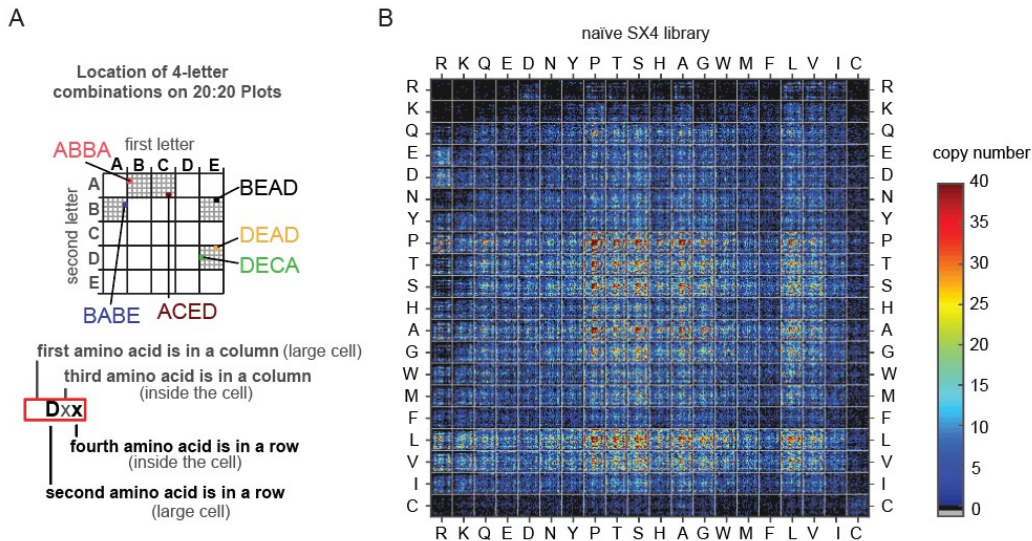
We analyzed the phage libraries before and after amplification and only retained sequences that showed no significant enrichment after phage amplification without selection (Figure 2-1C). The above differential enrichment (DE) analysis also compared the screen of glycosylated and unmodified libraries and discarded sequences that were equally enriched with and without

glycosylation (Figure 2-1D). The latter filtering strategy has previously shown to be effective in discovery of peptides that bind to GBP in synergy with the glycan.<sup>20, 22, 26</sup>

**Table 2- 1.** Raw copy numbers from deep sequencing analysis.

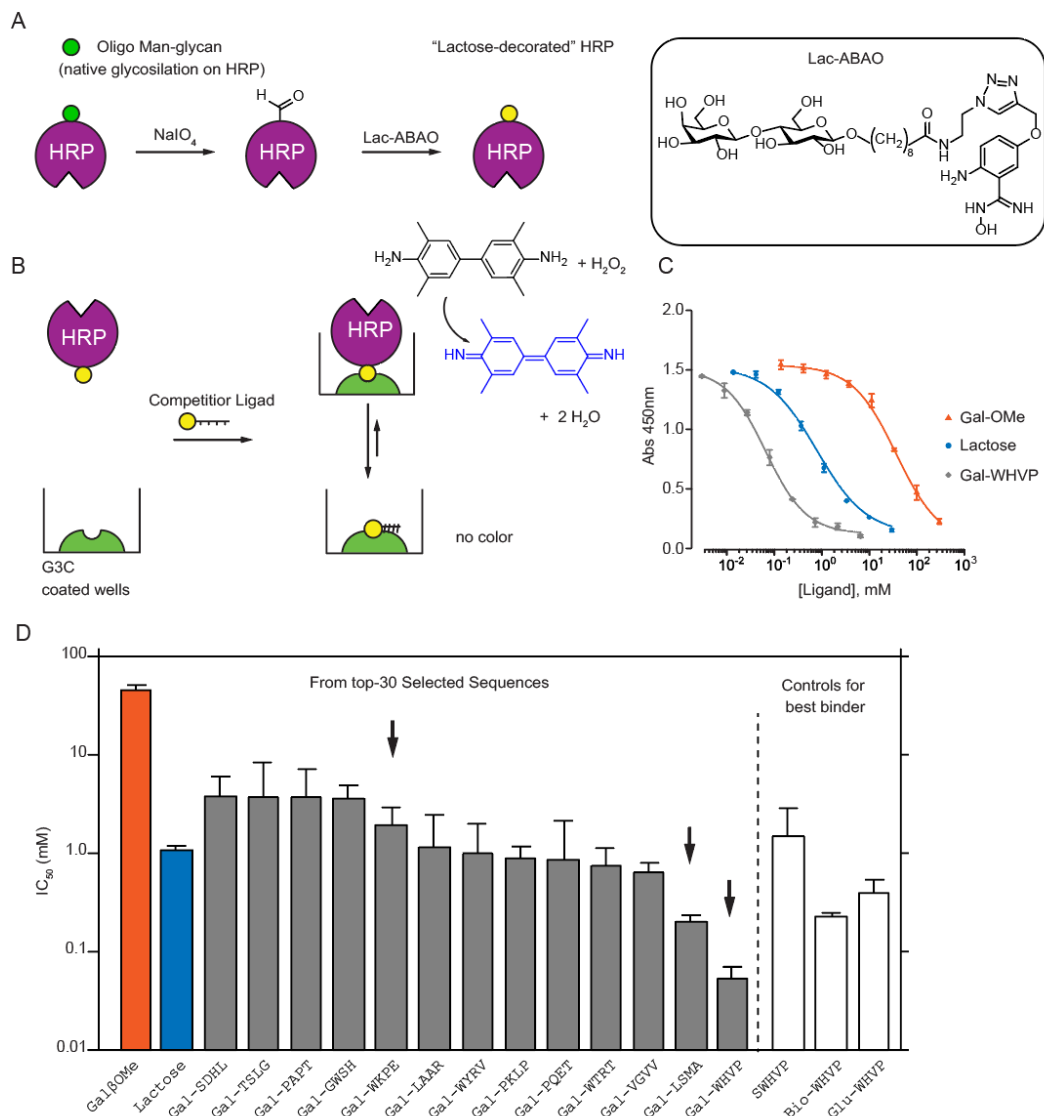
Sequence	N1	N2	<GP:G3C>			<P:G3C>			<GP:CTB>			<GP:GmA>			<GP:amp>			ratios				
			r1	r2	r3	r1	r2	r3	r1	r2	r3	r1	r2	r3	r1	r2	r3	R12	R13	R14	R15	R0
Gal-PFSGH	0	2	2841	1747	2097	3	0	7	0	0	0	0	0	0	3	4	1	377	1045	1062	317	1569
Xyl-ALRV	0	0	1841	1133	1177	0	0	0	0	0	0	0	0	0	0	1	0	1033	642	653	528	1478
Rha-IWVR	0	0	1429	956	1185	2	0	0	0	0	0	0	0	0	0	0	0	538	565	574	604	1141
Gal-HAHK	6	1	1788	1181	1321	8	19	27	0	0	0	0	0	0	7	7	7	57	673	684	90	965
Gal-WKPE	0	2	1143	752	807	0	0	0	0	0	0	0	0	0	3	0	5	679	422	429	117	915
Gal-PAPT	4	6	980	652	825	0	0	0	0	0	0	0	0	0	5	4	4	626	389	396	78	840
Gal-LSMA	1	0	1069	716	717	1	0	0	0	0	0	0	0	0	0	0	0	466	390	396	417	837
Gal-SDHL	2	2	812	485	630	0	0	0	0	0	0	0	0	0	0	0	0	486	302	307	323	725
Gal-PQGS	0	2	772	521	554	0	0	0	0	0	0	0	0	0	0	0	0	466	290	294	310	695
Rha-PQLP	2	0	723	520	563	0	0	0	0	0	0	0	0	0	1	2	2	459	286	290	114	624
Gal-HRVD	1	0	912	495	580	2	0	0	0	0	0	0	0	0	1	4	0	291	306	311	130	540
Gal-WIRT	2	0	845	513	632	1	0	0	2	0	2	0	0	0	0	3	0	372	134	317	175	536
Gal-WYRV	1	0	617	411	471	0	0	0	0	0	0	0	0	0	0	9	2	379	236	240	56	510
Gal-GWSH	1	2	676	382	449	0	0	0	0	0	0	0	0	0	3	0	2	375	233	237	92	509
Glu-FHVR	2	2	881	649	590	7	7	0	0	0	0	0	0	0	2	3	2	96	332	338	108	495
Gal-LAAR	4	1	772	461	524	1	0	1	0	0	0	0	0	0	6	5	4	260	273	277	49	470
Glu-SIYG	0	0	478	394	342	0	0	0	0	0	0	0	0	0	0	0	0	309	192	195	206	461
Gal-WHVP	1	0	708	438	558	3	0	0	0	0	0	0	0	0	2	2	1	212	268	272	110	450
Gal-YAAL	7	2	865	541	587	9	11	9	0	0	0	0	0	0	12	8	2	47	310	315	42	447
Rha-ISTM	1	1	623	423	487	4	2	4	0	0	0	0	0	0	0	0	0	89	242	246	259	440
Gal-TSLG	0	0	492	330	366	0	0	0	0	0	0	0	0	0	0	1	0	300	187	190	154	430
Gal-WQWK	2	0	675	455	512	2	1	0	0	0	0	0	0	0	2	3	11	208	258	263	41	425
Gal-VGVV	5	3	610	446	495	3	2	4	0	0	0	0	0	0	0	1	0	98	246	250	202	417
Gal-POET	0	0	457	354	329	0	0	0	0	0	0	0	0	0	0	0	1	290	180	183	139	411
Rha-TTML	4	0	495	304	305	0	0	0	0	0	0	0	0	0	0	0	0	274	171	173	183	410
Gal-GANE	2	0	515	312	284	0	0	0	0	0	0	0	0	0	0	0	0	274	170	173	182	409
Rha-MHVK	0	2	484	315	380	0	0	0	0	0	0	0	0	0	0	4	2	299	186	189	67	405
Gal-FKLP	0	0	474	327	342	0	0	0	0	0	0	0	0	0	1	1	0	289	180	183	119	404
Xyl-KQIV	2	0	438	317	304	0	0	0	0	0	0	0	0	0	0	0	0	268	166	169	178	399
Gal-AWTP	2	1	445	282	324	0	0	0	0	0	0	0	0	0	0	0	0	265	165	167	176	395
Gal-SAAM	0	0	413	304	316	0	0	0	0	0	0	0	0	0	0	0	0	263	164	166	175	392
Gal-NLGS	0	0	404	247	363	0	0	0	0	0	0	0	0	0	0	0	0	259	161	164	172	387
Gal-QVWH	0	2	423	290	292	0	0	0	0	0	0	0	0	0	0	0	0	253	157	160	168	378
Gal-MQEK	2	2	555	346	378	0	0	1	0	0	0	0	0	0	6	1	3	238	199	202	49	374
Gal-WSLQ	1	2	669	406	477	1	7	8	0	0	0	0	0	0	0	4	4	62	242	246	69	358
Xyl-DASR	4	2	404	291	236	0	0	0	0	0	0	0	0	0	0	0	0	233	145	147	155	347
Rha-ATPL	0	0	368	254	270	0	0	0	0	0	0	0	0	0	0	0	0	226	140	143	150	337
Xyl-SA-H	0	0	351	246	275	0	0	0	0	0	0	0	0	0	0	0	0	222	138	140	148	331
Glu-HSWY	0	3	403	284	288	0	0	0	0	0	0	0	0	0	3	6	3	247	153	156	33	331
Rha-PFYM	0	0	384	249	242	0	0	0	0	0	0	0	0	0	0	0	0	219	136	138	145	326
Rha-AHMM	2	2	438	271	265	0	0	0	0	0	0	0	0	0	3	3	4	242	151	153	37	326
Gal-SWGH	6	1	602	351	457	8	10	6	0	0	0	0	0	0	1	5	4	40	221	224	54	322
Xyl-FELR	1	0	460	345	315	2	0	0	0	0	0	0	0	0	1	0	1	168	176	179	111	322
Glu-LDHK	0	1	316	275	237	0	0	0	0	0	0	0	0	0	0	0	0	212	132	134	141	317
Gal-IMAV	3	0	554	327	435	6	3	2	0	0	0	0	0	0	2	0	2	71	207	210	92	317
Gal-IWPH	5	5	581	367	418	6	9	2	0	0	0	0	0	0	5	6	4	52	214	217	39	312
MaL-WQHQ	1	0	351	291	219	0	0	0	0	0	0	0	0	0	0	1	0	218	135	138	111	312
Rha-GAYQ	1	0	452	298	348	2	0	0	0	0	0	0	0	0	0	0	3	165	173	176	86	309
Xyl-TTHP	0	0	494	300	304	5	6	3	0	0	0	0	0	0	0	0	0	49	170	172	181	306
Gal-TAHQ	0	1	405	222	211	0	0	0	0	0	0	0	0	0	0	0	0	205	127	130	136	306
<b>Total</b>	<b>522</b>	<b>345</b>	<b>1257</b>	<b>788</b>	<b>740</b>	<b>1666</b>	<b>741</b>	<b>665</b>	<b>446</b>	<b>420</b>	<b>420</b>	<b>417</b>	<b>462</b>	<b>423</b>	<b>489</b>	<b>508</b>	<b>395</b>	<b>(reads x1000)</b>				

Intersect of the three DE-analyses yielded 1061 glycopeptides that satisfied all criteria (Figure 2-1E). The p-values were calculated using normalized reads and two-tailed, unequal variance Student t-test. The selected sequences of top 50 filtered sequences are shown on Table 2-1. Last four columns describe the ratios between normalized, average copy numbers; R0 is a vector sum of all the ratios, the list was sorted by R0.



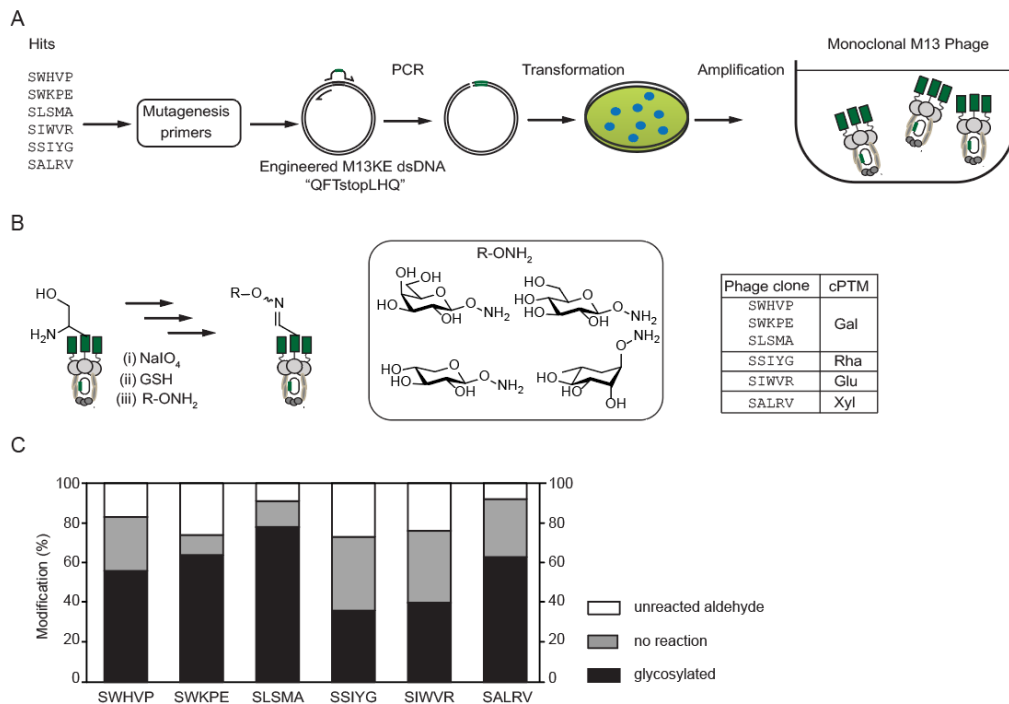
**Figure 2-3.** Sequence composition of peptides displayed in serine-terminated phage libraries SX4. (A) Model representation of a 5x5 letter plot with 25x25=625 pixels that describes the location of 4 letter combinations of the letters A, B, C, D and E. (B) In the 20:20 plots each peptide sequence in the library is presented as a unique pixel in a specific location.

Examining the ratio of the six encoded carbohydrate modifications in naïve and selected phage libraries showed that the Gal:Glu:Xyl:ManL:Rha:ManS sub-populations, changed from 5:3:2:3:4:1 ratio in naïve (Figure 2-1C) to 6:4:1:5:3:1 in libraries selected on G3C and to 11:4:2:2:5:1 ratio after application of DE analysis. In top 30-50 glycopeptides, more than 60% of the hits population were galactose-bearing tetrapeptide combinations (Figure 2-2D). Many of the discovered sequences featured tryptophan residues in the first position immediately linked to the glycan, indicating that the indole functionality might be synergistically contributing to binding. Although tryptophan was present in many of the selected sequences emerged from the screenings experiments, this residue was not found preferentially enriched in the sequence composition of the naïve library (Figure 2-3). Similar synergistic binding of indole was observed in previous GE-FBD approaches.<sup>20, 22</sup>



**Figure 2-4.** Validation of binding of selected glycopeptides in ELISA-like inhibition assay against immobilized G3C. (A) Chemical modification on the native HRP to yield HRP-Lactose conjugate (HRP-Lac). (B) Representation of the competition assay with the HRP-Lac against G3C. (C) Typical inhibition curves obtained in competitive assay with glycopeptides. (D) The affinity of selected G3C ligands measured as IC<sub>50</sub>. Sequences pointed with arrows were analyzed on phage-based assays.

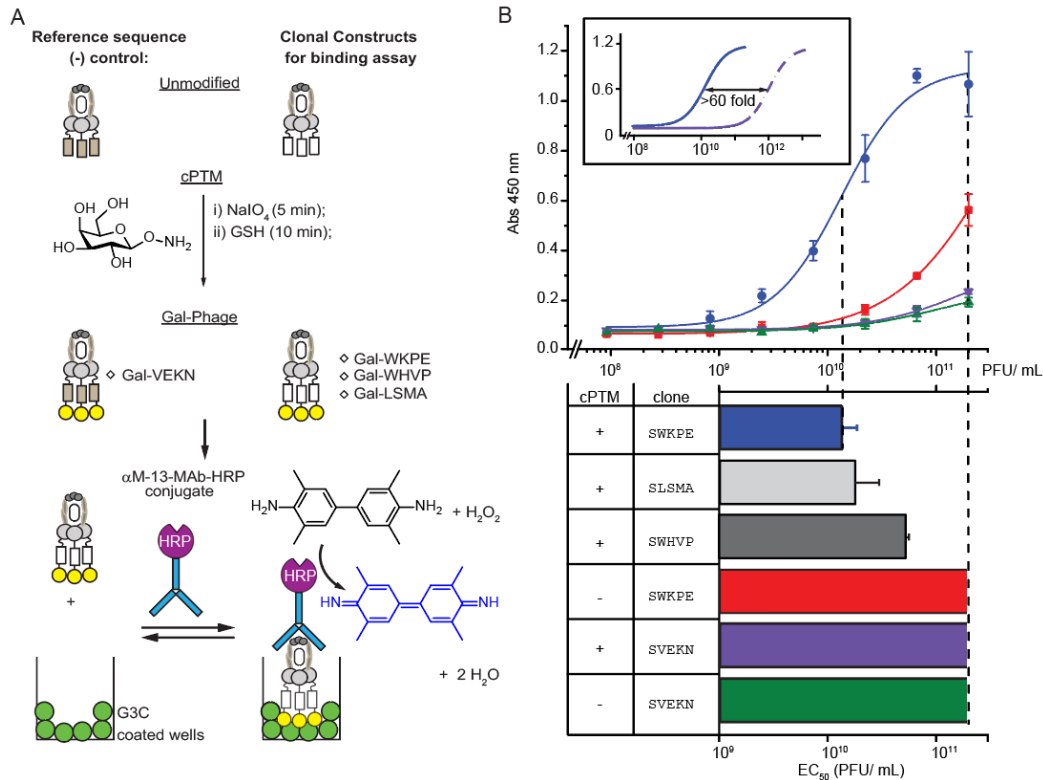
We then proceeded to synthesis and validation of these sequences in an assay that tested inhibition of G3C and horseradish peroxidase conjugated to Lactose (Lac-HRP).



**Figure 2-5.** Production single-clone phage-carbohydrate conjugates. (A) Site directed mutagenesis procedure to generate six clonal constructs. (B) The phage clones were decorated with aminoxy derivatives of Gal, Rha, Glu and Xyl using aldehyde-oxime chemistry. (C) Conversion of glycosylated phage clones estimated from biotin capture experiments.

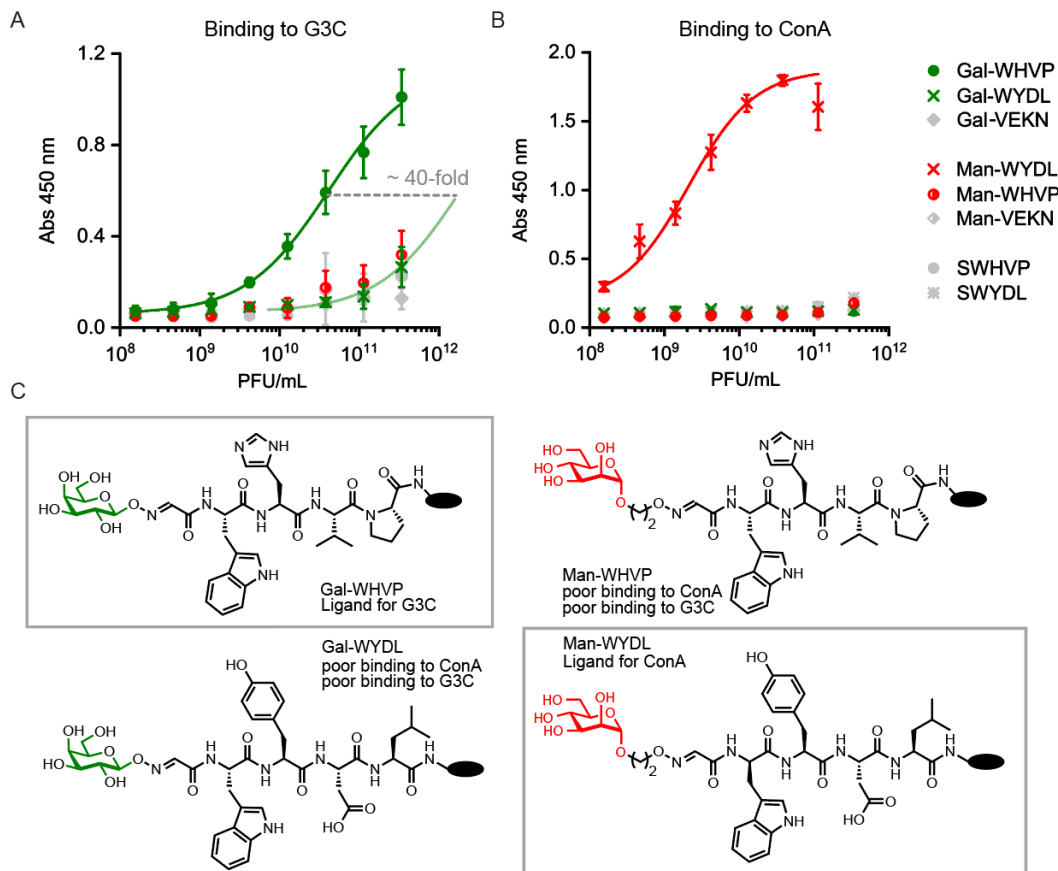
We tested the ability of 13 synthetic soluble Gal-peptides predicted by DE analysis to inhibit binding of Lac-HRP to G3C immobilized on the surface of microtiter well plates and observed an average millimolar  $IC_{50}$  for 11 out of 13 peptides. Two Gal-peptide conjugates Gal-LSMA and Gal-WKPE exhibited a significant difference ( $IC_{50} < 1$  mM) from the rest (Figure 2-4D). Based on these observations we selected three sequences for further validation.

We used cloning by PCR-mutagenesis to produce the hits. Each clone was amplified and chemically glycosylated with the corresponding glycan (Figure 2-5).



**Figure 2-6.** Binding of glycopeptide-phage constructs to immobilized G3C. **(A)** Diagram of ELISA-phage assay, that measures the ability of glycosylated phages to bind on G3C coated wells with a general representation of the glycosylation of the phage clones using oxime chemistry. **(B)** Concentration dependent curves of ELISA-phage assay, comparing quantitatively the ability of different glycosylated phages clones to bind G3C coated wells. Lower panel: estimated binding potency of selected glycosylated phage clones showed as EC<sub>50</sub> values.

The binding assay of the glycopeptides Gal-WHVP, Gal-LSMA and Gal-WKPE displayed on phage, showed that those ligands exhibited 17, 48 and 63-fold enhancement in binding to G3C when compared to unrelated galactosylated peptide Gal-VEKN, displayed on phage (Figure 2-6). The later construct originated from phage that displayed SVEKN sequence not selected in our experiment and commonly employed in our group as a control<sup>26, 42, 47</sup>. The sequence WKPE synergized with the Galactose in binding to G3C because the phage displaying aglycone peptide fragment SWKPE had 28-fold weaker potency than Gal-WKPE.



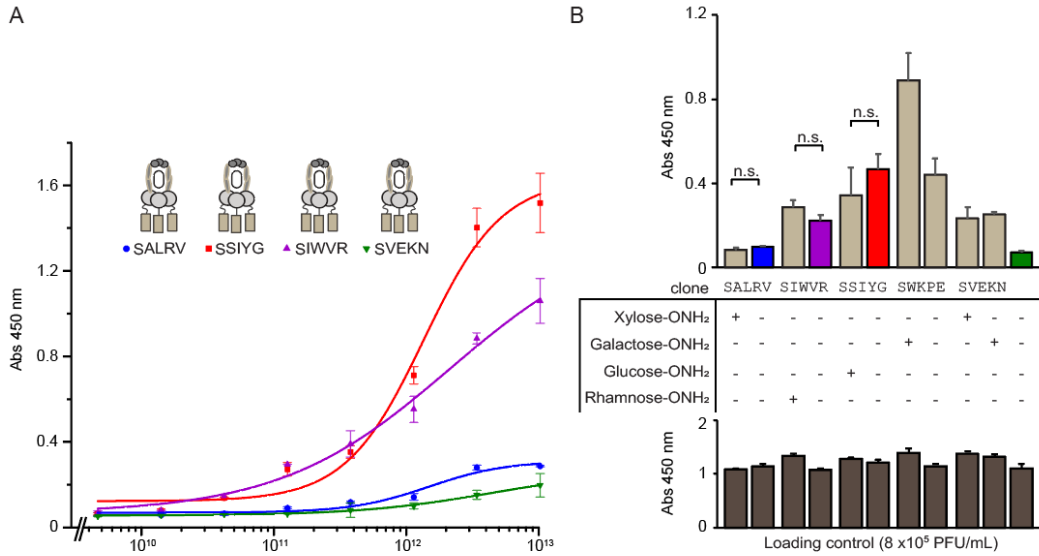
**Figure 2-7.** Binding of clonal phage-monosaccharide conjugates to immobilized target analyzed by phage-ELISA assay. (A) Binding of selected sequences to G3C: Gal-WHVP showed the strongest binding affinity to G3C whereas Man-WHVP and the aglycon WHVP were equally weak binders. (B) Binding curves of selected ligands, to ConA. (C) Structure of monoclonal phage constructs of the ligands Gal-WHVP, Gal-WYDL, Man-WHVP and Man-WYDL. The affinity of the ligand Gal-WHVP not selected in ConA screening was undistinguishable from control sequence Man-VEKN.

We also observed synergy in binding to G3C for the sequences SLSMA and SWHVP when decorated with Gal (Figure 2-7).

### 2.2.2. Validation of selected ligands derived from non-Galactose scaffolds

We were intrigued to observe phages that displayed clonal sequences corresponding to Rhamnose, Xylose and Glucose modified peptides, emerged among the top-30 hits in panning on G3C (Figure 2-1F).





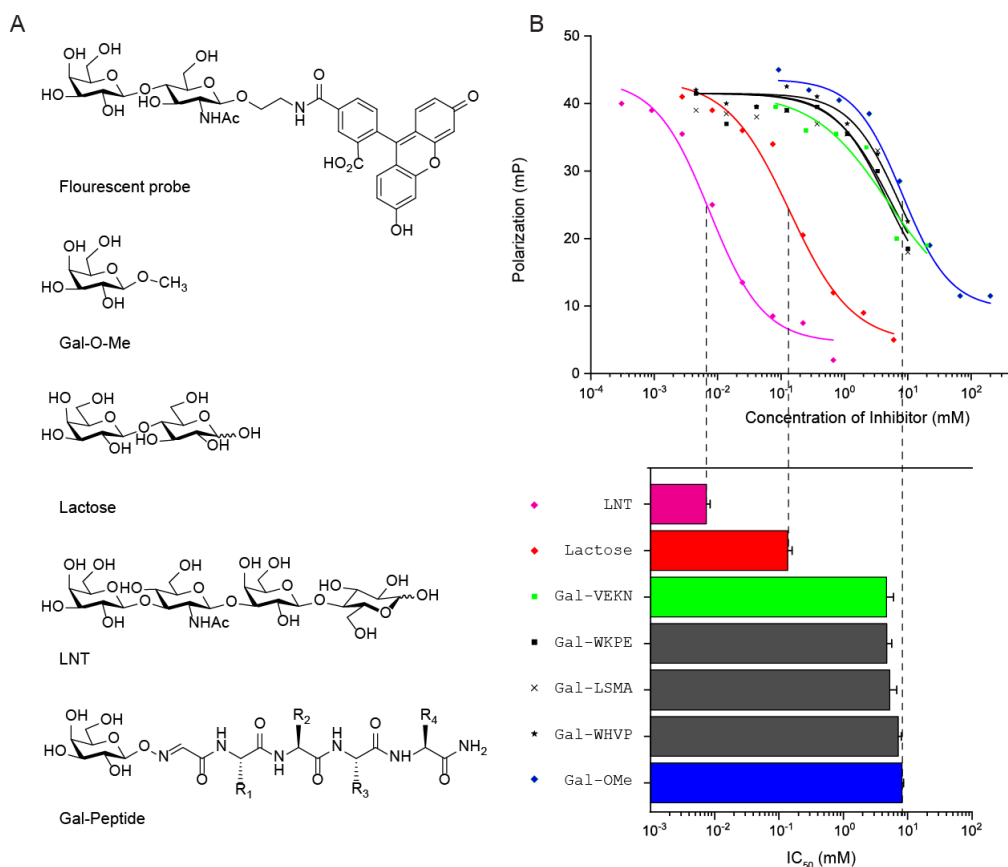
**Figure 2-8.** Discovered combinations Xylose Glucose and Rhamnose with tetramer peptide exhibited no binding to G3C. **(A)** The concentration dependent binding curves of the no glycosylated phage clones SALRV, SSIYG, and SIWVR compared to the reference clone SVEKN. **(B)** Binding of Clonal construct of Glu-SIYG, Xyl-ALRV and Rha-IWVR compared to the corresponding non-glycosylated clones.

Testing of the clonal constructs: SALRV, SSIYG and SIWVR for binding to immobilized G3C showed that sequences SSIYG and SIWVR had higher affinity than the reference SVEKN peptide even in the absence of the corresponding carbohydrate modification (Figure 2-8A). To assess whether glycan and peptide act in synergy, we compared binding of Xyl-ALRV, Glu-SIYG and Rha-IWVR and their aglycons (SALRV, SSIYG and SIWVR) displayed on phage to G3C coated wells. It was important to use wells coated individually with monoclonal phage in parallel, to control for identical loading of each construct (Figure 2-8B). In these conditions, none of the tested clones showed significant enhancement when unmodified and glycosylated phage were compared. Based on these observations, we concluded that Xylose, Glucose and Rhamnose did not contribute significantly to binding of ALRV, SIYG or IWVR

peptides and these fragments were likely selected due to interaction of peptide with protein.

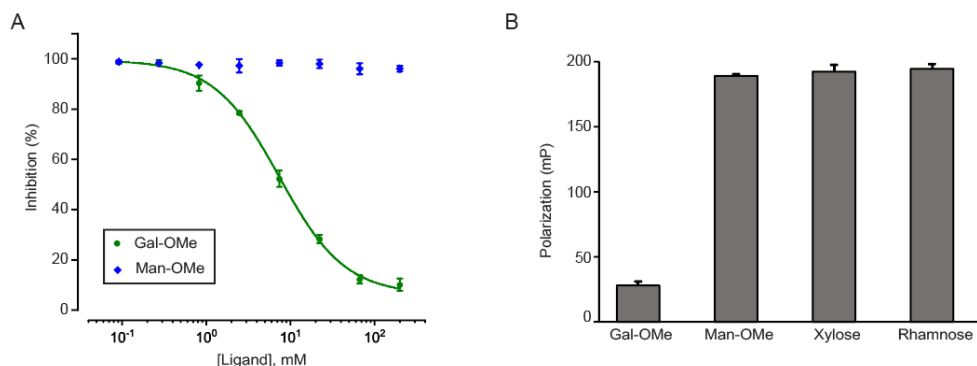
### 2.2.3. Validation of selected leads as monovalent G3C ligands

To evaluate the inhibitory potency of the monovalent glycopeptide ligands Gal-LSMA, Gal-WKPE and Gal-WHVP, we tested these ligands in a soluble inhibition assay. When taken beyond the primary inhibition screen, the  $IC_{50}$  of all tetramer peptides was indistinguishable from  $IC_{50}$  of Galactose or galacto-peptide constructed by using random SVEKN sequence (Figure 2-9B).



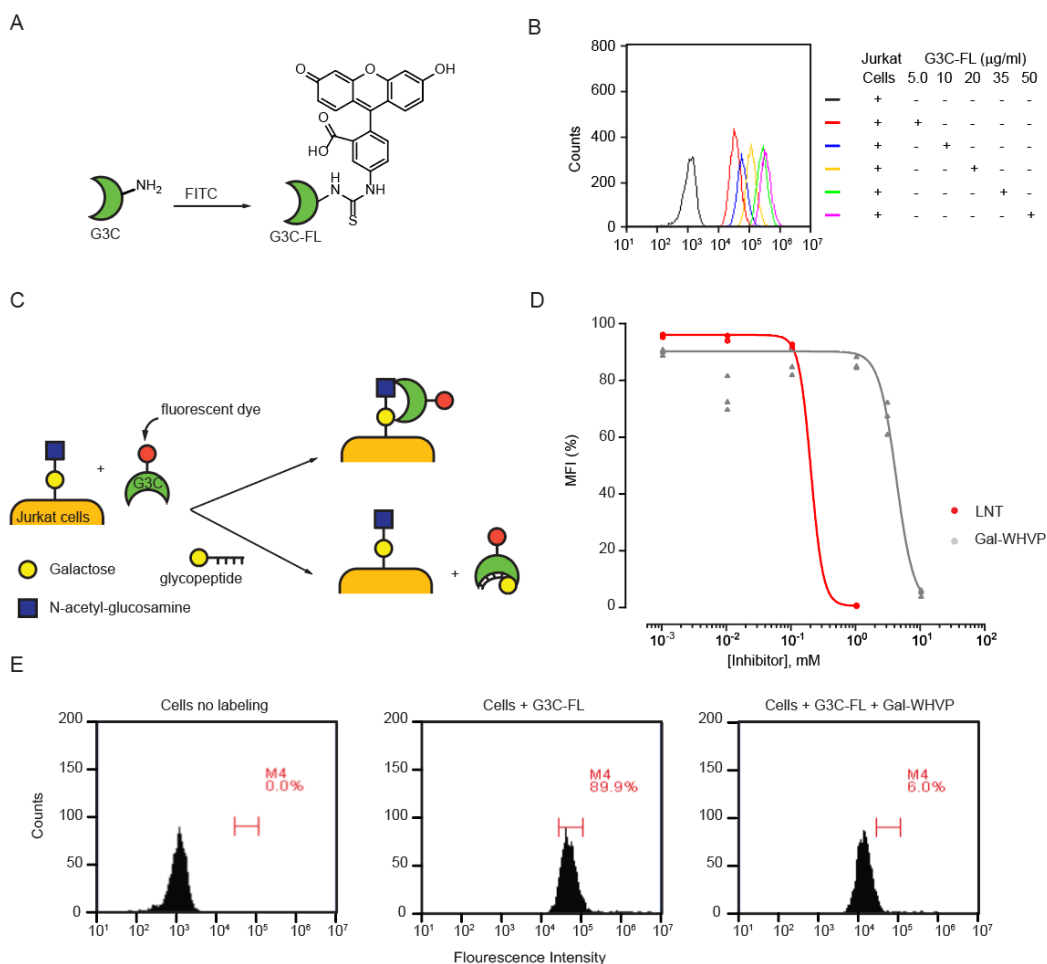
**Figure 2-9.** Fluorescence polarization inhibition assay for the analysis of the binding potency of selected leads in monovalent presentation. **(A)** Structure of the N-acetyl-Lactosamine (LacNAc) conjugated to fluorescein used as a probe to test each Gal-peptide synthetic construct. **(B)** The inhibitory potency of selected hits measured as  $IC_{50}$  against fluorescent LacNAc binding to G3C in solution.

We also measured inhibition by Rha and Xyl and observed no inhibition up to 100 mM glycan (Figure 2-10). It is thus not surprising that Rha and Xyl serve as poor guiding fragments for GE-FBD.



**Figure 2-10.** Inhibition of the binding of fluorescent LacNac to G3C by simple carbohydrates fragments. (A) Galactose showed concentration dependent inhibitory activity ( $IC_{50}$  7.8 mM) whereas Mannose showed weak  $IC_{50}$  >200 mM binding to G3C. (B) Fluorescence polarization measurements of wells containing 200 mM Galactose, Xylose, Rhamnose, and Mannose. None of the glycans but Galactose, displaced the binding of the probe LacNac to G3C.

A striking discrepancy between binding to surface-immobilized G3C (Figure 2-6), inhibition of binding of soluble Lac-HRP to surface-immobilized G3C (Figure 2-4), and inhibition of binding of Lac-FITC to G3C in solution (Figure 2-9), was unexpected. An observed millimolar  $IC_{50}$  of monovalent Gal-WHVP was further corroborated in flow cytometry study; Gal-WHVP inhibited binding of soluble fluorescently labeled G3C to Jurkat cells with  $IC_{50}$  of 4 mM (Figure 2-11).



**Figure 2-11.** Cell membrane binding of fluorescent Galectin-3(G3C-FL) on Jurkat cells. (A) G3C chemically conjugated to fluorescein to produce G3C-FL. (B) The fluorescence intensity measured at increasing concentrations of G3C-FL. (C) Cells incubated with G3C-FL, in the presence and the absence of Galectin-3 inhibitor were analyzed by flow cytometry. (D) Dose response curve of the effect of Gal-WHVP on the fluorescence intensity of the cell population. (E) Inhibition of the binding of G3C-FL to cell membrane in the presence of Gal-WHVP (10 mM).

We have concluded this study with a two-fold observation: (i) GE-FBD effectively discovers multivalent phage displayed Gal-peptides but is less effective in discovery of glycopeptides that contain fragments of higher than 100 mM affinity. (ii) When taken out of context in which the peptides were discovered (immobilized G3C), all tested peptides lose the synergy observed in assays that

use immobilized G3C. These findings corroborate previous observations that ligand binding properties of immobilized galectin-3 on the surface are different from ligand binding properties of the same galectin-3 in solution.

Still we believe that WHVP and WKPE could serve as promising starting points of design of multivalent inhibitors and further screens of focused libraries *e.g.* WHVP(X)<sub>n</sub>, where X<sub>n</sub> is a randomized peptide fragment.

### **2.3. Conclusions**

GE-FBD previously discovered synergistic combinations of strong (<1  $\mu$ M) modest (50-200  $\mu$ M) and weak fragment (8 mM). In this report we started from galactose fragment that has 10-100 millimolar affinity for G3C and discovered clones that showed increased affinity to G3C when decorated with galactose. The potency of each glycopeptide combination tested as soluble monovalent inhibitor was underwhelming. Although we were not able to identify glycopeptide ligands starting from fragments of weak affinity (> 100 mM), our study does not prove that Xyl-peptide and Rha-peptide hits do not exist in our screen. An exhaustive testing of all predicted sequences could uncover a rare hit; however, the very benefit of reproducible, and robust selection is to maximize success rate of discovery of active ligands in selected results and minimize the need for large number of ligands in exhaustive testing. These and previous observations map the potential fundamental physical limitations of GE-FBD and the ability of fragments of very weak affinity to steer selection landscape towards successful discovery of ligands in which the fragment and peptide sequences act

in synergy. It is tempting to propose that reproducibility and success rate of discovery in GE-FBD scales with affinity of the fragment.

We believe that machine learning approaches that analyze deep-sequencing results more thoroughly than a canonical differential enrichment analysis might be one of the effective ways to bypass the difficulties associated with low rate of success in ligand discovery. For analysis to provide better results, the library needs to contain both true positive and true negative internal controls, both active and inactive. Ligands identified in this screen can be added to future libraries and tracking of the ligands throughout the selection can yield valuable information for training of more advanced analysis models.

Observed steering of selection results towards ligands that fit to multivalent interactions with immobilized presentation of the target, reinforce the need for development of approaches that exploit soluble/monovalent targets rather than immobilized “pseudo-multivalent” targets. A fundamental challenge is catastrophic decrease in selection recovery when format of selection is “soluble” and “monovalent”. For this reason, soluble target and monovalent panning are rarely applied in selections that employ naïve library. Rather it is used in focused libraries that contain large copy number of potent ligands.<sup>111</sup> However, a radically different approach that combine selection of non-covalent interactions with reversible-covalent or irreversible covalent capture of the complex might serve as effective approaches to yield robust monovalent selection and discovery rate suitable for expedite discovery from naïve libraries.<sup>112-113</sup>

## **2.4. Experimental procedures**

### **2.4.1. Materials and general information**

All assays unless otherwise specified, were performed in MOPS buffer (50 mM MOPS, 150 mM NaCl, 2 mM CaCl<sub>2</sub> pH 7.4). Solutions used for phage work were sterilized by filtration through 0.22 µm filters. RP-HPLC purifications were performed on Waters HPLC system equipped with a 1525 EF binary pump, a FlexInject manual injector (dual mode) and a 2489 tunable UV detector. A SymmetryPrep<sup>TM</sup> C18 semi-preparative column (19 × 50 mm, particle size 5 µm, pore size 100 Å) was used for all purification at a typical flow rate of 12 mL/min. Dr. Pavel Kitov (University of Alberta) synthesized and characterized Lac-ABAO probe, as well as all hydroxylamine derivatives of galactose, glucose, xylose, rhamnose and mannose used in this study. Dr Zjay Tu (Academia Sinica, Taiwan) synthesized and characterized Lactosyl-fluorescein derivatives FPP-1. The silent barcode phage libraries were cloned by Katryna Tjhung following the methodology described below. HRP/Anti M13 monoclonal antibody and the mannose-binding protein concanavalin-A (ConA) were purchased from Sigma-Aldrich (#GE27-9471-01 and #C2010 respectively). Recombinant Cholera toxin subunit B(CTB) 103 aminoacids long, MW: 11645.3 Da was kindly donated by Dr. Bruce Turnbull (University of Leeds, UK). The carbohydrate recognition domain of the human Gal3 (G3C), residues 107–250 (MW: 16 327 Da) used in this study was provided by Dr. Christopher Cairo (University of Alberta, CA). The glycan LNT was ordered from Elictyl (#GLY010) Aminoxy-biotin was purchased from Cayman Chemical (#10009350). All plate-based measurements (absorbance A<sub>450</sub> nm and fluorescence polarization) were carried out in a

Cytation5 system (Biotek Instruments, Winooski VT, USA). Purification of aminoxy glycans and their precursors was accomplished with an automated chromatograph (CombiFlash® Rf, Teledyne ISCO, USA). HRMS-ESI spectra were recorded on Agilent 6220 TOF mass spectrometer using either positive or negative ionization mode.

#### **2.4.2. Phage Libraries construction**

The production of each silently encoded variant was described in our previous report<sup>22</sup>. The DNA libraries were purchased from Integrated DNA Technologies as anti-sense strand of the DNA oligonucleotide libraries. Oligos were extended by 15U Klenow fragment (#EP0051, Thermo Scientific). Double-stranded DNA libraries were purified by standard ethanol precipitation, resuspended in DNase-free water and treated with KpnI and EagI (1µL/mg of DNA) FastDigest restriction enzymes (#FD0524 and #FD0334, Thermo Scientific) for 20 min at 37 °C. Digested DNA was purified by 2% E-Gel® SizeSelect™ gel (Life Technologies). M13KE bacteriophage dsDNA was isolated from a single phage clone originating from the Ph.D.-12 phage display library (#E8111L New England BioLabs) using the GeneJET Plasmid Miniprep Kit (#K0502, Thermo Scientific) and similarly treated with KpnI and EagI FastDigest restriction enzymes. The resulting DNA was purified by 0.7% agarose gel purification followed by gel extraction by GeneJET Gel Extraction Kit (#K0691 Thermo Scientific). A 1:30 molar ratio of cut M13KE vector and library duplex was ligated by 400 U of T4 DNA ligase (M0202S New England BioLabs) at 16 °C overnight. The resulting DNA was purified by ethanol precipitation and re-



suspended in DNase-free H<sub>2</sub>O and transformed into commercially available F(+) TG1 electrocompetent cells (#605021 Lucigen, Middleton, WI, USA) using the Gene Pulser Xcell™ Electroporation System, and 2 mm GenePulser/MicroPulser Electroporation Cuvettes (all from BioRad). Transformed cells were allowed to recover in the provided recovery media for 15 min at 37 °C, 225 rpm before inoculation into 25 mL LB media and further amplification for 4.5 hours. Phage were collected by PEG/NaCl precipitation from supernatant of culture, quantified, and stored in MOPS buffer, pH 7.4.

### **2.4.3. Production of glycopeptide displayed libraries**

The production of each silently encoded variant was performed following the methodology described in our previous report.<sup>22</sup> Each library was conjugated with aminoxy glycan following published protocol.<sup>42</sup> A 10<sup>11</sup> PFU/mL solution of barcoded libraries was oxidized with 0.6 mM sodium periodate for 5 min on ice, in the dark and quenched with 5 mM glutathione (e.g. add 1 µL of 50 mM glutathione) for 10 min at room temperature (RT). A 5 µL aliquot of the oxidized phage was then modified with an equivalent volume of aminoxy-biotin (2 mM solution in 200 mM anilinium acetate) and incubated for 1 hour at RT. To the remaining oxidized phages, we added 95 µL of the corresponding aminoxy-monosaccharide (2 mM solution in 200 mM anilinium acetate) and incubated for 1 hour at rt. The conjugation efficiency of the hydroxylamine modifier was determined by biotin pulse-chase methodology previously described<sup>42</sup> using streptavidin-coated magnetic beads (Dynabeads, Life Technologies). Briefly, each phage solutions were diluted to 10<sup>5</sup> PFU/mL in blocking solution (0.1 % BSA in

MOPS buffer, pH 7.3). Following incubation with 5  $\mu$ L of washed magnetic beads for 15 min at rt with shaking. Beads were captured using a magnetic rack and the supernatants phage titer was then estimated. We purified the phage by dialysis against MOPS buffer at 4 °C (1 mM, pH 7.3, three buffer changes over 18 h, 10K MW cut-off). Mannose long linker (ManL), Galactose (Gal), Xylose (Xyl), Glucose (Glu), Mannose short linker (ManS) and Rhamnose (Rha) libraries were combined in a 1:1:1:1:1:1 ratio based on PFU counts, and  $10^9$  PFU of combo in 100  $\mu$ L of binding buffer (2% w/v BSA, 0.2% w/v Tween-20, 20 mM MOPS, pH 7.4) were used for panning on the following day.

#### **2.4.4. Panning of modified phage libraries**

We coated designated wells of a flat bottom 96-well Costar plate with 100  $\mu$ L of 10  $\mu$ g/mL G3C, ConA, and CTB in triplicate. The plate was sealed, incubated overnight at 4 °C, blocked for 1 h with a blocking solution (2% w/v BSA, in MOPS buffer, pH 7.4) and washed using a 405TM Touch Microplate Washer (BioTek) as follows: 300  $\mu$ L of wash buffer (0.1% w/v Tween-20, in MOPS, pH 7.4) followed by a 5s shake and 30 s soak, repeated for a total of 10 cycles. Glycopeptide library or unmodified peptide library of  $10^9$  PFU each was added to Gal-3 coated wells, incubated for 1 h at RT, washed using the BioTek washing protocol (see above). Adding 100  $\mu$ L of elution buffer (0.2 M glycine-HCl, pH 2.2, 0.1% w/v BSA) for 9 minutes and neutralizing with 1 M Tris-HCl, pH 9.1 allowed to detach phages from the selection wells. The eluted phage solution was quantified by plaque forming assay. Eluted phages were amplified for 4.5 h in 3 mL of LB supplemented with a 1:100 dilution of log phase *E. coli*

K12 ER2738 (New England BioLabs). We isolated single-stranded DNA (ssDNA) from the amplified phage using the QIAprep Spin M13 Kit (QIAGEN).

#### 2.4.5. Deep-sequencing analysis

To convert phage DNA to Illumina-compatible short double-stranded DNA (dsDNA) by PCR, we combined ~150 ng phage ssDNA with 1 mM dNTPs, 0.5  $\mu$ M forward (F) and reverse (R) primers, 0.5  $\mu$ L Phusion High Fidelity DNA polymerase in 1x PCR buffer (New England BioLabs) in a total volume of 50  $\mu$ L. The sequence of F and R primers is shown below, **XXXX** denotes four-nucleotide-long barcodes used to trace multiple samples in one Illumina sequencing experiment.

Forward (5'  $\rightarrow$  3') :

CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTCCGATCT**XXXX**CCCTTCTATTCTCACTC

Reverse (5'  $\rightarrow$  3') :

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**XXXX**ACAGTTTCGGCCGA

The temperature cycling protocol was as follows: 95  $^{\circ}$ C for 30 s, followed by 25 cycles of 95  $^{\circ}$ C for 10 s, 60.5  $^{\circ}$ C for 15 s and 72  $^{\circ}$ C for 30 s, final extension at 72  $^{\circ}$ C for 5 min and then hold at 4  $^{\circ}$ C until the samples are retrieved from the thermocycler. The resulting dsDNA with Illumina compatible adapters was sequenced using the Illumina HiSeq platform (The Donnelly Sequencing Centre at The Donnelley Centre for Cellular and Biomolecular Research, University of Toronto).

Raw FAQST data was processed using MATLAB scripts described in Tjhung *et al.*, 2015. Complete suite of MATLAB scripts is available as part of the supporting information. Volcano analysis was performed similarly to

procedure previously described<sup>20</sup> specifically, to identify significantly enriched hits, we compared selected clones against G3C by contrasting carbohydrate-modified versus non-modified populations. We also compared the enriched populations observed on G3C selection, against ConA and CTB control selections. To discard any preferentially amplified clones, the highly enriched clones from G3C selection were compared against a similar population of phages that were grown in bulk culture but without been initially selected against any target. Representative raw data form the volcano analysis of panning against G3C are available in Appendix A-2 and A-3.

#### 2.4.6. Production of Monoclonal phages

Phage clones SALRV, SSIYG, SIWVR, SLSMA, SWHVP and SWKPE were produced using Site-directed PCR mutagenesis. Primers sequences are described below:

Name	Sequence (5' -> 3')
SWHVP-Fwd.	CCTTTCTATTCTCACTCTTCTTGGCATGTGCCGGGTGGAGGTTTCGGCCGAAACTGTTGAA
SIWVR-Fwd.	CCTTTCTATTCTCACTCTTCTATTTGGGTGCGTGGTGGAGGTTTCGGCCGAAACTGTTGAA
SWKPE-Fwd.	CCTTTCTATTCTCACTCTTCTTGGAAACCGGAAGGTGGAGGTTTCGGCCGAAACTGTTGAA
SLSMA-Fwd.	CCTTTCTATTCTCACTCTTCTGTCTATGGCGGGTGGAGGTTTCGGCCGAAACTGTTGAA
SALRV-Fwd.	CCTTTCTATTCTCACTCTTCTGCGCTGCGTGTGGTGGAGGTTTCGGCCGAAACTGTTGAA
SSIYG-Fwd.	CCTTTCTATTCTCACTCTTCTTCTATTTATGGAGGTGGAGGTTTCGGCCGAAACTGTTGAA
SIWVR-Rev	TTCAACAGTTTCGGCCGAACCTCCACCACGCACCCAAATAGAAGAGTGAGAATAGAAAGG
SWKPE-Rev	TTCAACAGTTTCGGCCGAACCTCCACCTTCCGGTTTCCAAGAAGAGTGAGAATAGAAAGG
SALRV-Rev	TTCAACAGTTTCGGCCGAACCTCCACCAACACGCAGCGCAGAAGAGTGAGAATAGAAAGG
SSIYG-Rev	TTCAACAGTTTCGGCCGAACCTCCACCTCCATAAATAGAAGAAGAGTGAGAATAGAAAGG
SLSMA-Rev	TTCAACAGTTTCGGCCGAACCTCCACCCGCATAGACAGAGAAGAGTGAGAATAGAAAGG
SWHVP-Rev	TTCAACAGTTTCGGCCGAACCTCCACCCGCACATGCCAAGAAGAGTGAGAATAGAAAGG

The plasmid template was obtained M13KE-derived phage with the sequence 5'- CAG TTT ACG TAG CTG CAT CAG -3' which translates to QFTstopLHQ when propagated in 10G F' *E. coli* allowing that only those cells that incorporated the recombinant vector will grow. PCR products were digested with DpnI to remove template dsDNA and transformed into *E. coli* strain 10G F'.

The resulting plaques were clonally isolated, amplified and validated by Sanger sequencing to ensure accuracy.

#### **2.4.7. Binding assay of monoclonal phage to immobilized G3C**

A solution of G3C (10 µg/mL, 50 µL/well) in PBS buffer (11.9 mM phosphates, 137 mM sodium chloride, 2.7 mM potassium chloride, pH 7.4) was used to coat a polystyrene plate (Costar #3369). In parallel control wells were loaded with 50 µL of the monoclonal phage solutions. The plate was sealed and incubated at 4 °C overnight. In a separate non-binding surface plate (Corning #3641), a 3-fold serial dilution was performed for the solutions of glycopeptide-phage conjugates. The coating solution of G3C were then discarded and the wells were washed with the washing solution (3 x 200 µL, PBS buffer containing 0.1% (v/v) Tween-20). The mixture of the diluted phage solution (50 µL/well) was transferred to the plate coated with protein. The plate was incubated at RT for 1 h and washed with the washing solution (3 x 200 µL). A 50 µL solution of HRP-conjugated anti-M13 antibody (1:5000 dilution of stock) was added to each well and incubated at RT for 1 h. The solution was discarded, and the plate was washed with the washing solution (3 x 200 µL). To develop the color signal, 50 µL of TMB substrate (#34028, Thermo-Scientific) was added to each well. After 5 min incubation, 1 M phosphoric acid (50 µL) was added to quench the colorimetric assay. All the experiments were performed in triplicates. The color developed was read at 450 nm with a 96-wells plate reader.

#### **2.4.8. Fluorescence polarization assays to test the binding of glycopeptides to soluble G3C**

The fluorescein-lactosamine (Lac-FITC) probe was resuspended in PBS to obtain a concentration of 20  $\mu\text{M}$ . For direct binding the starting solution of G3C (80  $\mu\text{M}$ ) in PBS buffer was prepared as seriated 3-fold serial dilutions and loaded into a black 386-well plate (Perkin Elmer #110200365) to final volume of 19  $\mu\text{L}$ /well. Each well was supplemented with 1  $\mu\text{L}$  of the probe solution to get a final concentration of 1  $\mu\text{M}$  of fluorescent conjugate in 20  $\mu\text{L}$  total volume per well. The plate was incubated for 10 min in the dark at room temperature under slow shaking. Fluorescence polarization was measured at room temperature using a Cytation5 plate reader ( $\lambda_{\text{Ex}} = 485 \text{ nm}$ ,  $\lambda_{\text{Em}} = 528 \text{ nm}$ ). For the inhibition assay, 20  $\mu\text{L}$  of protein (2.9  $\mu\text{M}$ , PBS) and fluorescent probe (1  $\mu\text{M}$ ) at fixed concentrations were mixed with the inhibitor solution to make series of 3-fold dilutions. The plate was placed in the dark for 10 min and fluorescence polarization was measured on the Cytation5 as described above. All the data analysis and curve fitting were implemented using Origin software package (OriginLab, Massachusetts, USA).

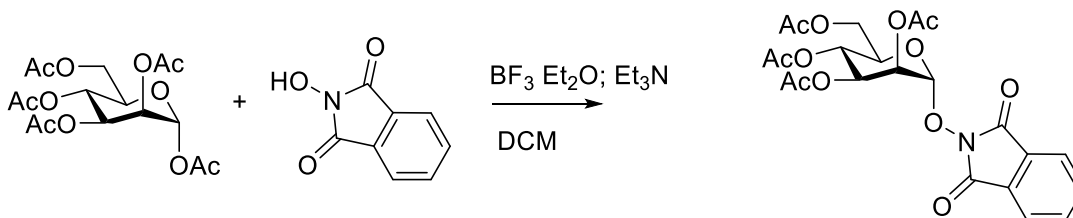
#### **2.4.9. Inhibition of G3C-FL binding to Jurkat cells**

Jurkat E6-1, (American Type Culture Collection), were maintained in complete RPMI (RPMI 1640, Invitrogen Life Technologies) supplemented with 2  $\text{mM}$  glutamine, 10% FBS, and 10  $\text{mM}$  HEPES) in 5%  $\text{CO}_2$  at 37  $^\circ\text{C}$ . Experimental procedure to test the binding of G3C-FL, to Jurkat cells was adapted from previously published protocol.<sup>114</sup> Initially  $5 \times 10^5$  cells were

incubated for 30 min at 4 °C with increasing concentrations of G3C-FL in PBS, to analyse the binding (Figure S8B). To minimize non-specific binding, we used G3C-FL at 5 µg/mL for inhibition-binding assays. Experiments were performed at 4°C to minimize cell death. For Gal-WHVP/LNT competition-binding studies, cells were incubated with 5 µg/mL of G3C-FL in the presence of increasing concentrations of inhibitor. Before fluorescence detection cells were washed two times with PBS. Flow cytometry data were acquired using a BD Biosciences Accuri C6 Flow Cytometer and analyzed with BD Accuri C6 software.

#### 2.4.10. Synthesis of anomeric hydroxylamine conjugates of monosaccharides

Synthesis of Mannose-hydroxylamine with 2-carbon linker has been described in our previous report.<sup>20</sup> We adjusted the procedure to generate five additional monosaccharide-hydroxylamine conjugates as described below



#### Synthesis of per-acetylated glycosides of *N*-hydroxyphthalimide

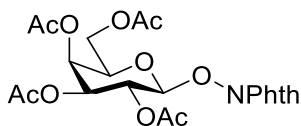
##### *N*-(2,3,4,6-Tetra-*O*-acetyl- $\alpha$ -D-mannopyranosyloxy)phthalimide

To a solution of 1,2,3,4,6-penta-*O*-acetyl-mannopyranose (0.38 g, 1 eq., 2.33 mmol), *N*-hydroxyphthalimide (456 mg, 1.2 eq., 2.8 mmol) and triethylamine (325 µL, 2.33 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (20 mL) BF<sub>3</sub>·Et<sub>2</sub>O (1.47 mL, 11.6 mmol) was added and the reaction was stirred at room temperature for 16 h. After

aqueous work-up, the residue was purified by silica-gel chromatography (AcOEt/Hexane, 50/50) to give compound 3 as white solid (0.95 g, 84%).

$^1\text{H}$  NMR (498 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.84 - 7.90 (m, 2 H, arom.) 7.76 - 7.84 (m, 2 H, arom) 5.72 (dd,  $J$  3.2 Hz,  $J$  1.7 Hz, 1 H, H-1) 5.36 - 5.48 (m, 3H, H-2,H-3,H-4) 4.89-4.97 (m, 1 H, H-5), 4.30 (dd, 1 H,  $J$  12.7 Hz,  $J$  4.6 Hz, H-6a), 4.22 (dd, 1 H,  $J$  12.7 Hz,  $J$  2.3 Hz, H-6b), 2.21 (s, 3 H, OAc), 2.11 (s, 3 H, OAc), 2.09 (s, 3 H, OAc), 2.04 (s, 3 H, OAc).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  = 170.59, 169.72, 169.44, 162.86, 134.79, 128.73, 123.80, 102.53, 70.87, 68.81, 67.24, 65.19, 61.87, 20.76, 20.73, 20.73, 20.61. HRMS-ESI calcd. for  $\text{C}_{22}\text{H}_{23}\text{NNaO}_{12}$  ( $\text{M}+\text{Na}$ ) $^+$ : 516.1112, found: 516.1116.

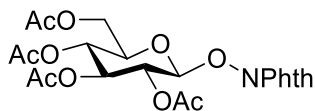
### ***N*-(2,3,4,6-Tetra-*O*-acetyl- $\beta$ -D-galactopyranosyloxy phthalimide**



$^1\text{H}$  NMR (498 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.82 - 7.88 (m, 2 H, arom.) 7.75 - 7.82 (m, 2 H, arom.) 5.64 (m, 2 H, H-1, H-4), 5.56 (dd, 1 H,  $J$  11.2 Hz,  $J$  3.3 Hz, H-2), 5.29 (dd, 1 H,  $J$  11.2 Hz,  $J$  3.9 Hz, H-3), 5.17 (t, 1 H,  $J$  6.5 Hz, H-5), 4.30 (dd, 1 H,  $J$  11.2 Hz,  $J$  6.6 Hz, H-6a), 4.03 (dd, 1 H,  $J$  11.3 Hz,  $J$  6.5 Hz, H-6b), 2.27 (s, 3 H, OAc), 2.17 (s, 3 H, OAc), 2.09 (s, 3 H, OAc), 2.05 (s, 3 H, OAc).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  = 170.71, 170.46, 170.07, 169.76, 163.00, 134.75, 128.80, 123.73, 102.04, 68.51, 67.82, 67.00, 66.70, 61.29, 20.76, 20.74, 20.65, 20.61. HRMS-ESI calcd. for  $\text{C}_{22}\text{H}_{23}\text{NNaO}_{12}$  ( $\text{M}+\text{Na}$ ) $^+$ : 516.1112, found: 516.1114.

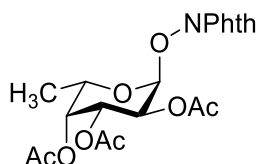


### ***N*-(2,3,4,6-Tetra-*O*-acetyl- $\beta$ -D-glucopyranosyloxy)phthalimide**



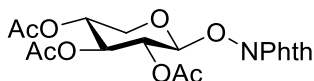
$^1\text{H}$  NMR (498 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.86 - 7.90 (m, 2 H, arom.), 7.77 - 7.84 (m, 2 H, arom.), 5.23 - 5.35 (m, 4 H) 5.11 - 5.15 (m, 1 H) 4.36 (dd,  $J$  12.3, 4.9 Hz, 1 H) 4.16 (dd,  $J$  12.4, 2.7 Hz, 2 H) 3.76 - 3.82 (m, 2 H) 2.21 (s, 3 H, OAc), 2.07 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 2.04 (s, 3 H, OAc).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  = 170.60, 170.13, 169.52, 169.29, 162.62, 134.80, 128.77, 123.86, 105.14, 72.47, 72.40, 69.65, 68.18, 61.81, 20.71, 20.69, 20.62, 20.57. HRMS-ESI calcd. for  $\text{C}_{22}\text{H}_{23}\text{NNaO}_{12}$  ( $\text{M}+\text{Na}$ ) $^+$ : 516.1112, found: 516.1112.

### ***N*-(2,3,4-Tri-*O*-acetyl- $\beta$ -L-rhamnopyranosyloxy)phthalimide**



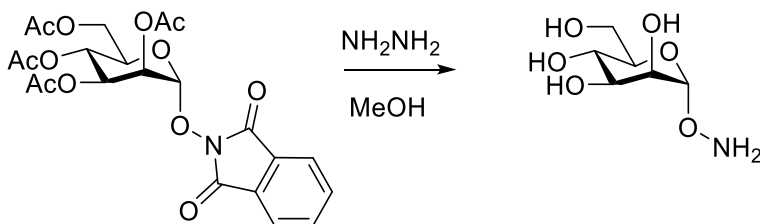
$^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.85-7.91 (m, 2 H, arom.), 7.77-7.83 (m, 2 H, arom), 5.72 (dd, 1 H,  $J$  3.3 Hz,  $J$  1.7 Hz, H-2), 5.41 (dd, 1 H,  $J$  9.9 Hz,  $J$  3.5 Hz, H-3), 5.37 (d, 1 H,  $J$  1.1 Hz, H-1), 5.18 (t, 1 H,  $J$  10.1 Hz, H-4), 4.81 (dd, 1 H,  $J$  10.1 Hz,  $J$  6.3 Hz, H-5), 2.21 (s, 3 H, OAc), 2.12 (s, 3 H, OAc), 2.04 (s, 3 H, OAc), 1.27 (d, 3 H,  $J$  6.3 Hz, H-6).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  = 170.00, 169.81, 169.52, 163.00, 134.68, 128.81, 123.76, 102.53, 70.19, 68.99, 68.83, 67.53, 20.82, 20.78, 20.65, 17.17. HRMS-ESI calcd. for  $\text{C}_{20}\text{H}_{21}\text{NNaO}_{10}$  ( $\text{M}+\text{Na}$ ) $^+$ : 458.1058, found: 458.1060.

## ***N*-(2,3,4-Tri-*O*-acetyl- $\beta$ -D-xylopyranosyloxy)phthalimide**



$^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.83-7.90 (m, 2 H, arom.), 7.75-7.82 (m, 2 H, arom.), 5.31 (d, 1H,  $J$  4.2 Hz, H-1), 5.26 (dd, 1 H,  $J$  6.9 Hz, H-2), 5.22 (t, 1 H,  $J$  5.9 Hz, H-3), 5.03 (m, 1 H, H-4), 4.68 (dd, 1 H,  $J$  13.0 Hz,  $J$  3.8 Hz, H-5a), 3.68 (dd, 1 H,  $J$  13.0,  $J$  4.2 Hz, H-5b), 2.19 (s, 3 H, OAc), 2.12 (s, 3 H, OAc), 2.11 (s, 3 H, OAc).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  = 169.85, 169.61, 169.24, 163.07, 134.68, 128.84, 123.74, 103.94, 68.67, 68.39, 67.19, 62.08, 20.85, 20.76, 20.70. HRMS-ESI calcd. for  $\text{C}_{19}\text{H}_{19}\text{NNaO}_{10}$  ( $\text{M}+\text{Na}$ ) $^+$ : 444.0901, found: 444.0903.

## **Deprotection of per-acetylated glycosides of *N*-hydroxyphthalimide**



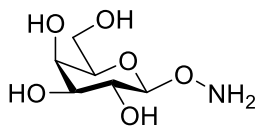
## ***O*-( $\alpha$ -D-mannopyranosyl)hydroxylamine**

To a solution of acetylated mannose phthalimide derivative (200 mg, 0.405 mmol) in methanol (4 mL), hydrazine hydrate (0.3 mL) was added. The mixture was stirred overnight at room temperature and copious precipitate has formed. The mixture was concentrated, taken up in hot methanol, silica gel (~2 g) was added and the mixture was concentrated. The residue was applied on silica

gel column as a dry pack and eluted with gradient of DCM-Methanol (up to 50%) to give the title compound as white solid (61 mg, 77%).

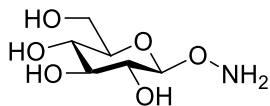
$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.89 (s, 1 H, H-1), 3.93 (s, 1 H, H-2), 3.87 (d, 1 H,  $J$  11.9 Hz, H-6a), 3.75 (dd, 1 H,  $J$  12.7 Hz,  $J$  3.3 Hz, H-6b), 3.59-3.68 (m, 3 H, H-3, H-4, H-5).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 103.19, 72.87, 70.59, 68.83, 66.60, 60.86. HRMS-ESI calcd. for  $\text{C}_6\text{H}_{14}\text{NO}_6$  ( $\text{M}+\text{H}$ ) $^+$ : 196.08156, found: 196.0816.

### ***O*-(- $\beta$ -D-galactopyranosyl)hydroxylamine**



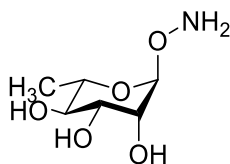
$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.50 (d, 1 H,  $J$  8.07 Hz, H-1), 3.91 (d, 1 H,  $J$  2.75 Hz, H-4), 3.67-3.82 (m, 3 H, H-5, H-6a, H-6b), 3.65 (dd, 1 H,  $J$  9.90,  $J$  3.3 Hz, H-3), 3.47 - 3.54 (m, 1 H, H-2).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 105.60, 75.15, 72.86, 69.39, 68.64, 61.09. HRMS-ESI calcd. for  $\text{C}_6\text{H}_{14}\text{NO}_6$  ( $\text{M}+\text{H}$ ) $^+$ : 196.08156, found: 196.0818.

### ***O*-(- $\beta$ -D-glucopyranosyl)hydroxylamine**



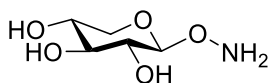
$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.55 (d, 1 H,  $J$  8.3 Hz, H-1), 3.92 (dd, 1 H,  $J$  12.3 Hz,  $J$  1.8 Hz, H-6a), 3.72 (dd, 1 H,  $J$  12.3 Hz,  $J$  6.1 Hz, H-6b), 3.43-3.52 (m, 2 H, H-3, H-5), 3.37 (t, 1 H,  $J$  9.5 Hz, H-4), 3.29 (t,  $J$  8.8 Hz, H-2).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 105.03, 75.88, 75.83, 71.70, 69.56, 60.76. HRMS-ESI calcd. for  $\text{C}_6\text{H}_{14}\text{NO}_6$  ( $\text{M}+\text{H}$ ) $^+$ : 196.08156, found: 196.0820.

### ***O*-(- $\beta$ -L-rhamnopyranosyl)hydroxylamine**



$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.84 (s, 1H, H-1), 3.94 (dd, 1 H,  $J$  3.12 Hz,  $J$  1.5 Hz, H-2), 3.67-3.75 (m, 1 H, H-5), 3.62 (dd,  $J$  9.7 Hz,  $J$  3.5 Hz, H-3), 3.42 (t, 1 H, H-4), 1.29 (d,  $J$  6.2 Hz, H-6).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 103.08, 72.05, 70.33, 68.94, 68.74, 16.73. HRMS-ESI calcd. for  $\text{C}_6\text{H}_{13}\text{NNaO}_5$  ( $\text{M}+\text{Na}$ ) $^+$ : 202.0686, found: 202.0684.

### ***O*-(- $\beta$ -D-xylopyranosyl)hydroxylamine**



$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.49 (d, 1 H,  $J$  8.0 Hz, H-1), 3.97 (dd, 1 H,  $J$  11.6 Hz,  $J$  5.3 Hz, H-5a), 3.60 (td, 1 H,  $J$  9.7 Hz,  $J$  5.5 Hz, H-4), 3.43 (t, 1 H,  $J$  9.2 Hz, H-3), 3.22-3.36 (m, 2 H, H-2, H-5b).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 105.79, 75.78, 71.55, 69.17, 65.08. HRMS-ESI calcd. for  $\text{C}_5\text{H}_{11}\text{NNaO}_5$  ( $\text{M}+\text{Na}$ ) $^+$ : 188.0529, found: 188.0527.

#### **2.4.11. Solid phase peptide synthesis**

Rink Amide AM resin (200 mg, 0.91 mmol/g, 0.18 mmol) was weighed into a Poly-Prep® chromatography column. The column was set up on a vacuum manifold. The manifold was equipped with a three-way stopcock that allows draining of the solvent by vacuum filtration and agitation of the resin by nitrogen bubbling.  $\text{CH}_2\text{Cl}_2$  (3 mL) was added to the dried resin for swelling. After 15 min,

the solvent was drained by vacuum aspiration. The resin was washed with DMF (3 mL) and the protective Fmoc group was cleaved with 20% (v/v) piperidine in DMF (3 mL) for 1 min. The treatment was repeated for 10 min using fresh 20% (v/v) piperidine in DMF (3 mL). The resin was washed with DMF (4 × 3 mL). Fmoc-protected amino acid (0.73 mmol, 4 eq.) in DMF (1 mL) and HBTU (276 mg, 0.73 mmol, 4 eq.) in DMF (1 mL) was added to the resin followed by N,N-diisopropylaminoethylamine (DIPEA, 0.25 mL, 1.46 mmol, 8 eq.). After 30 min of agitation with nitrogen, the reagents were removed by vacuum aspiration and the resin was washed with DMF (4 × 3 mL). The Fmoc-deprotection, amide coupling, and washing steps were repeated consecutively as described above to elongate the peptide sequence. After final Fmoc-deprotection, the resin was washed with DMF (5 × 3 mL), followed by CH<sub>2</sub>Cl<sub>2</sub> (5 × 3 mL). The resin was left on the manifold for 10 min to dry under the vacuum. A cleavage cocktail containing TFA/H<sub>2</sub>O/phenol/triisopropylsilane [3 mL, 85/5/5/5 (v/v/w/v)] was added to the resin. The column was left on a rocker for 2 h to cleave the peptide then the solution was collected, and the resin was rinsed with TFA (1 mL). The combined cleavage mixture was added dropwise to ice cold diethyl ether (20 mL) in a 50 mL centrifuge tube. The mixture was incubated on ice for 30 min then centrifuged for 5 min at 3000 rpm. Supernatant was discarded, and the precipitates were resuspended in cold diethyl ether (10 mL). The centrifugation and washing steps were repeated 2 times. The precipitates were air-dried and then left under vacuum overnight. Typical yield: 50–150 mg.

#### 2.4.12. Purification of peptides

Crude peptide (40 mg) was dissolved in DMF (0.25 mL) and 0.1% aqueous TFA (0.25 mL). The solution was injected into a semi-preparative RP-HPLC system. A gradient of solvent A (MQ water, 0.1% (v/v) TFA) and solvent B (MeCN, 0.1% (v/v) TFA) was run at a flow rate of 12 mL/min as shown below.

Time (min)	Eluent B (%)
0	5
2	5
26	35
27	100
29	100
30	5

The fractions containing target peptides were identified using mass spectrometry either by MALDI-TOF or ESI LCMS. MeCN was removed by evaporation under reduced pressure. The aqueous solution was lyophilized to yield the TFA-containing peptide. To remove TFA counterion the peptides were dissolved in 50 mM  $\text{NH}_4\text{CO}_3$  and freeze-dried to obtain the peptide as a white solid (20–32 mg).

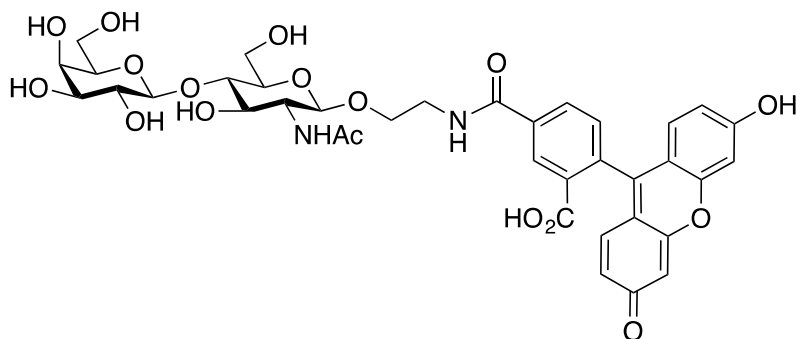
#### 2.4.13. Representative example of synthesis of Gal-peptide conjugates

The monosaccharide-peptide conjugates were synthesized from purified unprotected peptides following the methodology previously described<sup>20</sup>. Example of the procedure for  $\text{H}_2\text{N-SWHVP-CONH}_2$  peptide: the peptide (5.7 mg, 10  $\mu\text{mol}$ , 1 eq.) was dissolved in DMF (0.25 mL) followed by the addition of 200 mM MOPS (0.25 mL, pH 7.0). The solution was added to a 1.5-mL microcentrifuge tube containing solution of sodium periodate (2.6 mg, 12  $\mu\text{mol}$ , 1.2 eq.). The

reaction mixture was incubated for 10 min at RT. To quench the oxidation, we added glutathione (37 mg, 120  $\mu$ mol, 12 eq.) and mixed rapidly to ensure the dissolution of glutathione. After incubation for 10 min at RT, 2-(aminooxy) ethyl  $\beta$ -D-galactopyranoside (4.6 mg, 25  $\mu$ mol, 2.1 eq.) dissolved in 200 mM anilinium acetate (0.5 mL, pH 4.7) was added to the quenched solution. The oxime ligation was carried out for 30 min at RT. The reaction mixture was injected into a semi-preparative RP-HPLC system (section 2.4.10). Purified product was lyophilized to yield the final glycopeptide as a white fluffy powder (40–70% isolated yield). The purity of the product was determined with an analytical UPLC-MS system using a C18 column (Phenomenex Kinetex 1.7  $\mu$ m EVO C18, 2.1 $\times$ 50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min. Analytical data for galactoside conjugates can be found in Appendix A-1 to A-13

#### 2.4.14. Synthesis of FPP-1

##### 2-(fluorescein-5/6-yl-carbonylaminoethyl)]- $\beta$ -D-galactopyranosyl-(1-4)-2-acetamido-2-deoxy- $\beta$ -D-galactopyranoside



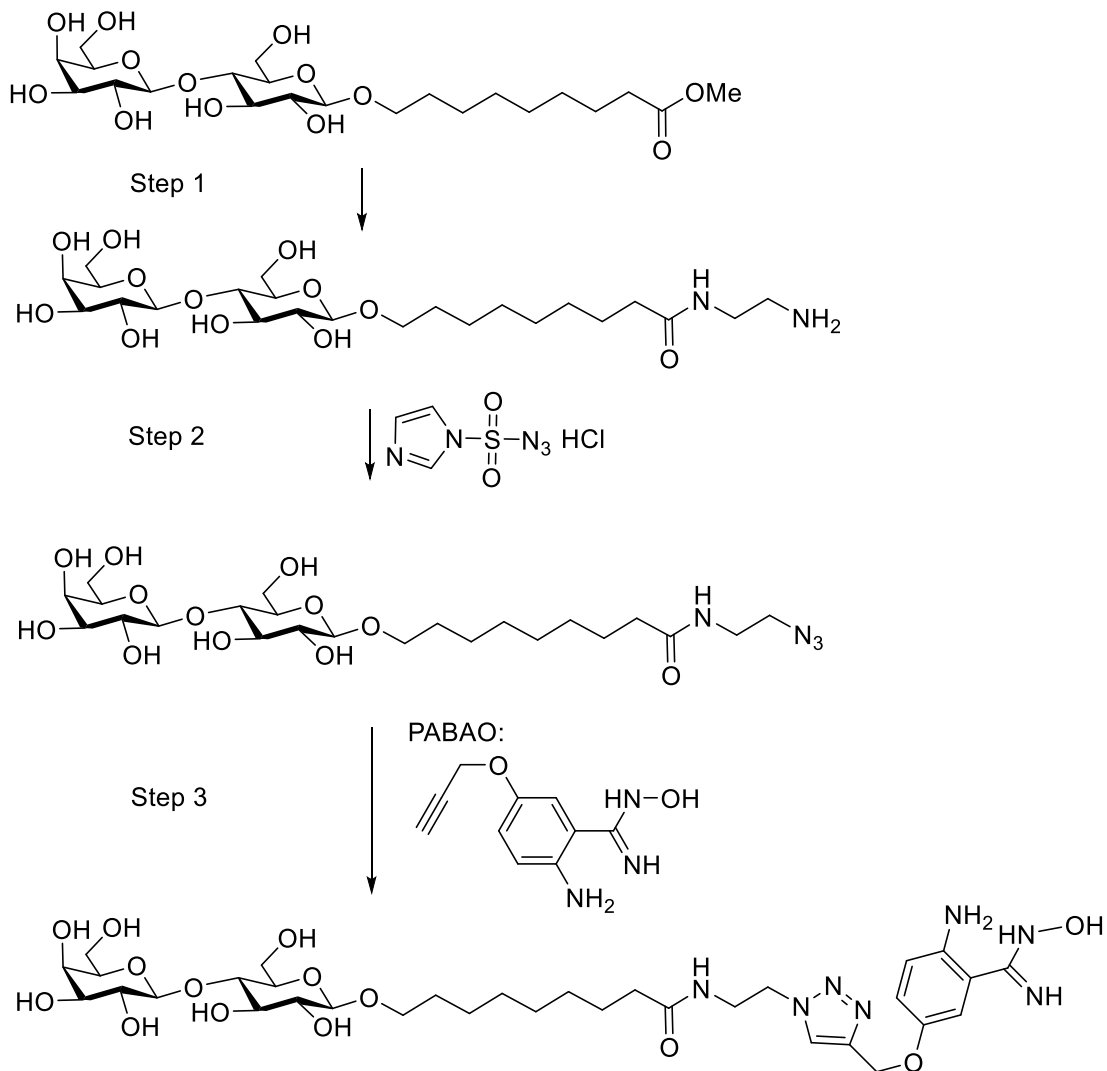
To a solution of aminoethyl N-acetyllactosamine (4.3 mg, 10  $\mu$ mol) in 0.1 M NaHCO<sub>3</sub>(aq) (1 mL) was added the 0.5 mL dry DMSO solution of 5/6

carboxyfluorescein NHS ester (9.6 mg, 20  $\mu\text{mol}$ ) at room temperature under nitrogen atmosphere. After stirring at room temperature for 1 h, the resulting residue was purified by column chromatography (EtOAc/MeOH/H<sub>2</sub>O, 6:3:1, v/v/v) followed by size-exclusion column (GE Sephadex<sup>TM</sup> LH-20, MeOH) to give the desired product (6.3 mg, 80%) as a red solid.  $R_f$  0.35 (EtOAc/MeOH/H<sub>2</sub>O, 6:3:1, v/v/v);  $[\alpha]^{25}_D$  -41.7 ( $c$  0.12, H<sub>2</sub>O); <sup>1</sup>H NMR of 5-isomer (500 MHz, D<sub>2</sub>O)  $\delta$  = 8.43 (s, 1H, ArH), 8.04-7.96 (m, 2H, ArH), 7.63 (s, 1H, ArH), 7.21-7.15 (m, 2H, ArH), 6.75-6.69 (m, 3H, ArH), 4.50 (d,  $J$  = 9.1 Hz, 1H, H1), 4.36 (d,  $J$  = 7.5 Hz, 1H, H1'), 4.08-3.49 (m, 16H, H2, H3, H4, H5, H6ab, H2', H3', H4', H5', H6ab', OCH<sub>2</sub>, CH<sub>2</sub>N), 1.76 (s, 3H, COCH<sub>3</sub>); <sup>1</sup>H NMR of 6-isomer (500 MHz, D<sub>2</sub>O)  $\delta$  = 8.22 (s, 1H, ArH), 8.04-7.96 (m, 1H, ArH), 7.41 (d,  $J$  = 7.9 Hz, 1H, ArH), 7.21-7.15 (m, 2H, ArH), 6.75-6.69 (m, 4H, ArH), 4.58 (d,  $J$  = 7.6 Hz, 1H, H1), 4.43 (d,  $J$  = 7.7 Hz, 1H, H1'), 4.08-3.49 (m, 16H, H2, H3, H4, H5, H6ab, H2', H3', H4', H5', H6ab', OCH<sub>2</sub>, CH<sub>2</sub>N), 1.81 (s, 3H, COCH<sub>3</sub>); HRMS (ESI-TOF):  $m/z$  calcd for C<sub>37</sub>H<sub>40</sub>N<sub>2</sub>O<sub>17</sub>:  $[M+H]^+$  785.2400; found: 785.2388.

#### 2.4.15. Synthesis of labelling reagent Lac-ABAO

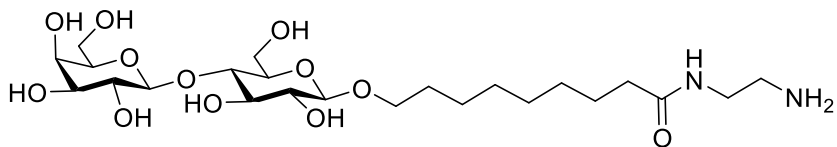
Synthesis of Lac-ABAO was performed in three steps as described below





### Step 1

#### 9-(2-aminoethylamino)-9-oxononyl 4-O-(β-D-galactopyranosyl)-β-D-glucopyranoside

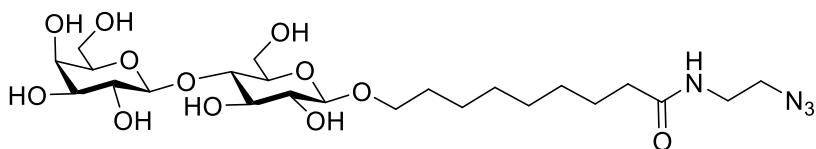


The 8-methoxycarbonyloctyl lactoside (700 mg, 1.36 mmol) was dissolved in neat ethylenediamine (3 mL) and stirred at 80 °C overnight. The mixture was concentrated, co-evaporated with water (3 times), dissolved in water

and freeze-dried.  $^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.55 (d, 1H,  $J$  7.9 Hz, H-1), 4.52 (d, 1 H,  $J$  7.9 Hz, H-1'), 4.05 -3.65 (m, 12 H, H-3,H-4, H-5, H-6a, H-6b, H-3', H-4', H-5', H-6'a, H-6'b,  $\text{CH}_2$ ), 3.62 (dd,  $J$  9.7 Hz,  $J$  8.07 Hz, H-2'), 3.30-3.40 (m, 3 H, H-2,  $\text{CH}_2$ ), 2.84 (t, 2 H,  $J$  6.2 Hz,  $\text{CH}_2$ ), 2.30-2.35 (m, 2 H,  $\text{CH}_2$ ), 1.68 (dt, 4 H,  $J$  14.7 Hz,  $J$  7.2 Hz,  $\text{CH}_2$ ), 1.33 - 1.47 (m, 8 H,  $\text{CH}_2$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 178.73, 103.92, 103.01, 85.21, 79.42, 76.34, 75.74, 75.46, 73.83, 73.52, 71.95, 71.71, 69.54, 62.00, 61.10, 42.00, 40.70, 36.78, 29.66, 29.21, 29.17, 29.02, 26.22, 25.90. HRMS-ESI calcd. for  $\text{C}_{23}\text{H}_{45}\text{N}_2\text{O}_{12}$  ( $\text{M}+\text{H}$ ) $^+$ : 541.2967, found: 541.2969.

## Step 2

### 9-(2-azidoethylamino)- 9-oxononyl 4-O-( $\beta$ -D-galactopyranosyl)- $\beta$ -D-glucopyranoside

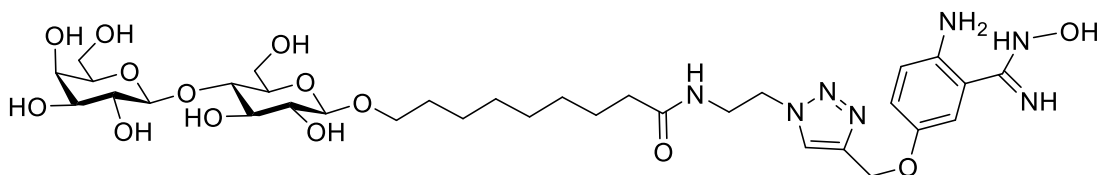


To the residue 1H-imidazole-1-sulfonyl azide (572 mg, 2 eq.) in water (3 mL) and  $\text{CuSO}_4$  (2 mg) was added, pH was adjusted to 8 with  $\text{NaHCO}_3$  solution. The mixture was stirred overnight then concentrated and chromatographed on silica gel (DCM-MeOH (up to 30%)) to give azido derivative (450 mg, 61%).  $^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 4.53 (m, 2 H, H-1, H-1'), 3.58-4.08 (m, 11 H, H-3, H-4, H-5, H-2', H-3', H-4', H-5',  $\text{CH}_2\text{CO}$ ,  $\text{CH}_2\text{NH}$ ), 3.48 (m, 4 H, H-6a, H-6b, H-6'a, H-6'b), 3.37 (m, 1 H, H-2), 2.33 (t,  $J$  7.2 Hz, 2 H,  $\text{CH}_2$ ), 1.59-1.74 (m, 4 H,  $\text{CH}_2$ ), 1.38 (br. s., 12 H,  $\text{CH}_2$ ).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  = 178.63, 103.92, 103.01, 79.44, 76.34, 75.74, 75.46, 73.86, 73.52, 71.96, 71.71, 69.55, 62.02,

61.13, 51.10, 39.61, 36.76, 29.69, 29.24, 29.20, 29.03, 26.20, 25.91. HRMS-ESI calcd. for C<sub>23</sub>H<sub>42</sub>N<sub>4</sub>NaO<sub>12</sub> (M+Na)<sup>+</sup>: 589.2691, found: 589.2692

### Step 3 (Lac-ABAO)

***N*-(2-(4-((4-amino-3-(*N*-hydroxycarbamimidoyl) phenoxy)methyl)-1*H*-1,2,3-triazol-1-yl)ethyl)-9-[4-*O*-(β-*D*-galacto-pyranosyl)-β-*D*-glucopyranosyloxy] nonanamide**



To a solution of Lac-azide (112 mg, 0.2 mmol) and 2-amino-*N*-hydroxy-5-(prop-2-ynyloxy)benzimidamide (PABAO)<sup>115</sup> (42 mg, 0.2 mmol) in 2 mL water CuSO<sub>4</sub> solution (100 uL, 10 mM) and sodium ascorbate (100 uL, 50 mM) were added. The reaction mixture was heated to 80 °C to dissolve starting material. Once dissolved the reaction was completed by TCL (DCM-MeOH=60:40). The product migrates lower than the starting material. The mixture was purified by HPLC on C-18 column in gradient water-MeCN (up to 30%) to give the title product (107 mg, 77%). <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O) δ = 8.07 (s, 1 H, triazole), 7.06 (s, 1 H, arom.), 6.89-7.05 (m, 2 H, arom.), 5.21 (s, 2 H, CH<sub>2</sub>), 4.61 (t, 2 H, *J* 5.41 Hz, CH<sub>2</sub>N), 4.50 (d, 1 H, *J* 7.8 Hz, H-1), 4.48 (d, 1 H, *J* 7.9 Hz, H-1'), 4.06-3.56 (m, 12 H, H-3, H-4, H-5, H-6a, H-6b, H-2', H-3', H-4', H-5', H-6'a, H-6'b, CH<sub>2</sub>), 3.30-3.39 (m, 1H, H-2), 2.15 (t, 2 H, *J* 7.2 Hz, CH<sub>2</sub>), 1.56-1.67 (m, 2 H, CH<sub>2</sub>) 1.40-1.50 (m, 2 H, CH<sub>2</sub>), 1.05-1.38 (m, 8 H, CH<sub>2</sub>).

$^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta = 178.31, 155.12, 151.15, 144.37, 140.29, 126.32, 119.93, 119.59, 116.69, 103.94, 103.04, 79.47, 76.34, 75.71, 75.46, 73.82, 73.53, 71.94, 71.66, 69.54, 63.22, 62.00, 61.12, 50.47, 39.73, 36.58, 29.72, 29.28, 28.95, 26.23, 26.00$ . HRMS-ESI calcd. for  $\text{C}_{33}\text{H}_{54}\text{N}_7\text{O}_{14}$  ( $\text{M}+\text{H}$ ) $^+$ : 772.3723, found: 772.3737

#### 2.4.16. Conjugation of horse radish peroxidase with Lac-ABAO

Horseshoe peroxidase (#77332, Sigma Aldrich) was resuspended (3.3 mg in 500  $\mu\text{L}$ ) in PBS buffer (11.9 mM phosphates, 137 mM sodium chloride, 2.7 mM potassium chloride, pH 7.4) and was mixed with a solution of  $\text{NaIO}_4$  (200  $\mu\text{L}$ , 10 mM in PBS). The mixture was incubated for 10 min and then the excess of  $\text{NaIO}_4$  was removed using 0.5 mL Amicon centrifuge filter (#UFC501096, Sigma-Aldrich) with PBS (3 times x 3 mL). The residue (~150  $\mu\text{L}$ ) was diluted to 300  $\mu\text{L}$  volume with sodium acetate buffer (100 mM, pH 4.5) and the labeling reagent Lac-ABAO (3 mg) was added. After 1 h, the excess label was removed on Amicon centrifuge filter and the protein solution in PBS was stored at  $-20^\circ\text{C}$  for further use. The protein concentration was estimated using the formula:

$$\text{Labeled protein} \left( \frac{\text{mg}}{\text{mL}} \right) = \frac{A_{280}(\text{G3C}_L) - [A_{495}(\text{G3C}_L) \times A_{280}(\text{free Label}) / A_{\text{max}}(\text{free Label})]}{\frac{A_{280}(\text{G3C}) \times \epsilon_{\text{ProtParam}}}{A_{280}(\text{G3C}_d)}}$$

- $\text{G3C}_L$ : labeled G3C
- $\text{G3C}_d$ : denatured G3C in 6M guanidine hydrochloride
- $\epsilon_{\text{ProtParam}}$ : extinction coefficient of G3C based on the primary sequence.

The absorbances  $A_{\text{wavelength}}$  were measured by Nanodrop™ spectrophotometer (Thermo-Scientific).

#### **2.4.17. Conjugation of G3C to Fluorescein isothiocyanate**

An aliquot of 50  $\mu\text{L}$  of FITC (#F7250, Sigma Aldrich) solution in DMSO (10 mg/mL) was added to 1 mL of G3C (10 mg/mL) in 100 mM sodium bicarbonate buffer pH 8.8. The reaction mix was gently stirred in the dark, at room temperature, for 90 min. Then the buffer was exchanged to PBS by centrifugation (5-times x 5 mL) using Amicon filter described on the previous section. The labelled-protein solution was protected from light and stored in PBS, at  $-20^{\circ}\text{C}$  for further use. The protein concentration was estimated by Nanodrop™ as described on the previous section.

## **Chapter 3: Selection of monovalent Galectin-3 ligands from phage-displayed glycopeptide libraries.**

### **3.1. Introduction**

Galectins are well known for their roles in human pathology being correlated with many degenerative diseases like cancer, cardiovascular disease and fibrosis.<sup>91, 116-117</sup> The development of potent galectin-specific inhibitors could enable novel therapeutic modalities focused on selective targeting of galectins.<sup>96, 118</sup> Specifically galectin-3 (Gal-3) inhibitors had proved to be beneficial in the treatment of chronic inflammation conditions such as pulmonary<sup>119</sup> and liver fibrosis.<sup>120-121</sup>

The development of glycopeptide inhibitors, could accelerate the development of combined therapies involving galectins inhibition with existing drugs, since the peptide portion could be further functionalized and coupled to a second pharmacophore against the same, or different target. Lin and co-workers,<sup>106</sup> demonstrated the potential advantages of combined therapies using *ex-vivo* studies. In cells from papillary thyroid cancer refractory to chemotherapy, they observed that Gal-3 inhibition induced concomitant apoptosis of tumor cells in combination with doxorubicin. In this study, the combined approach was synergistic and improved the chemosensitivity of cancer cells. In a similar way, Kouo and co-workers,<sup>107</sup> showed that the inhibition of Gal-3 combined with CD8+ cell-directed immunotherapies resulted in tumor-specific responses in patients with pancreatic ductal adenocarcinoma.

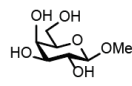
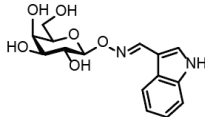
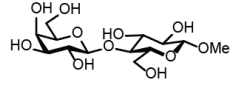
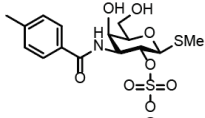
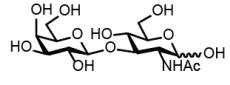
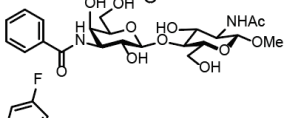
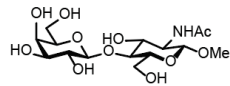
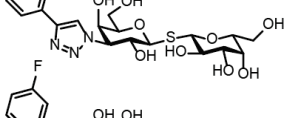
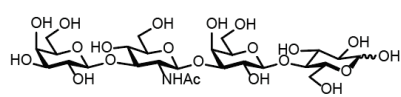
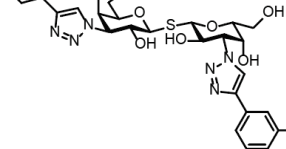
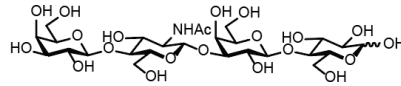
### 3.2. Previous work on the development of galectin ligands

Galectin proteins in general bind  $\beta$ -galactosides containing glycans with relatively weak affinity, in the high micromolar to millimolar range. Each galectin carbohydrate recognition domain (CRD) distinguishes different types of glycan ligands showing its highest affinity to different glycan structures.<sup>92,122-123</sup> Conversely, Gal-3 binds  $\beta$ -galactosides (NL1 to NL6, Table 3-1) with different affinities, showing tight associations with repeating N-acetyllactosamine (Gal $\beta$ 1-4GlcNAc) or Gal $\beta$ 1-3GlcNAc units.<sup>105,124-125</sup> Furthermore, the binding of Gal-3 to N-acetyllactosamine is enhanced by the presence of Galb1-3, GalNAca1-3 or Fuca1-2 substituents on the galactose residues.<sup>126-127</sup> Based on these observations, several research efforts had been directed towards the development of potent galectin binders based on  $\beta$ -galactosides derivatives. Synthetic 3,3'-diamido- or 3,3'-diazolyl-derivatization of thiodigalactoside (TDG) cores (SL4 and SL5, Table 3-1), exhibited high binding affinity towards galectins, with C2-symmetrical TDG derivatives among most potent Gal-3 inhibitors reported to date<sup>105, 128-130</sup> ( $K_D = 18$  nM). The TDG core laid the foundations for the development of selective glycan ligands towards galectin family members Gal-3 and galectin-1(Gal1). Specifically, C3 derivatives with either phenyltriazolyl<sup>104</sup> or thiazol-2-yltriazole<sup>131</sup> showed exclusive binding with single-digit nanomolar affinity to Gal3 or galectin-1 (Gal-1) respectively.

Crystallographic studies had shown that aromatic substituents of TAZTDG likely stack intermolecularly with adjacent arginine 144, providing  $\pi$ -cation interactions that accounts for its enhanced binding affinity when compared

with unsubstituted  $\beta$ -galactoside glycans. Further functionalization of C3' on TAZTDG could improve its binding affinity by engaging multiple arginine-arene interactions at the Gal-3 CRD. Co-crystallization of Gal-3 with the ligand TD131 discovered interactions between guanidino groups of Arg144 and Arg186 forming tandem arginine- $\pi$  associations that could account for the 13-fold higher binding potency of TD131 compared to TAZTDG with single cation- $\pi$  interaction.<sup>128</sup>

**Table 3- 1**  $K_D$  values of natural and synthetic ligands with galectin-3

natural ligands	$K_D$ ( $\mu$ M)	synthetic ligands	$K_D$ ( $\mu$ M)
	$4.4 \times 10^3$		$1.8 \times 10^2$
	$2.2 \times 10^2$		$8.7 \times 10^1$
	$1.2 \times 10^2$		$6.7 \times 10^0$
	$5.2 \times 10^1$		$8.4 \times 10^{-1}$
	$9.7 \times 10^{-1}$		$1.4 \times 10^{-2}$
	$6.5 \times 10^{-1}$		

In this chapter, we developed a method for using oxime linked TDG derivatives as fusion with bioactive peptides, that can be further extended to therapeutic proteins and antibodies. Fusion should not only retain binding activity but boost it while potentially contributing to the specificity of the combined ligand.

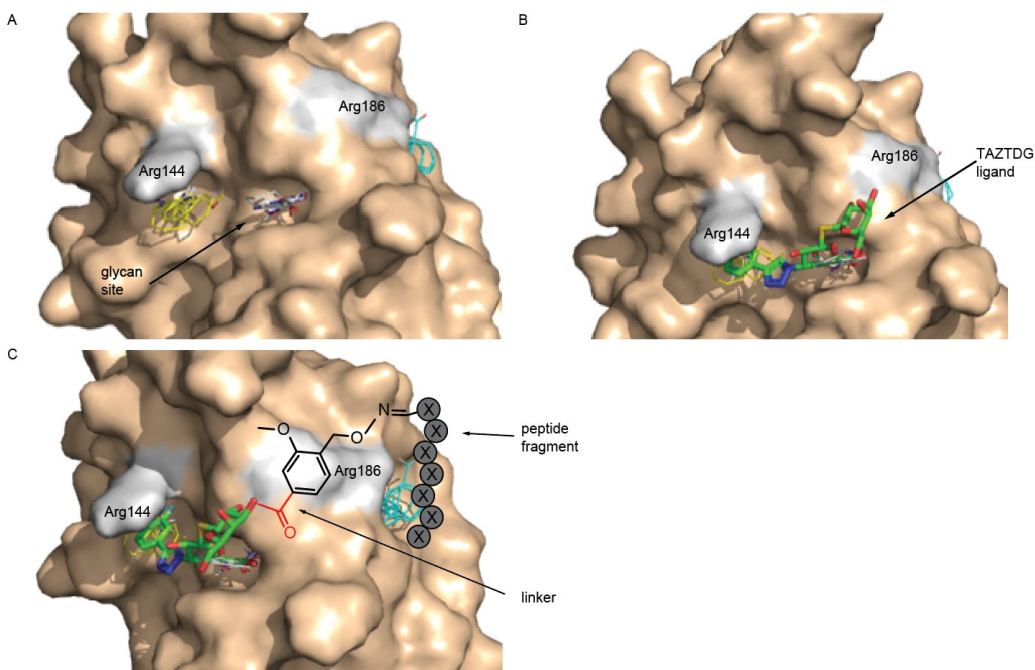


### **3.3. Results and Discussion**

#### **3.3.1. Design of TAZTDG-derived phage displayed libraries**

We capitalized on GE-FBD approach<sup>25-26,88</sup> to implement a screening system for the discovery of glycopeptide ligands that bind to Gal-3. Selection operates over a high-diversity repertoire of glycopeptide combinations displayed on M13 viral particles. Glycopeptide conjugates resulted from chemical linkage of aldehyde terminated heptamer peptides with 3-deoxy-3-(4-[*m*-fluorophenyl]-1H-1,2,3-triazol-1-yl)-thio-digalactoside (TAZTDG) glycan fragment;<sup>128</sup> *via* an  $\alpha$ -nucleophile handle installed on the glycan.

In this methodology TAZTDG fragment directs the interactions between glycan-peptide library and Gal3 CRD, to the vicinity of the binding site where the screening process can identify peptide sequences that bind near the glycan. These interaction could result in positive contributions to the overall binding affinity.



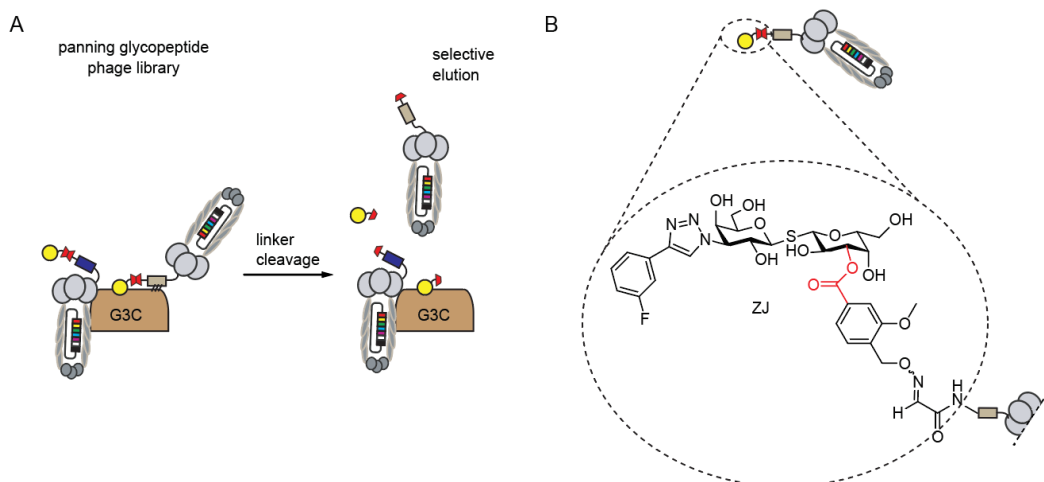
**Figure 3-1.** Study of the binding hot-spots near the glycan binding site of galectin-3 using FTMap. (A) Clusters of probes at “binding pockets” identified by FTMap. (B) Superimposed image of the TAZTDG-galectin-3 complex (PDB: 5h9r) with the probe clusters generated from FTMap. (C) Representation of C3’ derivatization on TAZTDG. The benzoyl substituent and oxime conjugated peptide cartoon were manually introduced to the structure.

Using FTMap, a computational solvent mapping server,<sup>132-134</sup> we identified putative binding “hot spots” on the surface of Gal-3 (Figure 3-1A). We hypothesized that most probable synergistic binding benefits will result from the interaction of the peptide fragment on the ligands with the binding site hot spots that are close to the glycan binding site (Figure 3-1B). To reach the hot spots, a peptidic “elongation” of the TAZTDG ligand, might be advantageous.

We built our GE-FBD strategy starting from a custom synthesized TAZTDG glycan (referred as ZJ hereafter), functionalized with a benzoyl radical at C3’ position. ZJ could benefit the overall binding from a second arginine- $\pi$  interactions at residue Arg186 as described in section 3.2 (Figure 3-1C). The

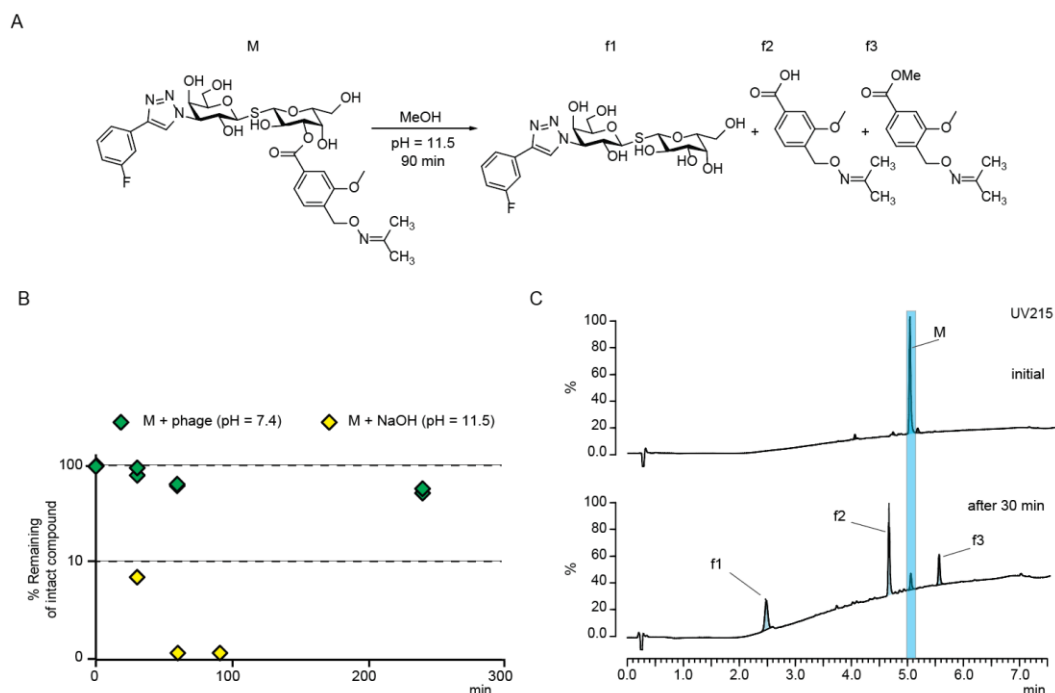
modification introduces a terminal amino-oxy handle to allow biorthogonal conjugation to serine terminated phage displayed peptide libraries.<sup>42</sup> The peptide portion effectively extend the contact surface of the glycan ligand with the protein. This extension also allows to screen for peptide-fragment combinations with synergistic contributions to Gal3 binding, so it that could result in an overall enhancement of affinity.

Successful screenings for of glycopeptide library against GBPs, depends on the feasibility of extracting only the true target-binders among the high diversity phage displayed library of conjugates. During selection, random peptide libraries can have multiple interactions with non-target components like polymer beads<sup>135</sup>, streptavidin<sup>136</sup>, biotin<sup>137</sup> and blocking agents.<sup>138</sup> Traditional elution procedures in high salt or acidic conditions will likely result in low signal-to-noise ratios due to the presence of many “ligands” that do not have a real affinity to the target. To minimize the recovery of non-specific binders we introduced an ester group between the glycan and the peptide fragment. Ester bond cleavage, allowed selective separation of phage from carbohydrate fragment by alkaline hydrolysis of the 3'-O-Aryloyl linker during the elution step (Figure 3-2A,B).



**Figure 3-2.** Design of a TAZTDG cleavable linker. Alkaline hydrolysis of the TAZTDG-peptide phage conjugate conveniently separates the glycan “warhead” in every glycopeptide in the phage library. (D) Representation of selective elution of true binders, from selection experiments via linker cleavage

To verify that ester hydrolysis did not compromise the availability of ZJ-peptide phage conjugates during selection screenings, we incubated ZJ glycan with phage solution at  $10^{11}$  PFU/mL concentration of viral particles, and tested whether the ester hydrolysis occurs at significant extent under the conditions of phage binding buffer. For these studies we capped the reactive amino-oxy handle of ZJ glycan with acetone to minimize side reaction byproducts and used LCMS to monitor the disappearance of the capped compound (Figure3-3A).



**Figure 3-3.** Stability study of ester group on ZJ glycan. (A) Alkaline hydrolysis of C3' O-aryloyl linker. (B) Time dependent degradation of esters in the presence of phage solution, or basic conditions. (C) HPLC traces following the alkaline hydrolysis of the ester sampled at the beginning of the reaction and after 30 min. The M, f1, f2, and f3 labels identified the retention times for the hydrolysis products of M, f3 correspond to the transesterification product with methanol.

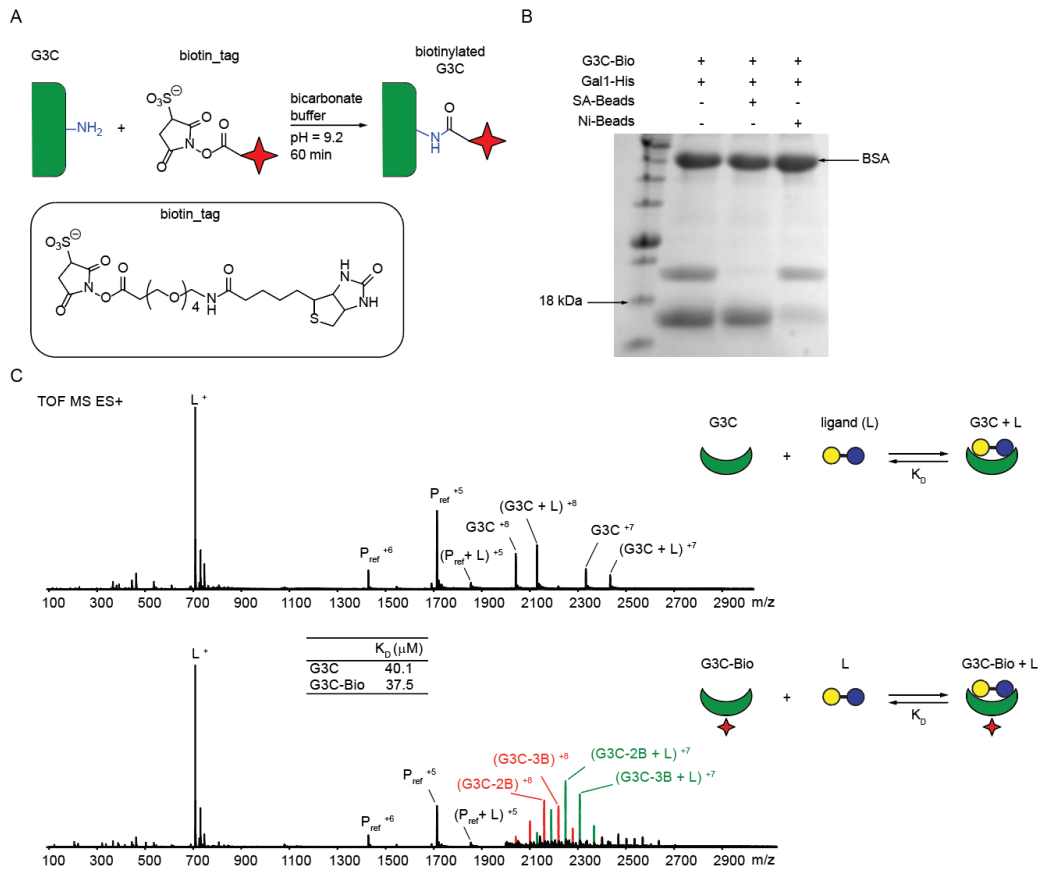
After 5 hours incubation at pH 7.4, 70~80 % of ZJ glycan remained intact in the phage mixture (Figure 3-3B). Under alkaline conditions, ester hydrolysis occurred faster, and after 30 min about 80 % of the initial ZJ concentration was mostly converted into hydrolysis products. We also observed appreciable transesterification after 30 min reaction due to the presence of methanol as a solvent in the reaction mixture (Figure 3-3C).

### 3.3.2. Solution-phase panning with affinity bead capture

We intended to implement a panning protocol in which the phage library was reacted with the target in solution, followed by affinity capture of the target-phage complexes onto an affinity matrix (bead) specific for the target protein. In

solution panning can result in improved accessibility of the putative ligand binding site to phage-displayed peptides, as well as avoiding partial denaturation of the target on a plastic surface.<sup>139-140</sup> The affinity pull down of target-phage complexes, requires the target to be labeled with an appropriate affinity tag. We used NHS-PEG4-Biotin (15-biotinlamino-4,7,10,13-dioxanonanoic acid N-hydroxysuccinimidyl ester) amine reactive probe, as labeling reagent (Figure3-4A), to install biotin tags to the C-terminal fragment (residues 107–250) of human galectin-3 (G3C). SDS-PAGE electrophoresis validated, in solution pull down of biotinylated G3C (G3C-Bio) from a protein mixture containing unlabeled bovine serum albumin (BSA) and polyhistidine tagged galectin-1(Gal1-His). Furthermore, using either chelated-nickel or streptavidin coated beads we could selectively pull down Gal1-His or chemically labeled G3C-Bio respectively (Figure3-4B).

Although covalent labeling of the target with biotin is a widely used methodology on phage display selections,<sup>141-143</sup> it typically results in a heterogeneous mixture of proteins, which differ in the number and location of the labels.<sup>144</sup> Covalent labeling could also have deleterious effects on the ligand binding affinity of the target. Therefore, prior to embark in the screening campaigns we performed a careful characterization of the labeling and functionality of G3C-Bio.



**Figure 3-4.** Chemical ligation of biotin tag and mass-spectrometry characterization of the protein G3C. (A) Conjugation of biotin to primary amine residues in G3C. (B) SDS-PAGE of target-specific pull down of labeled protein using affinity beads. (C) Mass spectrometry validation of the ligand binding affinity of galectin-3. Labeled G3C compared to the unlabeled protein showed similar affinity towards a positive control ligand LNT. Charged states +8 and +7 on G3C-Bio species are highlighted in red and green respectively. The signals corresponding to 2 (G3C-2B) and 3 (G3C-3B) biotin residues per protein are indicated in brackets.

Optimized reaction conditions for the protein labeling chemistry permitted to minimize heterogeneity resulting from the biotin ligation. Using 2.5 equivalents of labeling reagent yielded a homogeneous mixture of labeled G3C with most abundant species bearing 2 or 3 biotin labels per protein molecule as confirmed by ESI-MS measurements. The signals corresponding to 2 (G3C-2B) and 3(G3C-3B) biotin residues per protein were the most abundant ions species

on charged states +7 and +8 (Figure3-4C). We then analyzed the glycan binding affinity of G3C-Bio compared to unmodified G3C. For binding measurements we used, previously reported ESI-MS binding assay.<sup>145</sup> Briefly, the ligand affinity was calculated from the ratio of the abundances of the ion species corresponding to ligand-bound (G3C+L) and free protein (G3C). The measurements were calibrated using a control protein as a reference (Pref) and the glycan LNT as control ligand (L). The experiments confirmed that the binding affinity of G3C towards LNT was not affected by covalent labeling of the proteins.

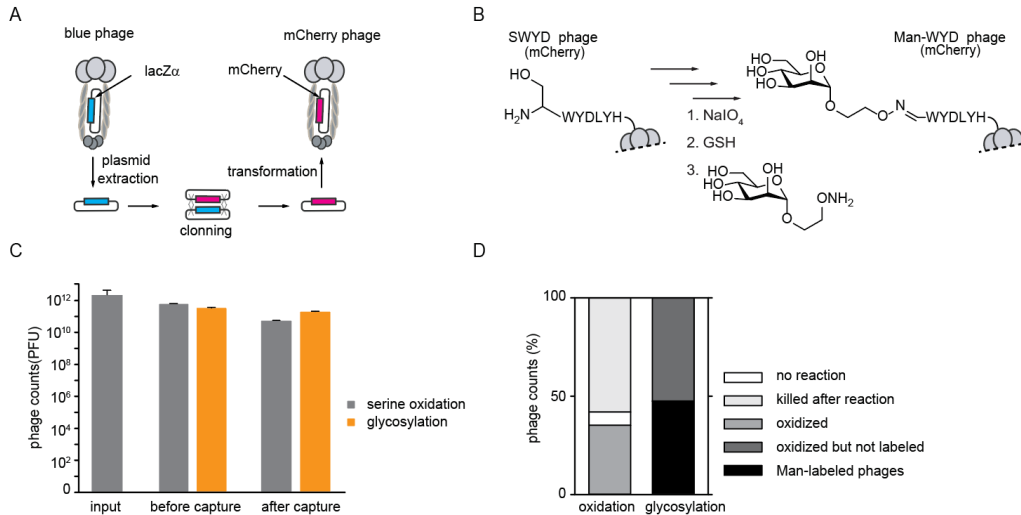
### **3.3.3. Probing reproducibility of discovery with color-coded phage clones**

Previous phage display screenings of glycopeptide libraries on the mannose binding lectin concanavalin A (ConA) described the discovery of the glycopeptide Man-WYDLHF as a potent ligand of ConA.<sup>20,22</sup> In Chapter 2 we showed that a phage displayed clone bearing the reported mannose-peptide conjugate binds selectively, with high affinity to ConA, our data was consistent with another reports of a similar phage displayed construct.<sup>25</sup> Based on these observations, a high recovery of the sequence Man-WYDLHF from ConA coated beads in pull-down phage display screenings can be used as a validation of the selection process.

To test our hypothesis, we genetically engineered a phage that transduces mCherry fluorescent reporter<sup>146</sup> (Figure3-5A). The red fluorescent color allowed to distinguish monoclonal phage from the canonical phage library that constitutively express lacZ $\alpha$  reporter<sup>147</sup> producing blue-colored plaques on agar plates supplemented with X-Gal. Into that red reporter we then cloned peptide



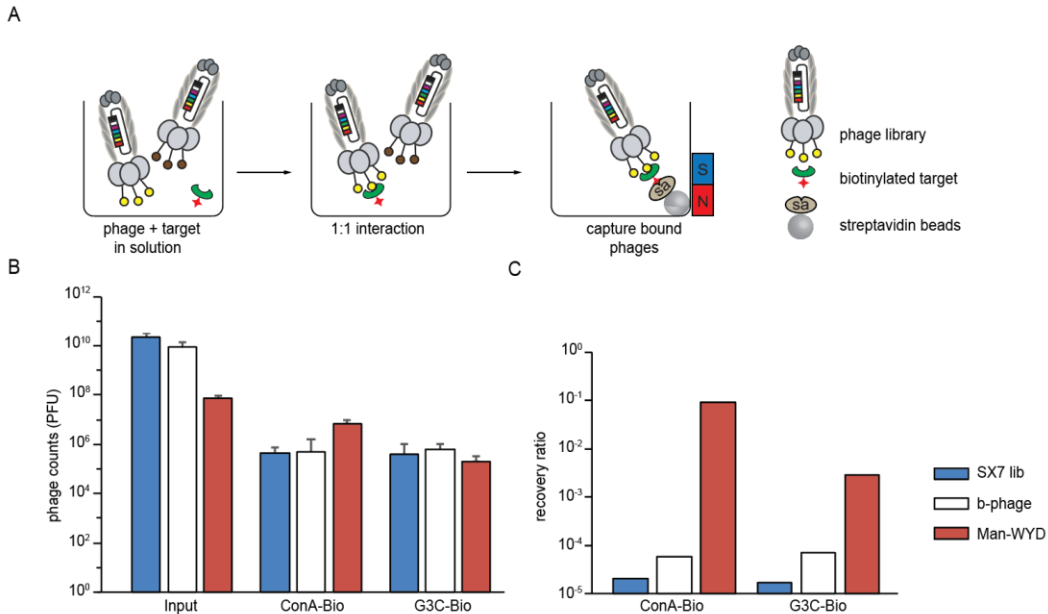
sequence SWYDLHF at the N-terminus site the phage coat protein pIII. To finally convert this construct into a positive ConA-selection control, we chemically decorated the SWYDLHF sequence with mannose glycan.



**Figure 3-5.** Development of color-coded phage clones introduced as screening positive-controls (A) Cloning a phage clone that transduces mCherry reporter. (B) Chemical modification on phage clone to install mannose glycan. (C) Phage counts in total plaque forming units (PFU) prior to chemical modification (input), and from supernatants before and after biotin pulse-chase capture assays. (D) Quantification of the % of yield after each chemical modification step.

We installed the amino-oxy mannose derivative into the displayed peptide, starting from oxidation of the N-terminal serine residue of the peptide, to obtain glyoxal-displaying phage (Figure3-5B). We characterized each modification step by separated biotin pulse-chase capture assays (see Chapter 2, section 2.2.1). In short, we biotinylated the aldehyde-terminated phages using aminoxy-biotin probe, before and after glycan ligation. Then, we incubated both samples with streptavidin-coated magnetic beads and titered the supernatants before and after capture with beads. The difference between the fractions of biotinylated clones in both populations indicated the efficiency of modification. Unexpectedly, we

observed that SWYDLHF clone was sensitive to serine oxidation step (Figure3-5C). Our experiments showed 60% drop in infectivity after oxidation step, compared to the input population, whereas after oxidation we successfully glycosylated ~50% of the phage-aldehydes (Figure3-5D).



**Figure 3-6.** Validation of the reproducibility of discovery in pull-down screenings using selection control phage clone. (A) Scheme of in solution panning with affinity bead capture. (B) Phage counts from selection experiments. The library composition can be traced by color. (C) Recovery ratios observed from beads coated with biotinylated ConA and G3C. The phage displayed Man-WYDLHF (Man-WYD) used as a selection control was preferentially recovered from ConA screening experiments.

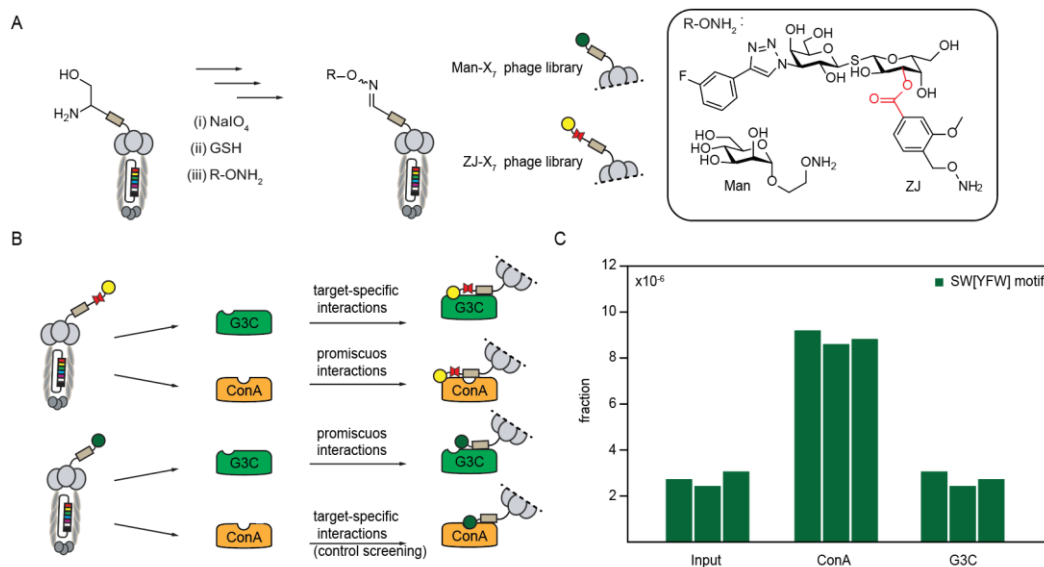
In model panning experiments with biotinylated GBPs (ConA-Bio and G3C-Bio) proteins were allowed to react with phage library displaying random 7mer peptide fragment (SX7), in solution. Next, using streptavidin-coated magnetic beads we pulled down the phage-ConA and phage-G3C conjugates (Figure3-6A). We spiked the phage library with 10<sup>8</sup> PFUs of the color tracer Man-WYDLHF (Man-WYD) and quantified the phage populations by plaque forming assay on agar plates. To reduce the target unrelated binding interactions of the

phage library to all the components of the screening system we introduced a mutant phage as a blocking agent in 1:1 ration to the phage library. We designed the blocking phage (b-phage) with a mutated display region that express no peptide on pIII. The b-phage is unable to be PCR amplified under standard phage library conditions, making it “invisible” during DNA sequencing analysis. The b-phage lacks the lacZ $\alpha$  reporter yielding colorless (white) plaques when propagated on agar plated host cells. Expanded color palette of phages: red, blue and white, allowed for fast and easy quantification of the output from selection experiments on G3C and ConA coated beads, using plaque forming assay on agar plates (Figure 3-6B). In concordance to previous studies on ConA, we observed recovery of Man-WYD, with 30-fold superiority in ConA screenings compared to G3C. Random peptide library SX7 or irrelevant b-phage did not show preferential enrichment on any of the protein targets (Figure 3-6C). These observations demonstrated the functionality of the screening system in our experimental conditions.

#### **3.3.4. Solution-phase screenings of glycopeptide libraries**

For selection screenings, we synthesized phage displayed libraries with a diversity of  $\sim 10^8$  random heptamer peptides conjugated to mannose (Man-X<sub>7</sub>) or ZJ (ZJ-X<sub>7</sub>) glycans. First, we oxidized serine terminated peptides libraries (SX<sub>7</sub>), with sodium periodate, to achieve the corresponding glyoxal at N-terminus. The glyoxal handle was then functionalized with hydroxylamine derivatives of either mannose or ZJ.

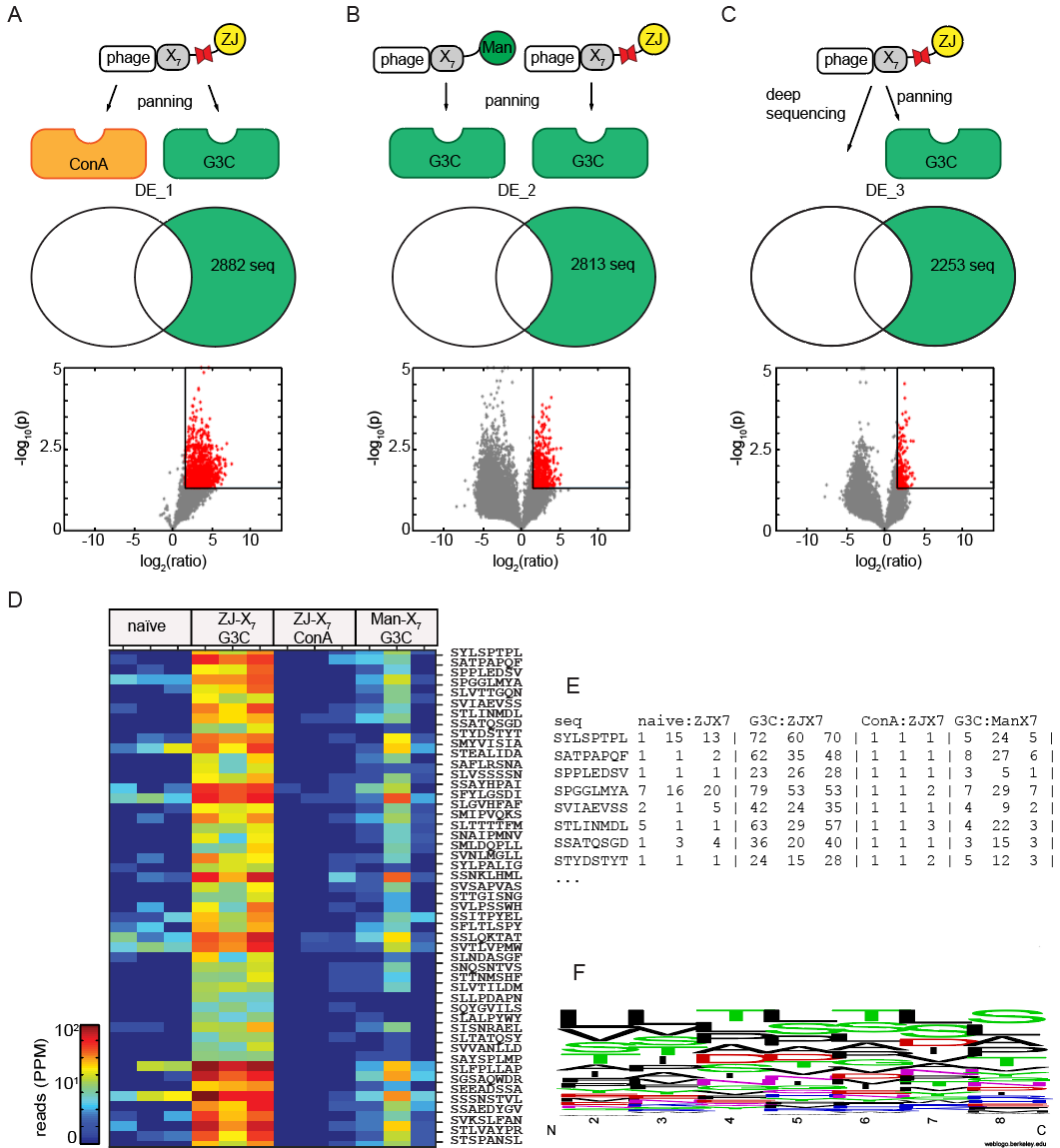
Using biotin pulse-chase technique, we evaluated that 40-60% of the phage library was successfully glycosylated (Figure3-7A). We then reacted the modified phage libraries with biotinylated targets in solution in a two by two combinatorial set up (Figure3-7B). Control screenings of ZJ and Man modified libraries against ConA and G3C respectively, permitted to discard unproductive interactions from selection experiments. We further validated the selection system by testing whether Man- $X_7$  panning on ConA could “re-discover” the Man-WYD motif. After in solution incubation of the phage libraries with targets we then captured target-phage complexes by adding streptavidin-coated beads to pull down biotinylated targets. To reduce noise in the recovery of glycopeptide that bind to G3C, we washed the beads thoroughly and detached phages from ZJ- $X_7$ , by alkaline elution via ester hydrolysis (see section 3.3.1).



**Figure 3-7.** Screenings of chemically-modified phage libraries against galectin-3. (A) Chemical modification of phage to generate Man- $X_7$  and ZJ- $X_7$  libraries. (B) Panning of glycopeptide libraries on G3C with parallel control screening on ConA. (C) Fraction of unique sequences that contain SW[YFW] motif in each selection screen.

Analysis of sequencing data from DNA of eluted phages, confirmed that SW[YFW] motif was preferentially enriched among the sequences recovered when Man-X7 library was selected against ConA. The same library against G3C target, showed that the fraction of unique sequences containing SW[YFW] motif was similar to that of the naïve library before selection (Figure 3-7C).

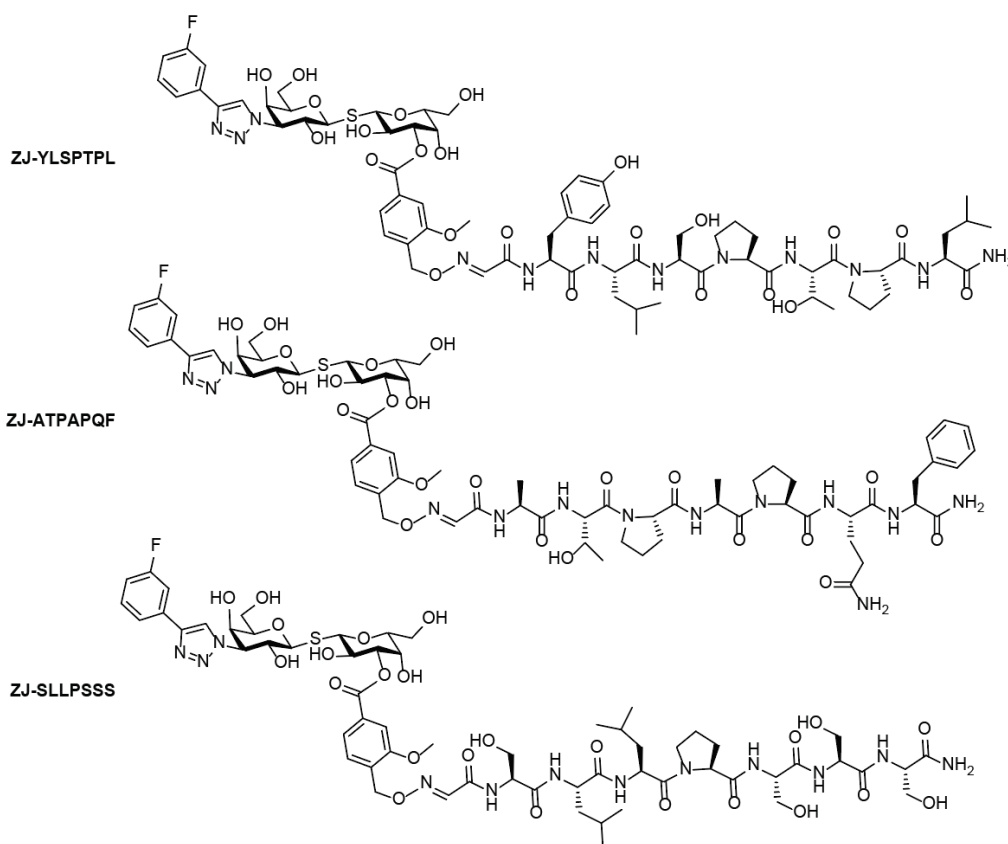
To identify enriched sequences from G3C selection experiments we used published MatLab scripts<sup>20, 22, 26, 109, 148</sup> that perform differential enrichment (DE) analysis. DE analysis implements *t*-test hypothesis testing to identify sequences with higher copy number on selection (G3C) experiments compared to control with a significance of  $p < 0.05$ . DE analysis identified 2882 (DE\_1) and 2253 (DE\_3) sequences, that were more than 3-fold enriched from panning on G3C when ConA or naïve ZJ-X<sub>7</sub> were used as control. Likewise, we found 2813 (DE\_2) sequences enriched on G3C only if decorated with ZJ glycan and not with mannose (Figure 3-8 A to C). Further filtering of DE sequences:  $DE_1 \cap DE_2 \cap DE_3$  discovered a total 646 hit sequences. We selected as hits sequences that showed consistently more than 3-fold enrichment across all selection controls (Figure 3-8D, E). We then searched for consensus patterns on the sequences of the top-50 hits, using WebLogo application.<sup>149</sup> We discovered a highly conserved serine (S) after the chemical conjugation site (N-terminal serine) indicating that this fragment might be important for binding to G3C (Figure 3-8F). Encouraged by these results, we selected the motif SLLPSSS identified from LOGO analysis and the sequences YLSPTPL and ATPAPQF for further validations.



**Figure 3-8.** Selection of phage libraries (A-C) Libraries of SX7 peptides displayed on M13 phage decorated with ZJ or Mannose through oxime ligation incubated with either G3C or ConA. (D) Top 50 from 646 total hits ( $p < 0.05$ ,  $R > 3$ ). (E) Example of the raw number of reads obtained from Illumina sequencing for some of differentially enriched sequences. (F) Hits identified from the intersection set led to the discovery of consensus sequences from LOGO analysis.

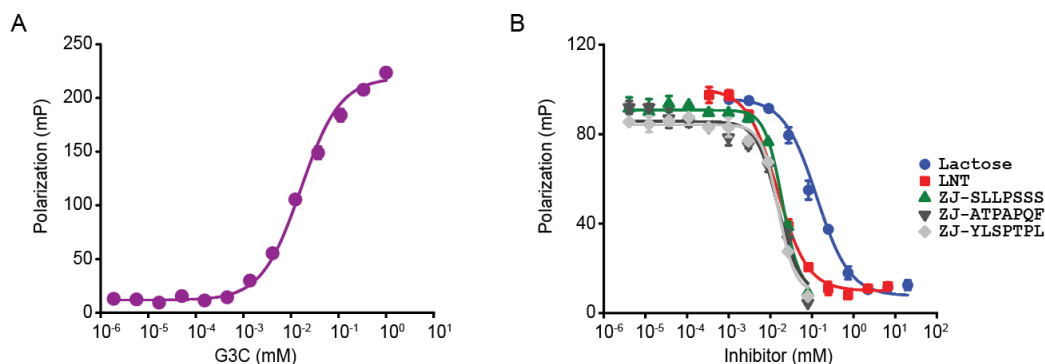
### 3.3.5. Validation of selected glycopeptides

To validate the binding activity of the selected hits (Scheme 3-1) we synthesized and tested the glycopeptides for their ability to compete with the fluorescent probe Lac-FITC (described on Chapter 2, section 2.4.12) for binding to G3C in solution.



**Scheme 3-1.** Structure of selected glycopeptides tested for G3C binding affinity in fluorescence polarization assays.

We started fluorescence polarization (FP) binding studies by establishing the binding isotherm of G3C in a presence of 1.0  $\mu\text{M}$  of fluorescent probe. Under these conditions, we determined the concentration of G3C that produced 50% response of the possible dynamic range (Figure 3-9A). Based on direct binding results, we implemented FP competition experiments.



**Figure 3-9.** Fluorescence polarization assays to validate the binding affinity of selected glycopeptides. (A) Polarization response of different concentrations of G3C incubated with the fluorescent probe FPP1 at room temperature. (B) Concentration dependent responses of the binding of competitive inhibitors to G3C at room temperature.

All synthetic glycopeptides showed IC<sub>50</sub> values in the range 10-20  $\mu$ M which is in the same order of that of the natural G3C ligand LNT (IC<sub>50</sub> = 14.6  $\pm$  0.9  $\mu$ M). In other words, the G3C binding potency of the TAZTDG containing glycopeptides is comparable to the binding potency of LNT. Our findings are consistent with previous studies of galectin-3 binding affinity to LNT (K<sub>D</sub> = 0.9  $\mu$ M)<sup>125</sup> and TAZTDG (K<sub>D</sub> = 0.8  $\mu$ M).<sup>128</sup> Further validations are needed to fully characterize ZJ-peptide conjugates as G3C inhibitors. Systematic alanine (A) mutations throughout all X positions on the 7<sup>mer</sup> peptide portion will provide valuable insights as to which residues are critical for binding affinity.

Although synthetic inhibitors based TDG core had shown superior binding potency to G3C (SL4, Table 3-1),<sup>129</sup> the ligands derived from glycopeptide phage selections represent a new class of inhibitors in which the peptide portion could further functionalized with ease. This class of ligands opens the possibility for the exploration of genetical fusions between bioactive peptides



identified in our studies with existing therapeutic proteins. Such studies could fuel the development of enhanced combined therapies that invoke galectin-3 inhibition.

### **3.4. Conclusions**

In summary with the use of GE-FBD technology, starting from a repertoire of  $\sim 10^8$  diverse glycan-peptide combinations, we discovered a novel class of G3C inhibitors with more than 10-fold improvements in binding affinity compared to native ligand lactose.

We also demonstrated that GE-FBD efforts starting from fragments with high (mM) affinity to the target yields monovalent ligands. Similar conclusion was obtained on previous GE-FBD that explored different selections with fragments of different affinities.<sup>22</sup> This observation complements previous chapter conclusion for weak (mM) fragments as starting point for GE-FBD campaigns that steers the selection to potent but multivalent ligands.

This work calls for further studies on the selection of the starting fragment on GE-FBD efforts. A deeper understanding of these effects on GE-FBD requires systematic galectin-binding studies of multiple glycan fragments displayed on phage.

### **3.5. Experimental procedures**

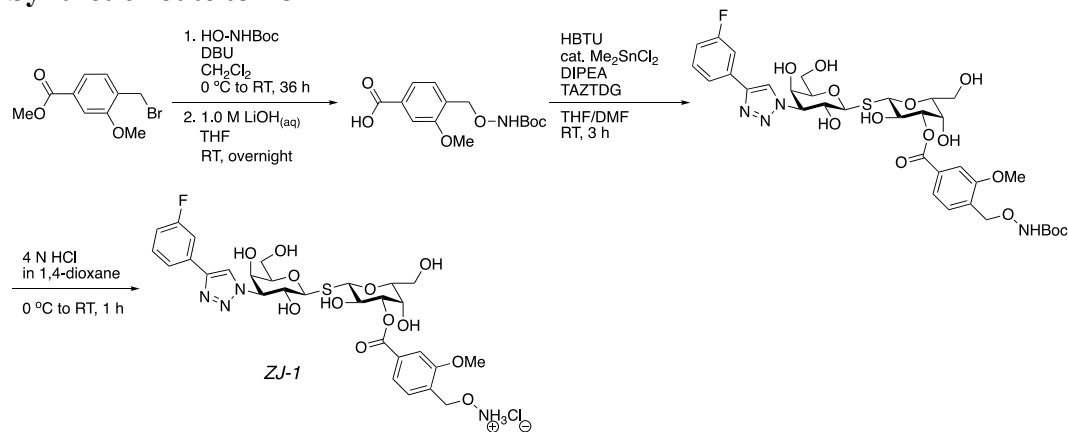
#### **3.5.1. Materials and general information**

All assays unless otherwise specified, were performed in MOPS buffer (50 mM MOPS, 150 mM NaCl, 2 mM CaCl<sub>2</sub> pH 7.4). Solutions used for phage work were sterilized by filtration through 0.22  $\mu$ m filters. All plate-based

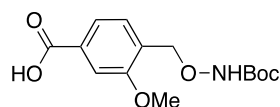
measurements (absorbance  $A_{450\text{ nm}}$  and fluorescence polarization) were carried out in a Cytation5 system (Biotek Instruments, Winooski VT, USA). RP-HPLC purifications were performed on Waters HPLC system equipped with a 1525 EF binary pump, a FlexInject manual injector (dual mode) and a 2489 tunable UV detector. A SymmetryPrep<sup>TM</sup> C18 semi-preparative column (19 × 50 mm, particle size 5  $\mu\text{m}$ , pore size 100 Å) was used for all peptide purifications at a typical flow rate of 12 mL/min. The mannose-binding protein concanavalin-A (ConA) was purchased from Sigma-Aldrich (#C2010). The carbohydrate recognition domain of the human Gal3 (G3C), residues 107–250 (MW: 16 327 Da) used in this study was provided by Dr. Christopher Cairo (University of Alberta, CA). The glycan ZJ was synthesized by Dr ZJay Tu (Academia Sinica, Taiwan) following the methodology described below. The glycan LNT was ordered from Elictyl (#GLY010) Aminoxy-biotin was purchased from Cayman Chemical (#10009350). High definition mass-spectrometry measurements were carried out in a Cytation5 system (Biotek Instruments, Winooski VT, USA). Purification of aminoxy glycans and their precursors was accomplished with an automated chromatograph (CombiFlash® Rf, Teledyne ISCO, USA). HRMS-ESI spectra were recorded on Agilent 6220 TOF mass spectrometer using either positive or negative ionization mode. Characterization of glycopeptides was performed with UPLC-MS using a C18 column (Phenomenex Kinetex 1.7  $\mu\text{m}$  EVO C18, 2.1×50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min.

### 3.5.2. Synthesis of ZJ glycan

#### Synthetic route to ZJ



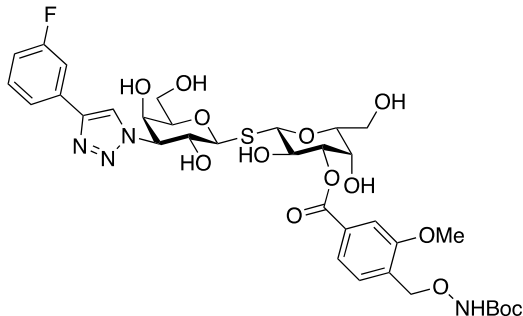
#### 4-([(tert-butoxycarbonyl)aminoxy]methyl)-3-methoxybenzoic acid



To a solution of commercially available methyl 4-(bromomethyl)-3-methoxybenzoate (500 mg, 1.0 equiv., 1.93 mmol) and *N*-Boc hydroxylamine (643 mg, 2.5 equiv., 4.83 mmol) in dry dichloromethane (10 mL) was added dropwise the 1,8-Diazabicyclo[7.1.1]undec-7-ene (DBU) (0.29 mL, 1.0 equiv. 1.93 mmol) at 0 °C under N<sub>2</sub> atmosphere. The ice bath was removed, and the resulting mixture was allowed to stir at room temperature for another 36 h. The reaction mixture was diluted with dichloromethane (10 mL) and washed with ice cold water. The combined organic layer was dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo* to afford the amber crude, which was pass through a pad of silica gel (hexanes/EtOAc, 4:1, v/v) to remove the residual *N*-Boc hydroxylamine. The desired product was obtained as a colorless syrup after concentration and was then proceeded for next step.

To the solution of above compound in 10 mL of THF was added the freshly prepared 1.0 M aqueous LiOH (10 mL) at room temperature. The resulting mixture was allowed to stir at room temperature overnight. Amberlite IR-120 [H<sup>+</sup>] resin was added to quench the reaction, to adjust the pH value to 6 and then filtered. The filtrate was concentrated, co-evaporated with ethanol twice and kept under high vacuum to afford the titled compound (436 mg, 76% yield over 2 steps) as a white solid in analytical purity.  $R_f = 0.2$  (hexanes/EtOAc, 1:2, v/v); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta = 7.70$  (d,  $J$  7.9 Hz, 1 H, ArH), 7.58 (br, 1 H, NH), 7.55 (s, 1 H, ArH), 7.48 (d,  $J$  7.9 Hz, 1 H, ArH), 4.98 (s, 2H, ArCH<sub>2</sub>), 3.90 (s, 3H, OCH<sub>3</sub>), 1.50 (s, 9H, t-Bu) ppm; <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta = 171.2, 157.8, 157.4, 130.9, 130.5, 130.3, 122.7, 111.7, 82.3, 73.0, 55.8, 28.4$  ppm; HRMS (ESI-TOF):  $m/z$  calcd for C<sub>14</sub>H<sub>19</sub>NO<sub>6</sub>Na: 320.1105 [ $M+Na$ ]<sup>+</sup>; found: 320.1105.

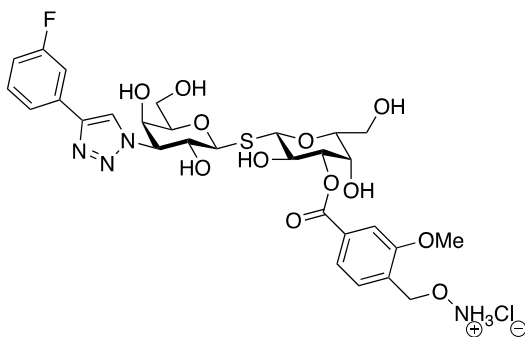
### TAZTDG-O3-[4-(NBoc-aminooxymethyl)-3-methoxy]-Benzoate



To a solution of above synthesized hydroxylamine-containing benzoic acid derivative (45 mg, 1.5 equiv. 0.15 mmol), HBTU (57 mg, 1.5 equiv., 0.15 mmol) in THF/DMF (1.0 mL, 9:1, v/v, 0.1 M) was sequentially added the DIPEA (35  $\mu$ L, 2.0 equiv., 0.20 mmol), Me<sub>2</sub>SnCl<sub>2</sub> (2.2 mg, 10 mol%, 10.0  $\mu$ mol) and TAZTDG (50 mg, 1.0 equiv., 0.10 mmol) at room temperature under N<sub>2</sub> atmosphere. The resulting mixture was stirred at room temperature for 3 h. Upon completion (monitored by TLC), the solvent was evaporated, and the resulting residue was then purified by column chromatography (CHCl<sub>3</sub>/MeOH, 4:1, v/v) to give the desired product as a white solid (56 mg, 72% yield). R<sub>f</sub>=0.5 (CHCl<sub>3</sub>/MeOH, 4:1, v/v); <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  = 8.51 (s, 1 H, triazolyl H), 7.74 (dd, *J* 7.8, 1.3 Hz, 1 H, ArH), 7.68 (d, *J* 1.0 Hz, 1 H, ArH), 7.66 (dt, *J* 7.8, 1.0 Hz, 1 H, ArH), 7.60 (m, 1 H, ArH), 7.52-7.42 (m, 2 H, ArH), 7.07 (td, *J* 8.4, 2.2 Hz, 1 H, ArH), 5.07 (dd, *J* 9.6, 3.2 Hz, 1 H, H<sub>3</sub>), 4.97 (d, *J* 9.6 Hz, 1 H, H<sub>1</sub>'), 4.93-4.88 (m, 4 H, H<sub>1</sub>, H<sub>3</sub>', ArCH<sub>2</sub>), 4.45 (t, *J* 10.1 Hz, 1 H, H<sub>2</sub>'), 4.22 (d, *J* 3.2 Hz, 1 H, H<sub>4</sub>), 4.16 (t, *J* 9.6 Hz, 1 H, H<sub>2</sub>), 4.15 (d, *J* 2.9 Hz, 1 H, H<sub>4</sub>'), 3.90 (s, 3 H, OCH<sub>3</sub>), 3.91-3.70 (m, 6 H, H<sub>6ab</sub>', H<sub>6ab</sub>, H<sub>5</sub>, H<sub>5</sub>'), 1.46 (s, 9 H, t-Bu) ppm; <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD):  $\delta$  = 167.9, 166.4, 163.9, 159.5, 159.3, 147.7, 134.9, 134.8, 133.0, 132.4, 132.3, 131.8, 131.3, 123.6, 123.0, 116.3, 116.1,

113.8, 113.6, 112.8, 86.7, 85.9, 82.6, 81.8, 81.2, 79.7, 74.0, 70.3, 69.8, 69.5, 69.02, 68.99, 63.2, 56.7, 31.2, 29.1 ppm; HRMS (ESI-TOF):  $m/z$  calcd for C<sub>34</sub>H<sub>44</sub>FN<sub>4</sub>O<sub>14</sub>S: 783.2553 [M+H]<sup>+</sup>; found: 783.2567.

## ZJ



To the above synthesized TAZTDG-O3-([4-(NBoc-aminoxy)methyl]-3-methoxy)-benzoate (24 mg, 1.0 equiv., 30  $\mu$ mol) in a test tube reactor was added the 0.3 mL of 4 N HCl(1,4-dioxane) (commercially available reagent) at 0 °C under N<sub>2</sub> atmosphere. The ice bath was removed, and the resulting white slurry was allowed to stir at room temperature for 1 h. 5 mL of ethyl ether was added to the reaction mixture and the resulting white precipitates was filtered. The collected solid was further washed with ethyl ether and chloroform and kept under high vacuum to provide the desired product (15 mg, 70% yield) in an analytical purity. R<sub>f</sub>=0.3 (CHCl<sub>3</sub>/MeOH, 4:1, v/v); <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  =8.53 (s, 1 H, triazolyl H), 7.79 (dd, *J* 7.8, 1.3 Hz, 1 H, ArH), 7.77 (d, *J* 1.2 Hz, 1 H, ArH), 7.66 (d, *J* 7.8 Hz, 1 H, ArH), 7.60 (dt, *J* 10.1, 1.4 Hz, 1 H, ArH), 7.50-7.43 (m, 2 H, ArH), 7.07 (td, *J* 8.6, 2.0 Hz, 1 H, ArH), 5.14 (s, 2 H, ArCH<sub>2</sub>), 5.08 (dd, *J* 9.8, 3.2 Hz, 1 H, H<sub>3</sub>), 4.97 (d, *J* 9.5 Hz, 1 H, H<sub>1</sub>'), 4.93-4.88 (m, 2 H, H<sub>1</sub>, H<sub>3</sub>'), 4.43 (t, *J* 10.0 Hz, 1 H, H<sub>2</sub>'), 4.23 (d, *J* 3.1 Hz, 1 H, H<sub>4</sub>), 4.15 (d, *J* 2.9 Hz, 1 H,

H4'), 4.14 (t, *J* 9.8 Hz, 1 H, H2), 3.96 (s, 3 H, OCH3), 3.91-3.70 (m, 6 H, H6ab', H6ab, H5, H5'), 1.46 (s, 9 H, t-Bu) ppm; <sup>13</sup>C NMR (125 MHz, CD3OD):  $\delta$  = 167.5, 166.1, 164.2, 159.8, 147.4, 134.8, 134.39, 134.33, 132.5, 132.4, 132.36, 128.2, 123.9, 123.3, 123.0, 116.6, 116.4, 113.9, 113.7, 113.4, 86.6, 85.8, 81.7, 81.3, 81.2, 79.9, 73.2, 70.3, 69.8, 69.78, 69.0 (x2), 63.2, 56.9 ppm; HRMS (ESI-TOF): *m/z* calcd for C<sub>29</sub>H<sub>36</sub>FN<sub>4</sub>O<sub>12</sub>S: 683.2029 [M+H]<sup>+</sup>; found: 683.2037.

### 3.5.3. Study of ZJ ester hydrolysis.

A 1.7 mL plastic tube was prepared with 0.7 mg of pure solid ZJ. The solid was dissolved in 0.5 mL of H<sub>2</sub>O/MeOH 95:5 v/v to give 2mM solution. To this solution, 10  $\mu$ L of acetone were added to quench the hydroxylamine group. Then the sample was equally distributed into two tubes. One tube was incubated with equal volume of phage solution containing 10<sup>11</sup> PFU in MOPS pH7.4 at room temperature. In parallel, the remaining tube was distributed into three reaction vessels and incubated in aqueous solution of NaOH at pH 10, 11.5 and 12. Hydrolysis of ZJ was followed by UPLC-MS. Samples were taken at *t* = 0, 30, 60, 90 and 240 min of incubation in either alkali or phage solution. Samples from incubation were injected (4  $\mu$ L injection) in a C18 column (Phenomenex Kinetex 1.7  $\mu$ m EVO C18, 2.1 $\times$ 50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min.

### 3.5.4. Chemical conjugation of biotin to glycan binding proteins

For NHS-ester lysine conjugation, 1 mL of protein target G3C (similar procedure was used for ConA) dissolved in PBS pH 7.4 at concentration of 10

mg/ml was transferred into 1 mL of Carbonate-Bicarbonate buffer 50mM, pH 9.2 using Amicon filter (Sigma, # UFC5010) with 10 kDa MW cut off. The protein solution was then mixed with 3.6 mg of lactose to obtain a final concentration of 10 mM. The mixture was then reacted with 2 equivalents of EZ-Link NHS-PEG4-Biotin(#271611ChemPep) which were dissolved, directly into the protein solution.

The reaction was incubated 1h at RT. The non-reacted NHS-PEG4-Biotin was removed by filtration into 200 mM ammonium acetate aqueous solution, pH 6.8 using an Amicon filter. Collected protein samples were stored at -20 °C until further use. The labeled-protein concentration was estimated using the formula described on Chapter 2, Section 2.4.16.

### **3.5.5. ESI-MS binding studies**

To perform glycan binding nanoESI tips were produced from borosilicate capillaries (1.0 mm o.d., 0.78 mm i.d.) pulled to ~2  $\mu\text{m}$  outer-diameter using a P-1000 micropipette puller (Sutter Instruments, Novato, CA). To perform nanoESI, a platinum wire was inserted into the solution in the tip and a voltage of ~1.0 kV was applied. A cone voltage of 20 V was used, and the source block temperature was maintained at 60 °C. The Trap and Transfer voltages were 3 V and 1 V, respectively; argon was used in the Trap and Transfer ion guides at pressures of  $2.22 \times 10^{-2}$  mbar and  $3.36 \times 10^{-2}$  mbar, respectively. The helium chamber preceding the traveling wave IMS (TWIMS) device was maintained at 7.72 mbar. The IMS parameters were as follows: 2 mL min<sup>-1</sup> Trap gas flow rate; 180 mL min<sup>-1</sup> helium cell gas flow rate; 90 mL min<sup>-1</sup> ion mobility gas flow rate; 3



V Trap voltage; 700 m s<sup>-1</sup> IMS wave velocity; 36 V IMS wave height. All IMS measurements were carried out using nitrogen as the mobility gas, at a pressure of 3.41 mbar. Data acquisition and processing were carried out using MassLynx (v4.1).

The oligosaccharide ligand (L) affinities for free and labeled G3C (referred to here as P<sub>*i*</sub>, where *i* is the number of labels added) were measured using the direct ESI-MS assay. Briefly, the affinity (association constant, K<sub>a,*i*</sub>) is calculated from the ratio (*R<sub>i</sub>*) of the measured abundances (*Ab*) of the gaseous L-bound and free P<sub>*i*</sub> ions, which is taken to be equal to the corresponding concentration ratio in solution, eq 1:

$$R_i = \frac{[P_iL]}{[P_i]} = \frac{Ab(P_iL)}{Ab(P_i)} \quad (1)$$

Where required, prior to calculating *R<sub>i</sub>*, the mass spectrum was corrected for the occurrence of nonspecific glycan-GBP binding during the ESI process using the reference protein method.<sup>16</sup>

The expression for K<sub>a,*i*</sub>, in terms of *R<sub>i</sub>*, is given by eq 2:

$$K_{a,i} = \frac{[P_iL]}{[P_i][L]} = \frac{R_i}{[L]} \quad (2)$$

When multiple forms of P<sub>*i*</sub> are present in the same solution, the ratio of affinities (e.g., K<sub>a,*i*</sub>/K<sub>a,*i-1*</sub>) for two different forms of P (e.g., P<sub>*i-1*</sub> and P<sub>*i*</sub>) can be determined from the corresponding abundance ratio (e.g., *R<sub>i</sub>*/*R<sub>i-1</sub>*), eq 3:

$$\frac{[P_iL]/[P_i]}{[PL_{i-1}]/[P_{i-1}]} = \frac{Ab(P_iL)/Ab(P_i)}{Ab(PL_{i-1})/Ab(P_{i-1})} = \frac{R_i}{R_{i-1}} = \frac{K_{a,i}}{K_{a,i-1}} \quad (3)$$

$K_{a,i}/K_{a,i-1}$  is independent of the protein and ligand concentrations. Consequently, it is not necessary to know the concentration of individual  $P_i$  produced by labeling in order to establish their relative affinities.

### 3.5.6. Construction of blocking phage

The Blocking phage is a M13 derivative contains silent mutations within the regions recognized by the Illumina primer and, thus, not amenable to either PCR amplification nor Illumina sequencing. The M13 Blocking phage was constructed using the NEB builder (NEB#E5510). Amplicon 1 was constructed from M13 dsDNA PCR amplified with primers P1 and P2:

```
Name:      Sequence (5' ->3')
P1 CAGAAAATTCATTTACTAACGTCTGGAA
P2 AAAGGAACAACATAAGGAATTGCG
P3 TATTTCGCAATTCTTTAGTTGTTTCCTTTGTACAGCCATAGTGCGGAGACCGTGGAAAGTTGTTTTCAGCAAAACCC-
CA
P4 TAAATGAATTTTCTGTA
```

The same M13 vector was then amplified using primers P3 and P4 to yield amplicon 2. PCR was performed using 50 ng phage dsDNA with 1 mM dNTPs, 0.5  $\mu$ M primers, 0.5  $\mu$ L Phusion High Fidelity DNA polymerase in 1x PCR buffer (NEB #B0518S ) in a total volume of 50  $\mu$ L. The temperature cycling protocol was performed as follows: a) 98 °C 3 min, b) 98 °C 30 s, c) 60 °C 30 s, d) 72 °C 4 min s, e) repeat b) - d) for 35 cycles, f) 72 °C 10 min, g) 4 °C hold. PCR Amplicons 1 and 2 were treated with Dpn1(NEB #R0176S), gel purified, ligated using NEBuilder Hifi assembly, DNA assembly was then carried out following by mixing 100 ng of vector, 4ng insert, 10  $\mu$ L of NEBuilder Hifi DNA assembly master mix, and deionized H<sub>2</sub>O up to a total volume of 20  $\mu$ L according

to manufacturer recommendations. Ligated DNA was transformed into E.coli XL1 Blue by electroporation and propagated overnight at 37 °C. The overnight culture was then centrifuged to separate bacteriophage from host cells, incubated with 5% PEG-8000, 0.5 M NaCl for 8 h at 4 °C, followed by 15 min centrifugation at 13000 g to concentrate released phage. PEG precipitated phage were resuspended in PBS-Glycerol 50% and stored at -20 °C.

### **3.5.7. Construction of phage clone SWYD that express mCherry fluorescent protein.**

We produced filamentous phage vector that contains the gene for fluorescent protein mCherry cloned in place of the lacZ $\alpha$  fragment. Vector containing mCherry was donated by Dr. Robert Campbell (University of Alberta). Genetic construct was built by HiFi NEB builder ligation of fluorescent protein (FP) fragment and M13 fragment. Vector pBAD-mCherry was used as the source for the mCherry insert. FP fragment was PCR amplified following the protocol described on section 3.6.6 using primers P5 and P6, whereas the M13 fragment was PCR amplified from M13vector using primers P7 and P8.

Name:	Sequence (5' ->3')
P5	GCGGATAACAATTTACACAGGAAACAGCTATGGTGAGCAAGGGCGAG
P6	TTAAATTTTTGTAAATCAGCTCATTTTTACTTGTACAGCTCGTCCA
P7	AAAATGAGCTGATTTAACAAAAATTTAA
P8	AGCTGTTTCCTGTGTGAAAT
P9	TTAAGACTCCTTATTACGCAGTA
P10	CCTTCTATTCTCACTCGAGCTGGTATGATCTGTATCATGGTGGAGGTTCGGCC
P11	TTTTTTCACGTTGAAAATCTC
P12	CGAGTGAGAATAGAAAGGTAC

The cloning of SWYDLYH peptide was conducted as follows. The insert fragment was amplified by PCR following the described protocol using

primers P9 and P10. The vector fragment was PCR amplified using primers P11 and P12. PCR fragment were processed using NEBuilder Hifi DNA assembly kit using steps described above. The resulting ligated DNA was transformed into electrocompetent cells *E.coli* SS320 (Lucigen). Assembled DNA was transformed into competent cells by electroporation and propagated overnight at 37 °C. The overnight culture was then centrifuged to separate bacteriophage from host cells, incubated with 5% PEG-8000, 0.5 M NaCl for 8 h at 4 °C, followed by 15 min centrifugation at 13000 g to concentrate released phage. PEG precipitated phage were resuspended in PBS-Glycerol 50% and stored at -20 °C.

### **3.5.8. Serine-X<sub>7</sub> phage library amplification**

The phage library used in this project was amplified as follows, from stock of *N-SerX<sub>7</sub>* phage displayed library (diversity:  $3 \times 10^8$ ).<sup>150-151</sup> 200µL of log phase *E. coli* K12 ER2738 were added to 20 ml of LB broth in 125 mL culture flask and incubated at 37 °C until OD<sub>600nm</sub>: 0.2. To this culture, 10<sup>10</sup> PFU from *N-SerX<sub>7</sub>* phage library were added and incubated for 30 mins at 37 °C with shaking. After incubation, the solution transfer was transferred to pre-warm 180 mL of LB and incubated for 3 hours. The culture was then centrifugated 6000 g for 10 min. The supernatant was transferred to a new 250 mL centrifuge bottle, PEG 8000 and NaCl were added to a final concentration of 5% PEG 8000 and 0.5 M of NaCl. The bottle was incubated overnight at 4°C. The PEG-precipitated centrifuge bottle was centrifugated at 20,000×g for 30 mins at 4°C to collect the phage pellet. The pellet was resuspended with 2 mL 1×PBS and then transferred into a 10K MW

dialysis tube and dialyzed in 1×PBS overnight at 4°C. After dialysis phages were stored in glycerol until further use.

### **3.5.9. Production of phage displayed glycopeptide conjugates**

Phage clone transducing mCherry SWYDLYH (section 3.6.6), and *N*-SerX<sub>7</sub> phage library (section 3.6.7), were chemically modified to produce Man-WYDLYH, Man-X7 and ZJ-X7 phage displayed glycopeptide conjugates. First the N-terminus serine displayed on phage pIII was oxidized to convert it into aldehyde that was then reacted with the corresponding aminoxy glycans to yield the desired phage displayed glycopetides. The chemical conjugation protocol, and biotin capture technique for quantification of efficiency of glycosylation were done following published protocol<sup>42</sup> described in Chapter 2 section 2.4.3.

### **3.5.10. Panning with fluorescent selection controls**

Biotinylated ConA and G3C proteins (10 µg per experiment) were prepared in separated 1.7 mL plastic tubes. Biotinylated proteins were incubated with a mixture of Man-WYDLYH ( $9 \times 10^7$  PFU/mL), *N*-SerX<sub>7</sub> phage library ( $3 \times 10^{10}$  PFU/mL) and blocking phage ( $9 \times 10^9$  PFU/mL) in MOPS buffer, total volume 0.5 mL. Three technical replicates were performed for each phage-mixture/target pair. After 90 min, room temperature incubation, 30 µL of streptavidin-coated magnetic beads (Promega, #Z5481) were added to each replicate and incubated with gentle shaking for 15 min. The tubes were the placed in magnetic rack to capture the beads. The efficiency of capture of biotinylated targets was confirmed by SDS-PAGE of the supernatant remaining after beads depletion (Figure 3-4B). Captured beads were washed 5 times using 1 mL of 0.1

% tween in HEPES buffer in each wash. After the final wash, the beads were resuspended in 30  $\mu$ L of elution buffer (0.2 M glycine-HCl, pH 2.2, 0.1% w/v BSA) for 9 minutes and neutralized with 6  $\mu$ L 1 M Tris-HCl, pH 9.1 to detach phages from beads. The eluted phage solution was quantified by plaque forming assay.

### **3.5.11. Solution-phase panning with ZJ-X<sub>7</sub> and Man-X<sub>7</sub> glycopeptide phage libraries**

The selection experiments were performed using KingFisher™ Duo Prime Purification System, 96 Deepwell plate (ThermoFisher, #95040450) was used to contain the solutions and 12-tip Deepwell magnetic comb was used to transfer beads (ThermoFisher, #97003500). Portions 30  $\mu$ L of streptavidin-coated magnetic beads (Promega, #Z5481) were transferred into the 96 Deepwell plate the beads were then washed in 1mL of 0.1% Tween-20 in MOPS (v/v) for 30 seconds on gentle shaking. The beads were then incubated in blocking buffer (1 mL, MOPS buffer + 2% BSA (w/v),) for 1.5 h at RT to reduce nonspecific binding. In parallel, biotinylated ConA and G3C proteins (10  $\mu$ g per experiment) were transferred into the 96 Deepwell plate. Biotinylated proteins were incubated with either of two selection inputs: Man-X<sub>7</sub> ( $3 \times 10^{11}$  PFU/mL) phage library and blocking phage ( $7 \times 10^{11}$  PFU/mL); or ZJ-X<sub>7</sub> ( $5 \times 10^{11}$  PFU/mL) library and blocking phage ( $7 \times 10^{11}$  PFU/mL) in MOPS buffer. Four selection-target combinations with three technical replicates were performed for a total of 12 selection screenings. After blocking, the beads were transferred into the selection wells. After 15 min capture the beads were washed twice as described above.

After the washes the beads were transferred into 1.7 mL microcentrifuge tubes containing 1mL MOPS and subsequently placed on a magnet stand to collect the beads on bottom of the tube. After capture, the beads were resuspended in 30  $\mu$ L of 10 mM Tris, pH 8.5. Hexane (30  $\mu$ L) of was added, and the mixture was shaken at 3000 rpm for 10 min to extract phage DNA. Hexane was evaporated for 10 min on a heat block at 70 °C. The extracted phage ssDNA was subjected to PCR according to the procedure described below.

### **3.5.12. DNA amplification and sequencing**

To convert phage DNA to Illumina-compatible short double-stranded DNA (dsDNA) by PCR, we followed the PCR routine, and used the primers described on Chapter 2 section 2.4.3. Sequencing was performed using the Illumina NextSeq platform (Molecular Biology Service Unit, University of Alberta).

### **3.5.13. Differential enrichment analysis**

Raw FASTQ were downloaded from Illumina cloud service and processed using MATLAB scripts described in Chapter 2 section 2.4.3. Identification of significantly enriched sequences from deep-sequencing data was performed similarly to described procedure<sup>20, 22</sup> and the algorithm is written in one MATLAB script entitled *DE\_analysis.m* available in Appendix B17. In short, for each sequence we calculated the average normalized frequency at which it appeared in each sequencing set and then the ratio as average frequency in test experiment divided by an average frequency in the control experiment. We also calculated the p-value between the replicates of test and replicates of control

experiment using a two-sided unequal variance t-test. Hit sequences were defined as those that had  $p < 0.05$  and ratio  $\geq 3$  for all control experiments and were presented as heat map and volcano plots (Figure 3-9A,B).

#### **3.5.14. Synthesis of glycopeptides**

Synthesis and purification of glycopeptides was performed following the procedures described on Chapter 2 sections 2.4.9, 2.4.10 (solid phase peptide synthesis and purification) and section 2.4.11 (chemical glycosylation of peptides). Characterization of the glycopeptides was performed using a UPLC system equipped with a C18 column (Phenomenex Kinetex 1.7  $\mu\text{m}$  EVO C18, 2.1 $\times$ 50 mm) running with a gradient of water/acetonitrile with 0.1% formic acid from 98/2 at 0 min to 40/60 at 5 min under a flow rate of 0.5 mL/min. Characterization data is available in Appendix B1-B3.

#### **3.5.15. Fluorescence polarization assays to test the binding of glycopeptides to soluble G3C**

The fluorescein-lactosamine (Lac-FITC) probe was resuspended distilled water to obtain a concentration of 20  $\mu\text{M}$ . For direct binding the starting solution of G3C at  $\sim 2$  mM in MOPS buffer was prepared as serial 3-fold serial dilutions and loaded into a black 386-well plate (Perkin Elmer #110200365) to final volume of 19  $\mu\text{L}$ /well. Each well was supplemented with 1  $\mu\text{L}$  of the probe solution to get a final concentration of 1  $\mu\text{M}$  of fluorescent conjugate in 20  $\mu\text{L}$  total volume per well. The plate was incubated for 10 min in the dark at room temperature under slow shaking. Fluorescence polarization was measured at room temperature using a Cytation5 plate reader ( $\lambda_{\text{Ex}} = 485$  nm,  $\lambda_{\text{Em}} = 528$  nm). For the



inhibition assay, 20  $\mu\text{L}$  of protein (1.6  $\mu\text{M}$ , MOPS) and fluorescent probe (1  $\mu\text{M}$ ) at fixed concentrations were mixed with each inhibitor solution to make series of 3-fold dilutions. The plate was placed in the dark for 10 min and fluorescence polarization was measured on the Cytation5 as described above. All the data analysis and curve fitting was implemented using Origin software package (OriginLab, Massachusetts, USA)

## **Chapter 4: Unsupervised study of the interactions between glycan and glycan binding proteins in Genetically Encoded Fragment Based Discovery**

### **4.1. Introduction**

In previous chapters, we systematically tested whether GE-FBD approaches can give rise to specific and selective glycopeptide ligands for galectin-3. Each campaign started from the first, critical decision-making step: selection of a fragment(s) from which GE-FBD starts. When more than one fragment was used, silent encoding was employed to incorporate and track up to six monosaccharide units conjugated to peptide libraries. To maximize the success of future GE-FBD campaigns for which the fragments might not be obvious it could be advantageous to identify what glycan fragments bind to the target and can serve as a starting point for GE-FBD. For the screening of the glycan fragments we introduced 1:1 correspondence between DNA and glycan using “silent barcodes” technology, analogous to that used in chapter 2. Expanding on this concept, we developed a genetically-encoded glycan array technology termed “Liquid Glycan Array” (LiGA) which is a collection of glycosylated M13 virions. We then demonstrated that LiGA technology can identify glycan fragments that interact with galectin-3, galectin-1, as well as other lectins in purified form, in cells *ex vivo*, and *in vivo*.

### **4.2. Liquid Glycan Array (LiGA) platform**

The Central Dogma of Biology, DNA → RNA → Protein, allows studying DNA, RNA and proteins using a unified tool of next-generation DNA sequencing.<sup>152</sup> This ability revolutionized and transformed all areas of biomedical

and life science in the last 20-30 years.<sup>153</sup> Investigation of carbohydrates in biology cannot rely on DNA sequencing directly. Akin to a DNA microarray used in 1990s, the glycan array made by printing carbohydrates on distinct locations on the glass surface,<sup>154</sup> is the workhorse tool for identification of carbohydrates that bind to a GBP by providing a high throughput approach to identifying glycan ligands.<sup>155-159</sup> The information provided by glycan array—a glycan binding profile for a receptor—is a critical starting point for GE-FBD and most downstream fundamental applications such as improved design of inhibitors,<sup>160</sup> biomarkers, vaccines<sup>161</sup> and therapeutics.<sup>162</sup> The DNA arrays have been largely replaced by de novo analysis of DNA by deep sequencing<sup>163</sup>. We sought to develop a technology that introduces the missing one-to-one correspondence between DNA sequence and carbohydrate structure and employ powerful deep sequencing approaches to identify glycan binding profiles of GBPs *in vitro* and in cells *ex vivo* and *in vivo*.

Soluble GBPs like galectins, or membrane-bound GBPs like Siglecs and C-type lectins, recognize glycans displayed on the surface of cells in a mobile, semi-fluidic environment surrounded by other glycan's and biomolecules. Traditional “solid” glycan arrays critically mimic the multivalent nature of these interactions.<sup>164-166</sup> To ensure molecular recognition, the carbohydrate ligands must match the orientation and spacing of the binding sites of GBPs. Hence, features of spatial presentation like surface loading-density on glycan arrays, have a profound impact on the interactions of glycan and GBPs. The density of glycan immobilized on surface, influences both, the affinity and specificity of binding in

lectins,<sup>154, 167-168</sup> enzymes,<sup>169-170</sup> and antibodies,<sup>161, 171</sup> furthermore; systematic variations in the glycan density of microarrays had been used to differentiate subpopulations of serum antibodies, that are not detectable using a single glycan density.<sup>161</sup> The solid, immobile design of the array however, fundamentally limits the ability to study the cross talk and dynamic competition between multiple glycans with a GBP. Solid format is also incompatible with investigation of GBPs on the surface of cells *in vitro* and *in vivo*.

Multivalent glycan-decorated liposomal nanoparticles have been extensively used to study interactions between glycans and cell surface GBPs *ex vivo* and *in vivo*.<sup>172-175</sup> Such technology unfortunately does not permit encoding or tracking of different glycan structures. Ligation of glycans to DNA molecule is possible,<sup>26</sup> but monovalent display on DNA cannot mimic the multivalent presentation of glycans. Multivalent presentation of glycans on micron-size Luminex® beads permits encoding of ~100 glycans but there is currently no evidence that arrays built on micron-size beads can be used to perform reliable cell-binding assays and injection of these arrays *in vivo* is not possible.

Ideal platform for glycan arrays should display glycans on a multivalent, monodisperse carrier of sub-micron size with DNA barcode hidden inside the carrier. This carrier must be stable, robust and non-interfering with binding assays. Hiding DNA inside the carrier should also protect the DNA message to avoid the undesired interactions with the GBP or enzymes that may degrade of the DNA message *ex vivo* or *in vivo*.

In this chapter, we describe a new array format that combines the

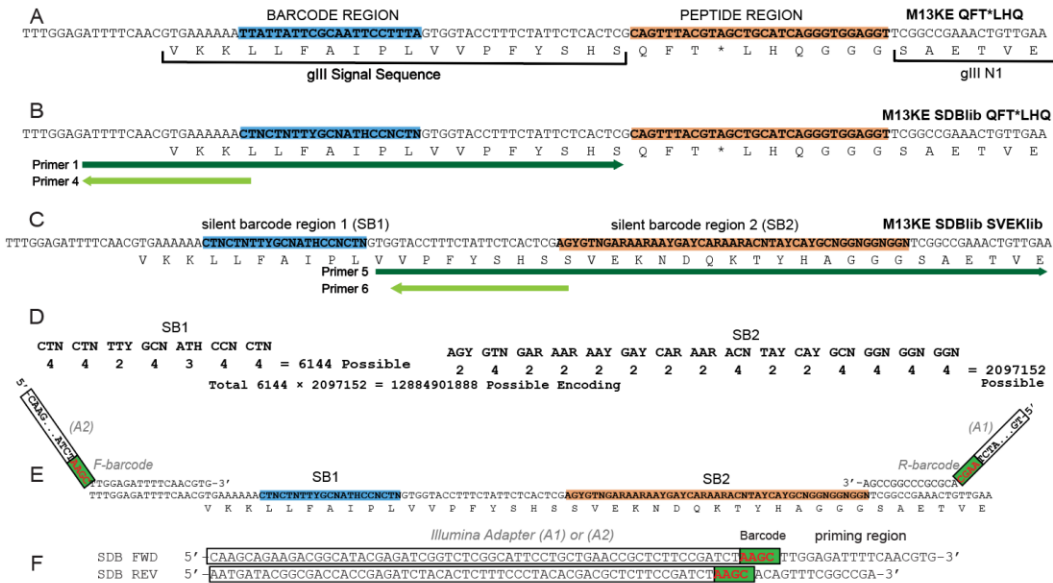
benefits of traditional “solid” or bead-based arrays, such as multivalent presentation while introducing DNA encoding and eliminating the weakness that hinder investigation of study interaction between glycans and GBP on the surface of intact cells, *ex vivo* and *in vivo*.

### **4.3. Results and discussion**

#### **4.3.1. Construction of the LiGA building block: glycan-phage conjugate**

As a platform that satisfies all the above requirements, we adopted M13 phage with silent DNA barcodes<sup>22</sup> inside of the phage genome. The M13 platform is scalable, portable, compatible with a wide range of cloning methods, as well as compatible with many chemical and enzymatic conjugation strategies in a variety of mixed aqueous-organic solvents.<sup>42</sup> In our design, glycans are chemically conjugated to the major coat protein of phage in a multivalent fashion at a density 200-1500 glycans per 700 nm long virion. Both the composition and presentation/density of glycan are encoded by “silent” DNA barcodes in phage genome.

We cloned silent double barcode (SDB) system in two locations of the phage genome proximal to the pIII cloning site of vector M13KE. The SDB design permitted constructing large silent barcode libraries and made analysis of SDB compatible with previously established deep-sequencing protocols<sup>109</sup> (Figure 4-1A,B).



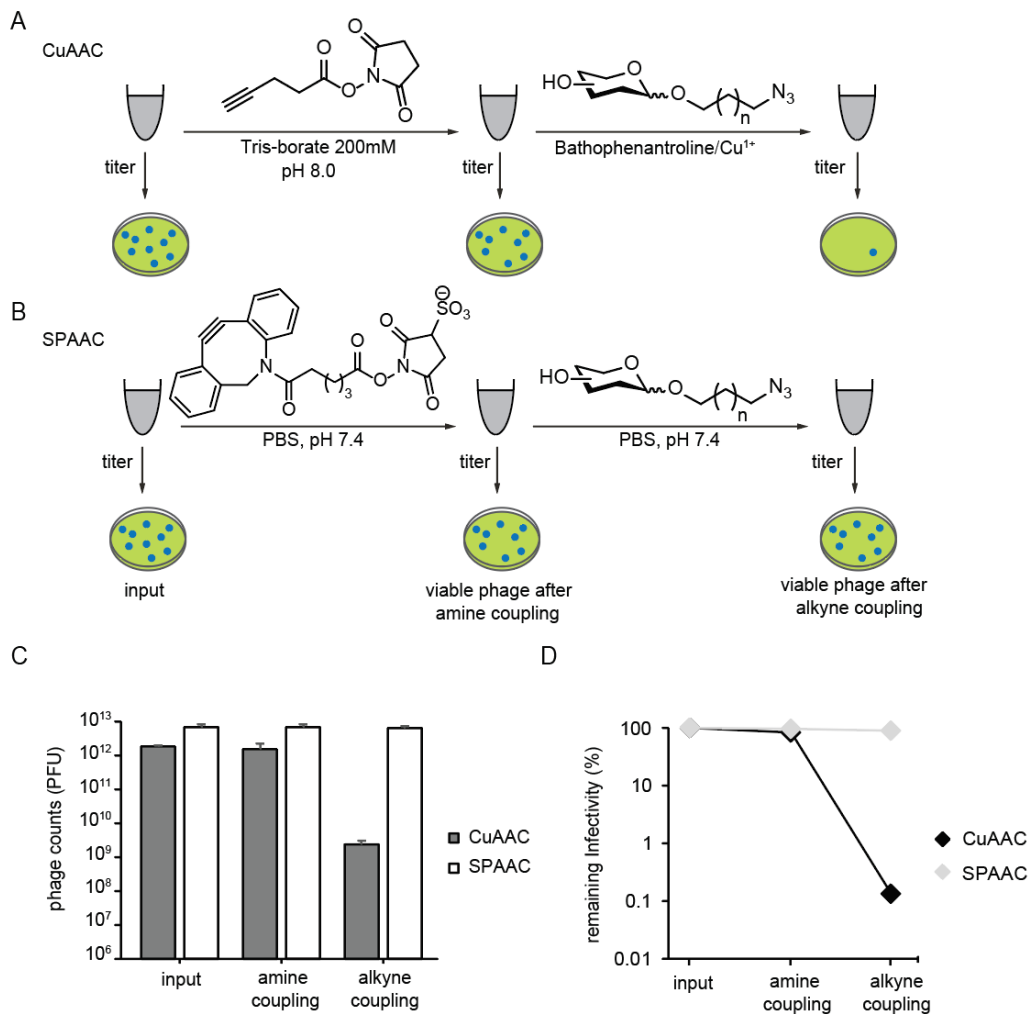
**Figure 4-1.** Cloning of SDB and Silent SVEK regions. (A) Location of SDB and Peptide region on initial template. (B) Primer 1 and 4 were used to convert M13KE QFT\* LHQ to M13KE SDBlib QFT\* LHQ. (C) Primer 5 and 6 were used to convert M13KE SDB QFT\* LHQ to M13KE SDBlib SVEKlib. (D) Summary of Degenerate sites in M13KE SDB QFT\* LHQ. (E) Location of Illumina Sequencing Primers. (F) Illumina Sequencing Primers.

After cloning of variable SB1 sequence and serial propagation of the library to remove unfit clones we selected 28 clones with SB1 sequences that are at least 3 units of Hamming (H) distance<sup>176</sup> apart. The H=3 spacing permitted correction of any point mutation arising during analysis by deep sequencing. Into these “silently” barcoded vectors, we cloned a DNA library encoding a unique peptide sequence. Encoding SB-2 by redundant DNA sequences, allowed to expand SDB library, adding an extra 2 million combinations. Taken together SB-1 and SB2 form a global library of  $10^{10}$  possible encoding combinations (Figure 4-1D). From that repertoire, we isolated and amplified 120 SDB phage clones that can be individually attached to glycan moieties (Appendix C1). A two-step modular design made it possible to change the nature of peptide sequence from inactive to active. For example, one can clone bioactive peptide purification tags

such as FLAG-tag<sup>177</sup> or HA-tag,<sup>178</sup> or inactive sequence SVEKNDQKTYHAGGG (SB-2 region, Figure 4-1C) which has no known biological activity and it has been used as control in multiple prior publications.<sup>26, 42, 47, 86, 179</sup>

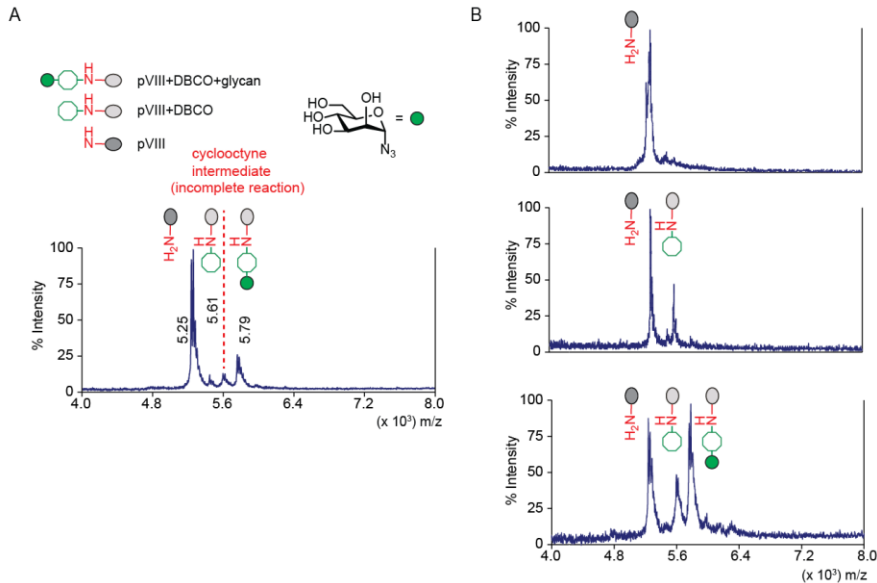
Glycans with alkylazido linker are common intermediates in oligosaccharide synthesis and approximately 100 different saccharides are readily available as part of the public Consortium for Functional Glycomics (CFG) collection.<sup>180</sup> This collection can be easily conjugated to phage particles via azido-alkyne cycloaddition. We tested both Cu-Activated Azido Alkyne Cycloaddition<sup>181</sup> (CuAAC) and Strain Promoted Azido Alkyne Cycloaddition<sup>182</sup> (SPAAC) using phage particles acylated by *N*-succinimidyl 4-pentynoate for CuAAC (Figure 4-2A) or dibenzocyclooctyne -*N*-hydroxysuccinimidyl ester (DBCO) for SPAAC (Figure 4-2B).

Acylation of the N-terminus of the major coat protein pVIII installed the alkyne handles (Figure 4-2C) followed by SPAAC ligation led to no detectable decrease in the number of infectious particles indicating that these modifications leave M13 particles intact. In contrast, CuAAC ligation reduced the observed number of infectious phages by a factor of 1000 (Figure 4-2C, D). While it should be possible to tune CuAAC conditions to minimize toxicity, we employed SPAAC as the glycosylation strategy in this report.



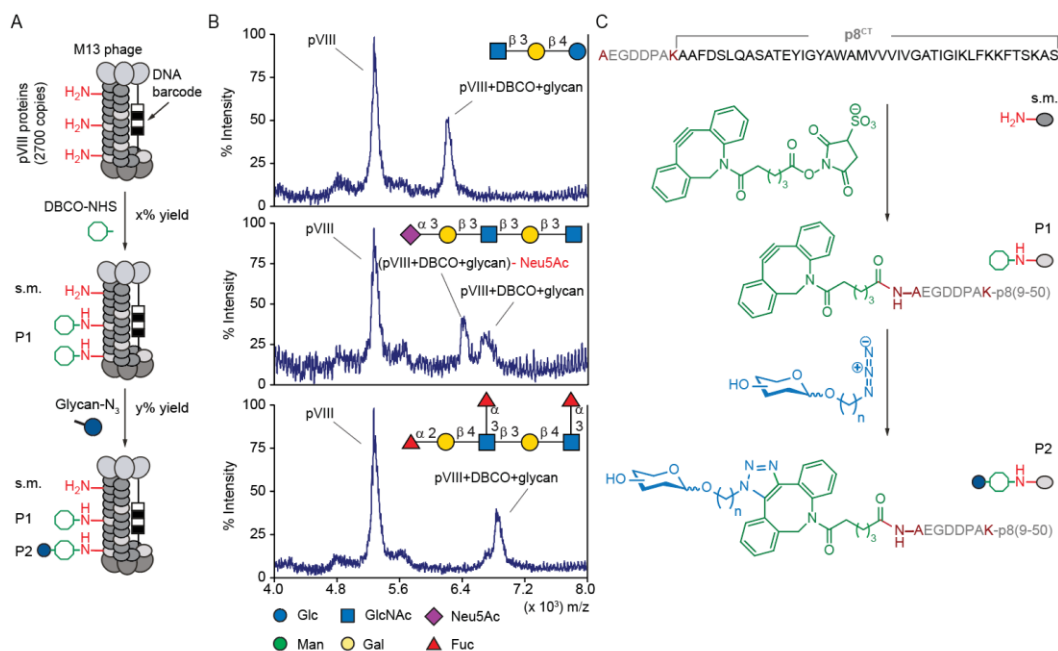
**Figure 4-2.** Chemical glycosylation of the phage surface using bi-functional tethers. (A) General steps of the procedure to incorporate the glycans via CuAAC. (B) Glycosylation of phages using copper-free (SPAAC) chemistry. (C) Phage titers as total counts of plaque-forming units (PFU) from chemical glycosylation experiments. (D) Phage viability results represented as percentage of remaining infectivity observed after first and second coupling steps amine and alkyne respectively, relative to the unmodified phage mixture (input).





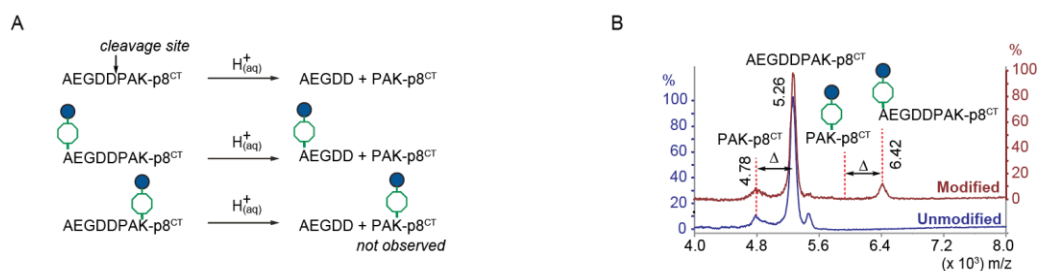
**Figure 4-3.** Chemical ligation of azido-glycans onto pVIII protein of M13 phage. (A) Two-step ligation reaction produced glycosylated pVIII product (P2), partially modified intermediate (P1) or unreacted pVIII protein (s.m.). (B) Sequence of pVIII protein, modification by DBCO linker, and ligation of azido glycan.

SPAAC allowed immobilization of structurally diverse glycans with azide handle to the pVIII (Figure4-3AB). MALDI-TOF analyses confirmed the conjugation of pVIII and served as a quality control (QC) step for every produced phage-glycan construct. We verified by MALDI, that chemical glycosylation was achieved successfully for most glycan structures (Figure4-3) with one exception: secondary anomeric azido glycans yielded suboptimal convergence leaving unreactive pVIII-DBCO even after prolonged reaction times (Figure4-3B). In all other cases mono, di, tri, tetra, penta, hexa, or higher saccharides with anomeric alkyl-azido linker were successfully incorporated into phages and confirmed by MALDI (Figure4-4). The ratio of the peak intensities in MALDI allowed to estimate the densities of the glycans on phage.



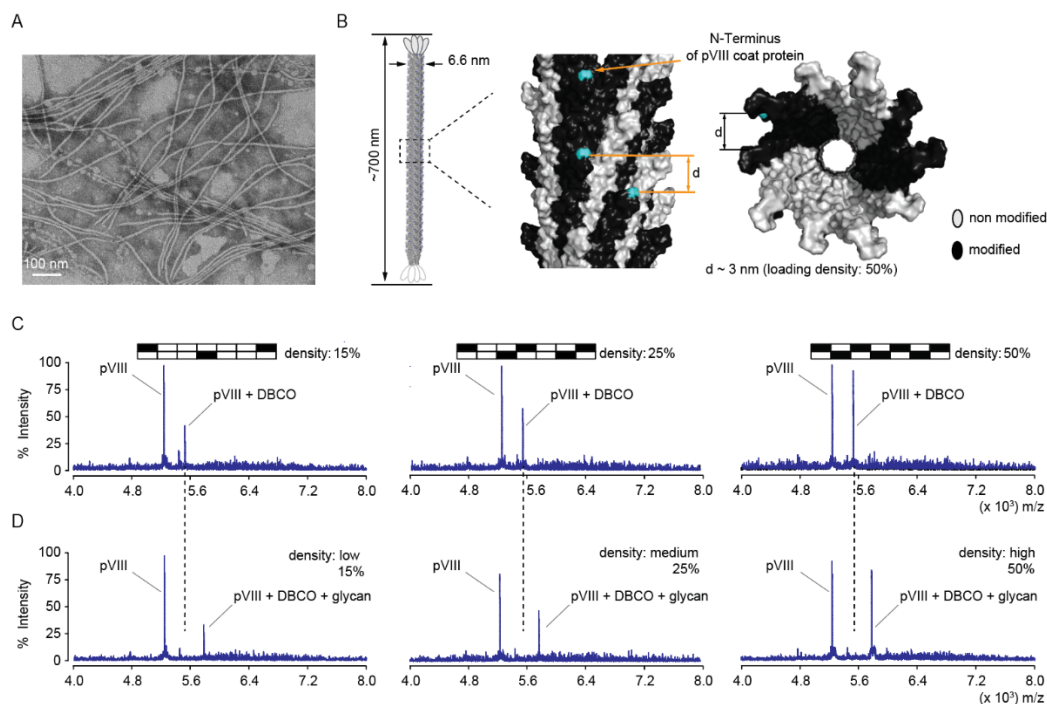
**Figure 4-4.** MALDI mass spectrometry characterization of chemical modifications on coat protein pVIII. (A) The ratio between unmodified and modified pVIII species monitored by MALDI determines the density of modification on phage. (B) Representative spectra of chemical modification of coat protein pVIII with glycans of different structural complexity. Oligosaccharides containing sialic acid (Neu5Ac) residues are known to be sensitive to spontaneous hydrolysis in acidic conditions.<sup>183</sup> MALDI of phage particles decorated with glycans containing terminal sialic acid contained two peaks (intact glycoconjugate and that with cleaved Neu5Ac).

To assess the regioselectivity of conjugation, we site specifically cleaved the Asp7-Pro8 bond in pVIII by TFA using a sinapinic acid matrix<sup>184</sup> and confirmed that the amide bond formed at N-terminal amine and not at the Lys10 of the pVIII sequence (Figure4-5A). Cleavage of the pVIII-DBCO-glycan conjugate in acidic conditions, confirmed that modification was site-specific for the N-terminus on pVIII (Figure4-5B).



**Figure 4-5.** Site-specific modification of pVIII at N-terminus. (A) Cleavage of Asp-Pro (D-P) bond in pVIII protein in acidic conditions. (B) MALDI analysis of pVIII modification products

There are 2700 copies of pVIII per phage virion,<sup>185</sup> the copy number of displayed glycan for 15, 25 and 50% correspond to 400, 700 and 1500. Considering the ordered packing with 5-fold symmetry of the N-termini of two pVIII (Figure 4-6A, B), the copy numbers and densities can be used to estimate an average spacing between glycans on the phage. Detecting lower than 1% modification density can be challenging by MALDI. Production of glycosylated particles with <30 glycans (~ 20 nm spacing) is challenging. We produced and tested conjugates with 2-50%. Controlling concentration of acylating reagent (DBCO) and reaction conditions, achieved a predictable range of densities of modifications on phage-surface. For example, the use of 0.5, 1.0 and 1.5 mM of DBCO, acylated 15, 25 and 50% of the total pVIII protein on phage as verified by MALDI (Figure 4-6C). The addition of the 2 mM azido-glycan led to quantitative consumption of strained alkyne and ligation of the glycan to pVIII coat protein, by the appearance of pVIII-DBCO-glycan peak in MALDI (Figure 4-6D) and concomitant disappearance of pVIII-DBCO signal (dashed line). Densities < 15% can be achieved by diluting DBCO or shortening reaction time. Densities of over 50% are feasible but were not produced because we noticed a decrease in long term stability of particles that contain >50% pVIII glycosylated.



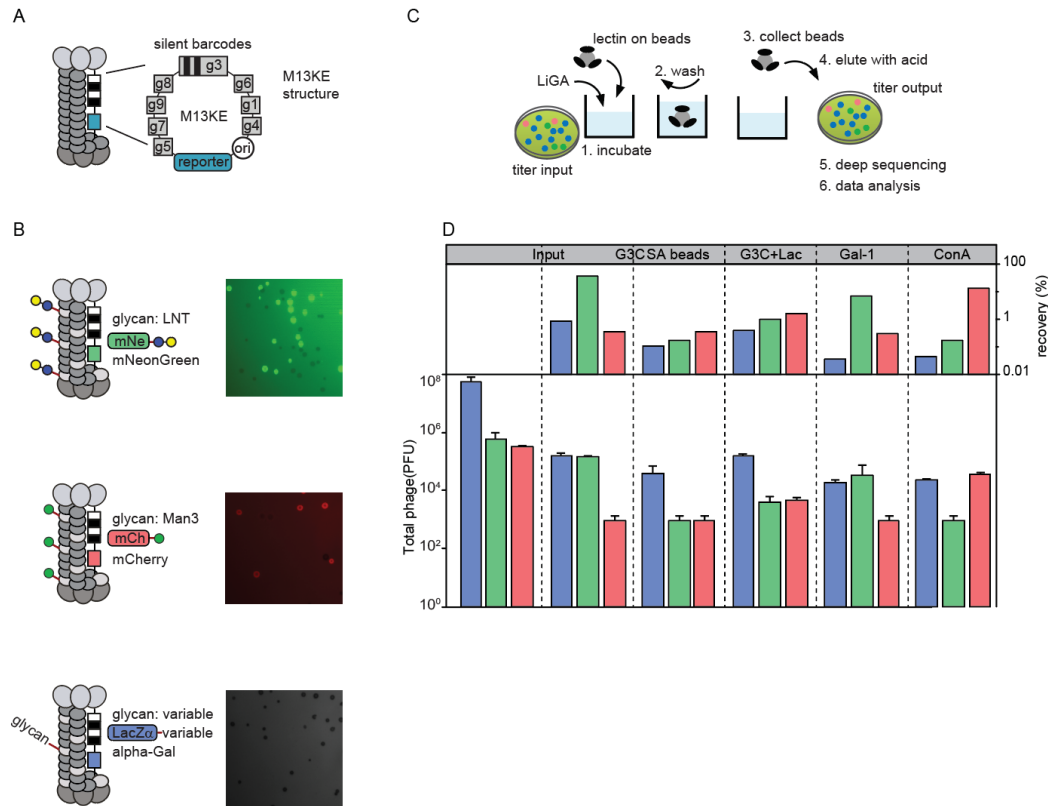
**Figure 4-6.** The loading density of glycans on the phage surface monitored by MALDI. (A) Electron microscopy image of the filamentous M13 virions. (B) The M13 is rod-shaped 1000 x 6.6 nm. On a 3D structure of the assembly of pVIII (PDBid: 2MJZ) adjacent N-terminus of a 50% modified clone are about 3 nm apart. (C) The incubation of phages with different concentration of DBCO-NHS ester, acylated the pVIII N-terminus to form pVIII-DBCO species. The ratio of pVIII : pVIII-DBCO signals allowed estimation of the density of modification from the total pVIII population. (D) Conversion of the pVIII-DBCO-glycan after the incubation phage-DBCO with azido-glycan.

#### 4.3.2. Screening LiGA against glycan binding proteins in vitro

LiGA, like any other display technique is agnostic to presentation of the target, and is compatible with targets immobilized on plates, beads, cells, or other carriers. To test binding interaction of glycan on phage with purified GBPs, we built an expanded array that contained 60 to 90 glycans at similar densities. We supplemented the array with tracer phage clones that transduce either galactosidase,<sup>147</sup> mNeonGreen<sup>186</sup> or mCherry<sup>146</sup> reporters into the *E. coli* host (Figure 4-7A,B). These tracer clones, when decorated with specific glycans allowed to track the performance of these glycans during optimization of binding

assays. Plating the output of the assay in bacterial agar overlay and counting blue, green-fluorescent or red-fluorescent plaques alleviated the need to sequence the sample during optimization (Figure 4-7C). For example, mCherry tracer was modified with the tri-mannoside  $\text{Man}\alpha\text{1-6}(\text{Man}\alpha\text{1-3})\text{Man}\alpha\text{-S6}$  (referred as Man3), and mNeonGreen tracer was decorated with  $\text{Gal}\beta\text{1-3GlcNAc}\beta\text{1-3Gal}\beta\text{1-4GlcNAc}\beta$  (referred as LNT). Panning of LiGA tracers on ConA coated beads preferentially enriched Man3 red tracer phage whereas the panning on galectin-1 and galectin-3 targets, enriched LNT-green tracer. The incubation of G3C coated beads with 60 mM Lactose abrogated the enrichment of LNT-green tracer (Figure 4-7D). Lectin-free streptavidin beads enriched neither of the tracer clones.

The extraction and deep sequencing of the phage DNA, from the beads coated with GBPs or lectin-free beads permitted to identification of target-specific interactions between the components of the glycan array and GBPs.



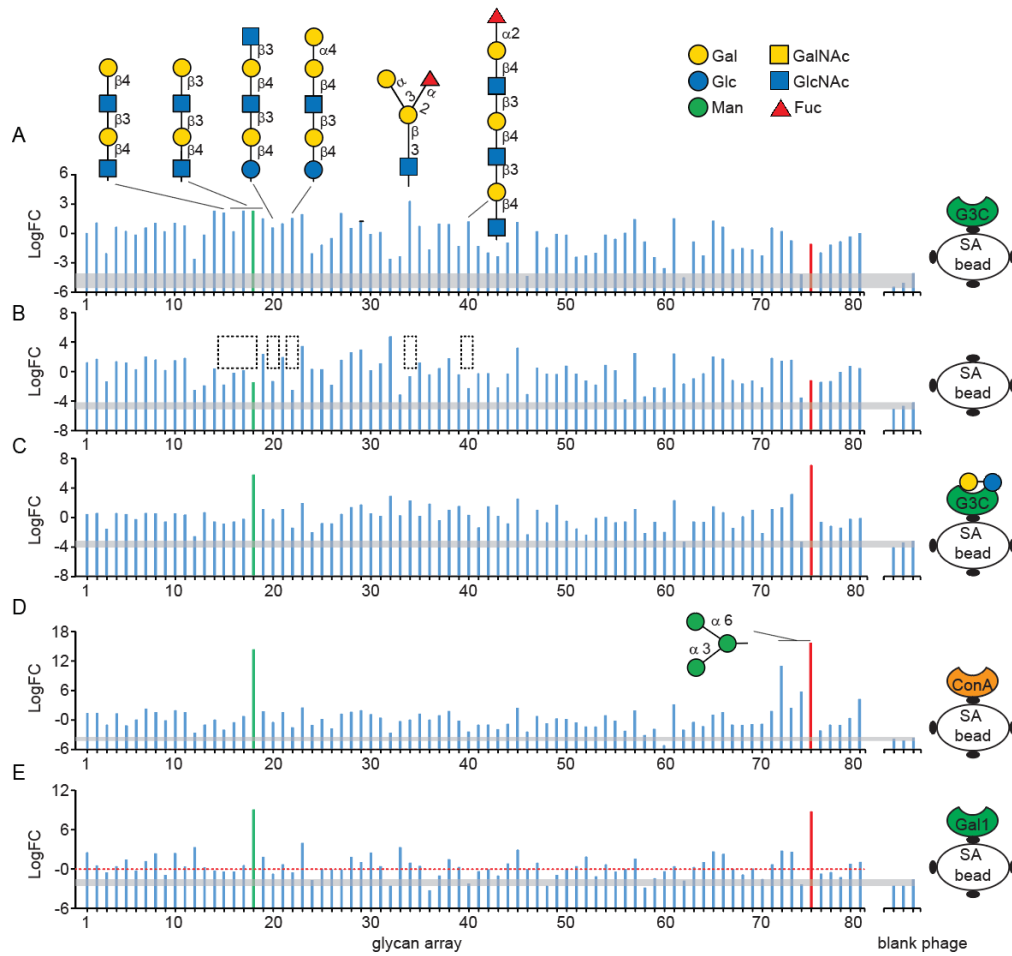
**Figure 4-7.** Screening LiGA against carbohydrate binding proteins. (A) Location of silent DNA barcode and colorimetric reporter in the phage. (B) Colorimetric control phages visualized by plaque forming assays. NeonGreen reporter was decorated with LNT, mCherry reporter was conjugated with mannose, and LacZ $\alpha$  reporter contained LiGA. (C) Scheme of panning experiments. (D) Phage counts from selection experiments of LiGA spiked with color-coded control ligands for Galectins and ConA.

Using edgeR<sup>187</sup> statistical software package we performed differential enrichment (DE) analysis of selections on G3C, G3C+Lac, Gal1 and ConA data sets, using the Input and SA-beads as controls. We employed the quasi-likelihood Fisher exact test designed to deploy exact statistical methods for multigroup experiments.<sup>187-188</sup> Prior to DE the reads of each identified glycan were converted to counts-per million (CPM) units by trimmed mean of M-values (TMM) normalization method<sup>189</sup> to minimize differences between experiment groups due to sequencing depth.

To obtain glycan binding profile for G3C we contrasted the data from G3C coated beads with the input, using lectin-free SA-beads as control. Additionally, LiGA contained four barcoded blank clones present at four different concentrations, serving as negative controls in every selection experiment represented in the set of bars separated to the left of each panel (Figure 4-8). We visualized the magnitude of DE changes as a log<sub>2</sub>-fold-change (Log<sub>2</sub>FC) for each glycan on LiGA identified with numbers 1 to 80; a list with the corresponding glycan structures present in LiGA is provided on Appendix C2. The height of the bars on the Log<sub>2</sub> FC indicates the enrichment of a glycan in the designated target relative to the input.

On G3C coated wells we identified N-acetyllactosamine (LN) motifs Type 1: (Galβ1-3GlcNacβ) and Type 2 (Galβ1-4GlcNacβ) preferentially enriched on G3C and not on empty beads specifically, tetra-saccharides Galβ1-4GlcNacβ1-3Galβ1-4Glc (LNT) and Galβ1-3GlcNacβ1-3Galβ1-4Glc (LNnT) were identified with > 4-fold enrichment (p < 0.05) on G3C coated wells.

We also observed preferential binding to G3C, in glycans with α1-2-fucosylation of terminal LN Type 2. Additionally, we found blood group B antigen Galβ1-3(Fucα1-2)Galβ1-3GlcNacβ1 specifically enriched on G3C selection. Similar to our observations LN type 1 and 2 motif had been reported on binding profile studies for galectin-3 using glycan arrays immobilized on glass surface.<sup>190-192</sup>



**Figure 4-8.** Glycan binding profiles of different carbohydrate binding proteins (glycan structures 1 to 80 are provided in Appendix C2). In each plot the Log<sub>2</sub>FC represents the enrichment of glycans compared to the input. (A) G3C. (B) Empty beads (SA-beads). (C) G3C blocked with lactose. (D) ConA and (E) Gal-1 The glycans decorated on phages expressing fluorescent proteins are highlighted in green (LNT-green) and red (Man3-red).

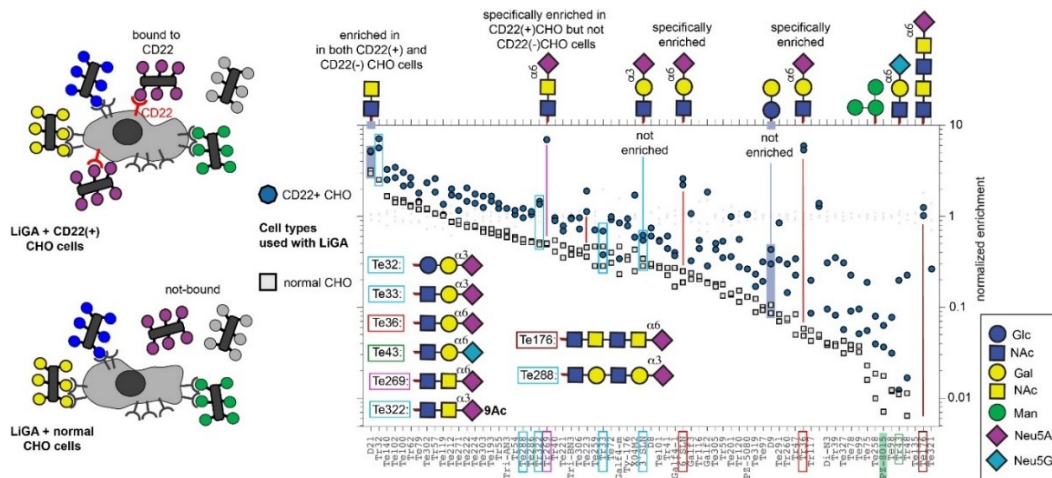
To test for robustness of the carbohydrates binding profiles on G3C using LiGA we, encoded the glycan LNT redundantly, into four phage clones, with different DNA barcodes including the previously described fluorescent tracer LNT-Green. From the sequencing we observed higher recovery of the green fluorescent tracer compared to the same glycan encoded on phage that do not transduce fluorescent signal. Thus, LNT-Green clone was highly enriched in selections against beads coated with G3C blocked with lactose (Figure 4-8C),



ConA (Figure 4-8D), and Gal-1(Figure 4-8E). Similarly, the fluorescent tracer Man3-red was also decorated into three additional non-fluorescent(blue) clones and showed >10-fold higher recovery compared to Man3 glycan on blue clones. Promiscuous recovery of colored tracers was not observed during panning experiments when the titer information showed strong target-specific recovery on both LNT-green and Man3-red clones (Figure4-7D). This observations led to two conclusions (1) introducing multiple clones of the same glycan structure on the LiGA array could help improve robustness of the screenings, and (2) the presence of fluorescent tag on the phage clone, skews the levels of recovery identified by DNA sequencing *via* presently unknown mechanism.

#### **4.3.3. LiGA platform on cell-based and *in vivo* studies**

To demonstrate the advantage of the LiGA methodology compared to canonical glass and bead arrays we tested the platform with target repertoire beyond purified proteins *in vitro*. We envisioned that similarly to phage-displayed peptide libraries, LiGA can be used to identify binders to cells, organs,<sup>193</sup> in animals,<sup>194</sup> and potentially even in humans.<sup>195-197</sup>

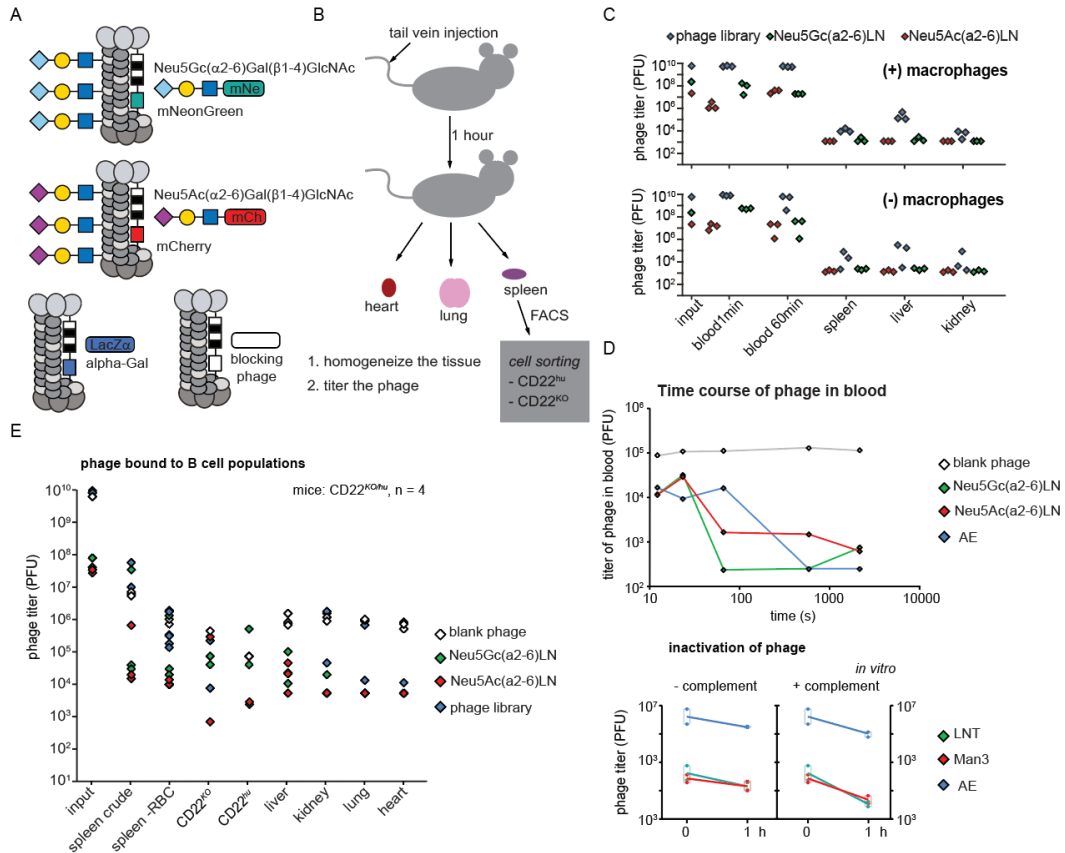


**Figure 4-9.** Analysis of glycan-binding preferences of CD22 expressed on the surface of CHO cells using LiGA. Enrichment was determined as a ratio of the fraction of glycan in LiGA before and after pulling down with cells.

We first validated LiGA in cell-based assays with chinese hamster ovary (CHO) cell lines stably transfected with the sialic acid binding lectin Siglec-2 (CD22). In separate experiments, CHO-CD22(+) and background CHO-CD22(-) cells were mixed with LiGA, incubated and washed by centrifugation. Cell pellet was boiled to release the DNA of the phage, which was PCR-amplified and sequenced (Figure4-10). Enrichment was determined as a ratio of the fraction of glycan in LiGA before and after pulling down with cells. We observed that  $\alpha$ 2-6 linked sialic acid glycans Tr36, Tr43, Tr269 and Te176 displayed specific enrichment on CD22(+) but not in CD22(-) cells whereas, closely related glycans and substitutional isomers with  $\alpha$ 2-3 sialic acid Tr32, Tr33, Tr322 and Te288 did not show significant enrichment on CD22(+) cells. These observations confirmed known binding preferences of CD22: glycans that display  $\alpha$ 2-6 linked sialic acid.

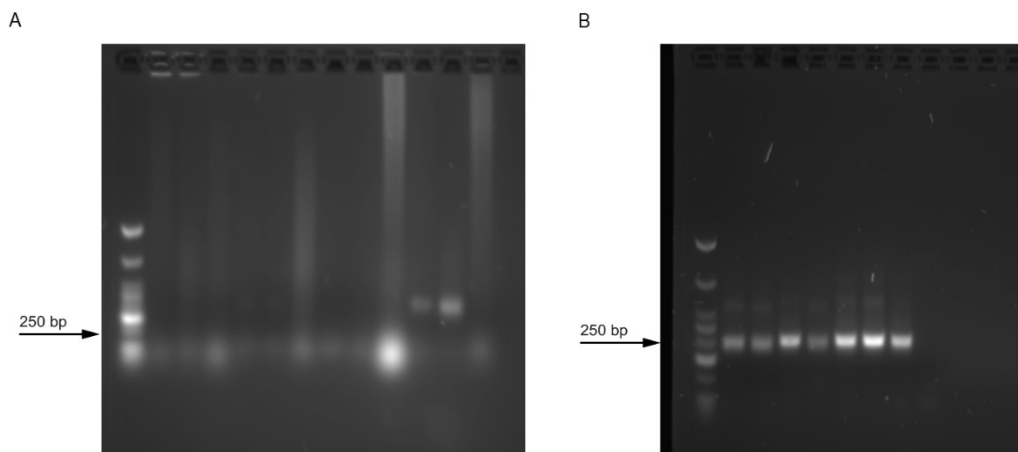
Encouraged by cell binding *ex vivo*, we performed an *in vivo* panning, which starts from injection of LiGA into the tail vein of (CD22<sup>hu/KO</sup>) transgenic

mice or wild-type, followed by harvesting of organs and specific immune cells 1h after injection (Figure 4-10A,B). Extraction of phage DNA from organs, and its sequencing requires calibration. We instead, used plaque forming assay to monitor green and red tracer clones conjugated to 2,6-Neu5Gc-LacNAc and 2,6-Neu5Ac-LacNAc. LacZ transducing phage was either an unresolvable mixture of unmodified phages, or phage modified with DBCO and capped with azido-ethanol (AE). We studied the binding of sialic acid decorated phages to Siglec-1 *in vivo*. Siglec-1 is highly expressed on macrophages but in our conditions, we observed no difference in the distribution of phages on animals with and without macrophage depletion (Figure 4-10C). Measurements of the clearance from blood, showed that glycosylated phage get eliminated at higher rate when compared to unmodified phages. Additionally, heat-inactivation of the complement system increased the recovery of phage from serum after one-hour incubation (Figure 4-10D). The blood clearance of phage particles mechanism appears to be mediated by complement but the data collected was insufficient to prove this hypothesis. Injecting LiGA into transgenic chimera mice expressing human CD22<sup>hu</sup> or CD22<sup>KO</sup> followed by cell sorting experiment allowed to track the population of phages binding to B cells expressing CD22<sup>hu</sup>. Although inconclusive, preliminary results showed a preferential recovery of Neu5Gc on CD22<sup>hu</sup> cells (Figure 4-10E) Albeit preliminary, this data demonstrated the feasibility of *in vivo* mapping glycan-receptor interactions using LiGA technology.



**Figure 4- 10.** Injection of model LiGA into mouse (B). Isolation of organs and (C) Effect of macrophages depletion on phage biodistribution. (D) Time course clearance of phages from blood. (E) Biodistribution of phage on transgenic mice.

We deep sequenced DNA material extracted from each organ and cell population. While we observed encouraging preliminary results in some samples, the extraction protocol of ssDNA from intact organ was not robust (Figure 4-11A): simple boiling of the sample, known to work with phage *in vitro* and phage bound to cells *ex vivo*; was poorly reproducible with intact organs *in vivo*. We subsequently shown that altered ssDNA extraction procedure can dramatically improve the recovery of DNA and reproducibility of sequencing (Figure 4-11B).



**Figure 4-11.** PCR amplification of phage DNA extracted from organs CD22<sup>ko</sup>/CD22<sup>hu</sup> chimera mouse. (A) Extraction by boiling the cell suspension at 90°C. (B) Extraction by alkaline hydrolysis.

To date we have tested binding of LiGA to organs, bacteria, and growing sheets of ice (to identify ice binding glycans). Results of these studies will be presented elsewhere.

#### 4.4. Conclusions

In conclusion we described the implementation of a novel and flexible methodology that enabled rapid, method for measuring glycan binding preference of target proteins *in vitro*, on cells *ex vivo*, and *in vivo*. A “liquid” array of genetically encoded carbohydrates permitted the discovery of effective GBP binding fragments; these fragments can serve as the basic building blocks to fuel future development of GE-FBD strategies.

Here we show that for general ligand design efforts that utilize phage displayed GE-FBD approaches, preliminary LiGA-based fragment profiling study, can identify promising structural motifs that are active in a phage display context, paving the way to accelerated success rates on GE-FBD.

We also demonstrated the versatility of the LiGA platform far beyond *in vitro* studies, with validations at cellular level and *in vivo* model studies.

LiGA offers flexibility in building arrays of custom compositions, ease of distribution, simplicity in analysis with the ability to embark on questions previously unattainable with canonical solid arrays or bead-based arrays. Additional possibilities include the use of LiGA for the unsupervised discovery of synergistic interactions between multiple glycans and GBPs.

## **4.5. Experimental procedures**

### **4.5.1. Materials and general information**

HEPES buffer contains 20 mM HEPES, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, pH=7.4. PBS buffer contains 137 mM NaCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.7 mM KCl, pH = 7.4. Solutions used for phage work were sterilized by filtration through 0.22 µm filters. The mannose-binding protein concanavalin-A (ConA) was purchased from Sigma-Aldrich (#C2010). Recombinant human Galectin-1 (Gal-1) 135 amino acids long, MW: 14715.7 Da was kindly donated by Dr. Hans C. Lin (Academia Sinica, Taipei, Taiwan). The carbohydrate recognition domain of the human Gal3 (G3C), residues 107–250 (MW: 16 327 Da) was provided by Dr. Christopher Cairo (University of Alberta, CA). The collection 78 of anomeric azido-glycans used in this study was donated by the glycan depository of the Consortium for Functional Glycomics at the Scripps Research Institute (La Jolla, CA, USA). The tether *N*-succinimidyl 4-pentynoate (NP) and the azidomethyl 1-thio-β-D-mannopyranoside (JB-1) were kindly donated by Dr D. Bundle (University of Alberta). The synthesis and characterization of the tether and JB-1

was reported by Bailey and coworkers.<sup>198</sup> MS-MALDI-TOF spectra were recorded on AB Sciex Voyager Elite MALDI, mass spectrometer equipped with MALDI-TOF pulsed nitrogen laser (337nm) (3ns pulse - up to 300  $\mu$ J/pulse) operating in Full Scan MS in either positive or negative ionization modes.

#### **4.5.2. Construction of silent distal barcoded (SDB) M13 phage library**

A library of silent double barcode-codons (SDB) in the phage genome proximal to the pIII cloning site was created using the Gibson Assembly cloning kit (NEB#E5510) purchased from New England Biolabs. The first SDB region was introduced into M13KE using PCR amplification followed by using NEBuilder HiFi DNA Assembly. Double stranded DNA from a phage based on M13KE clone containing the stuffer sequence CAGTTTACGTAGCTGCATCAGGGTGGAGGT equating to the peptide QFT\*LHQGGG was used as a template, with \* representing a stop codon. The insert fragment was PCR amplified using the primers P1 and P2 and the vector fragment was PCR amplified using primers P3 and P4:

Name Sequence (5' ->3') :

```
P1 GAGATTTTCAACGTGAAAAACTNCTNTTYGCNATHCCNCTNGTGGTACCTTTCTATTCTCA
P2 TTAAGACTCCTTATTACGCAGTA
P3 TTGCTAACATACTGCGTAATAAG
P4 TTTTTTCACGTTGAAAATCTC
P5 GTGGTACCTTCTATTCTCACTCGAGYGTNGARAARAAYGAYCARAARACNTAYCAYGCNGGNGGNGNTCGGCCGAAA
-CTGTTGAAAG
P6 CGAGTGAGAATAGAAAGGTAC
```

PCR was performed using 50 ng phage dsDNA with 1 mM dNTPs, 0.5  $\mu$ M primers, 0.5  $\mu$ L Phusion High Fidelity DNA polymerase in 1x PCR buffer (NEB #B0518S ) in a total volume of 50  $\mu$ L. The temperature cycling protocol

was performed as follows: a) 98 °C 3 min, b) 98 °C 30 s, c) 60 °C 30 s, d) 72 °C 4 min s, e) repeat b) - d) for 35 cycles, f) 72 °C 10 min, g) 4 °C hold. PCR amplified fragments were treated with restriction enzyme Dpn1(NEB #R0176S) and then gel purified. NEBuilder Hifi DNA assembly was then carried out following the manufacturer protocols by mixing 100 ng of vector, 4 ng insert, 10 µL of NEBuilder Hifi DNA assembly master mix, and deionized H<sub>2</sub>O up to a total volume of 20 µL. The resulting ligated DNA was transformed into *E.coli* K12 ER2738 and propagated overnight at 37 °C. The overnight culture was then centrifuged to separate bacteriophage from host cells. The cloning of SVEKNDQKTYHAGGG peptide was conducted as follows. The insert fragment was amplified by PCR following the described protocol using primers P5 and P2. The vector fragment was PCR amplified using primers P4 and P6. PCR fragment were processed using NEBuilder Hifi DNA assembly kit using steps described above. The resulting ligated DNA was transformed into electrocompetent cells *E.coli* SS320 (Lucigen). The resulting overnight culture was centrifuged to remove host cells and incubated with 5% PEG-8000, 0.5 M NaCl for 8 h at 4 °C, followed by 15 min centrifugation at 13000 g to concentrate released phage. PEG precipitated phage were resuspended in PBS-Glycerol 50% and stored at -20 °C. The SDB silent encoding in the SDB region produced a library of chemically identical phage with  $6.1 \times 10^3$  possible sequence combinations, and further cloning in of SVEK sequence results in  $2.1 \times 10^6$  possible sequence combinations. Combined the SDB-SVEK construct  $1.2 \times 10^{10}$  possible sequence combinations. We noted that culture of the mixture of these clones eliminated a few sequences



with grow disadvantage. A detailed analysis of eliminated clones is beyond this report but it can be performed by analysis of deep sequencing data describing SDB library. URL: <http://34.209.120.210>

### **4.5.3. SDB clone isolation and amplification**

The SDB-SVEK library described in the previous section was used to isolate each monoclonal silently encoded phage. A 10  $\mu$ L aliquot of phage was diluted and plated at a density of 100 plaques per plate. Single colonies were manually picked, and individually transferred into a clean 1.7 mL plastic tube containing 0.5 mL of PBS-Glycerol 50% and incubated at room temperature for 30 min. The tubes were then placed in 55°C heating block for 10 mins to inactivate any remaining bacterial cells. After the incubation, 20  $\mu$ L sample of each colony suspension was amplified for 4.5 h in 5 mL of LB supplemented with a 1:100 dilution of log phase *E. coli* K12 ER2738. After amplification the phage clones were collected from the culture supernatant by centrifugation at 4500 g for 10 min. Next, bacterial cell pellet and culture supernatant were processed separately. The supernatant was incubated with 5% PEG-8000, 0.5 M NaCl for 8 h at 4 °C, followed by 15 min centrifugation at 13000 g to precipitate the viral particles. The phages were resuspended into 1 mL PBS-Glycerol 50%, titered, and stored at -20 °C until further use. The bacterial cell pellet was processed for phage-DNA extraction using GeneJET Plasmid Miniprep kit (ThermoFisher, #K0502). For SDB identification, a sample of 400 ng of the phage DNA was submitted for Sanger sequencing at the Molecular Biology Service Unit (University of Alberta). We selected the clones who contained three base pair

substitutions from one another (i.e., Hamming distance (H)=3). H=3 permits correction of any point mutations that may have arisen during the analysis by deep sequencing.

#### **4.5.4. Construction of phage that transduce mCherry and mNeonGreen fluorescent proteins**

We produced filamentous phage vector that contains the gene for fluorescent protein mCherry and mNeonGreen cloned in place of the lacZ $\alpha$  fragment. Vectors containing mCherry and mNeonGreen genes were generous gifts from Dr. Robert Campbell (University of Alberta). Both constructs were built by HiFi NEB builder ligation of fluorescent protein (FP) fragment and M13 fragment. Vector pBAD-mCherry was used as the source for the mCherry insert, whereas Vector pBAD-mNeonGreen was used as a source for the mNeonGreen insert. FP fragments were PCR amplified following the protocol described on section 4.4.2 using primers P7 and P8, whereas the M13 fragment was PCR amplified from SDB vector using primers P9 and P10.

```
Name:      Sequence (5' ->3')
P7      GCGGATAACAATTTACACAGGAAACAGCTATGGTGAGCAAGGGCGAG
P8      TTTAAATTTTTGTTAAATCAGCTCATTTTTACTTGTACAGCTCGTCCA
P9      AAAATGAGCTGATTTAACAAAAATTTAA
P10     AGCTGTTTCCTGTGTGAAAT
```

Vector with SDB sequence CTTCTATTTGCTATTCCTCTA was used to produce M13 fragment for the mCherry construct, and vector with the SDB sequence CTTCTATTTGCTATTCCTCTA was used to produce M13 fragment for the mNeonGreen construct.

#### 4.5.5. Monitoring phage viability under CuAAC reaction

A tube containing 150  $\mu\text{L}$  of  $3.4 \times 10^{13}$  PFU/mL of phage clone SBD3 in PBS pH 7.4 was concentrated using 0.5 mL Amicon filter (Sigma, # UFC5010) with 10 kDa MW cut off, and resuspended in Tris-borate 200 mM pH 8.0, to a final volume of 150  $\mu\text{L}$ . The phage solution was now tittered by plaque forming assay and then divided into triplicates containing 50  $\mu\text{L}$  of phage solution. Each aliquot was reacted with NP (1mM final concentration) at room temperature 1 h. The reaction mix was then passed through Zeba Spin Desalting Columns 40K (Thermo-Fisher, #87766) to eliminate unreacted linker following these steps: first the column was washed with 0.5 mL PBS, then, the phage mix was loaded onto the resin and centrifuged at  $1000\times g$  for 1 min. After filtration, 10  $\mu\text{L}$  sample of phage mix was used for analysis by MALDI-TOF spectrometry and plaque forming assay. The CuAAC reaction was implemented following literature procedures<sup>199</sup> with minor modifications. Bathophenanthroline/ $\text{Cu}^{1+}$  catalyst was prepared as follows: glass vial (4 mL) was charged with 10 mg of  $\text{Cu}_2\text{SO}_4 \cdot 5\text{H}_2\text{O}$  and bathophenanthroline sulfonate (64.4 mg) (GFS Chemicals Inc.) dissolved in 0.2 M Tris HCl, pH 8.0 buffer (1 mL). Copper powder ( $\sim 50$  mg) was added; the vial was closed with a rubber septa and purged with argon. The vial was rotated for 2 h; the reduction of copper II to copper I by metallic copper was indicated by the appearance of a dark green color. To the filtered phage solution was added azido-glycan JB-1 up to a final concentration of 2 mM and  $\sim 1$ mg of copper powder. The phage solution was now degassed by 30 min sonication in argon atmosphere. The glycosylation reaction was then initiated by adding 3  $\mu\text{L}$  of

Bathophenanthroline/ $\text{Cu}^{1+}$  catalyst. The tube was closed under argon atmosphere and incubated 8h at 4 °C. The mixture was then passed through Zeba Spin column as described above. The viability of the recovered phages was analyzed by plaque forming assay and the chemical modification was analyzed by MALDI-TOF.

#### **4.5.6. Monitoring phage viability under SPAAC reaction**

A solution of 150  $\mu\text{L}$  of  $3.4 \times 10^{13}$  PFU/mL of phage clone SBD3 in PBS pH 7.4 was distributed by 50  $\mu\text{L}$  aliquots into three 1.7 mL plastic tubes. Each sample was reacted with DBCO linker (Sigma, #761524) at room temperature for 1 h at 1mM final concentration of linker. The reaction mix was then passed through Zeba Spin column to eliminate unreacted linker. After filtration 10  $\mu\text{L}$  sample mix was analyzed by MALDI-TOF spectrometry and plaque forming assay. The remaining phages were incubated 8 h at 4 °C with azido-glycan JB-1(2 mM final concentration). The mix was then filtered through Zeba Spin column as described above. The viability of the recovered phages was analyzed by plaque forming assay and the chemical modification analyzed by MALDI-TOF.

#### **4.5.7. Analysis of glycosylation of phage samples by MALDI-TOF MS**

Matrix-assisted laser desorption/ionization (MALDI) ionization was implemented in a sinapinic acid matrix.<sup>184</sup> The matrix was formed by deposition of two layers. Layer 1 was prepared as a 10mg/mL solution of sinapinic acid (Sigma, #D7927) in acetone-methanol (4:1). Layer 2 was prepared as 10 mg/mL solution of sinapinic acid in acetonitrile:water (1:1) with 0.1% TFA. The sample preparation was as follows: 2  $\mu\text{L}$  of phage solution was added to 4  $\mu\text{L}$  of layer2,

then a mixture of 1:1 layer1-layer2+phage was deposited on the MALDI inlet plate for measurements. To plot the spectrum and compute the percentages of pVIII modification we utilized a MATLAB script named plotONEmaldi.m available as Appendix C79.

#### **4.5.8. Quantification of phage-surface chemical modifications.**

To study the loading density of chemical modifications on the phage surface we utilized as starting materials: the phage clone SDB3, the linker DBCO and the azido-glycan JB-1 as a model system. A tube containing 180  $\mu$ L solution of phage clone SDB3 ( $\sim 10^{13}$  PFU/mL in PBS, pH 7.4) was distributed into 3 separated reaction tubes. The linker DBCO was then added into the reaction tubes to get the reactions at either 0.5 or 1.0 or 1.5 mM of DBCO. All the experiments were performed in triplicates. The amine and alkyne coupling reactions were conducted following the steps and conditions described in section 4.4.6. The reaction vessels were sampled after purifications with Zeba columns. Yields of pVIII-DBCO and pVIII-DBCO-glycan achieved under each experimental condition were measured by MALDI-TOF and computed using plotONEmaldi.m script.

#### **4.5.9. Regioselectivity experiments on pVIII**

Two tubes were prepared with 30  $\mu$ L of SDB3 phage clone ( $\sim 10^{13}$  PFU/mL). One tube was reacted with 1mM DBCO for 1 hr according to conditions described on section 4.4.6. After Zeba filtration the azido-glycan GalF4 (GalF $\beta$ 1-5GalF $\beta$ 1-5GalF $\beta$ 1-5GalF $\beta$ -S9) was added and the reaction mix was incubated for 8 h at 4  $^{\circ}$ C to allow azide-alkyne cycloaddition to occur. The sample

was then Zeba purified to remove unreacted glycan and the filtrate was then incubated in 70% (v/v) trifluoroacetic acid (TFA) for 1hr at RT. In parallel, a second tube containing 30  $\mu$ L of ( $\sim 10^{13}$  PFU/mL) unmodified SDB3 phage clone was incubated in 70% TFA 1hr at RT, as a control sample. After TFA incubation, both samples were analyzed by MALDI-TOF for the presence of modified and unmodified, Asp-Pro cleavage products.

#### **4.5.10. Building a phage displayed Liquid Glycan Array**

Monoclonal phage isolates carrying a single silent barcode were reacted with DCBO linker under the conditions described in section 4.4.6 and subsequently conjugated with the corresponding azido glycan following the Table 4-1. The amine coupling reaction was performed in 1mM of DCBO, to achieve  $\sim 25\%$  of pVIII modification. All chemical reactions were verified and quantified by MALDI as described on the previous sections. The MALDI traces are available as Appendix C3 to C78. After chemical glycosylation and characterization, each glycosylated SDB phage clone was then tittered by plaque forming assay and stored in 50% PBS-Glycerol at  $-20\text{ }^{\circ}\text{C}$ . The assembly of the LiGA array was performed by mixing equal quantities of all available glycan-phage conjugates into a single tube to form the master-mix. The quantity needed from each individual glycan-phage clone was estimated as  $N \times 10^6$  PFUs, where  $N$  is the number of experiments.

#### **4.5.11. Preparation of biotinylated lectins BioConA, BioG3C and BioGal-1**

For NHS-ester lysine conjugation, 1 mL of protein target G3C 10 mg/mL (similar procedure was used for ConA and Gal-1) was reacted with NHS-PEG4-Biotin following the procedure described on Chapter 3 Section 3.5.4.

#### **4.5.12. Screening LiGA against glycan binding proteins**

Streptavidin coated beads (30  $\mu$ L, Promega #Z5481) were mixed with 10  $\mu$ g, of biotinylated proteins BioConA, BioG3C and BioGal1 in 1 mL of HEPES buffer. As negative controls, 30  $\mu$ L of streptavidin beads were mixed in 1 mL of HEPES buffer into two groups. Control groups were incubated with either buffer (SA-beads), or 10  $\mu$ g of BioG3C in 60 mM lactose (G3C+Lac) per replicate. The microcentrifuge tubes were placed in a rotator overnight at + 4°C. The steps onwards were performed using KingFisher™ Duo Prime Purification System, 96 Deepwell plate (ThermoFisher, #95040450) was used to contain the solutions and 12-tip Deepwell magnetic comb was used to transfer beads (ThermoFisher, #97003500). Contents on each tube was transferred into the 96 Deepwell plate and the beads were then washed in 1mL, for 30 seconds, gentle shaking, with wash buffer (0.1% Tween-20 (v/v) in HEPES). The beads were then incubated in blocking buffer (1 mL, HEPES buffer + 2% BSA (w/v),) for 1 h at RT to reduce nonspecific binding. After blocking, the beads were transferred to HEPES buffer solution containing  $10^8$  PFU of LiGA per experiment, and incubated for 1 h and 30 min at RT. Then the beads were washed twice as described above. After the washes the beads were transferred into 1.7 mL microcentrifuge tubes containing 1mL HEPES and subsequently placed on a magnet to collect the beads on bottom

of the tube. Supernatant was then aspirated, and the beads were resuspended in 60  $\mu$ L Nuclease Free H<sub>2</sub>O. The solution was boiled at 90°C for 10 min, centrifuged at 21,000 x G for 5 min and 25  $\mu$ L of the supernatant was used to PCR for deep sequencing (see section 4.4.15).

#### **4.5.13. Screening LiGA against CHO cells expressing CD22**

For human Siglec-2 (CD22) positive cells, full length human CD22, in the pcDNA5/FRT vector, was stably transfected into Chinese Hamster Ovary (CHO) cell line through Flp-In™(ThermoFisher, #K601001) system under selection with 0.5 mg mL<sup>-1</sup> hygromycin-B (ThermoFisher) for two weeks. The cells were then passaged using 1 mM EDTA/PBS (ThermoFisher #13151014) as a dissociation solution and grown in DMEM-F12 medium (ThermoFisher, # 11330057) with 10% Fetal Bovine Serum (ThermoFisher, #26140079) and Penicillin-Streptomycin (ThermoFisher, #15140122). For CD22(-) negative control cells were untransfected CHO cells. Log phase cells were detached from culture flask using TrypLE (ThermoFisher, # 12605036) and resuspended in incubation buffer (20 mM HEPES, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, pH=7.4, with 1% BSA,) at 4x10<sup>6</sup> cells/mL and 125  $\mu$ L of cells was aliquots in to FACS tubes (Corning, #352054). LiGA solution was added to the cells (N x 10<sup>6</sup> pfu/mL, where N is the number of glycans). Cells and LiGA solution were incubated for 1 hour at + 4 °C. After incubation, 3 mL of wash buffer (20 mM HEPES, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, pH=7.4 with 0.1% BSA) was added into each tube containing cells with the squeeze bottle. Cells were harvested by centrifugation at 300 g for 5 min at +4 °C. The supernatant was discarded by inverting the tubes,



cell pellet resuspended in 1 mL of HEPES buffer, and transferred to a clean microcentrifuge tube. Cells were then spinned at 300 g for 5 min at +4 °C to collect cell pellets. Cells pellets were resuspended in 60 µL of nuclease free H<sub>2</sub>O and incubated at 98 °C for 10 min. The solution was centrifuged at 21000 g for 5 min and 25 µL of the supernatant was used as DNA template for PCR and submission to deep sequencing (see section 4.4.12)

#### **4.5.14. Panning LiGA in wild-type and CD22<sup>hu</sup> mice**

All the procedures and experiments involving animals were carried out using a protocol approved by the Health Sciences Laboratory Animal Services (HSLAS), University of Alberta. The protocol was approved as per the Canadian Council on Animal Care (CCAC) guidelines. All mice were maintained in pathogen-free conditions at the University of Alberta breeding facility. To screen the specific ligand for human CD22, we created a mixed chimeric mouse by lethally irradiating the recipient mice and then transplanting them with mixed bone marrow collected from mouse-CD22 knockout (mCD22KO) and human-CD22 transgenic (hCD22Tg) mice. Four weeks after transplantation, mice were bled to verify that the desired chimera had been established. Six weeks post transplantation; mice were injected intravenously with 0.2 mL of LiGA (1 x 10<sup>11</sup> pfu/mL). After 1hour post-injection mice were euthanized. Internal organs heart, liver, kidney, lung and spleen were collected and placed on individual petri dishes with DMEM (11965084). Each organ was homogenized using 75 mm frosted microscope slides (Thermo-Fisher, #12550143). Homogenized tissue was transferred into 5 mL of RBC\_lysis buffer (Amonium chloride 155 mM, EDTA

10mM in Tris-HCL 100 mM, pH 7.4) and incubated 1 min the filtered to remove any tissue debris. Cells were collected by centrifugation at 300 g for 5 min and resuspended in 0.4 mL PBS-BSA 0.1 %. A 2 mL sample from cell suspension was used to titer the amount of phage particles recovered from each specific organ by plaque forming assay. the cells were then resuspended in 0.3 mL of 50 mM Tris-HCl, 10 mM EDTA, pH 8.0 supplemented with 0.1mg/mL RNase-A. Then, 0.3 mL of lysis buffer (200 mM NaOH, 1% SDS) were added, with gentle mixing and incubated for 5 min at room temperature. To allow the precipitation of cellular debris and chromosomal DNA, 0.3 mL of precipitation buffer (2.55 M of potassium acetate, pH 4.8) was added. The preparation was then centrifugated at 15 000 g for 15 min at 4°C, the supernatant was loaded into a Qiagen miniprep spin column (Qiagen, # 27115). To elute DNA 0.06 mL of RNase free water was added to the center of the spin column and the column was centrifugated at 10 000 g for 1 min to collect the filtrate into a clean 1.7 mL tube. The collected DNA sample was stored at – 80 °C until further use

#### **4.5.15. Cell sorting experiments of LiGA in wild-type and CD22<sup>hu</sup> mice**

Spleen sample was crushed, and red blood cells were removed to isolates the splenocytes. Next these cells were stained with a cocktail of fluorescently labeled immune cell markers. And these labeled cells were then sorted into two populations - mCD22KO and hCD22Tg - by following the gating strategy described below. One million cells from each population were collected and further analysed using Illumina sequencing.

Gating strategy: First lymphocytes were identified using low forward scatter (FSC) and low side scatter (SSC). Then viable cells were identified by negative selection of propidium iodide. Next b-cells (CD19+ & B220+) were divided into CD45.1+(marker used in mCD22KO ) and CD45.2+ (marker used for hCD22Tg). Finally, the mCD22KO cell population were collected from CD45.1+hCD22-mCD22- zone and hCD22Tg cell population were collected from CD45.2+hCD22+mCD22-zone.

#### **4.5.16. Illumina sequencing**

To convert phage DNA to Illumina-compatible short double-stranded DNA (dsDNA) by PCR, we followed the PCR routine, and used the primers described on Chapter 2 section 2.4.3. The PCR products were then sequenced using the Illumina NextSeq service provided by 48Hour Discovery Inc. (Edmonton, AB). The resulting FASTAQ files on Illumina cloud were converted into a unique table of counts per experiment, according to procedures previously described.<sup>22, 26, 86</sup> The data was then deposited in a cloud-based server (<http://34.209.120.210>) maintained by 48Hour Discovery Inc.

#### **4.5.17. Data analysis**

Data analysis was implemented using Python 3.7.2 and R 3.5.2 programming languages. We utilized a code portfolio developed in-house, to analyze our specific set of Illumina sequencing results. The sequencing raw data was retrieved from the 48Hour Discovery Inc. cloud server. Each selection experiment was downloaded as an individual text file. The text file contained tab-separated data organized into several columns. The first column holds the names

(CFG-codes) of identified glycans while the other columns contained the raw number of Illumina counts/reads per glycan in all experimental replicates. The violin plots of *in vitro* experiments were generated using Python library for statistical data visualization Seaborn 0.9.0. Pairwise comparisons to test for DE glycans were performed using R script DE\_analysis.R (Appendix C80) according to the following pipeline: first, two individual text files corresponding to the desired comparison tuple, were consolidated into a single, comma separated value (csv) file, using the python script makeContrast.py (Appendix C81). Then the output csv file was loaded into the script DE\_analysis.R to run the comparison. The DE analysis was performed using Fisher exact test implemented as a function named *glmQLFit()* within the edgeR package invoked in the script. The steps described above were repeated as many times as there were relevant comparisons to make. The DE data was obtained as a plot that visualized the library size-adjusted log-fold change between compared libraries, against the average log-expression across those libraries. A list containing all the identified hits with its respective enrichments, and statistical significance scores was also generated.

## Chapter 5: Conclusions and outlook

### 5.1. Conclusions

This thesis describes the application of a novel technology for the discovery of ligands for inhibiting the carbohydrate binding protein galectin-3 (Gal3). Galectin family represents an important class of therapeutic targets, and the discovery of inhibitors for galectins traditionally involves rational design demanding deep structural knowledge of the binding site. Thus, the ligand development requires complex multistep chemical synthesis due to the structural complexity of glycans. Genetically-encoded fragment-based discovery (GE-FBD) could accelerate ligand discovery for glycan binding proteins by combining the strengths of genetically encoded libraries of peptides with the basic principles of fragment-based discovery.

In Chapter 2, I demonstrated that starting GE-FBD efforts from galactose fragment that has millimolar affinity for Gal3 discovered ligands that showed increased affinity to Gal3 when decorated with galactose. The underwhelming potency of each glycopeptide combination tested as soluble monovalent inhibitor marks the fundamental physical limitations of GE-FBD. It points out the inability of fragments of very weak affinity to direct selection screenings towards successful discovery of ligands in which the fragment and peptide sequences act in synergy.

In Chapter 3, I have demonstrated that a high diversity carbohydrate-peptide library in which the glycan fragment is a monovalent binder of Gal3, could be useful source for the discovery of monovalent ligands for Gal3.

Additionally, this study reinforced the advantages of approaches that exploit soluble/monovalent targets rather than immobilized “pseudo-multivalent” targets. Summarizing Chapter 2 and Chapter 3 observations, it's tempting to propose that reproducibility and success rate of discovery in GE-FBD scales with the binding affinity of the initial fragment.

In Chapter 4, I describe a “liquid” array technology, of genetically encoded carbohydrates that permitted the discovery of effective GBP binding fragments; these fragments can serve as the basic building blocks to fuel future development of GE-FBD strategies. For general ligand hunt efforts that utilize phage displayed GE-FBD approaches, preliminary profiling studies with arrays of initial fragments, can identify promising structural motifs that are active in phage display context, paving the way to accelerated success rates on GE-FBD. The described array technology offers flexibility of custom compositions, ease of distribution, simplicity in analysis with the ability to embark on molecular recognition questions beyond *in vitro* level, previously unattainable with canonical solid arrays or bead-based arrays.

## **5.2. Future directions**

As the promise of GE-FBD becomes progressively fulfilled, it is convenient to point the way to the future of an emerging technology, where possibilities are vast.

In Chapter 2, I demonstrated that silent encoding technology can be used to investigate specific parameter on GE-FBD, *e.g.* the initial binding affinity of the constant glycan fragment, however further characterization of the properties

of the covalent linkage between glycan and peptide fragments on GE-FBD has yet to be investigated. I envision that silent encoding technology could be employed to interrogate custom libraries of constant fragments with different anchoring/scars structures connecting fragment to peptide libraries. The GE-GBD campaign of Chapter 3 was relatively successful however further characterization of the binding potency of discovered ligands is still needed. To validate the functionality of the ligands cell-based assays of inhibition of Gal3 binding to cell surface can be performed on Jurkat cells. Likewise, the characterization of Gal3 binding selectivity of discovered ligands, needs to be addressed.

The glycan array technology described in Chapter 4 is a versatile emerging technology that requires more examples to demonstrate its general application. Currently, our lab has initiated several projects embarking on “liquid arrays” ventures such as the screening of arrays of Siglec lectins on cancer cells, and the use of glycan arrays to profile anti-glycan antibodies on human serum. In this field scar-less conjugation chemistries need to be explored. Efficient quantification of the screening results after elution of phage particles remains challenging and correlation between results from DNA sequencing and plaque forming assays is not always clear. Novel quantification strategies based on TaqMan™ multiplex qPCR can be implemented to improve and validate the robustness of selection screenings. I envision that, similar to initial fragment optimization, the liquid array technology will allow the further exploration of other open questions on GE-FBD like those related to the impact of employing a mixture of glycan fragments in selection screenings

## References

- (1) Bussel, J. B.; Kuter, D. J.; George, J. N.; McMillan, R.; Aledort, L. M.; Conklin, G. T.; Lichtin, A. E.; Lyons, R. M.; Nieva, J.; Wasser, J. S.; Wiznitzer, I.; Kelly, R.; Chen, C.-F.; Nichol, J. L., AMG 531, a Thrombopoiesis-Stimulating Protein, for Chronic ITP. *N. Engl. J. Med.* **2006**, *355* (16), 1672-1681.
- (2) Peng, S.-B.; Zhang, X.; Paul, D.; Kays, L. M.; Gough, W.; Stewart, J.; Uhlik, M. T.; Chen, Q.; Hui, Y.-H.; Zamek-Gliszczynski, M. J.; Wijnsman, J. A.; Credille, K. M.; Yan, L. Z., Identification of LY2510924, a Novel Cyclic Peptide CXCR4 Antagonist That Exhibits Antitumor Activities in Solid Tumor and Breast Cancer Metastatic Models. *Mol. Cancer Ther.* **2015**, *14* (2), 480-490.
- (3) Howell, S. M.; Fiacco, S. V.; Takahashi, T. T.; Jalali-Yazdi, F.; Millward, S. W.; Hu, B.; Wang, P.; Roberts, R. W., Serum Stable Natural Peptides Designed by mRNA Display. *Sci. Rep.* **2014**, *4*, 6008.
- (4) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G., Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009**, *5*, 502.
- (5) Ranganath, S.; Bhandari, A.; Avitahl-Curtis, N.; McMahon, J.; Wachtel, D.; Zhang, J.; Leitheiser, C.; Bernier, S. G.; Liu, G.; Tran, T. T.; Celino, H.; Tobin, J.; Jung, J.; Zhao, H.; Glen, K. E.; Graul, C.; Griffin, A.; Schairer, W. C.; Higgins, C.; Reza, T. L.; Mowe, E.; Rivers, S.; Scott, S.; Monreal, A.; Shea, C.; Bourne, G.; Coons, C.; Smith, A.; Tang, K.; Mandyam, R. A.; Masferrer, J.; Liu, D.; Patel, D. V.; Fretzen, A.; Murphy, C. A.; Milne, G. T.; Smythe, M. L.; Carlson, K. E., Discovery and Characterization of a Potent Interleukin-6 Binding Peptide with Neutralizing Activity In Vivo. *PLoS One* **2015**, *10* (11), e0141330.
- (6) Ishizawa, T.; Kawakami, T.; Reid, P. C.; Murakami, H., TRAP Display: A High-Speed Selection Method for the Generation of Functional Polypeptides. *J. Am. Chem. Soc.* **2013**, *135* (14), 5433-5440.
- (7) Guillen Schlippe, Y. V.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W., In Vitro Selection of Highly Modified Cyclic Peptides That Act as Tight Binding Inhibitors. *J. Am. Chem. Soc.* **2012**, *134* (25), 10469-10477.
- (8) Jencks, W. P., On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (7), 4046-4050.
- (9) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274* (5292), 1531-1534.
- (10) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H., Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discovery* **2016**, *15*, 605.
- (11) Bohacek, R. S.; McMartin, C.; Guida, W. C., The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **1996**, *16* (1), 3-50.



- (12) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L., Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52* (11), 2864-2875.
- (13) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R., DNA-Templated Organic Synthesis and Selection of a Library of Macrocycles. *Science* **2004**, *305* (5690), 1601-1605.
- (14) Halpin, D. R.; Harbury, P. B., DNA Display II. Genetic Manipulation of Combinatorial Chemistry Libraries for Small-Molecule Evolution. *PLoS Biol.* **2004**, *2* (7), e174.
- (15) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D., Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **2004**, *22* (5), 568-574.
- (16) Brenner, S.; Lerner, R. A., Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (12), 5381-5383.
- (17) Favalli, N.; Bassi, G.; Scheuermann, J.; Neri, D., DNA-encoded chemical libraries - achievements and remaining challenges. *FEBS Lett.* **2018**, *592* (12), 2168-2180.
- (18) Ahn, S.; Kahsai, A. W.; Pani, B.; Wang, Q.-T.; Zhao, S.; Wall, A. L.; Strachan, R. T.; Staus, D. P.; Wingler, L. M.; Sun, L. D.; Sinnaeve, J.; Choi, M.; Cho, T.; Xu, T. T.; Hansen, G. M.; Burnett, M. B.; Lamerdin, J. E.; Bassoni, D. L.; Gavino, B. J.; Husemoen, G.; Olsen, E. K.; Franch, T.; Costanzi, S.; Chen, X.; Lefkowitz, R. J., Allosteric "beta-blocker" isolated from a DNA-encoded small molecule library. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (7), 1708-1713.
- (19) Goodnow Jr, R. A.; Dumelin, C. E.; Keefe, A. D., DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discovery* **2016**, *16*, 131.
- (20) Ng, S.; Lin, E.; Kitov, P. I.; Tjhung, K. F.; Gerlits, O. O.; Deng, L.; Kasper, B.; Sood, A.; Paschal, B. M.; Zhang, P.; Ling, C.-C.; Klassen, J. S.; Noren, C. J.; Mahal, L. K.; Woods, R. J.; Coates, L.; Derda, R., Genetically Encoded Fragment-Based Discovery of Glycopeptide Ligands for Carbohydrate-Binding Proteins. *J. Am. Chem. Soc.* **2015**, *137* (16), 5248-5251.
- (21) Ven Chang, I.; Tsutsumi, H.; Mihara, H., Screening for concanavalin A binders from a mannose-modified  $\alpha$ -helix peptide phage library. *Mol. Biosyst.* **2017**, *13* (11), 2222-2225.
- (22) Tjhung, K. F.; Kitov, P. I.; Ng, S.; Kitova, E. N.; Deng, L.; Klassen, J. S.; Derda, R., Silent Encoding of Chemical Post-Translational Modifications in Phage-Displayed Libraries. *J. Am. Chem. Soc.* **2016**, *138* (1), 32-35.
- (23) Arai, K.; Tsutsumi, H.; Mihara, H., A monosaccharide-modified peptide phage library for screening of ligands to carbohydrate-binding proteins. *Bioorg. Med. Chem. Lett.* **2013**, *23* (17), 4940-4943.
- (24) Li, S.; Roberts, R. W., A Novel Strategy for In Vitro Selection of Peptide-Drug Conjugates. *Chem. Biol.* **2003**, *10* (3), 233-239.

- (25) Ng, S.; Bennett, N.; Gao, N.; Drickamer, K.; Derda, R., Genetically-encoded Fragment-based Discovery of Glycopeptide Ligands for DC-SIGN. *Bioorg. Med. Chem. Lett.* **2018**, *26* (19), 5368-5377.
- (26) Chou, Y.; Kitova, E. N.; Joe, M.; Brunton, R.; Lowary, T. L.; Klassen, J. S.; Derda, R., Genetically-encoded fragment-based discovery (GE-FBD) of glycopeptide ligands with differential selectivity for antibodies related to mycobacterial infections. *Org. Biomol. Chem.* **2018**, *16* (2), 223-227.
- (27) Hang, H. C.; Yu, C.; Kato, D. L.; Bertozzi, C. R., A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (25), 14846-14851.
- (28) Devaraj, N. K., The Future of Bioorthogonal Chemistry. *ACS Central Science* **2018**, *4* (8), 952-959.
- (29) Bertozzi, C. R., A Decade of Bioorthogonal Chemistry. *Acc. Chem. Res.* **2011**, *44* (9), 651-653.
- (30) Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40* (11), 2004-2021.
- (31) Gutmiedl, K.; Wirges, C. T.; Ehmke, V.; Carell, T., Copper-Free “Click” Modification of DNA via Nitrile Oxide–Norbornene 1,3-Dipolar Cycloaddition. *Org. Lett.* **2009**, *11* (11), 2405-2408.
- (32) Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G.-J., Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angew. Chem., Int. Ed.* **2008**, *47* (12), 2253-2255.
- (33) Blackman, M. L.; Royzen, M.; Fox, J. M., Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130* (41), 13518-13519.
- (34) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046-15047.
- (35) Saxon, E.; Bertozzi, C. R., Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287* (5460), 2007-2010.
- (36) Dürr, C.; Nothhaft, H.; Lizak, C.; Glockshuber, R.; Aebi, M., The Escherichia coli glycoprotein display system. *Glycobiology* **2010**, *20* (11), 1366-1372.
- (37) Denton, K. E.; Wang, S.; Gignac, M. C.; Milosevich, N.; Hof, F.; Dykhuizen, E. C.; Krusemark, C. J., Robustness of In Vitro Selection Assays of DNA-Encoded Peptidomimetic Ligands to CBX7 and CBX8. *SLAS Discovery* **2018**, *23* (5), 417-428.
- (38) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S., Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776-779.
- (39) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T., Staudinger Ligation: A Peptide from a Thioester and Azide. *Org. Lett.* **2000**, *2* (13), 1939-1941.

- (40) Jongkees, S. A. K.; Umemoto, S.; Suga, H., Linker-free incorporation of carbohydrates into in vitro displayed macrocyclic peptides. *Chem. Sci.* **2017**, *8* (2), 1474-1481.
- (41) Ichihara, O.; Barker, J.; Law, R. J.; Whittaker, M., Compound Design by Fragment-Linking. *Mol. Inform.* **2011**, *30* (4), 298-306.
- (42) Ng, S.; Jafari, M. R.; Matochko, W. L.; Derda, R., Quantitative Synthesis of Genetically Encoded Glycopeptide Libraries Displayed on M13 Phage. *ACS Chem. Biol.* **2012**, *7* (9), 1482-1487.
- (43) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B., One-step site-specific modification of native proteins with 2-pyridinecarboxaldehydes. *Nat. Chem. Biol.* **2015**, *11*, 326.
- (44) Saggiomo, V.; Lüning, U., On the formation of imines in water—a comparison. *Tetrahedron Lett.* **2009**, *50* (32), 4663-4665.
- (45) Dirksen, A.; Hackeng, T. M.; Dawson, P. E., Nucleophilic Catalysis of Oxime Ligation. *Angew. Chem., Int. Ed.* **2006**, *45* (45), 7581-7584.
- (46) Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R., A Pictet-Spengler ligation for protein chemical modification. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (1), 46-51.
- (47) Triana, V.; Derda, R., Tandem Wittig/Diels-Alder diversification of genetically encoded peptide libraries. *Org. Biomol. Chem.* **2017**, *15* (37), 7869-7877.
- (48) Devaraj, N. K.; Upadhyay, R.; Haun, J. B.; Hilderbrand, S. A.; Weissleder, R., Fast and Sensitive Pretargeted Labeling of Cancer Cells through a Tetrazine/trans-Cyclooctene Cycloaddition. *Angew. Chem., Int. Ed.* **2009**, *48* (38), 7013-7016.
- (49) Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L., Contemporary approaches to site-selective protein modification. *Nature Reviews Chemistry* **2019**, *3* (3), 147-171.
- (50) Gunnoo, S. B.; Madder, A., Chemical Protein Modification through Cysteine. *Chembiochem* **2016**, *17* (7), 529-553.
- (51) Kather, I.; Bippes, C. A.; Schmid, F. X., A Stable Disulfide-free Gene-3-protein of Phage fd Generated by In vitro Evolution. *J. Mol. Biol.* **2005**, *354* (3), 666-678.
- (52) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; van Opijnen, T.; Gao, J., Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (19), 6137-6145.
- (53) Uematsu, S.; Tabuchi, Y.; Ito, Y.; Taki, M., Combinatorially Screened Peptide as Targeted Covalent Binder: Alteration of Bait-Conjugated Peptide to Reactive Modifier. *Bioconjug. Chem.* **2018**, *29* (6), 1866-1871.
- (54) Diderich, P.; Bertoldo, D.; Dessen, P.; Khan, M. M.; Pizzitola, I.; Held, W.; Huelsken, J.; Heinis, C., Phage Selection of Chemically Stabilized  $\alpha$ -Helical Peptide Ligands. *ACS Chem. Biol.* **2016**, *11* (5), 1422-1427.
- (55) Ng, S.; Derda, R., Phage-displayed macrocyclic glycopeptide libraries. *Org. Biomol. Chem.* **2016**.

- (56) Jafari, M. R.; Lakusta, J.; Lundgren, R. J.; Derda, R., Allene Functionalized Azobenzene Linker Enables Rapid and Light-Responsive Peptide Macrocyclization. *Bioconjug. Chem.* **2016**, *27* (3), 509-514.
- (57) Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R., Discovery of Light-Responsive Ligands through Screening of a Light-Responsive Genetically Encoded Library. *ACS Chem. Biol.* **2014**, *9* (2), 443-450.
- (58) Kalhor-Monfared, S.; Jafari, M. R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. M.; Derda, R., Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chem. Sci.* **2016**, *7* (6), 3785-3790.
- (59) Fukunaga, K.; Hatanaka, T.; Ito, Y.; Minami, M.; Taki, M., Construction of a crown ether-like supramolecular library by conjugation of genetically-encoded peptide linkers displayed on bacteriophage T7. *Chemical Communications* **2014**, *50* (30), 3921-3923.
- (60) Young, T. S.; Schultz, P. G., Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J. Biol. Chem.* **2010**, *285* (15), 11039-44.
- (61) Horiya, S.; Bailey, J. K.; Temme, J. S.; Guillen Schlippe, Y. V.; Krauss, I. J., Directed Evolution of Multivalent Glycopeptides Tightly Recognized by HIV Antibody 2G12. *J. Am. Chem. Soc.* **2014**, *136* (14), 5407-5415.
- (62) Lim, R. K. V.; Li, N.; Ramil, C. P.; Lin, Q., Fast and Sequence-Specific Palladium-Mediated Cross-Coupling Reaction Identified from Phage Display. *ACS Chem. Biol.* **2014**, *9* (9), 2139-2148.
- (63) Tsao, M.-L.; Tian, F.; Schultz, P. G., Selective Staudinger Modification of Proteins Containing p-Azidophenylalanine. *Chembiochem* **2005**, *6* (12), 2147-2149.
- (64) Tian, F.; Tsao, M.-L.; Schultz, P. G., A Phage Display System with Unnatural Amino Acids. *J. Am. Chem. Soc.* **2004**, *126* (49), 15962-15963.
- (65) Trevors, J. T.; Cotter, C. M., Copper toxicity and uptake in microorganisms. *J. Ind. Microbiol.* **1990**, *6* (2), 77-84.
- (66) Li, J.; Dennehy, J. J., Differential Bacteriophage Mortality on Exposure to Copper. *Appl. Environ. Microbiol.* **2011**, *77* (19), 6878-6883.
- (67) Beech, J.; Saleh, L.; Frentzel, J.; Figler, H.; Correa, I. R., Jr.; Baker, B.; Ramsbacher, C.; Marshall, M.; Dasa, S.; Linden, J.; Noren, C. J.; Kelly, K. A., Multivalent site-specific phage modification enhances the binding affinity of receptor ligands. *Bioconjug. Chem.* **2015**, *26* (3), 529-36.
- (68) Hoesl, M. G.; Budisa, N., In Vivo Incorporation of Multiple Noncanonical Amino Acids into Proteins. *Angew. Chem., Int. Ed.* **2011**, *50* (13), 2896-2902.
- (69) Zhang, Y.; Park, K.-Y.; Suazo, K. F.; Distefano, M. D., Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* **2018**, *47* (24), 9106-9136.
- (70) Li, C.; Wang, L.-X., Chemoenzymatic Methods for the Synthesis of Glycoproteins. *Chem. Rev.* **2018**, *118* (17), 8359-8413.

- (71) Scholle, M. D.; Kriplani, U.; Pabon, A.; Sishtla, K.; Glucksman, M. J.; Kay, B. K., Mapping Protease Substrates by Using a Biotinylated Phage Substrate Library. *Chembiochem* **2006**, *7* (5), 834-838.
- (72) Nguyen, G. K. T.; Cao, Y.; Wang, W.; Liu, C. F.; Tam, J. P., Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew. Chem., Int. Ed.* **2015**, *54* (52), 15694-15698.
- (73) Weeks, A. M.; Wells, J. A., Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol.* **2017**, *14*, 50.
- (74) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A., Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004**, *126* (9), 2670-2671.
- (75) Wilson, H. D.; Li, X.; Peng, H.; Rader, C., A Sortase A Programmable Phage Display Format for Improved Panning of Fab Antibody Libraries. *J. Mol. Biol.* **2018**, *430* (21), 4387-4400.
- (76) Hess, G. T.; Cragolini, J. J.; Popp, M. W.; Allen, M. A.; Dougan, S. K.; Spooner, E.; Ploegh, H. L.; Belcher, A. M.; Guimaraes, C. P., M13 Bacteriophage Display Framework That Allows Sortase-Mediated Modification of Surface-Accessible Phage Proteins. *Bioconjug. Chem.* **2012**, *23* (7), 1478-1487.
- (77) Fuh, G.; Sidhu, S. S., Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. *FEBS Lett.* **2000**, *480* (2-3), 231-234.
- (78) Meyer, C.; Liebscher, S.; Bordusa, F., Selective Coupling of Click Anchors to Proteins via Trypsiligase. *Bioconjug. Chem.* **2016**, *27* (1), 47-53.
- (79) Liebscher, S.; Schöpfel, M.; Aumüller, T.; Sharkhuukhen, A.; Pech, A.; Höss, E.; Parthier, C.; Jahreis, G.; Stubbs, M. T.; Bordusa, F., N-Terminal Protein Modification by Substrate-Activated Reverse Proteolysis. *Angew. Chem., Int. Ed.* **2014**, *53* (11), 3024-3028.
- (80) Liebscher, S.; Kornberger, P.; Fink, G.; Trost-Gross, E.-M.; Höss, E.; Skerra, A.; Bordusa, F., Derivatization of Antibody Fab Fragments: A Designer Enzyme for Native Protein Modification. *Chembiochem* **2014**, *15* (8), 1096-1100.
- (81) Urban, J. H.; Moosmeier, M. A.; Aumüller, T.; Thein, M.; Bosma, T.; Rink, R.; Groth, K.; Zulle, M.; Siegers, K.; Tissot, K.; Moll, G. N.; Prassler, J., Phage display and selection of lanthipeptides on the carboxy-terminus of the gene-3 minor coat protein. *Nat. Commun.* **2017**, *8* (1), 1500.
- (82) Shi, Y.; Yang, X.; Garg, N.; van der Donk, W. A., Production of Lantipeptides in *Escherichia coli*. *J. Am. Chem. Soc.* **2011**, *133* (8), 2338-2341.
- (83) Ernst, B.; Magnani, J. L., From carbohydrate leads to glycomimetic drugs. *Nat Rev Drug Discov* **2009**, *8* (8), 661-77.
- (84) Aretz, J.; Baukman, H.; Shanina, E.; Hanske, J.; Wawrzinek, R.; Zapol'skii, V. A.; Seeberger, P. H.; Kaufmann, D. E.; Rademacher, C.,

- Identification of Multiple Druggable Secondary Sites by Fragment Screening against DC-SIGN. *Angew. Chem., Int. Ed.* **2017**, *56* (25), 7292-7296.
- (85) Jain, D.; Kaur K Fau - Sundaravadivel, B.; Sundaravadivel B Fau - Salunke, D. M.; Salunke, D. M., Structural and functional consequences of peptide-carbohydrate mimicry. Crystal structure of a carbohydrate-mimicking peptide bound to concanavalin A. *J. Biol. Chem.* **2000**, *275* (0021-9258 (Print)), 16098-16102.
- (86) Ng, S.; Bennett, N. J.; Schulze, J.; Gao, N.; Rademacher, C.; Derda, R., Genetically-encoded fragment-based discovery of glycopeptide ligands for DC-SIGN. *Bioorg. Med. Chem.* **2018**, *26* (19), 5368-5377.
- (87) Qi, X.; Xia, T.; Roberts, R. W., Acridine-N Peptide Conjugates Display Enhanced Affinity and Specificity to BoxB RNA Targets. *Biochemistry* **2010**, *49* (27), 5782-5789.
- (88) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; van Opijnen, T.; Gao, J., Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**.
- (89) Sandman, K. E.; Benner, J. S.; Noren, C. J., Phage Display of Selenopeptides. *J. Am. Chem. Soc.* **2000**, *122* (5), 960-961.
- (90) Boscher, C.; Dennis, J. W.; Nabi, I. R., Glycosylation, galectins and cellular signaling. *Curr. Opin. Cell Biol.* **2011**, *23* (4), 383-92.
- (91) Rabinovich, G. A.; Conejo-Garcia, J. R., Shaping the immune landscape in cancer by galectin-driven regulatory pathways. *J. Mol. Biol.* **2016**.
- (92) Thiemann, S.; Baum, L. G., Galectins and Immune Responses-Just How Do They Do Those Things They Do? *Annu. Rev. Immunol.* **2016**, *34*, 243-64.
- (93) Ahmad, N.; Gabius, H. J.; Sabesan, S.; Oscarson, S.; Brewer, C. F., Thermodynamic binding studies of bivalent oligosaccharides to galectin-1, galectin-3, and the carbohydrate recognition domain of galectin-3. *Glycobiology* **2004**, *14* (9), 817-25.
- (94) Leffler, H.; Barondes, S. H., Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian beta-galactosides. *J. Biol. Chem.* **1986**, *261* (22), 10119-26.
- (95) Saraboji, K.; Hakansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; Logan, D. T., The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry* **2012**, *51* (1), 296-306.
- (96) Blanchard, H.; Bum-Erdene, K.; Bohari, M. H.; Yu, X., Galectin-1 inhibitors and their potential therapeutic applications: a patent review. *Expert Opin. Ther. Pat.* **2016**, 1-18.
- (97) Tellez-Sanz, R.; Garcia-Fuentes, L.; Vargas-Berenguel, A., Human galectin-3 selective and high affinity inhibitors. Present state and future perspectives. *Curr. Med. Chem.* **2013**, *20* (24), 2979-90.

- (98) Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W. E.; Yagi, F.; Kasai, K., Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim. Biophys. Acta* **2002**, *1572* (2-3), 232-54.
- (99) Henrick, K.; Bawumia, S.; Barboni, E. A.; Mehul, B.; Hughes, R. C., Evidence for subsites in the galectins involved in sugar binding at the nonreducing end of the central galactose of oligosaccharide ligands: sequence analysis, homology modeling and mutagenesis studies of hamster galectin-3. *Glycobiology* **1998**, *8* (1), 45-57.
- (100) Jayaram, S.; Kapoor, S.; Dharmesh, S. M., Pectic polysaccharide from corn (*Zea mays* L.) effectively inhibited multi-step mediated cancer cell growth and metastasis. *Chem. Biol. Interact.* **2015**, *235*, 63-75.
- (101) Alonso-Plaza, J. M.; Canales, M. A.; Jimenez, M.; Roldan, J. L.; Garcia-Herrero, A.; Iturrino, L.; Asensio, J. L.; Canada, F. J.; Romero, A.; Siebert, H. C.; Andre, S.; Solis, D.; Gabius, H. J.; Jimenez-Barbero, J., NMR investigations of protein-carbohydrate interactions: insights into the topology of the bound conformation of a lactose isomer and beta-galactosyl xyloses to mistletoe lectin and galectin-1. *Biochim. Biophys. Acta* **2001**, *1568* (3), 225-36.
- (102) Ryan, C. M.; Mehlert, A.; Richardson, J. M.; Ferguson, M. A.; Johnson, P. J., Chemical structure of *Trichomonas vaginalis* surface lipoglycan: a role for short galactose (beta1-4/3) N-acetylglucosamine repeats in host cell interaction. *J. Biol. Chem.* **2011**, *286* (47), 40494-508.
- (103) Blanchard, H.; Yu, X.; Collins, P. M.; Bum-Erdene, K., Galectin-3 inhibitors: a patent review (2008-present). *Expert Opin. Ther. Pat.* **2014**, *24* (10), 1053-65.
- (104) Peterson, K.; Kumar, R.; Stenstrom, O.; Verma, P.; Verma, P. R.; Hakansson, M.; Kahl-Knutsson, B.; Zetterberg, F.; Leffler, H.; Akke, M.; Logan, D. T.; Nilsson, U. J., Systematic tuning of fluoro-galectin-3 interactions provides thiodigalactoside derivatives with single digit nM affinity and high selectivity. *J. Med. Chem.* **2017**.
- (105) Tejler, J.; Leffler, H.; Nilsson, U. J., Synthesis of O-galactosyl aldoximes as potent LacNAc-mimetic galectin-3 inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15* (9), 2343-5.
- (106) Lin, C. I.; Whang, E. E.; Donner, D. B.; Jiang, X.; Price, B. D.; Carothers, A. M.; Delaine, T.; Leffler, H.; Nilsson, U. J.; Nose, V.; Moore, F. D., Jr.; Ruan, D. T., Galectin-3 targeted therapy with a small molecule inhibitor activates apoptosis and enhances both chemosensitivity and radiosensitivity in papillary thyroid cancer. *Mol. Cancer Res.* **2009**, *7* (10), 1655-62.
- (107) Kouo, T.; Huang, L.; Pucsek, A. B.; Cao, M.; Solt, S.; Armstrong, T.; Jaffee, E., Galectin-3 Shapes Antitumor Immune Responses by Suppressing CD8+ T Cells via LAG-3 and Inhibiting Expansion of Plasmacytoid Dendritic Cells. *Cancer Immunol Res* **2015**, *3* (4), 412-23.

- (108) Croci, D. O.; Cerliani, J. P.; Dalotto-Moreno, T.; Mendez-Huergo, S. P.; Mascanfroni, I. D.; Dergan-Dylon, S.; Toscano, M. A.; Caramelo, J. J.; Garcia-Vallejo, J. J.; Ouyang, J.; Mesri, E. A.; Junttila, M. R.; Bais, C.; Shipp, M. A.; Salatino, M.; Rabinovich, G. A., Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* **2014**, *156* (4), 744-58.
- (109) Matochko, W. L.; Cory Li, S.; Tang, S. K. Y.; Derda, R., Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res.* **2014**, *42* (3), 1784-1798.
- (110) Derda, R.; Tang, S.; Li, S. C.; Ng, S.; Matochko, W.; Jafari, M., Diversity of Phage-Displayed Libraries of Peptides during Panning and Amplification. *Molecules* **2011**, *16* (2), 1776.
- (111) Kehoe, J. W.; Kay, B. K., Filamentous Phage Display in the New Millennium. *Chem. Rev.* **2005**, *105* (11), 4056-4072.
- (112) Shi, B.; Deng, Y.; Zhao, P.; Li, X., Selecting a DNA-Encoded Chemical Library against Non-immobilized Proteins Using a “Ligate–Cross-Link–Purify” Strategy. *Bioconjug. Chem.* **2017**, *28* (9), 2293-2301.
- (113) Denton, K. E.; Krusemark, C. J., Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *MedChemComm* **2016**, *7* (10), 2020-2027.
- (114) Stillman, B. N.; Hsu, D. K.; Pang, M.; Brewer, C. F.; Johnson, P.; Liu, F. T.; Baum, L. G., Galectin-3 and Galectin-1 Bind Distinct Cell Surface Glycoprotein Receptors to Induce T Cell Death. *J. Immunol.* **2006**, *176* (2), 778-789.
- (115) Kitov, P. I.; Vinals, D. F.; Ng, S.; Tjhung, K. F.; Derda, R., Rapid, Hydrolytically Stable Modification of Aldehyde-Terminated Proteins and Phage Libraries. *J. Am. Chem. Soc.* **2014**, *136* (23), 8149-8152.
- (116) Wu, C. K.; Su, M. Y.; Lee, J. K.; Chiang, F. T.; Hwang, J. J.; Lin, J. L.; Chen, J. J.; Liu, F. T.; Tsai, C. T., Galectin-3 level and the severity of cardiac diastolic dysfunction using cellular and animal models and clinical indices. *Sci. Rep.* **2015**, *5*, 17007.
- (117) Nanthakumar, C. B.; Hatley, R. J.; Lemma, S.; Gauldie, J.; Marshall, R. P.; Macdonald, S. J., Dissecting fibrosis: therapeutic insights from the small-molecule toolbox. *Nat Rev Drug Discov* **2015**, *14* (10), 693-720.
- (118) Blanchard, H.; Bum-Erdene, K.; Hugo, M. W., Inhibitors of Galectins and Implications for Structure-Based Design of Galectin-Specific Therapeutics. *Aust. J. Chem.* **2014**, *67* (12), 1763-1779.
- (119) Mackinnon, A. C.; Gibbons, M. A.; Farnworth, S. L.; Leffler, H.; Nilsson, U. J.; Delaine, T.; Simpson, A. J.; Forbes, S. J.; Hirani, N.; Gauldie, J.; Sethi, T., Regulation of transforming growth factor-beta1-driven lung fibrosis by galectin-3. *Am. J. Respir. Crit. Care Med.* **2012**, *185* (5), 537-46.
- (120) Traber, P. G.; Zomer, E., Therapy of Experimental NASH and Fibrosis with Galectin Inhibitors. *PLoS One* **2013**, *8* (12), e83481.
- (121) Traber, P. G.; Chou, H.; Zomer, E.; Hong, F.; Klyosov, A.; Fiel, M.-I.; Friedman, S. L., Regression of Fibrosis and Reversal of Cirrhosis in Rats



- by Galectin Inhibitors in Thioacetamide-Induced Liver Disease. *PLoS One* **2013**, *8* (10), e75361.
- (122) Leffler, H.; Carlsson, S.; Hedlund, M.; Qian, Y.; Poirier, F., Introduction to galectins. *Glycoconj. J.* **2004**, *19* (7-9), 433-40.
- (123) Barondes, S. H.; Cooper, D. N.; Gitt, M. A.; Leffler, H., Galectins. Structure and function of a large family of animal lectins. *J. Biol. Chem.* **1994**, *269* (33), 20807-10.
- (124) Salomonsson, E.; Carlsson, M. C.; Osla, V.; Hendus-Altenburger, R.; Kahl-Knutson, B.; Oberg, C. T.; Sundin, A.; Nilsson, R.; Nordberg-Karlsson, E.; Nilsson, U. J.; Karlsson, A.; Rini, J. M.; Leffler, H., Mutational tuning of galectin-3 specificity and biological function. *J. Biol. Chem.* **2010**, *285* (45), 35079-91.
- (125) Brewer, C. F., Thermodynamic binding studies of galectin-1, -3 and -7. *Glycoconj. J.* **2004**, *19* (7-9), 459-65.
- (126) Shams-Ud-Doha, K.; Kitova, E. N.; Kitov, P. I.; St-Pierre, Y.; Klassen, J. S., Human Milk Oligosaccharide Specificities of Human Galectins. Comparison of Electrospray Ionization Mass Spectrometry and Glycan Microarray Screening Results. *Anal. Chem.* **2017**, *89* (9), 4914-4921.
- (127) Arthur, C. M.; Rodrigues, L. C.; Baruffi, M. D.; Sullivan, H. C.; Heimburg-Molinaro, J.; Smith, D. F.; Cummings, R. D.; Stowell, S. R., Examining galectin binding specificity using glycan microarrays. *Methods Mol. Biol.* **2015**, *1207*, 115-131.
- (128) Hsieh, T. J.; Lin, H. Y.; Tu, Z.; Lin, T. C.; Wu, S. C.; Tseng, Y. Y.; Liu, F. T.; Hsu, S. T.; Lin, C. H., Dual thio-digalactoside-binding modes of human galectins as the structural basis for the design of potent and selective inhibitors. *Sci. Rep.* **2016**, *6*, 29457.
- (129) Salameh, B. A.; Cumpstey, I.; Sundin, A.; Leffler, H.; Nilsson, U. J., 1H-1,2,3-triazol-1-yl thiodigalactoside derivatives as high affinity galectin-3 inhibitors. *Bioorg. Med. Chem.* **2010**, *18* (14), 5367-78.
- (130) Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J., Structural and thermodynamic studies on cation- $\pi$  interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J. Am. Chem. Soc.* **2005**, *127* (6), 1737-43.
- (131) Peterson, K.; Collins, P. M.; Huang, X.; Kahl-Knutsson, B.; Essén, S.; Zetterberg, F. R.; Oredsson, S.; Leffler, H.; Blanchard, H.; Nilsson, U. J., Aromatic heterocycle galectin-1 interactions for selective single-digit nM affinity ligands. *RSC Advances* **2018**, *8* (44), 24913-24922.
- (132) Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **2015**, *5* (1750-2799 (Electronic)).
- (133) Kozakov, D.; Hall, D. R.; Chuang, G.-Y.; Cencic, R.; Brenke, R.; Grove, L. E.; Beglov, D.; Pelletier, J.; Whitty, A.; Vajda, S., Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (33), 13528-13533.

- (134) Mattos, C.; Hall, D.; Kozakov, D.; Beglov, D.; Chuang, G.-Y.; Landon, M. R.; Brenke, R.; Vajda, S., Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25* (5), 621-627.
- (135) Adey, N. B.; Mataragnon, A. H.; Rider, J. E.; Carter, J. M.; Kay, B. K., Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene* **1995**, *156* (1), 27-31.
- (136) Devlin, J.; Panganiban, L.; Devlin, P., Random peptide libraries: a source of specific protein binding molecules. *Science* **1990**, *249* (4967), 404-406.
- (137) Saggio, I.; Laufer, R., Biotin binders selected from a random peptide library expressed on phage. *Biochem. J.* **1993**, *293* (3), 613-616.
- (138) Menendez, A.; Scott, J. K., The nature of target-unrelated peptides recovered in the screening of phage-displayed random peptide libraries with antibodies. *Anal. Biochem.* **2005**, *336* (2), 145-157.
- (139) Murai, R.; Nogi, T.; Tateoka, K.; Sato, A., Affinity Selection of Peptide Binders with Magnetic Beads *via* Organic Phase Separation (MOPS). *Biol. Pharm. Bull.* **2015**, *38* (11), 1822-1826.
- (140) Zhang, M.-Y.; Dimitrov, D. S., Sequential antigen panning for selection of broadly cross-reactive HIV-1-neutralizing human monoclonal antibodies. *Methods Mol. Biol.* **2009**, *562*, 143-154.
- (141) Haque, A.; Tonks, N. K., The use of phage display to generate conformation-sensor recombinant antibodies. *Nat. Protoc.* **2012**, *7* (12), 2127-2143.
- (142) Koide, A.; Wojcik, J.; Gilbreth, R. N.; Reichel, A.; Piehler, J.; Koide, S., Accelerating phage-display library selection by reversible and site-specific biotinylation. *Protein engineering, design & selection : PEDS* **2009**, *22* (11), 685-690.
- (143) Scholle, M. D.; Collart, F. R.; Kay, B. K., In vivo biotinylated proteins as targets for phage-display selection experiments. *Protein Expr. Purif.* **2004**, *37* (1), 243-252.
- (144) Sadaghiani, A. M.; Verhelst, S. H. L.; Bogyo, M., Tagging and detection strategies for activity-based proteomics. *Curr. Opin. Chem. Biol.* **2007**, *11* (1), 20-28.
- (145) El-Hawiet, A.; Kitova, E. N.; Klassen, J. S., Quantifying Carbohydrate-Protein Interactions by Electrospray Ionization Mass Spectrometry Analysis. *Biochemistry* **2012**, *51* (21), 4244-4253.
- (146) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **2004**, *22*, 1567.
- (147) Karttunen, J.; Shastri, N., Measurement of ligand-induced activation in single viable T cells using the lacZ reporter gene. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (9), 3972-3976.
- (148) He, B.; Tjhung, K. F.; Bennett, N. J.; Chou, Y.; Rau, A.; Huang, J.; Derda, R., Compositional Bias in Naïve and Chemically-modified

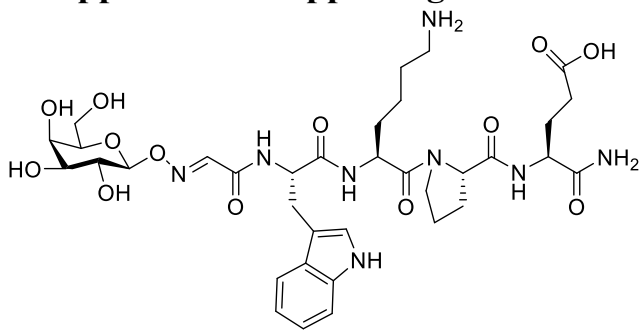
- Phage-Displayed Libraries uncovered by Paired-end Deep Sequencing. *Sci. Rep.* **2018**, *8* (1), 1214.
- (149) Doerks, T.; Copley, R. R.; Schultz, J.; Ponting, C. P.; Bork, P., Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* **2002**, *12* (1), 47-56.
- (150) Ng, S.; Tjhung, K. F.; Paschal, B. M.; Noren, C. J.; Derda, R., Chemical posttranslational modifications of phage-displayed peptides. In *Peptide Libraries*, Derda, R., Ed. Humana Press: New York, 2015; pp 155-172.
- (151) Noren, K. A.; Noren, C. J., Construction of High-Complexity Combinatorial Phage Display Peptide Libraries. *Methods* **2001**, *23* (2), 169-178.
- (152) Shendure, J.; Ji, H., Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135.
- (153) Lander, E. S., Array of hope. *Nat. Genet.* **1999**, *21* (1061-4036 (Print)), 3-4.
- (154) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; Razi, N.; Stevens, D. J.; Skehel, J. J.; van Die, I.; Burton, D. R.; Wilson, I. A.; Cummings, R.; Bovin, N.; Wong, C.-H.; Paulson, J. C., Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (49), 17033-17038.
- (155) van Kooyk, Y.; Rabinovich, G. A., Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **2008**, *9*, 593.
- (156) Varki, A., Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **2007**, *446*, 1023.
- (157) Stevens, J.; Blixt, O.; Tumpey, T. M.; Taubenberger, J. K.; Paulson, J. C.; Wilson, I. A., Structure and Receptor Specificity of the Hemagglutinin from an H5N1 Influenza Virus. *Science* **2006**, *312* (5772), 404-410.
- (158) Stevens, J.; Blixt, O.; Paulson, J. C.; Wilson, I. A., Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **2006**, *4*, 857.
- (159) Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R., Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat. Methods* **2005**, *2*, 817.
- (160) Malone, J. H.; Oliver, B., Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **2011**, *9* (1), 34.
- (161) Oyelaran, O.; Li, Q.; Farnsworth, D.; Gildersleeve, J. C., Microarrays with Varying Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies. *J. Proteome Res.* **2009**, *8* (7), 3529-3538.
- (162) Oyelaran, O.; McShane, L. M.; Dodd, L.; Gildersleeve, J. C., Profiling Human Serum Antibodies with a Carbohydrate Antigen Microarray. *J. Proteome Res.* **2009**, *8* (9), 4301-4310.
- (163) Fukui, S.; Feizi, T.; Galustian, C.; Lawson, A. M.; Chai, W., Oligosaccharide microarrays for high-throughput detection and

- specificity assignments of carbohydrate-protein interactions. *Nat. Biotechnol.* **2002**, *20*, 1011.
- (164) Donczo, B.; Kerekgyarto, J.; Szurmai, Z.; Guttman, A., Glycan microarrays: new angles and new strategies. *Analyst* **2014**, *139* (11), 2650-2657.
- (165) Park, S.; Gildersleeve, J. C.; Blixt, O.; Shin, I., Carbohydrate microarrays. *Chem. Soc. Rev.* **2013**, *42* (10), 4310-4326.
- (166) Rillahan, C. D.; Paulson, J. C., Glycan Microarrays for Decoding the Glycome. *Annu. Rev. Biochem.* **2011**, *80* (1), 797-823.
- (167) Song, X.; Xia, B.; Lasanajak, Y.; Smith, D. F.; Cummings, R. D., Quantifiable fluorescent glycan microarrays. *Glycoconj. J.* **2008**, *25* (1), 15-25.
- (168) 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), 6140-6148.
- (169) van Munster, J. M.; Thomas, B.; Riese, M.; Davis, A. L.; Gray, C. J.; Archer, D. B.; Flitsch, S. L., Application of carbohydrate arrays coupled with mass spectrometry to detect activity of plant-polysaccharide degradative enzymes from the fungus *Aspergillus niger*. *Sci. Rep.* **2017**, *7*, 43117.
- (170) Houseman, B. T.; Mrksich, M., Carbohydrate Arrays for the Evaluation of Protein Binding and Enzymatic Modification. *Chem. Biol.* **2002**, *9* (4), 443-454.
- (171) Grant, C. F.; Kanda, V.; Yu, H.; Bundle, D. R.; McDermott, M. T., Optimization of Immobilized Bacterial Disaccharides for Surface Plasmon Resonance Imaging Measurements of Antibody Binding. *Langmuir* **2008**, *24* (24), 14125-14132.
- (172) Frenz, T.; Grabski, E.; Durán, V.; Hozsa, C.; Stępczyńska, A.; Furch, M.; Gieseler, R. K.; Kalinke, U., Antigen presenting cell-selective drug delivery by glycan-decorated nanocarriers. *Eur. J. Pharm. Biopharm.* **2015**, *95*, 13-17.
- (173) Boks, M. A.; Bruijns, S. C. M.; Ambrosini, M.; Kalay, H.; van Bloois, L.; Storm, G.; Gruijl, T. d.; van Kooyk, Y., In situ Delivery of Tumor Antigen- and Adjuvant-Loaded Liposomes Boosts Antigen-Specific T-Cell Responses by Human Dermal Dendritic Cells. *J. Invest. Dermatol.* **2015**, *135* (11), 2697-2704.
- (174) Boks, M. A.; Ambrosini, M.; Bruijns, S. C.; Kalay, H.; van Bloois, L.; Storm, G.; Garcia-Vallejo, J. J.; van Kooyk, Y., MPLA incorporation into DC-targeting glycoliposomes favours anti-tumour T cell responses. *J. Control. Release* **2015**, *216*, 37-46.
- (175) Chen, W. C.; Sigal, D. S.; Saven, A.; Paulson, J. C., Targeting B lymphoma with nanoparticles bearing glycan ligands of CD22. *Leuk. Lymphoma* **2012**, *53* (2), 208-210.
- (176) Hamming, R. W., Error detecting and error correcting codes. *The Bell System Technical Journal* **1950**, *29* (2), 147-160.

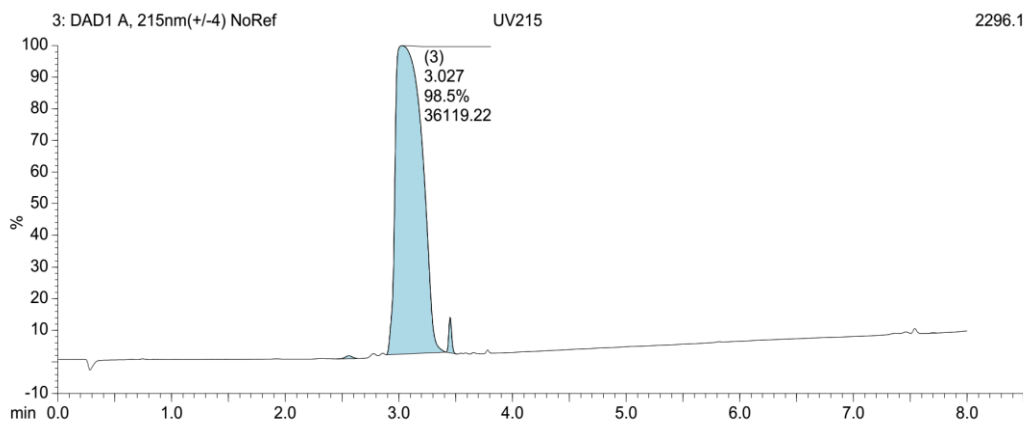
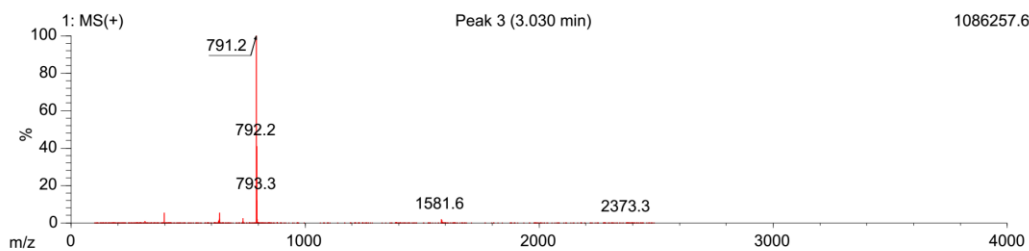
- (177) Einhauer, A.; Jungbauer, A., The FLAG<sup>TM</sup> peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **2001**, *49* (1), 455-465.
- (178) Field, J.; Nikawa, J.; Broek, D.; MacDonald, B.; Rodgers, L.; Wilson, I. A.; Lerner, R. A.; Wigler, M., Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* **1988**, *8* (5), 2159-2165.
- (179) Vinals, D. F.; Kitov, P. I.; Tu, Z.; Zou, C.; Cairo, C. W.; Lin, H. C.-H.; Derda, R., Selection of galectin-3 ligands derived from genetically encoded glycopeptide libraries. *Pept. Sci.* **2019**, *111* (1), e24097.
- (180) <http://www.functionalglycomics.org>.
- (181) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem., Int. Ed.* **2002**, *41* (14), 2596-2599.
- (182) Jewett, J. C.; Bertozzi, C. R., Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **2010**, *39* (4), 1272-1279.
- (183) Sonnenburg, J. L.; van Halbeek, H.; Varki, A., Characterization of the Acid Stability of Glycosidically Linked Neuraminic Acid. *J. Biol. Chem.* **2002**, *277* (20), 17502-17510.
- (184) Crimmins, D. L.; Mische, S. M.; Denslow, N. D., Chemical Cleavage of Proteins in Solution. *Current Protocols in Protein Science* **2005**, *41* (1), 11.4.1-11.4.11.
- (185) Smith, G. P.; Petrenko, V. A., Phage Display. *Chem. Rev.* **1997**, *97* (2), 391-410.
- (186) Shaner, N. C.; Lambert, G. G.; Chammas, A.; Ni, Y.; Cranfill, P. J.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Day, R. N.; Israelsson, M.; Davidson, M. W.; Wang, J., A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* **2013**, *10*, 407.
- (187) McCarthy, D. J.; Smyth, G. K.; Robinson, M. D., edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26* (1), 139-140.
- (188) Robinson, M. D.; Smyth, G. K., Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **2007**, *9* (2), 321-332.
- (189) Robinson, M. D.; Oshlack, A., A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11* (3), R25.
- (190) Noll, A. J.; Gourdine, J.-P.; Yu, Y.; Lasanajak, Y.; Smith, D. F.; Cummings, R. D., Galectins are human milk glycan receptors. *Glycobiology* **2016**, *26* (6), 655-669.
- (191) Stowell, S. R.; Arthur, C. M.; Mehta, P.; Slanina, K. A.; Blixt, O.; Leffler, H.; Smith, D. F.; Cummings, R. D., Galectin-1, -2, and -3 Exhibit Differential Recognition of Sialylated Glycans and Blood Group Antigens. *J. Biol. Chem.* **2008**, *283* (15), 10109-10123.

- (192) Leppänen, A.; Stowell, S.; Blixt, O.; Cummings, R. D., Dimeric Galectin-1 Binds with High Affinity to  $\alpha$ 2,3-Sialylated and Non-sialylated Terminal N-Acetylglucosamine Units on Surface-bound Extended Glycans. *J. Biol. Chem.* **2005**, *280* (7), 5549-5562.
- (193) Pasqualini, R.; Ruoslahti, E., Organ targeting In vivo using phage display peptide libraries. *Nature* **1996**, *380* (6572), 364-366.
- (194) Krag, D. N.; Fuller, S. P.; Oligino, L.; Pero, S. C.; Weaver, D. L.; Soden, A. L.; Hebert, C.; Mills, S.; Liu, C.; Peterson, D., Phage-displayed random peptide libraries in mice: toxicity after serial panning. *Cancer Chemother. Pharmacol.* **2002**, *50* (4), 325-332.
- (195) Shukla, G. S.; Krag, D. N.; Peletskaya, E. N.; Pero, S. C.; Sun, Y.-J.; Carman, C. L.; McCahill, L. E.; Roland, T. A., Intravenous infusion of phage-displayed antibody library in human cancer patients: enrichment and cancer-specificity of tumor-homing phage-antibodies. *Cancer Immunol. Immunother.* **2013**, *62* (8), 1397-1410.
- (196) Krag, D. N.; Shukla, G. S.; Shen, G.-P.; Pero, S.; Ashikaga, T.; Fuller, S.; Weaver, D. L.; Burdette-Radoux, S.; Thomas, C., Selection of Tumor-binding Ligands in Cancer Patients with Phage Display Libraries. *Cancer Res.* **2006**, *66* (15), 7724-7733.
- (197) Arap, W.; Kolonin, M. G.; Trepel, M.; Lahdenranta, J.; Cardó-Vila, M.; Giordano, R. J.; Mintz, P. J.; Ardel, P. U.; Yao, V. J.; Vidal, C. I.; Chen, L.; Flamm, A.; Valtanen, H.; Weavind, L. M.; Hicks, M. E.; Pollock, R. E.; Botz, G. H.; Bucana, C. D.; Koivunen, E.; Cahill, D.; Troncoso, P.; Baggerly, K. A.; Pentz, R. D.; Do, K.-A.; Logothetis, C. J.; Pasqualini, R., Steps toward mapping the human vasculature by phage display. *Nat. Med.* **2002**, *8*, 121.
- (198) Bailey, J. J.; Bundle, D. R., Synthesis of high-mannose 1-thio glycans and their conjugation to protein. *Org. Biomol. Chem.* **2014**, *12* (14), 2193-2213.
- (199) Lipinski, T.; Kitov, P. I.; Szpacenko, A.; Paszkiewicz, E.; Bundle, D. R., Synthesis and Immunogenicity of a Glycopolymer Conjugate. *Bioconjug. Chem.* **2011**, *22* (2), 274-281.

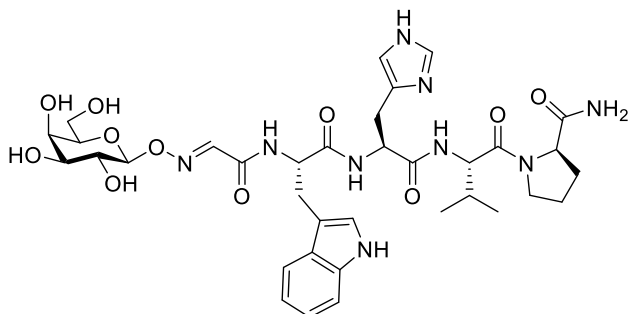
## Appendix A: Supporting information for Chapter 2



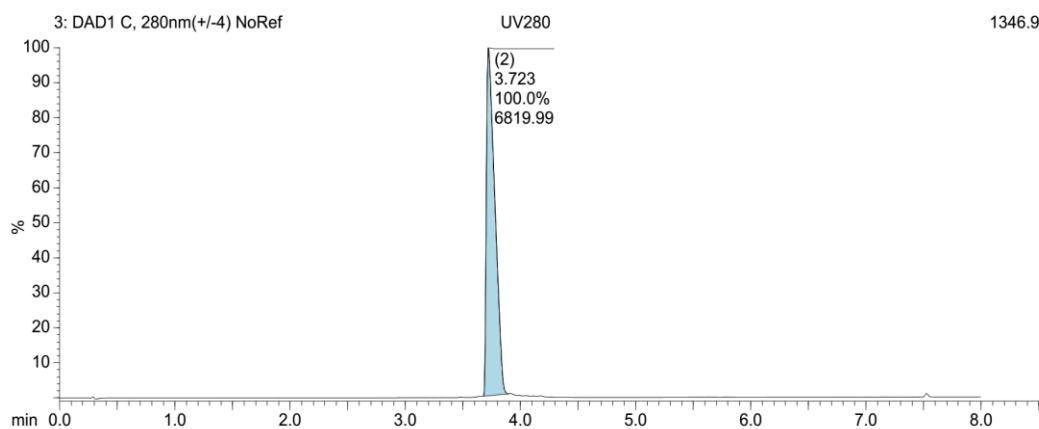
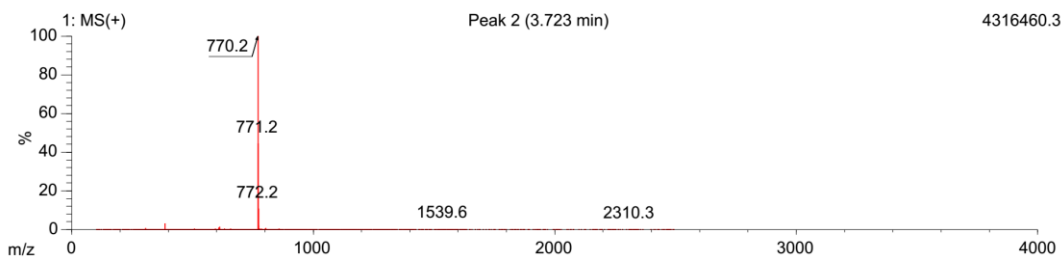
Gal-WKPE  
Chemical Formula: C<sub>35</sub>H<sub>50</sub>N<sub>8</sub>O<sub>13</sub>  
Molecular Weight: 790.83



### Appendix A-1 LC-MS characterization of Gal-WKPE glycopeptide

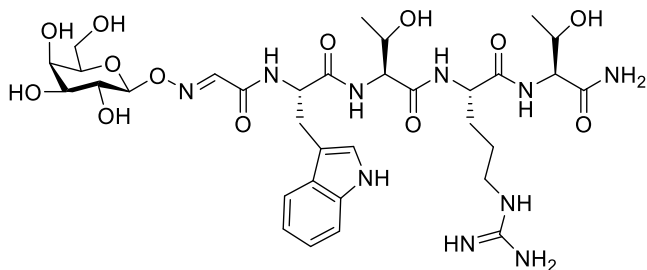


Gal-WHVP  
 Chemical Formula: C<sub>35</sub>H<sub>47</sub>N<sub>9</sub>O<sub>11</sub>  
 Molecular Weight: 769.81

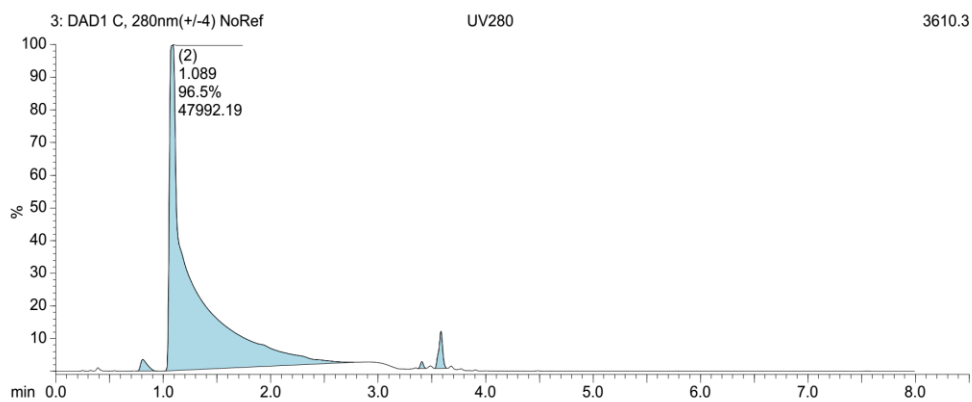
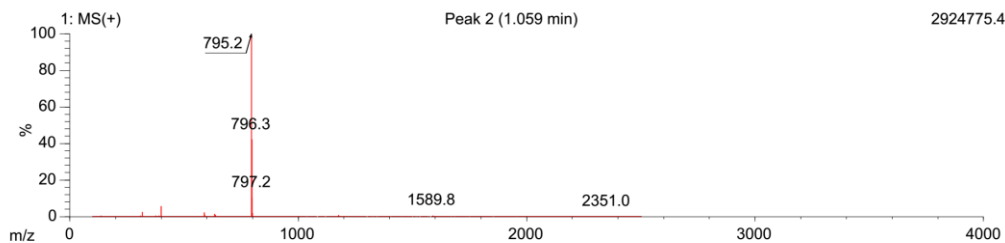


**Appendix A-2** LC-MS characterization of Gal-WHVP glycopeptide.

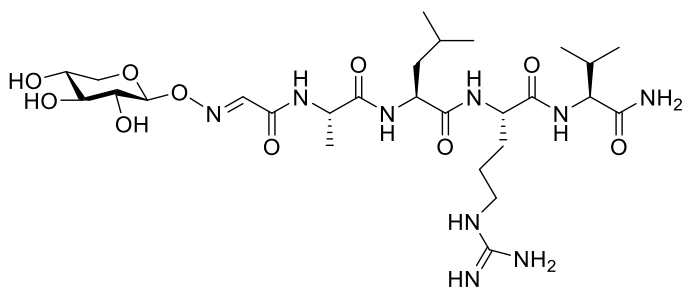




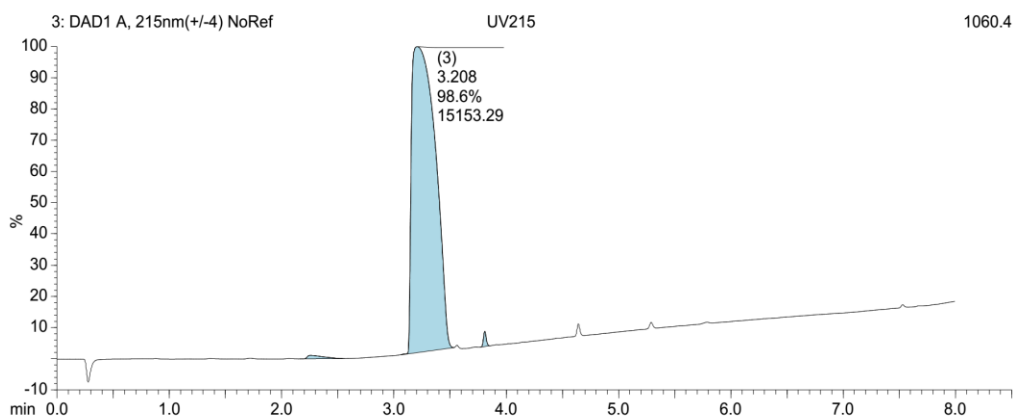
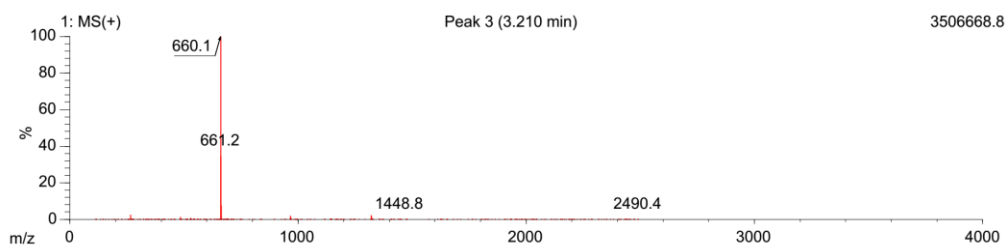
Gal-WTRT  
 Chemical Formula: C<sub>33</sub>H<sub>50</sub>N<sub>10</sub>O<sub>13</sub>  
 Molecular Weight: 794.82



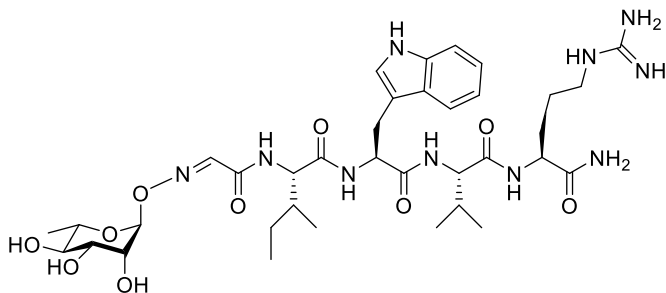
**Appendix A-3** LC-MS characterization of Gal-WTRT glycopeptide



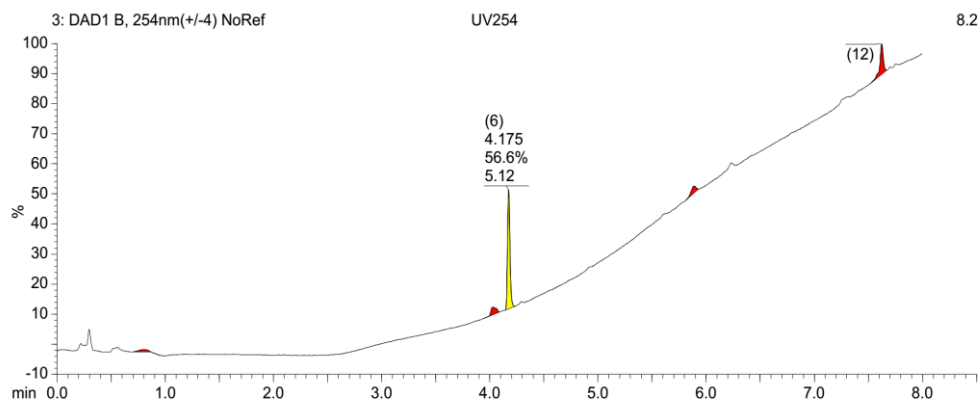
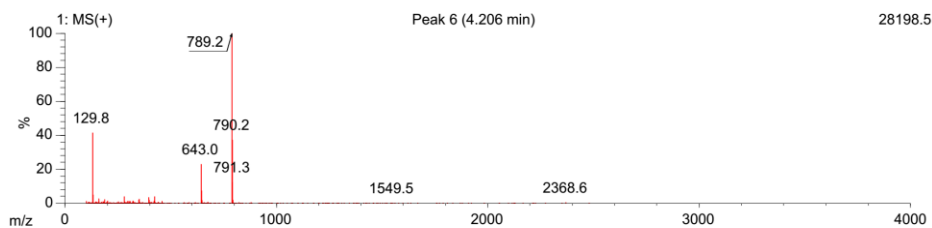
Xyl-ALRV  
 Chemical Formula: C<sub>27</sub>H<sub>49</sub>N<sub>9</sub>O<sub>10</sub>  
 Molecular Weight: 659.74



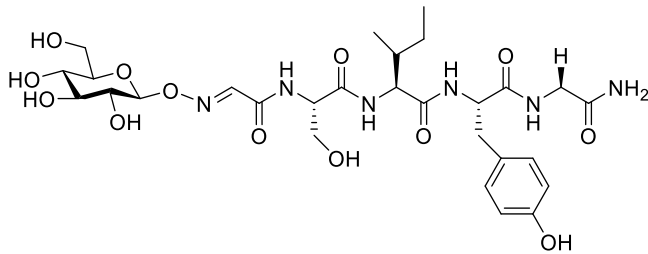
**Appendix A-4 LC-MS characterization of Xyl-ALRV glycopeptide**



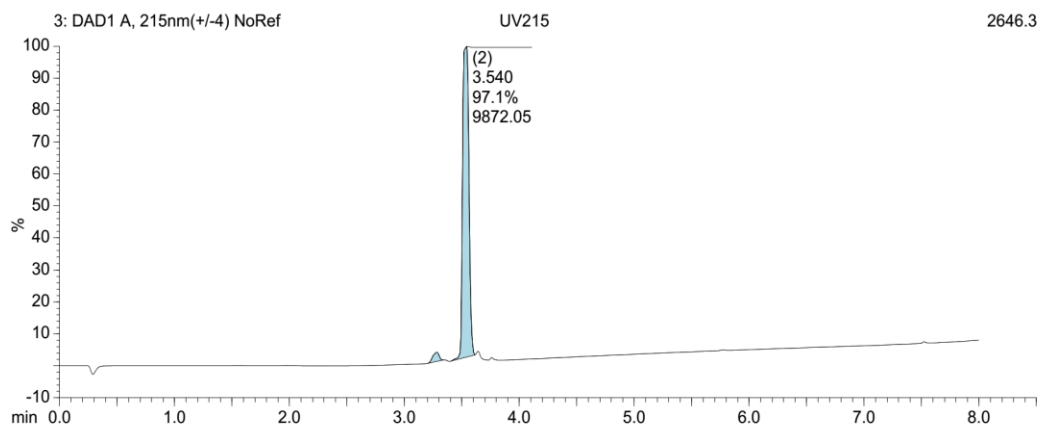
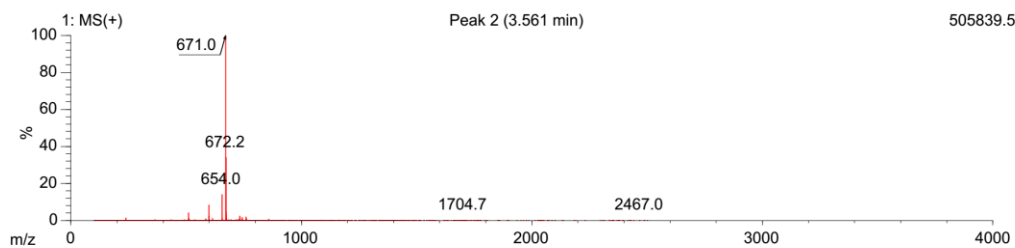
Rha-IWVR  
 Chemical Formula: C<sub>36</sub>H<sub>56</sub>N<sub>10</sub>O<sub>10</sub>  
 Molecular Weight: 788.90



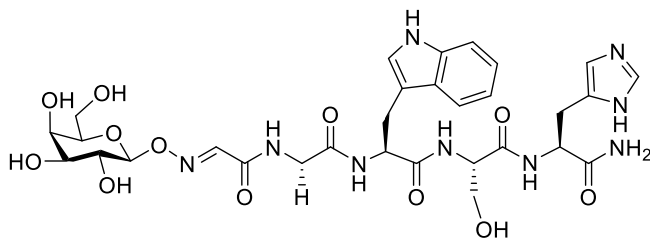
### Appendix A-5 LC-MS characterization of Rha-IWVR glycopeptide



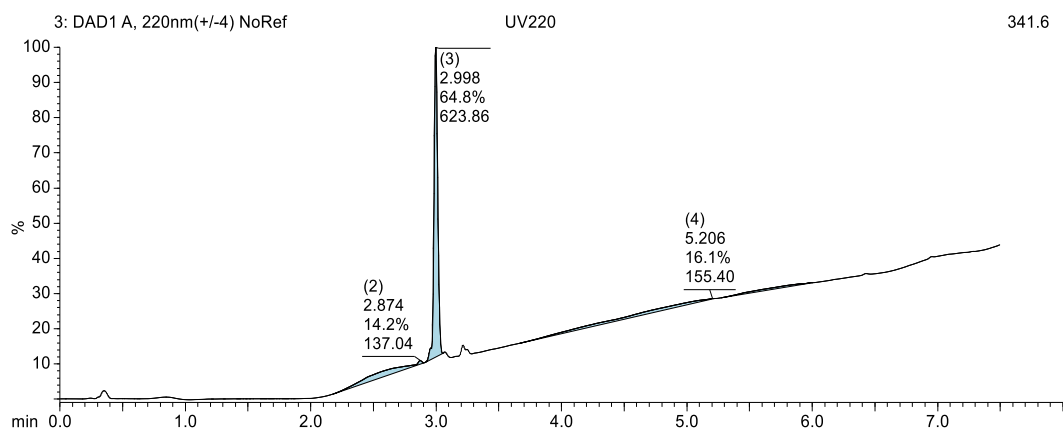
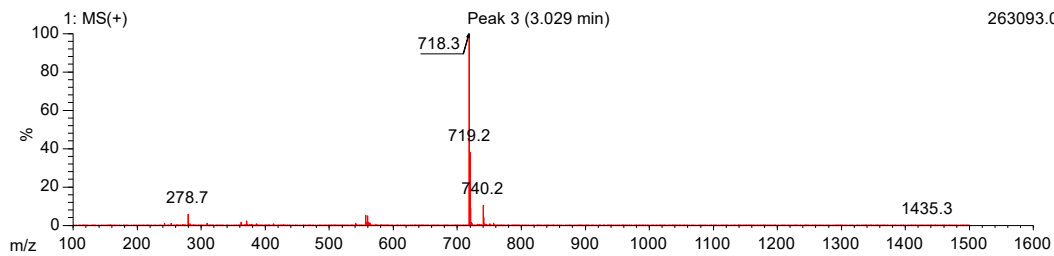
Glu-SIYG  
 Chemical Formula:  $C_{28}H_{42}N_6O_{13}$   
 Molecular Weight: 670.67



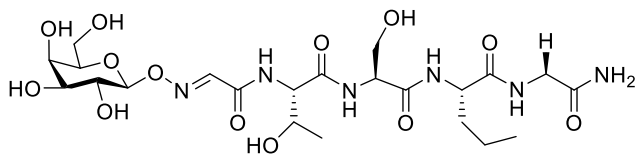
**Appendix A-6 LC-MS characterization of Glu-SIYG glycopeptide**



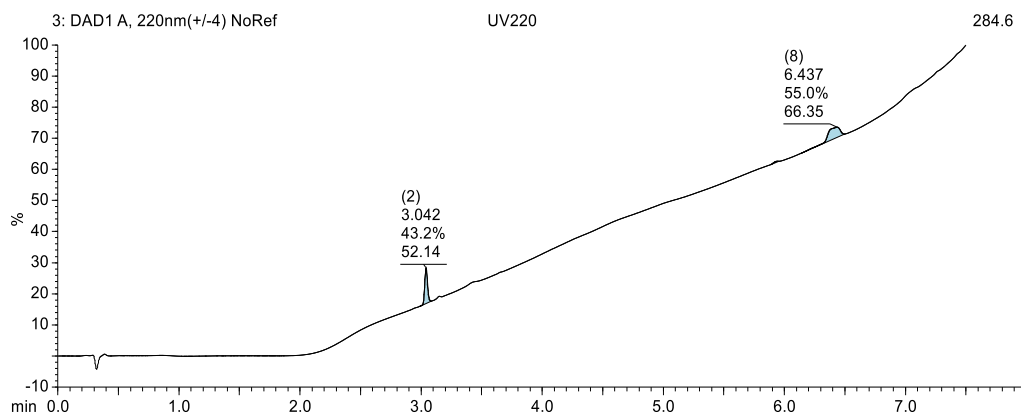
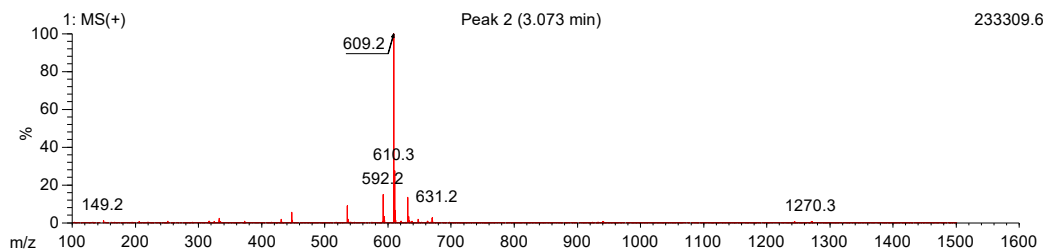
Gal-GWSH  
 Chemical Formula: C<sub>30</sub>H<sub>39</sub>N<sub>9</sub>O<sub>12</sub>  
 Molecular Weight: 717.69



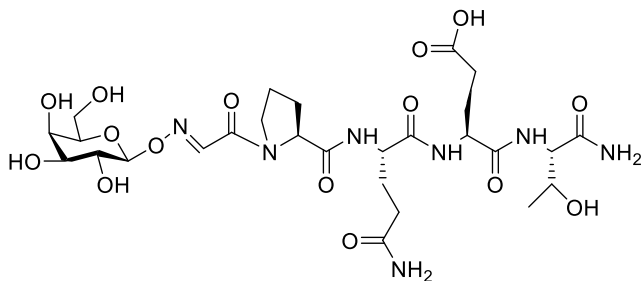
**Appendix A-7 LC-MS characterization of Gal-GWSH glycopeptide**



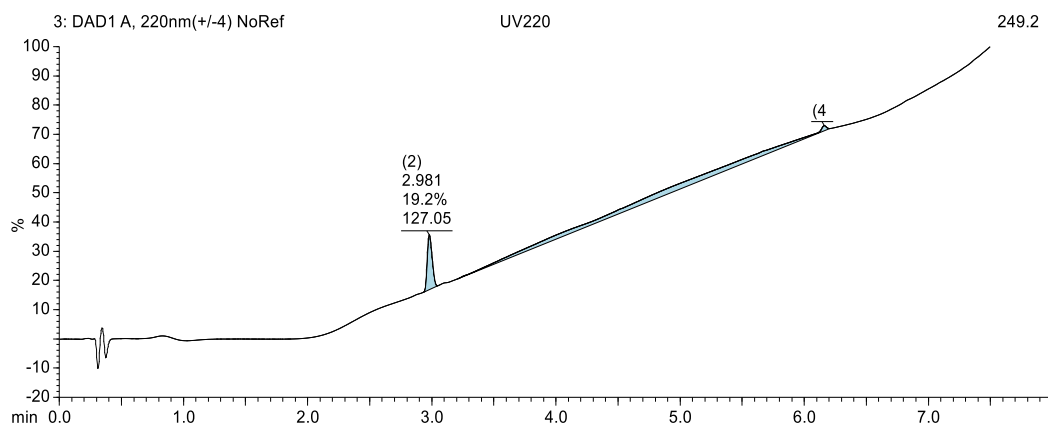
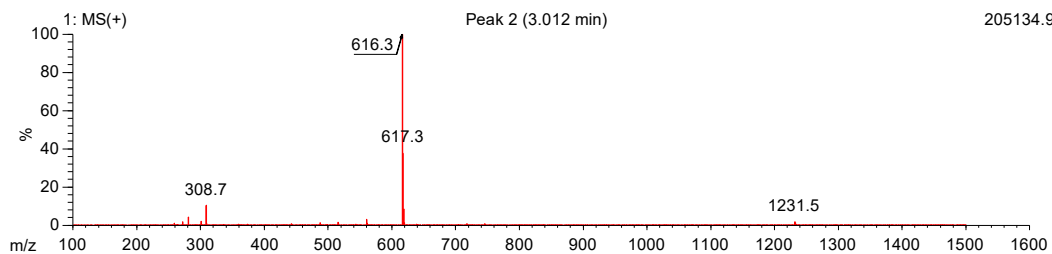
Gal-TSLG  
 Chemical Formula: C<sub>23</sub>H<sub>40</sub>N<sub>6</sub>O<sub>13</sub>  
 Molecular Weight: 608.60



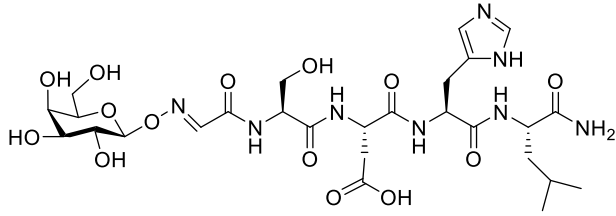
**Appendix A-8 LC-MS characterization of Gal-TSLG glycopeptide**



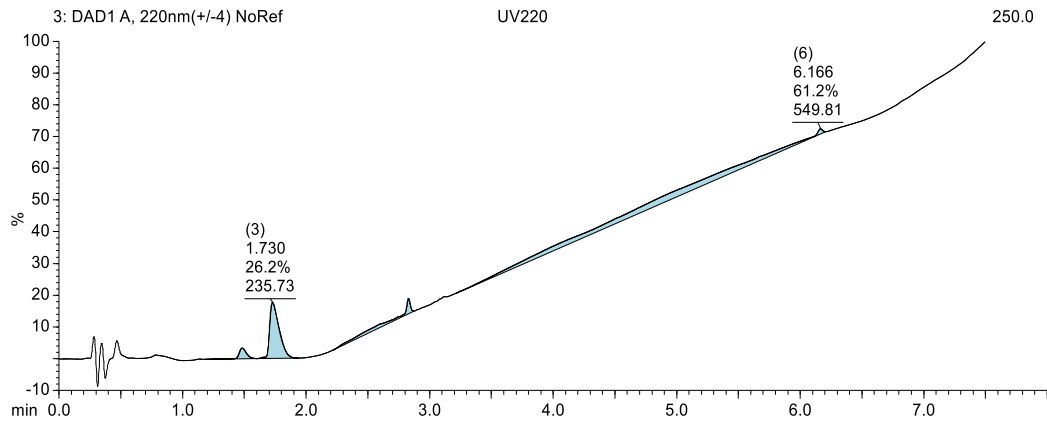
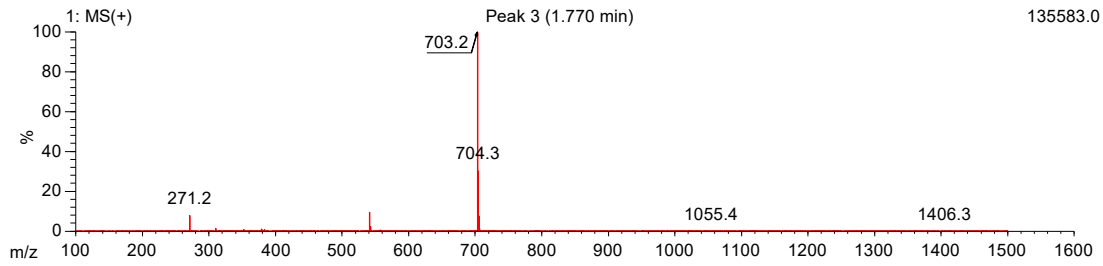
Gal-PQET  
 Chemical Formula: C<sub>27</sub>H<sub>43</sub>N<sub>7</sub>O<sub>15</sub>  
 Molecular Weight: 705.68



**Appendix A-9** LC-MS characterization of Gal-PQET glycopeptide

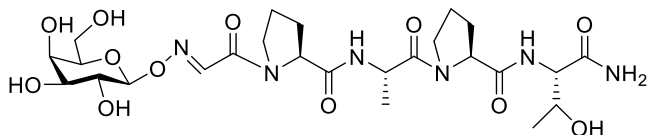


Gal-SDHL  
 Chemical Formula:  $C_{27}H_{42}N_8O_{14}$   
 Molecular Weight: 702.68

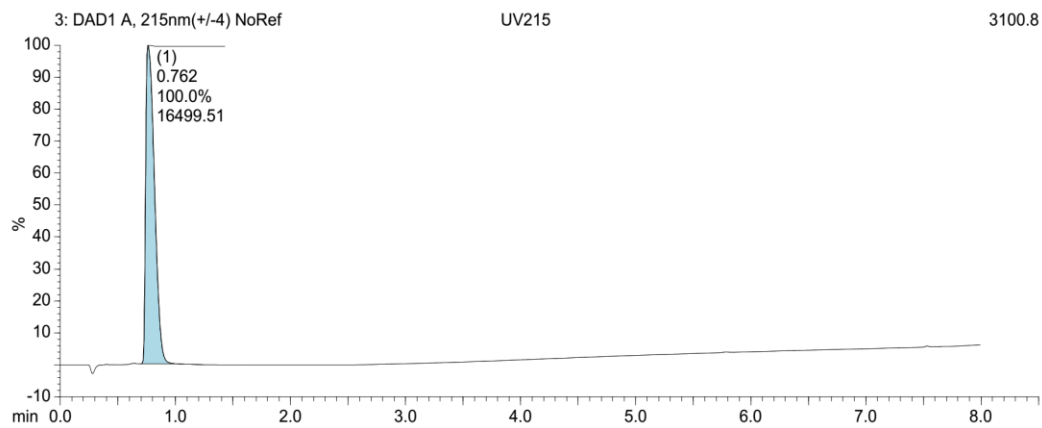
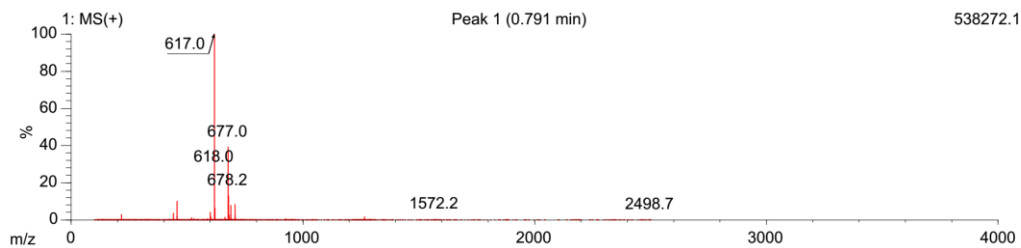


**Appendix A-10 LC-MS characterization of Gal-SDHL glycopeptide**

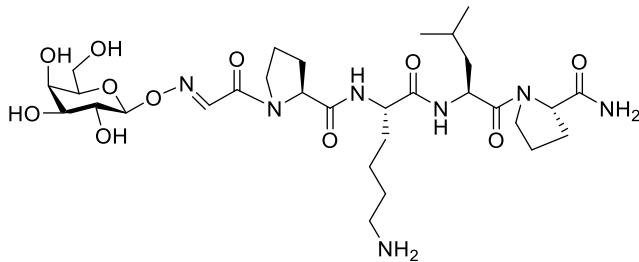




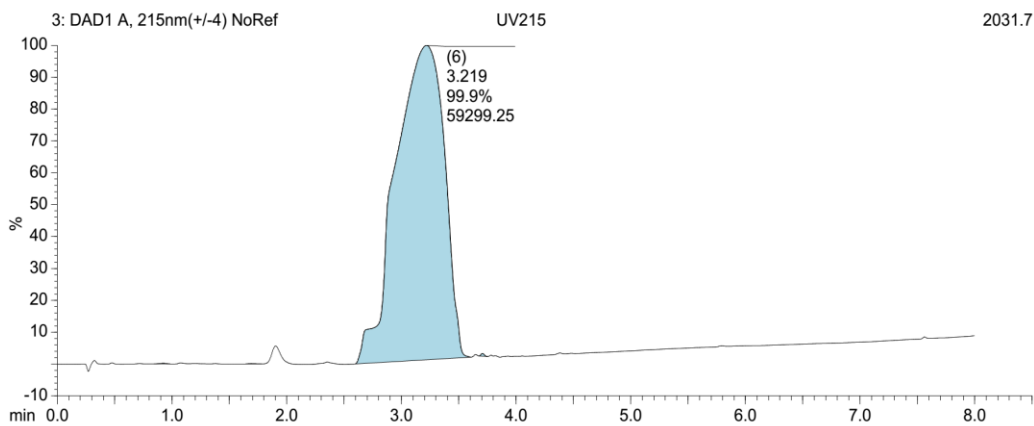
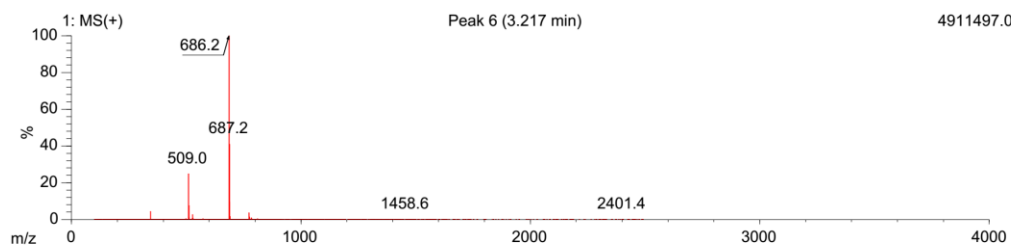
Gal-PAPT  
 Chemical Formula: C<sub>25</sub>H<sub>40</sub>N<sub>6</sub>O<sub>12</sub>  
 Molecular Weight: 616.63



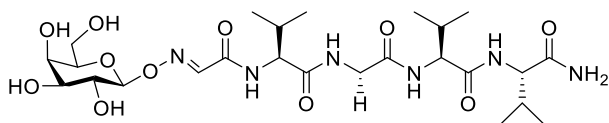
**Appendix A-11 LC-MS characterization of Gal-PAPT glycopeptide**



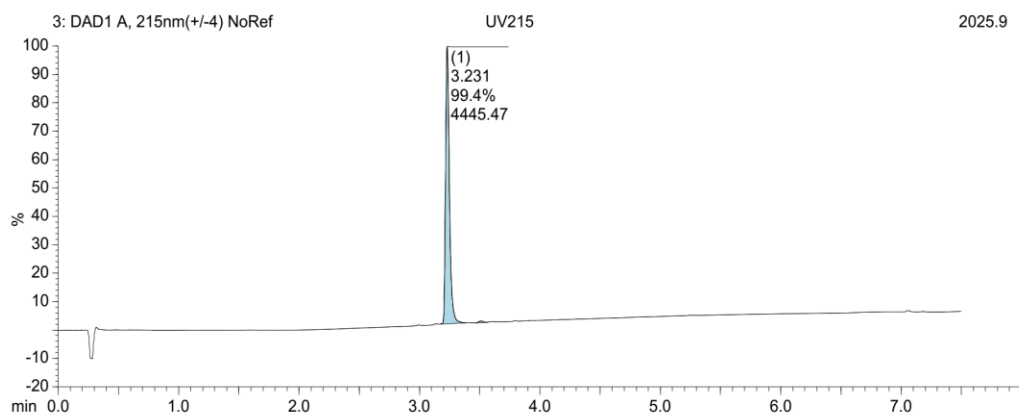
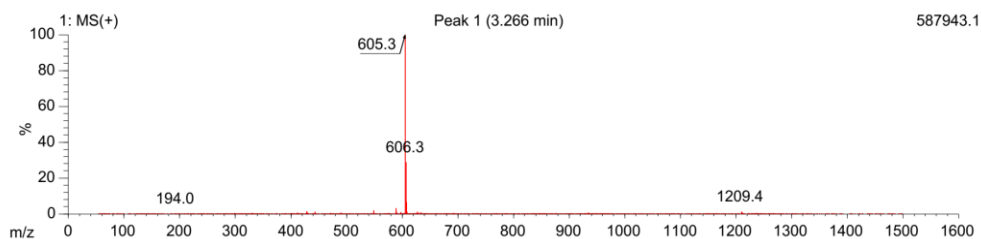
Gal-PKLP  
 Chemical Formula:  $C_{30}H_{51}N_7O_{11}$   
 Molecular Weight: 685.78



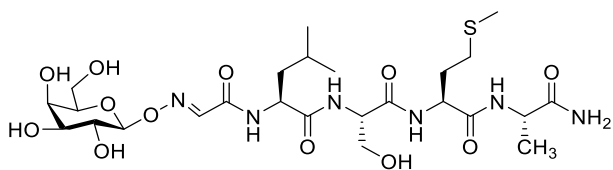
**Appendix A-12 LC-MS characterization of Gal-PKLP glycopeptide**



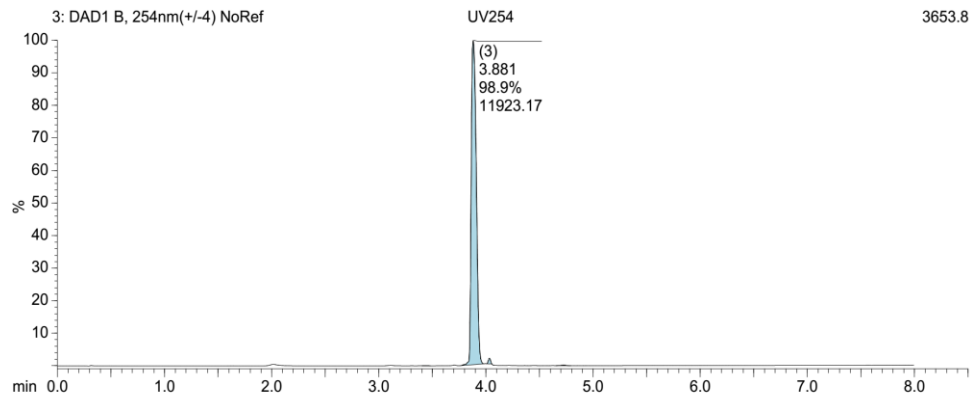
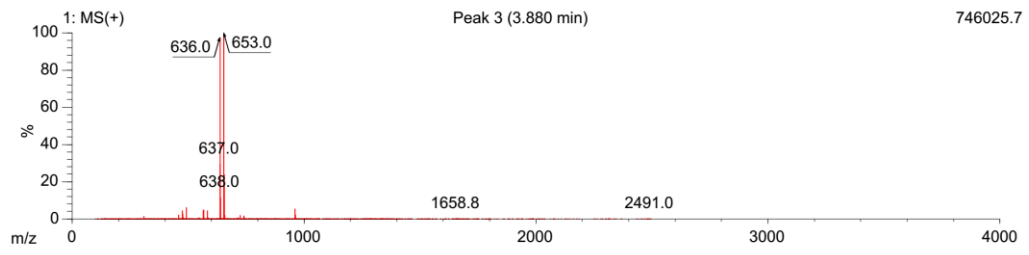
Gal-VGVV  
 Chemical Formula:  $C_{25}H_{44}N_6O_{11}$   
 Molecular Weight: 604.66



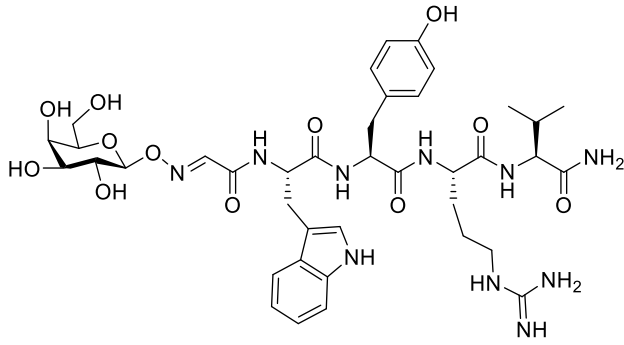
**Appendix A-13 LC-MS characterization of Gal-VGVV glycopeptide**



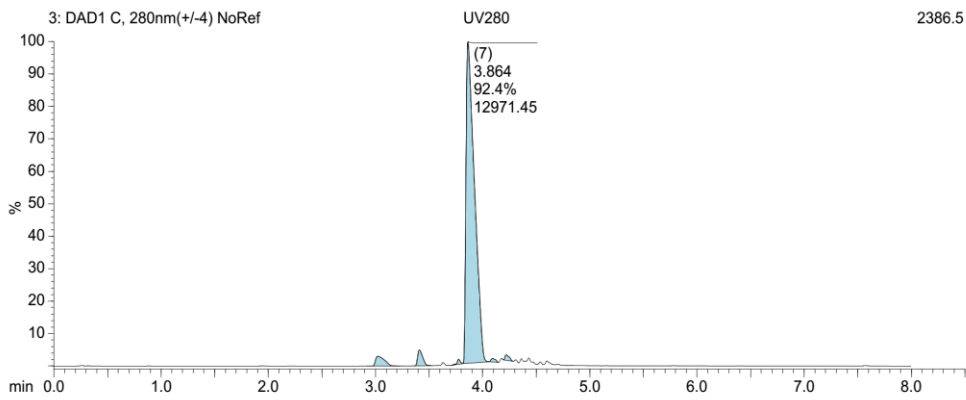
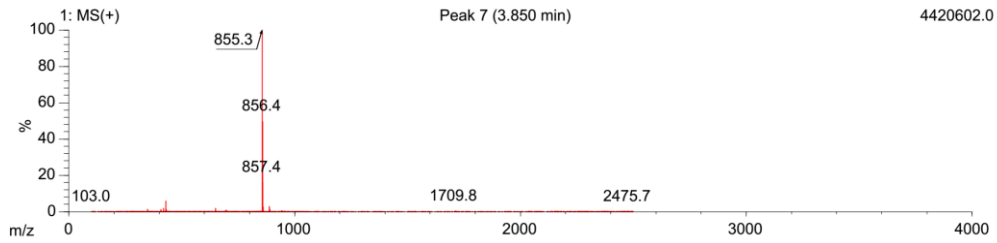
Gal-LSMA  
 Chemical Formula: C<sub>25</sub>H<sub>44</sub>N<sub>6</sub>O<sub>12</sub>S  
 Molecular Weight: 652.72



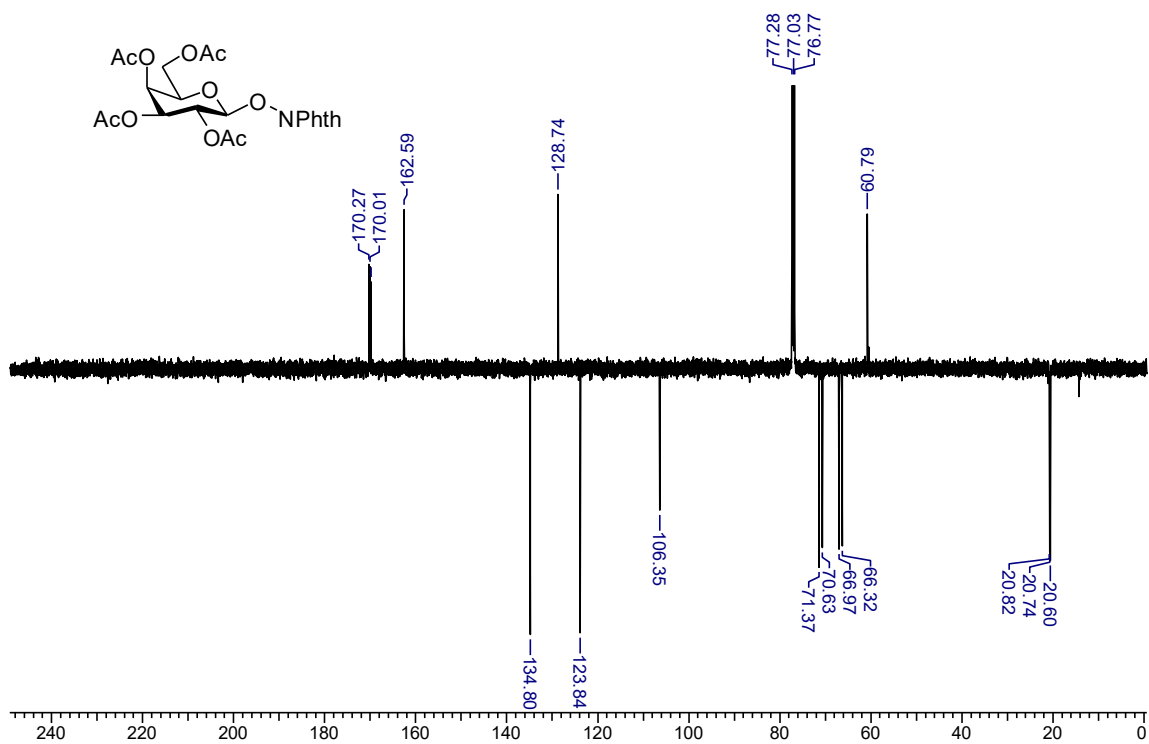
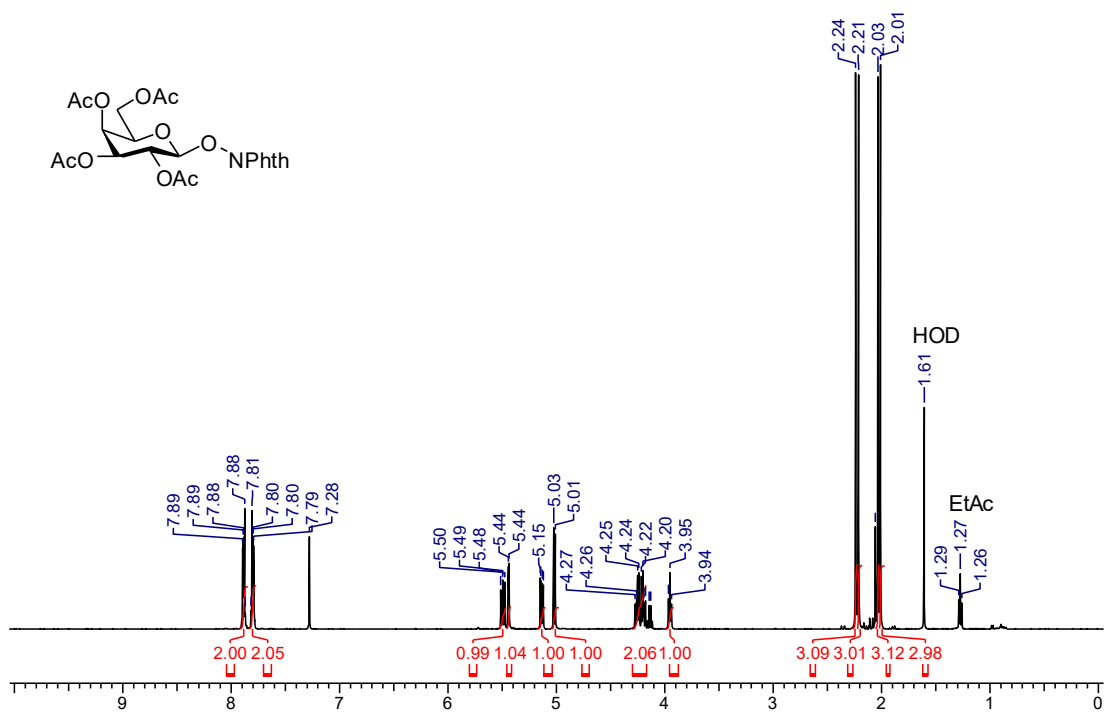
**Appendix A-14** LC-MS characterization of Gal-LSMA glycopeptide

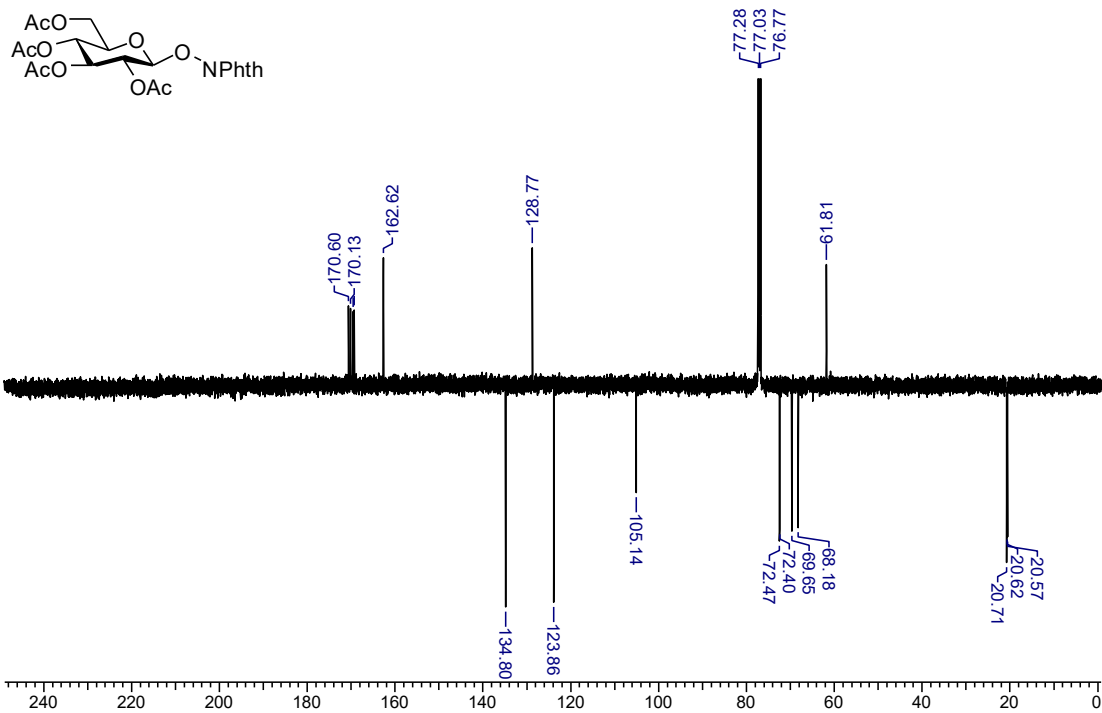
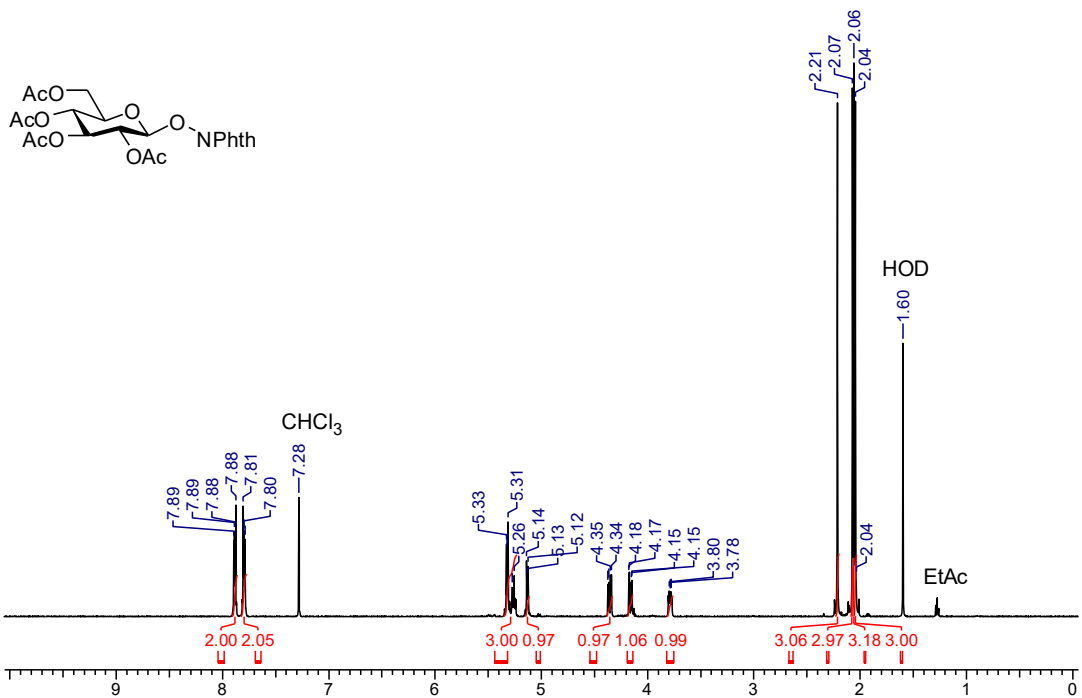


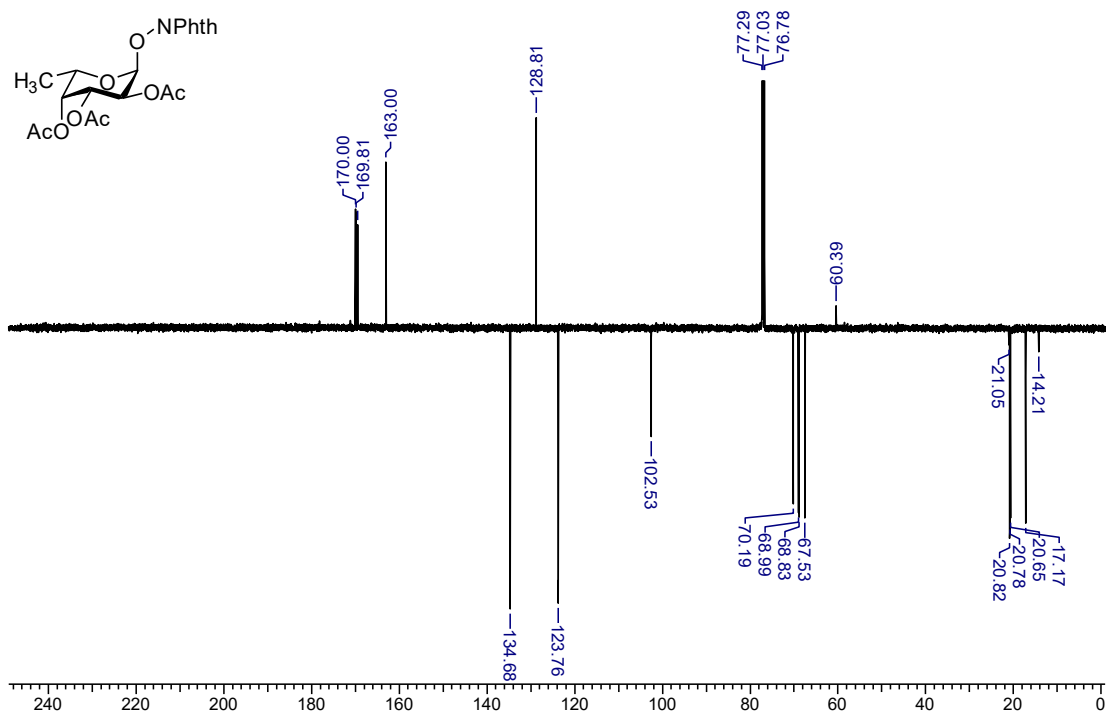
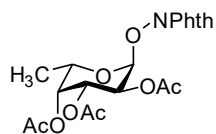
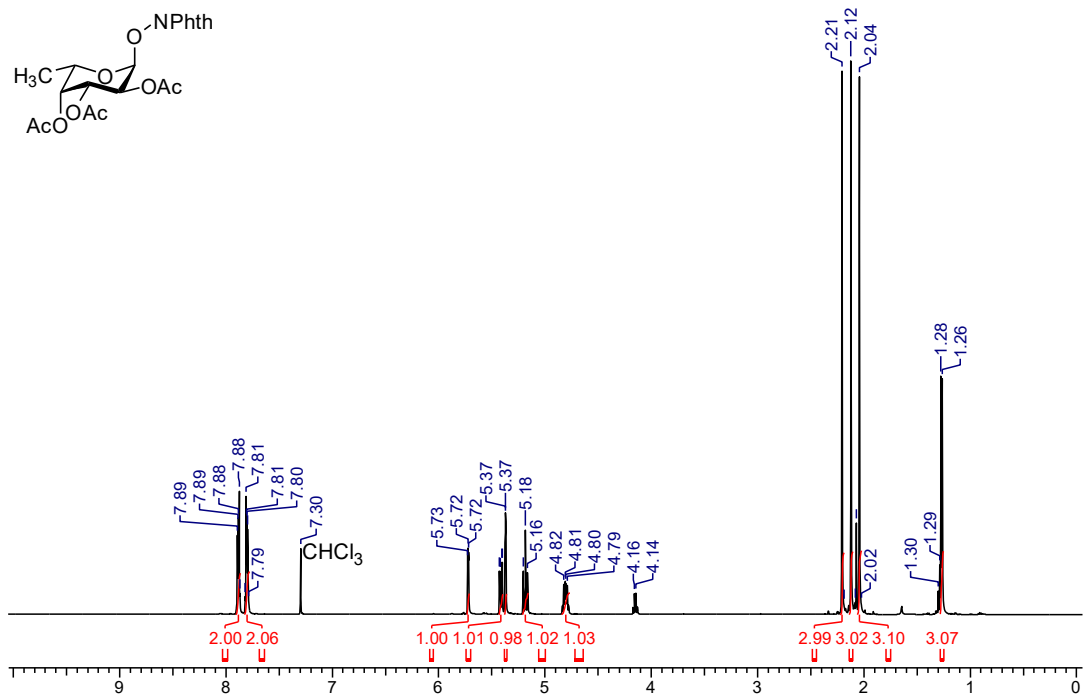
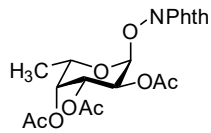
Gal-WYRV  
 Chemical Formula: C<sub>39</sub>H<sub>54</sub>N<sub>10</sub>O<sub>12</sub>  
 Molecular Weight: 854.92



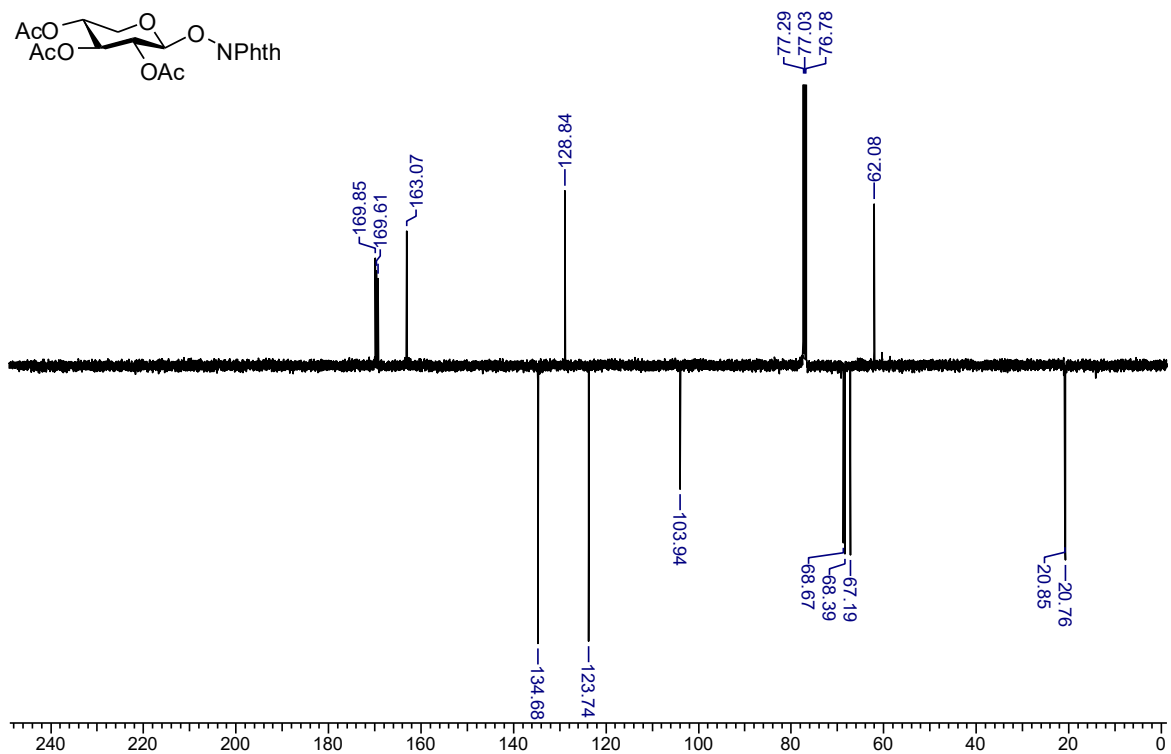
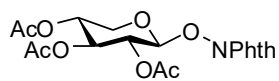
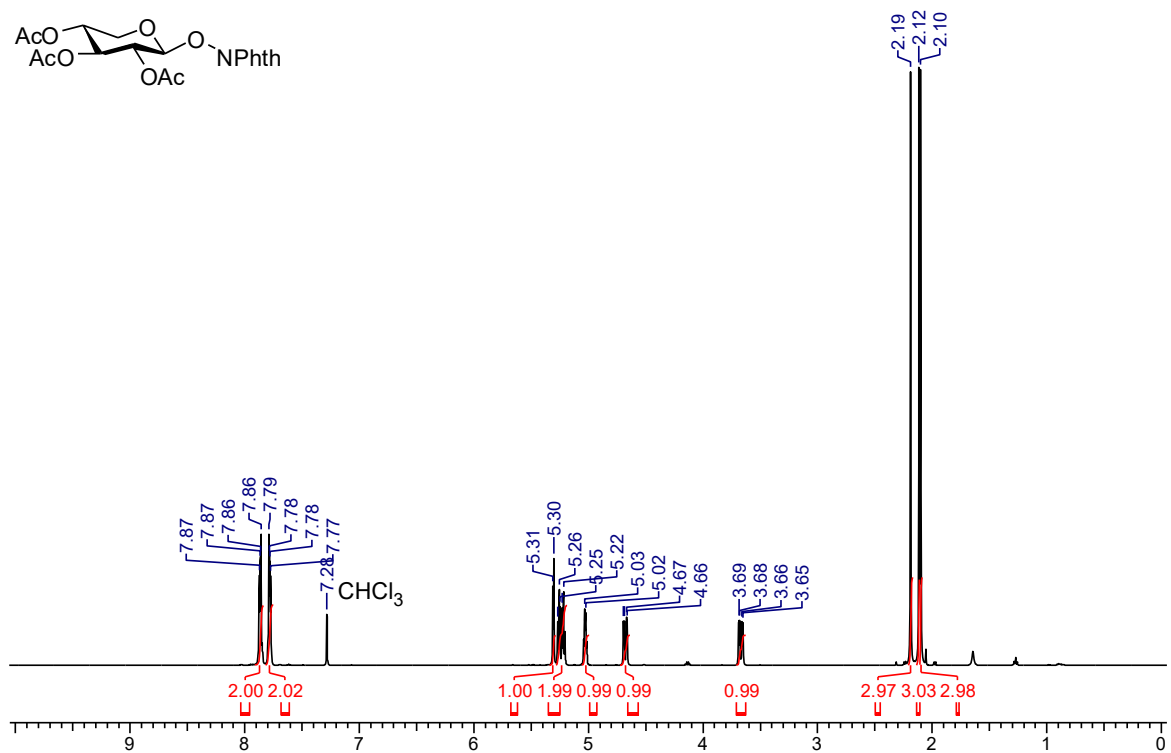
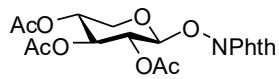
**Appendix A-15** LC-MS characterization of Gal-WYRV glycopeptide

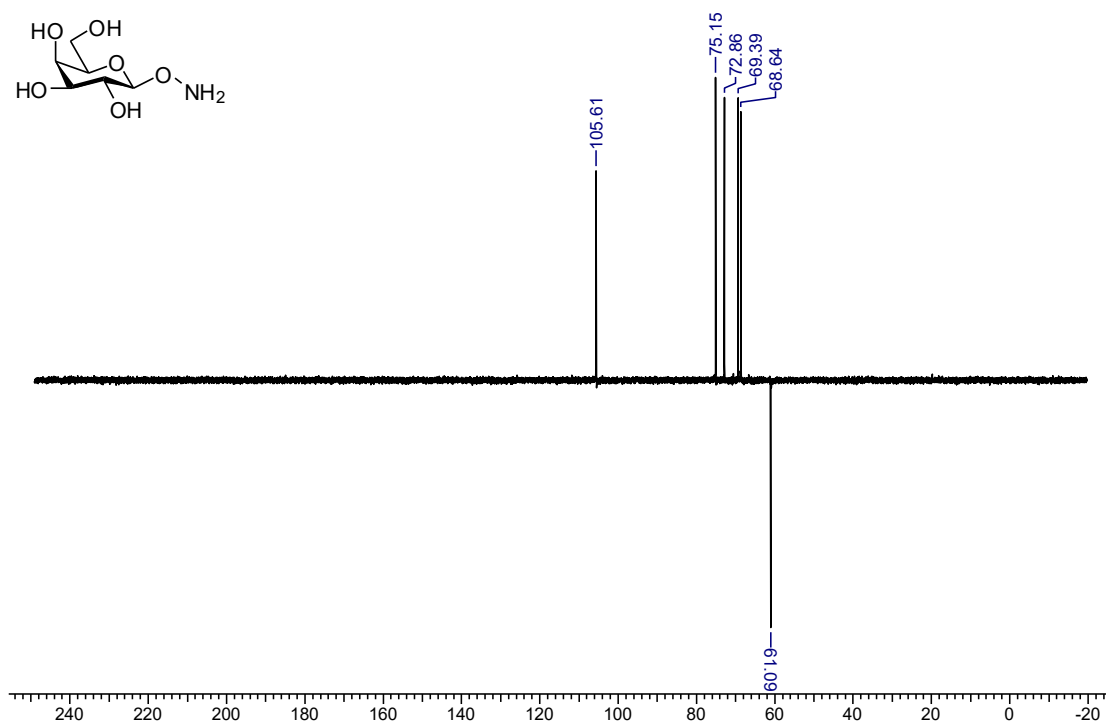
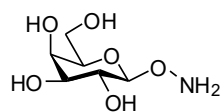
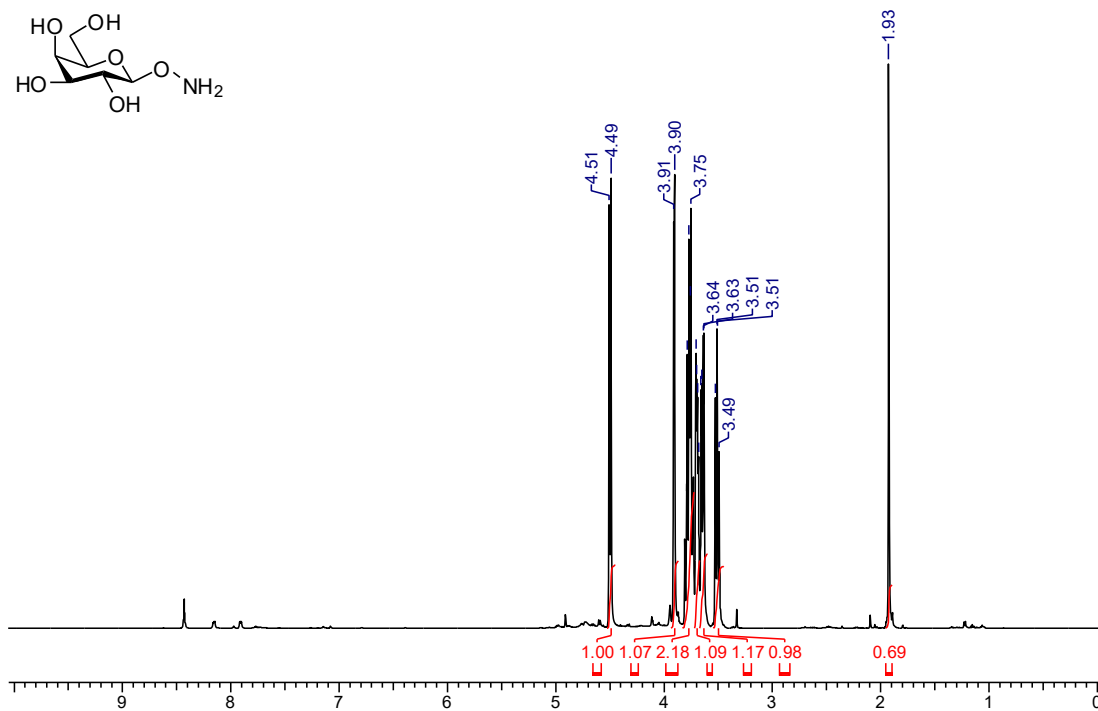
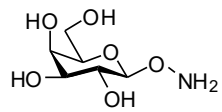


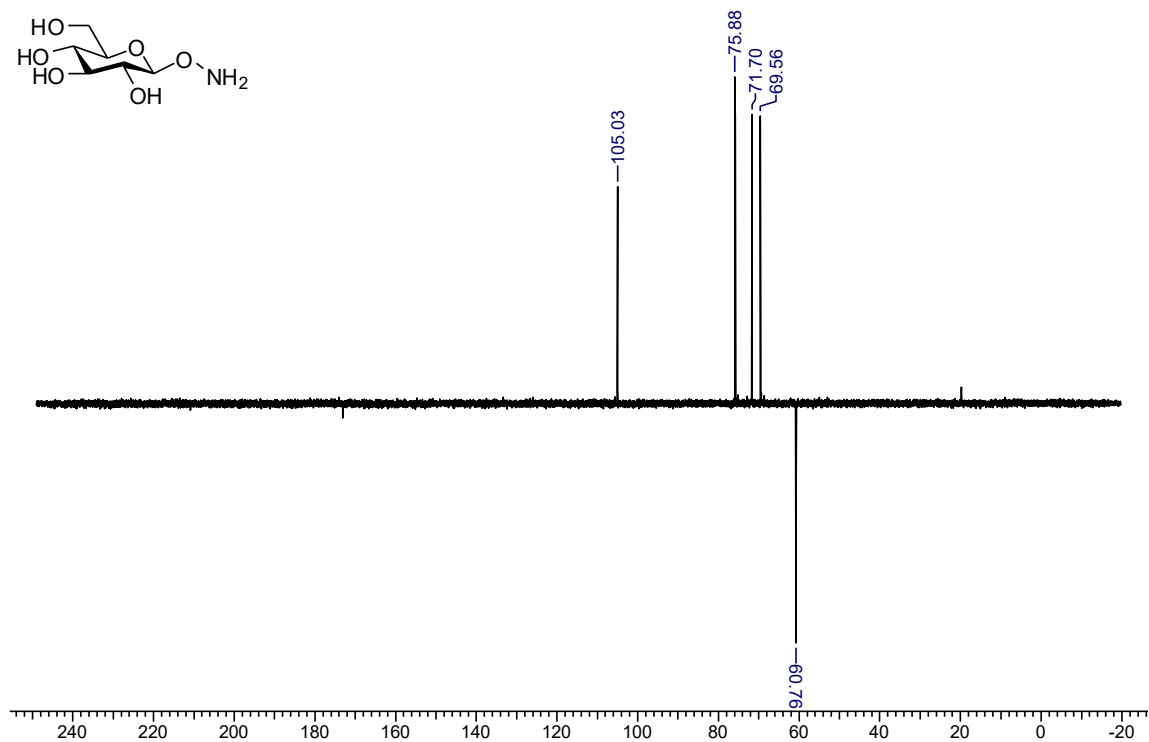
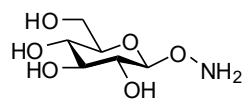
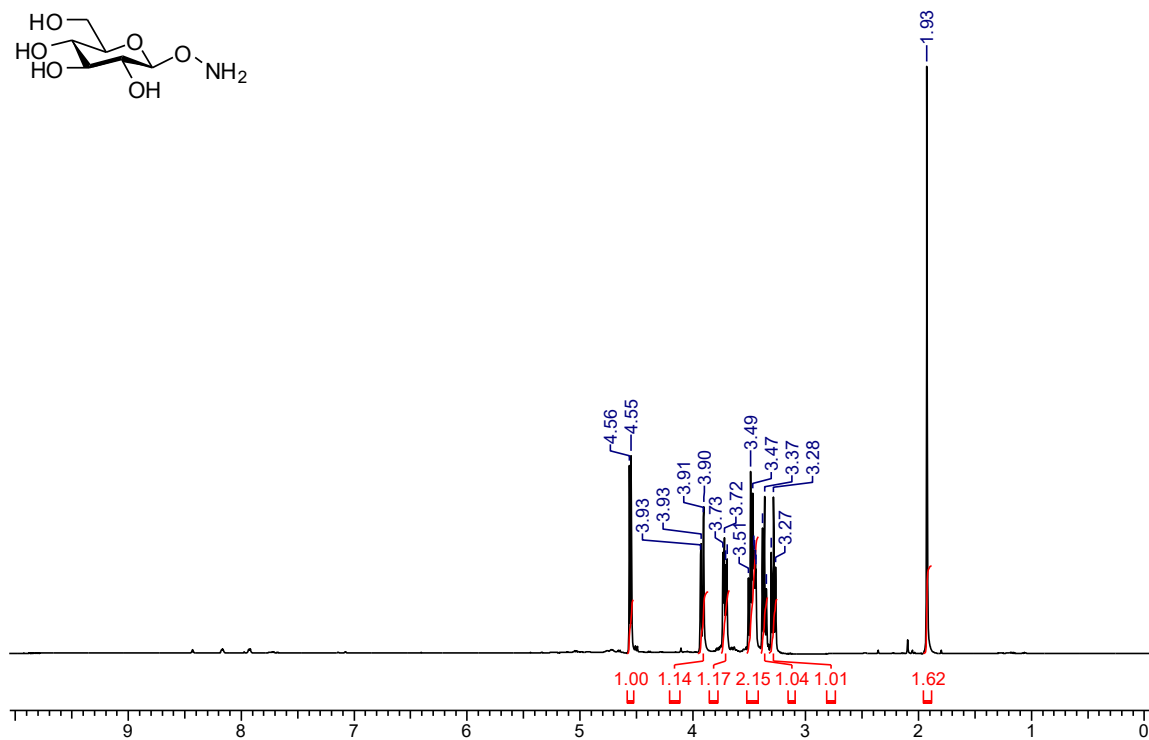
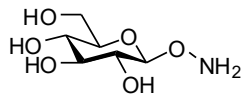


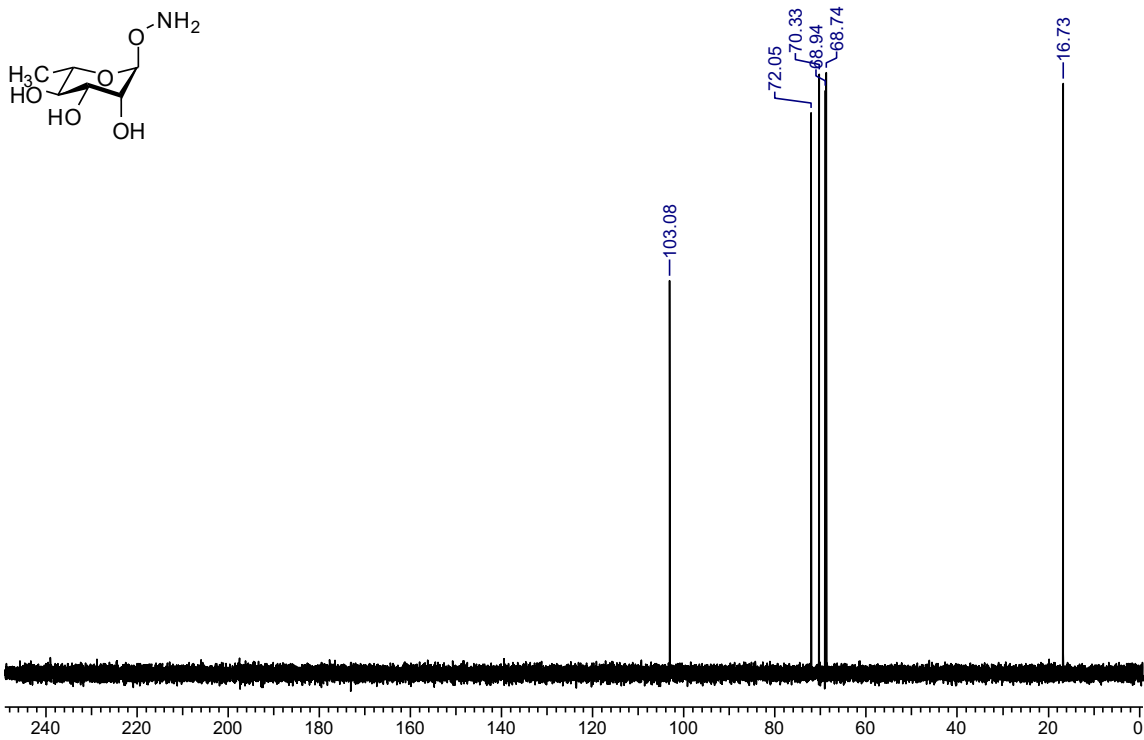
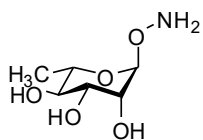
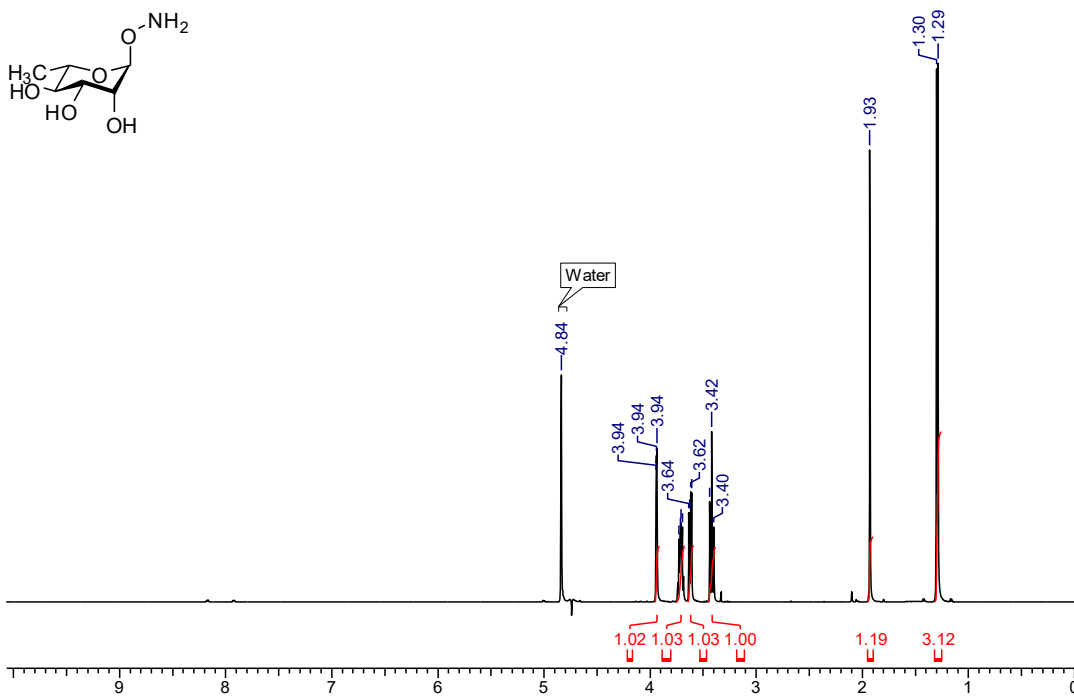
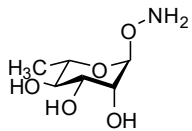


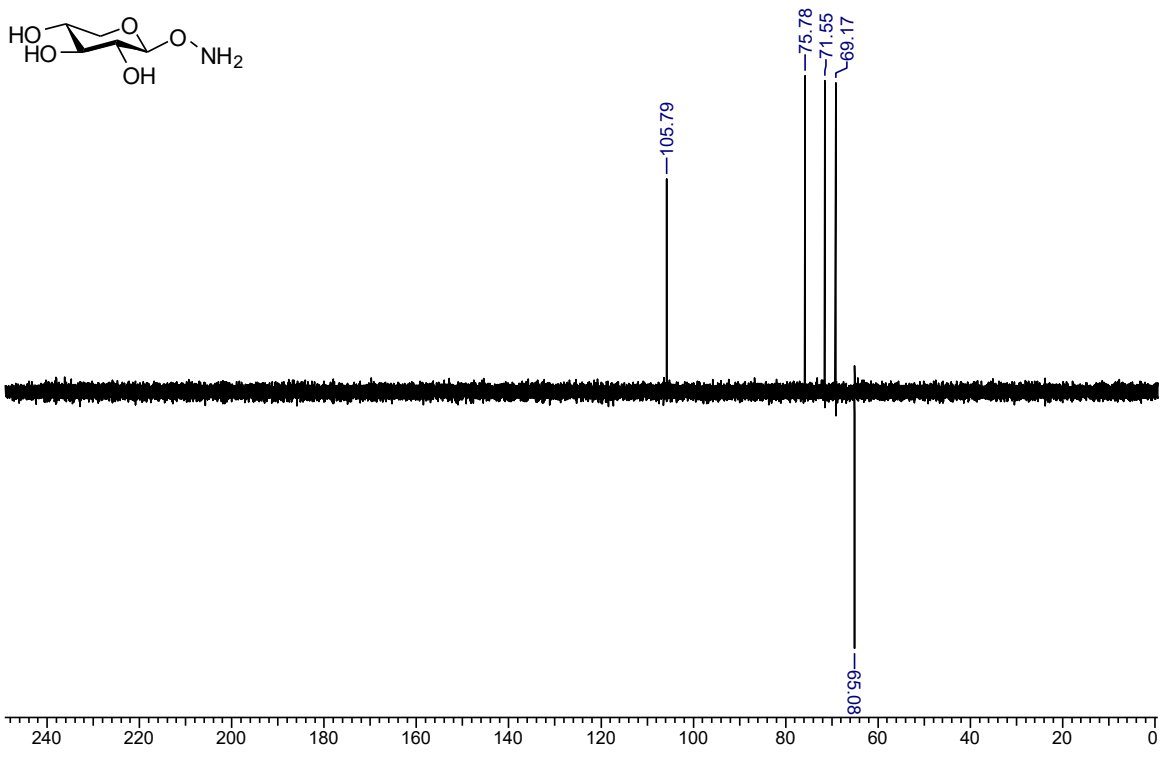
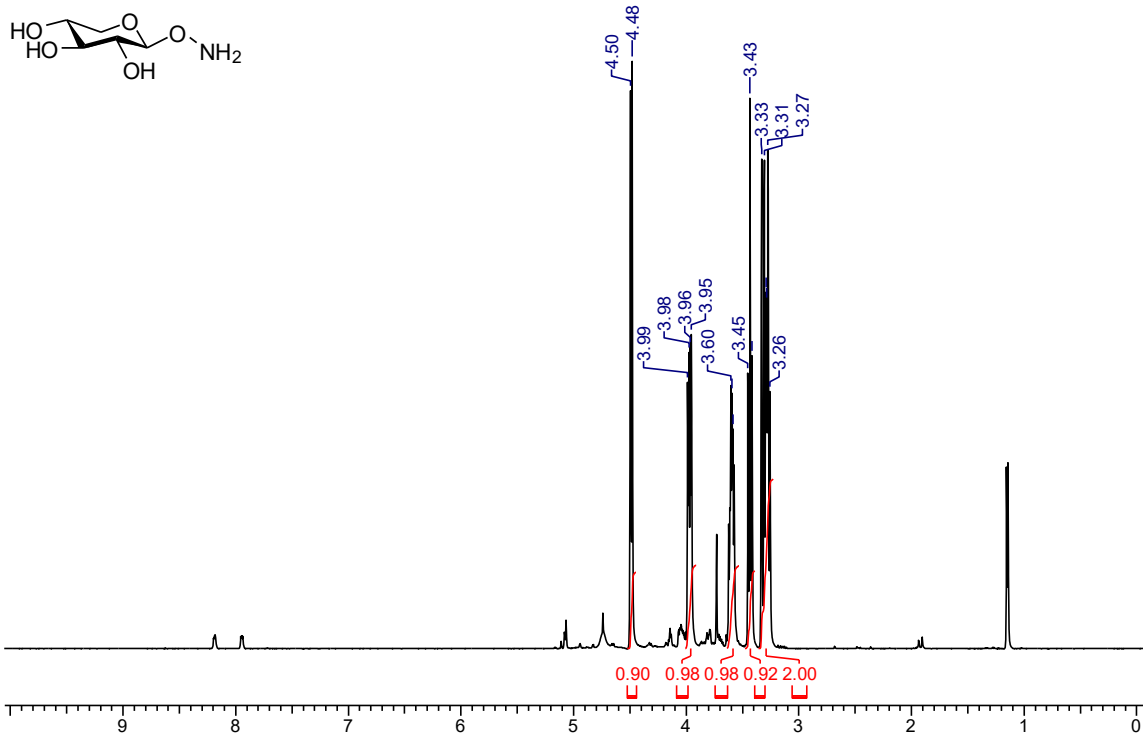


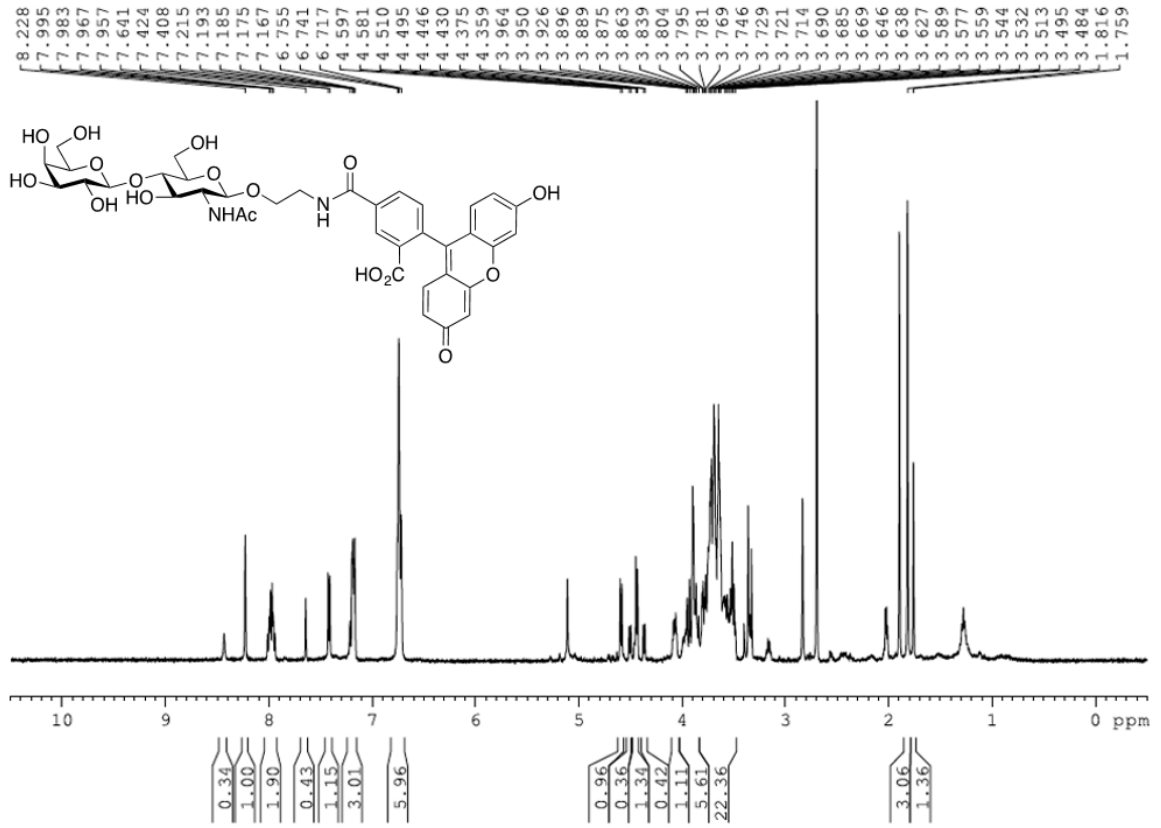


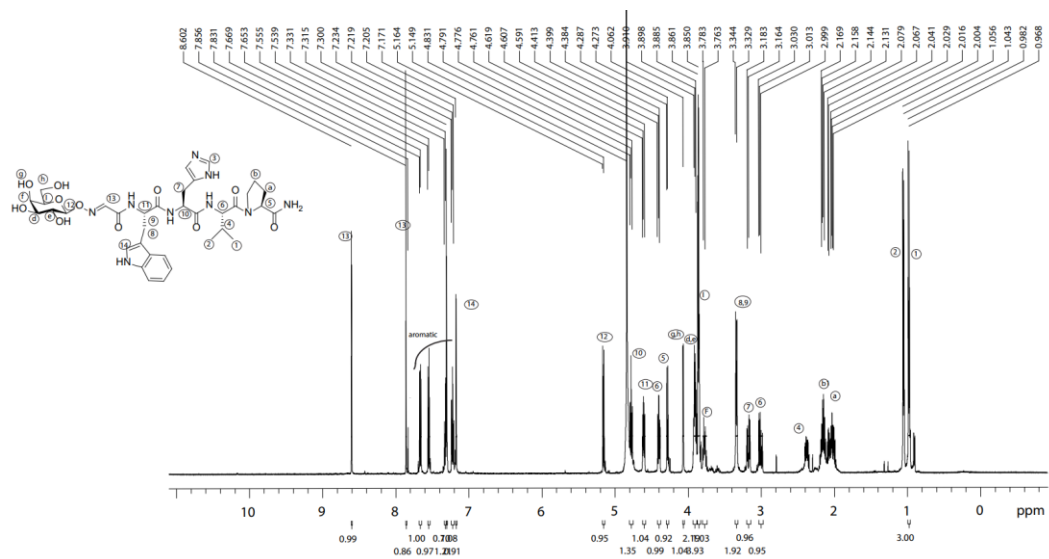


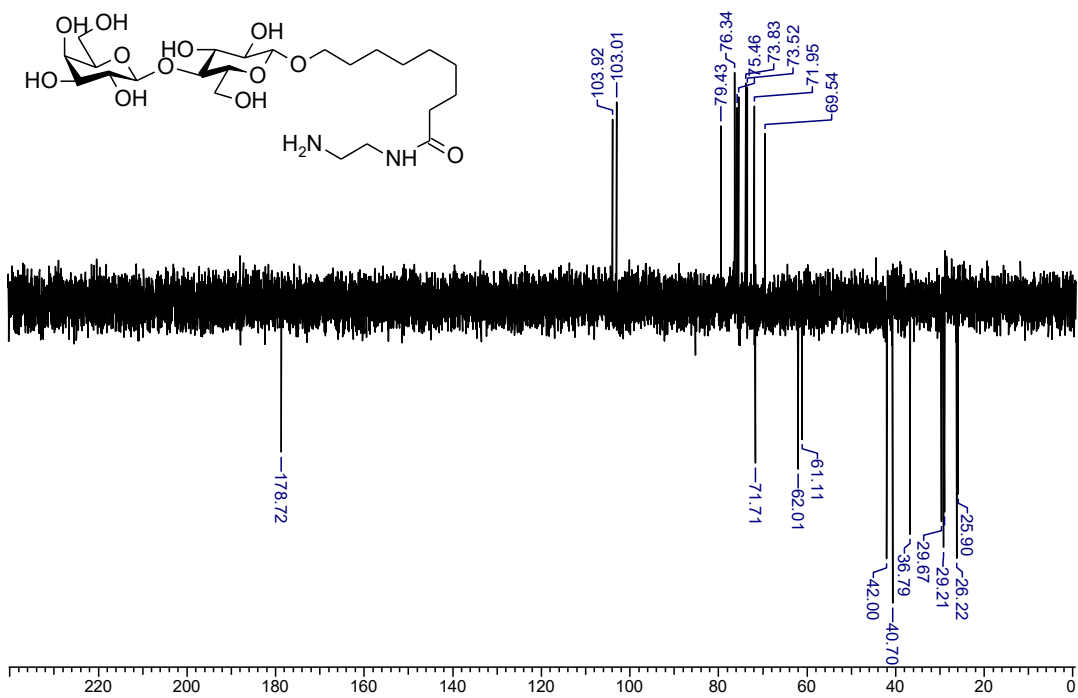
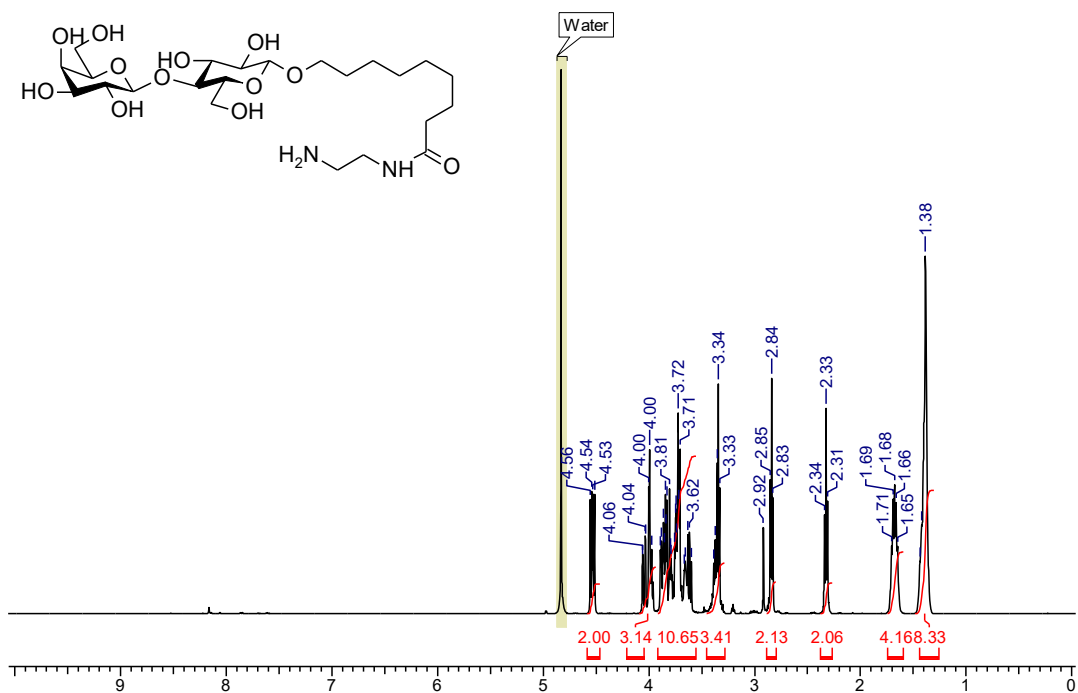




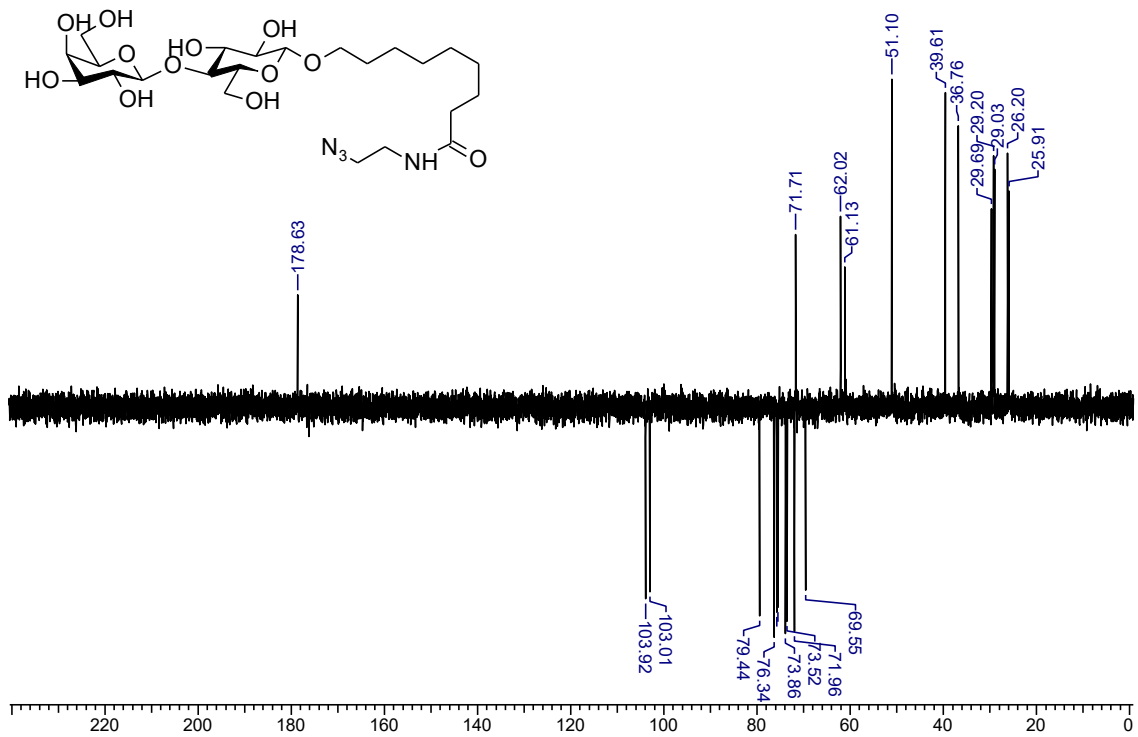
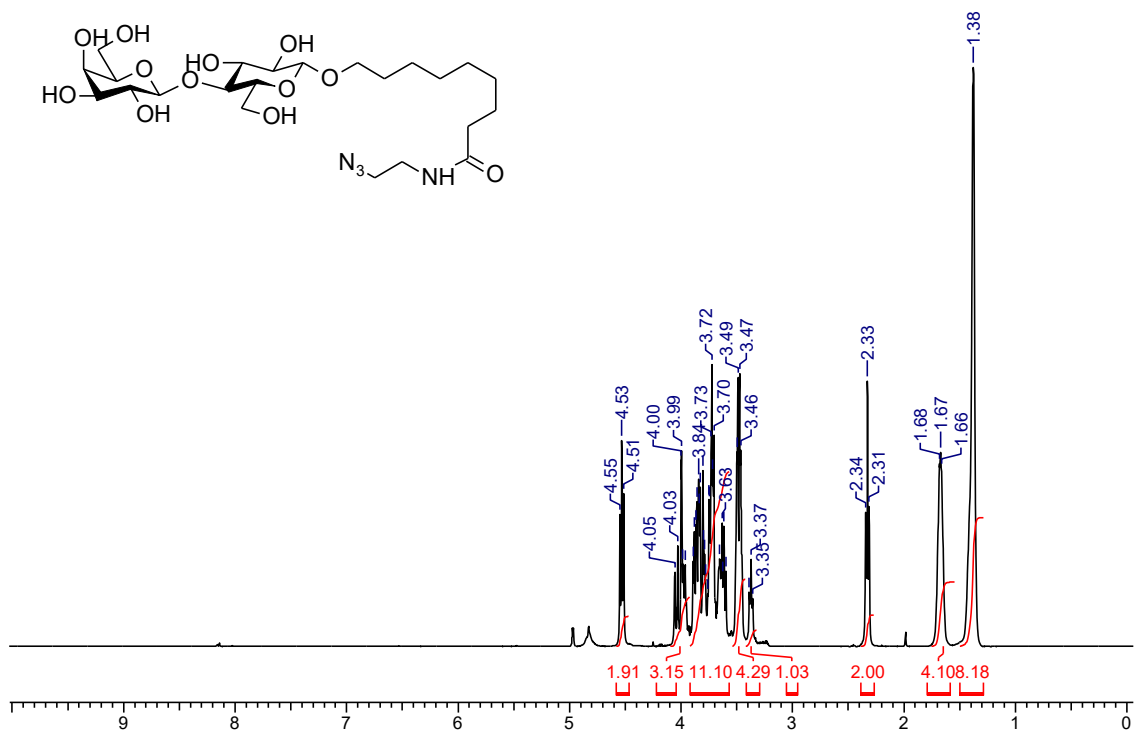


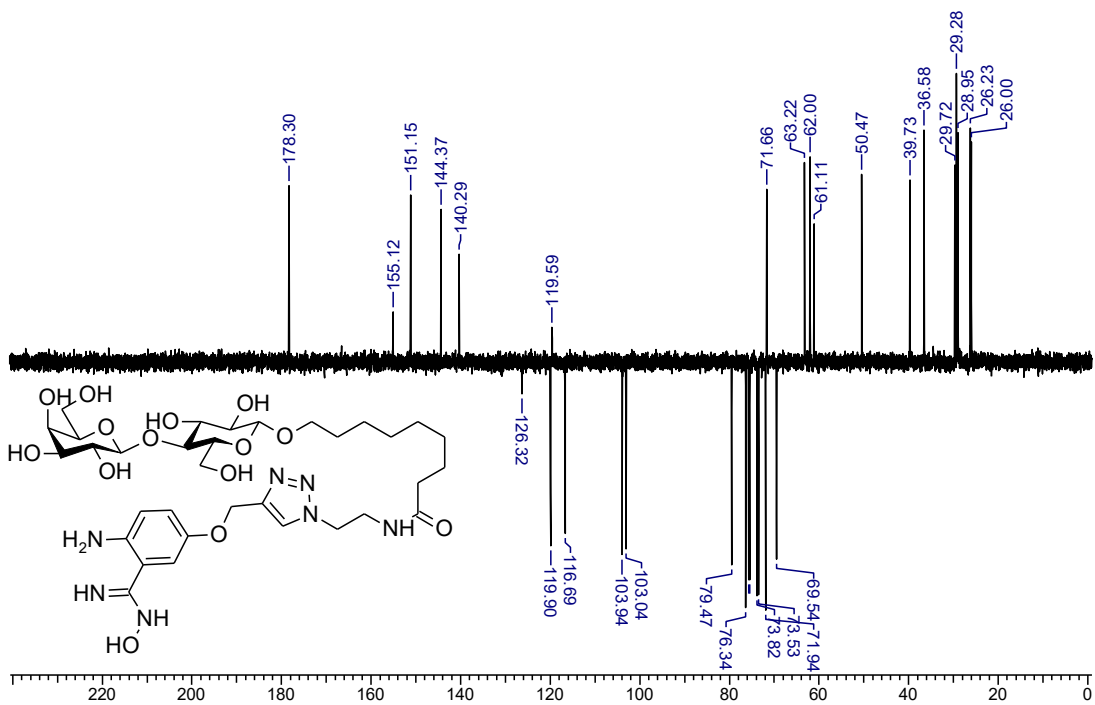
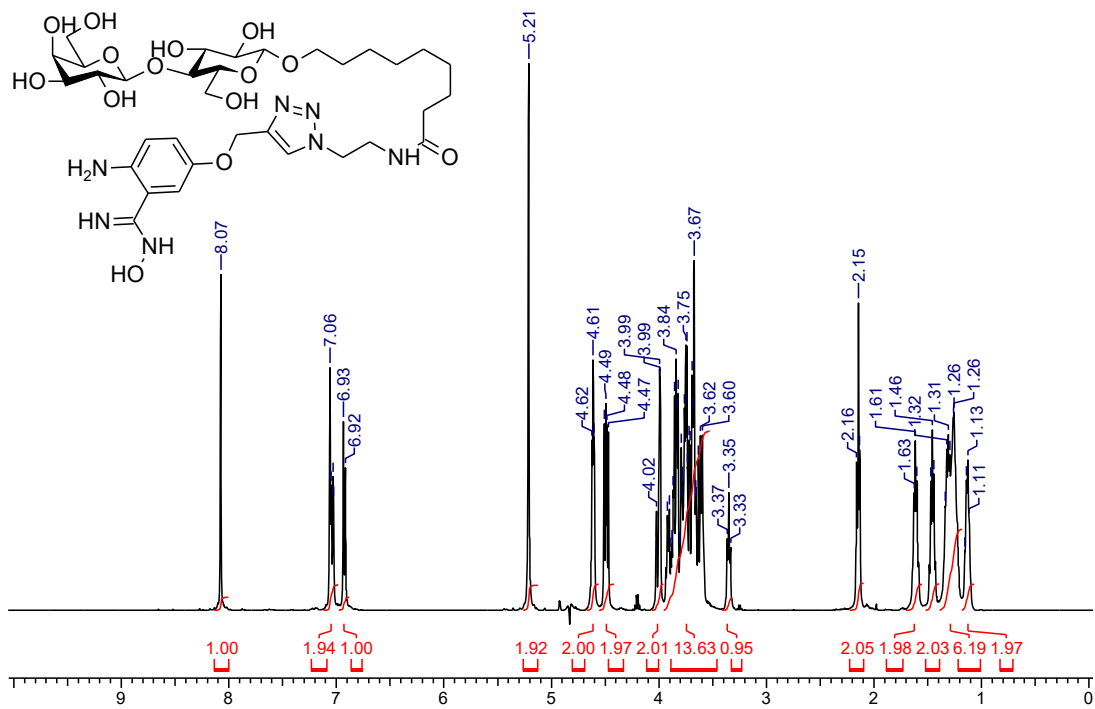




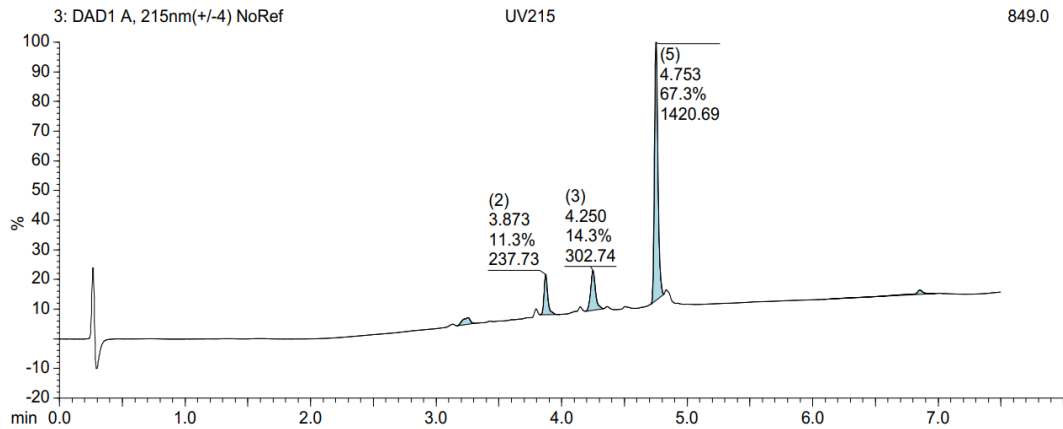
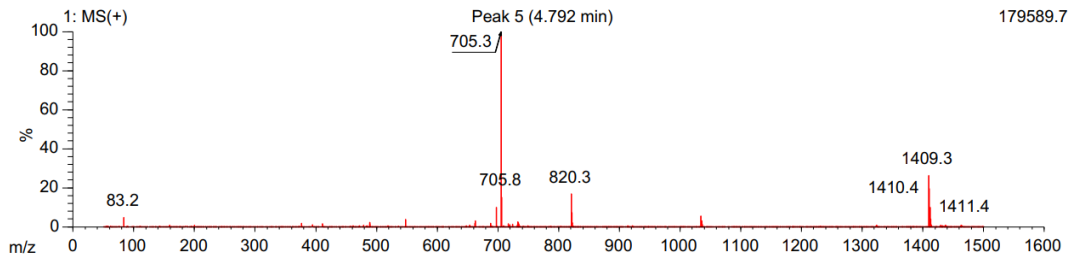
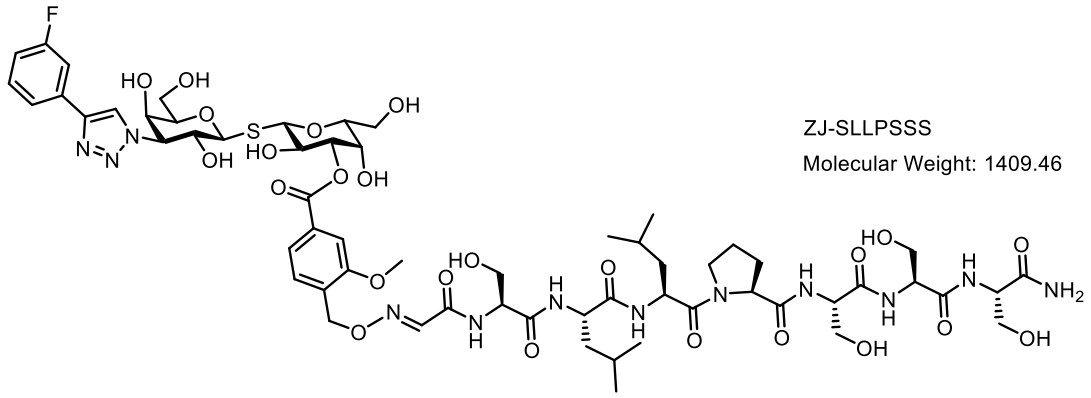




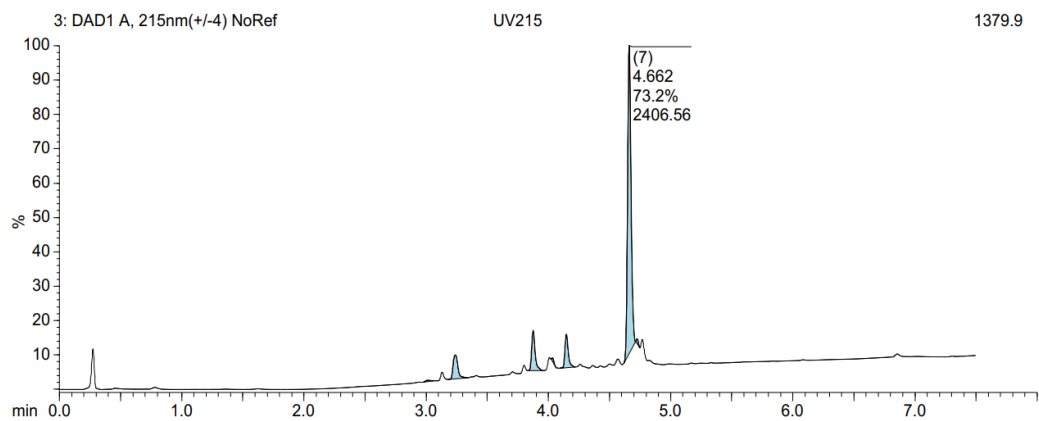
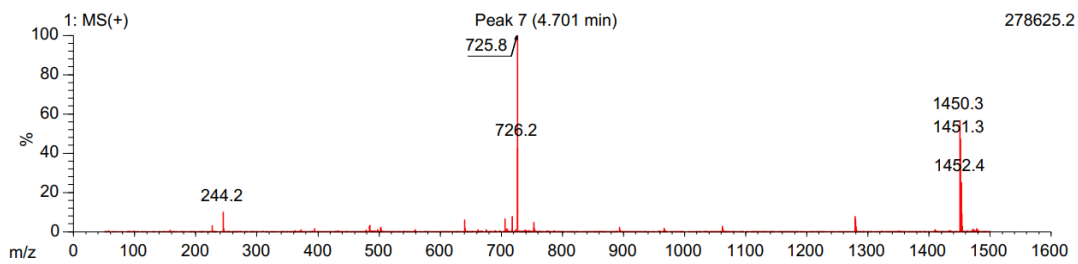
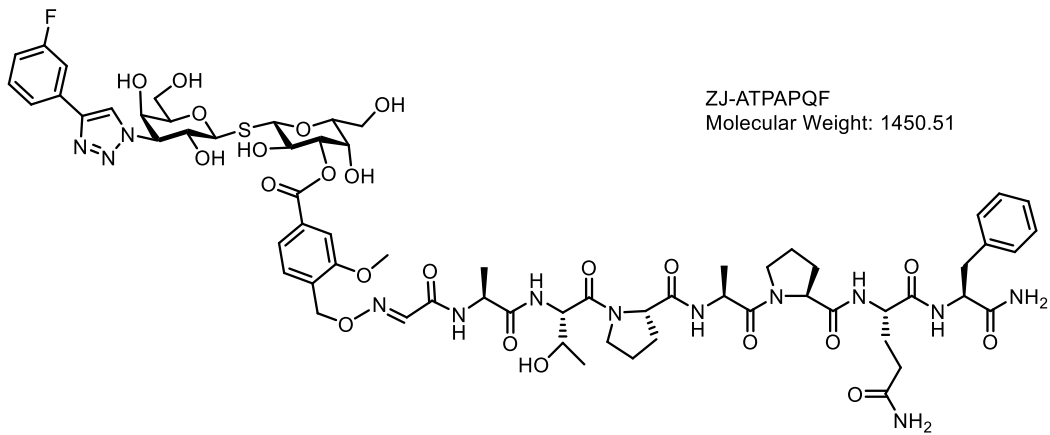




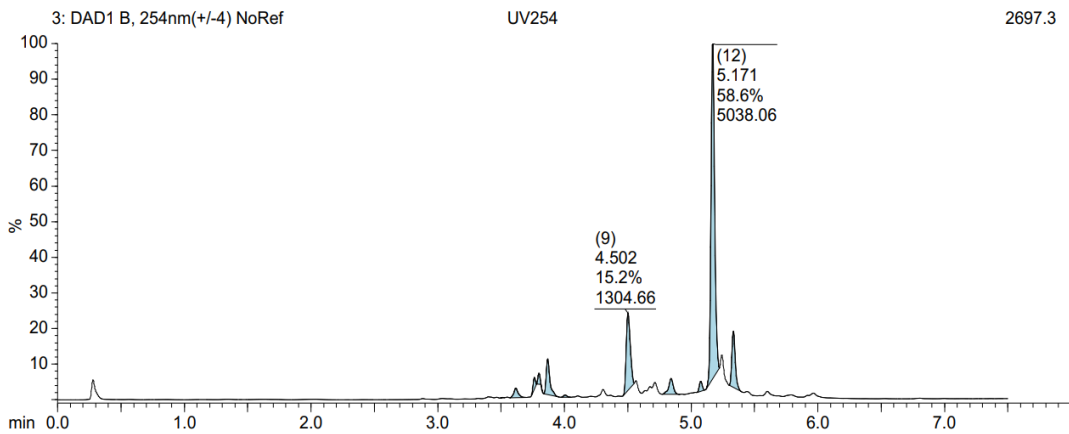
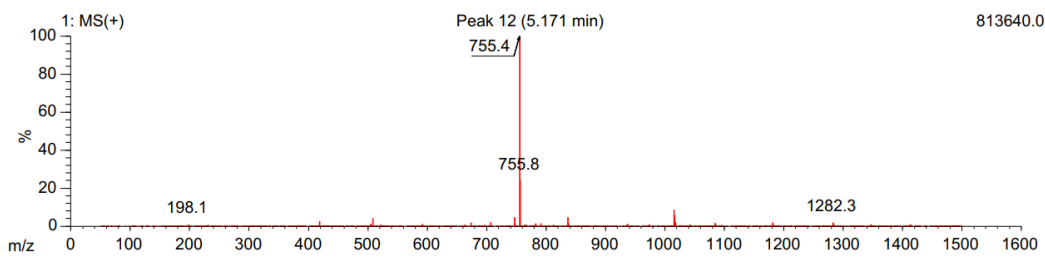
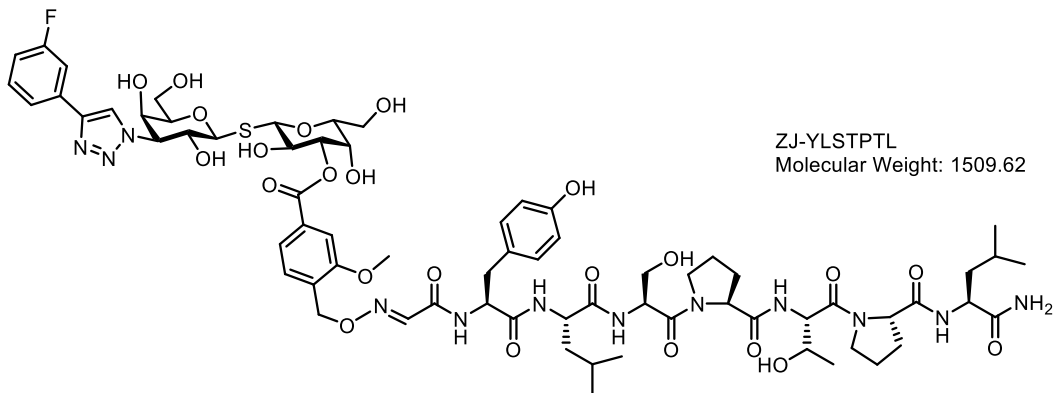
## Appendix B: Supporting information for Chapter 3



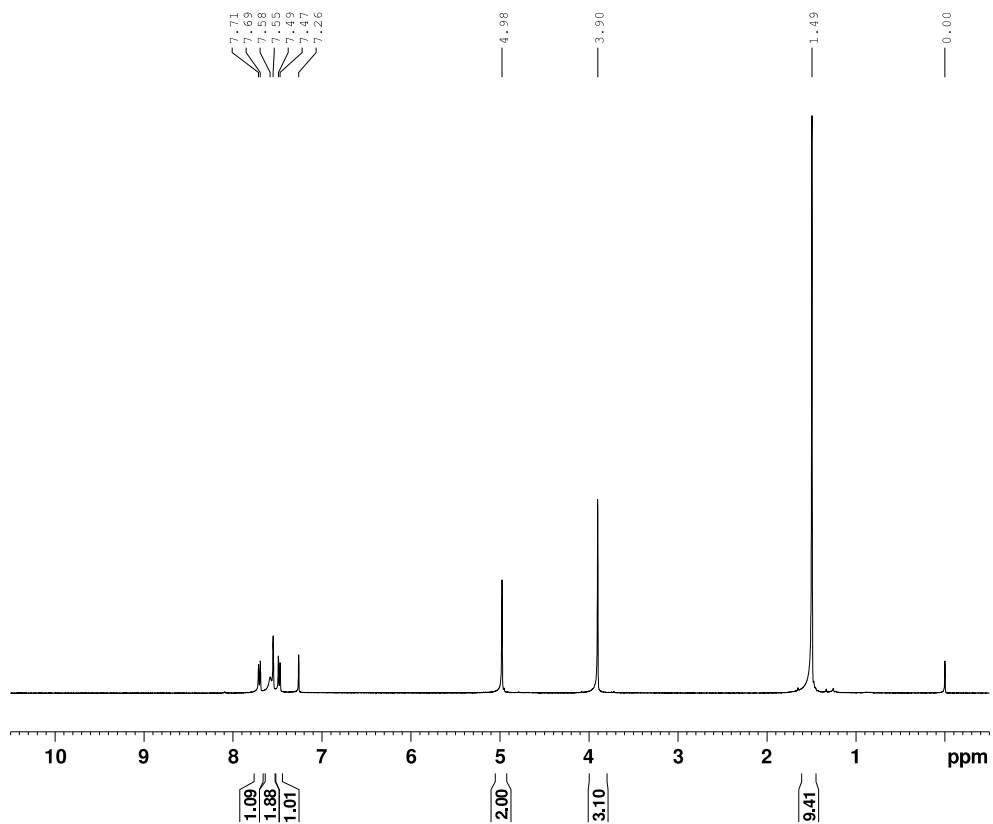
Appendix B-1 LC-MS characterization of ZJ-SLLPSS glycopeptide.



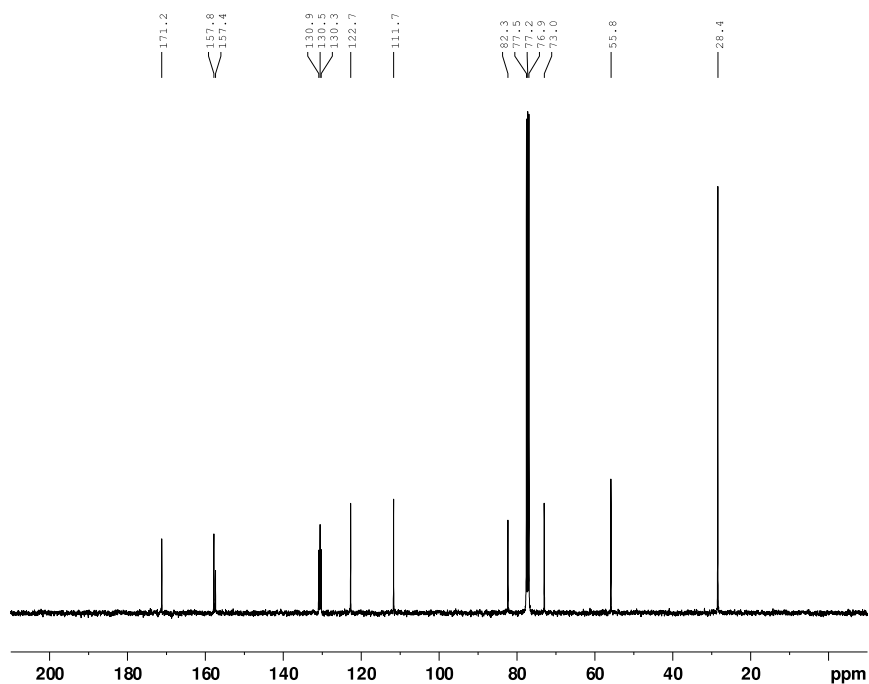
**Appendix B-2** LC-MS characterization of ZJ-ATPAPQF glycopeptide.



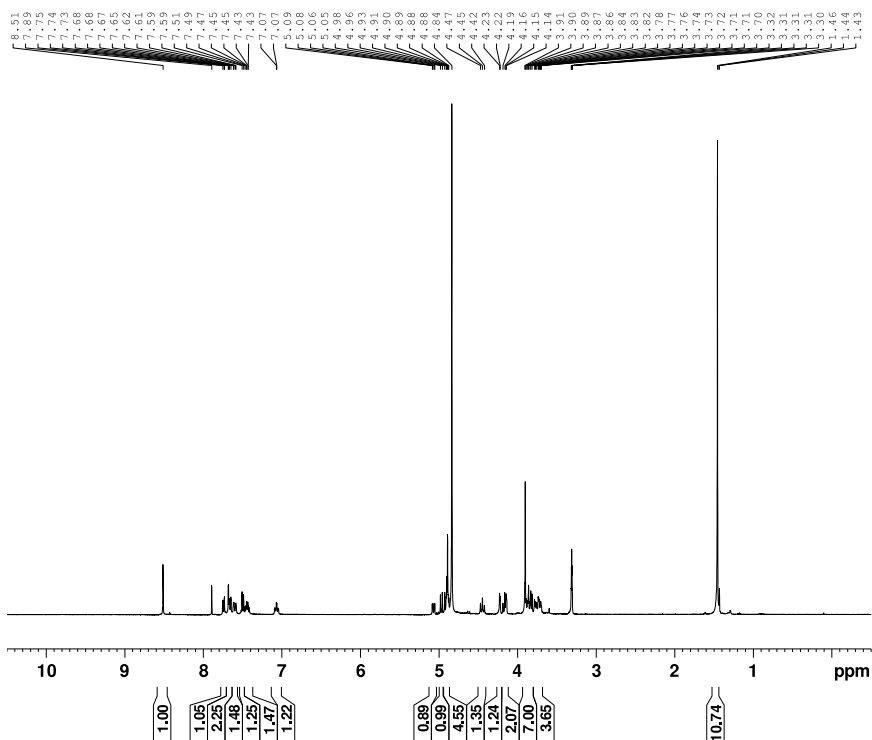
**Appendix B-3** LC-MS characterization of ZJ-YLSPTL glycopeptide



**Appendix B-4**  $^1\text{H}$  NMR spectrum of 4-{{*tert*-butoxycarbonyl}aminoxy}methyl}-3-methoxybenzoic acid

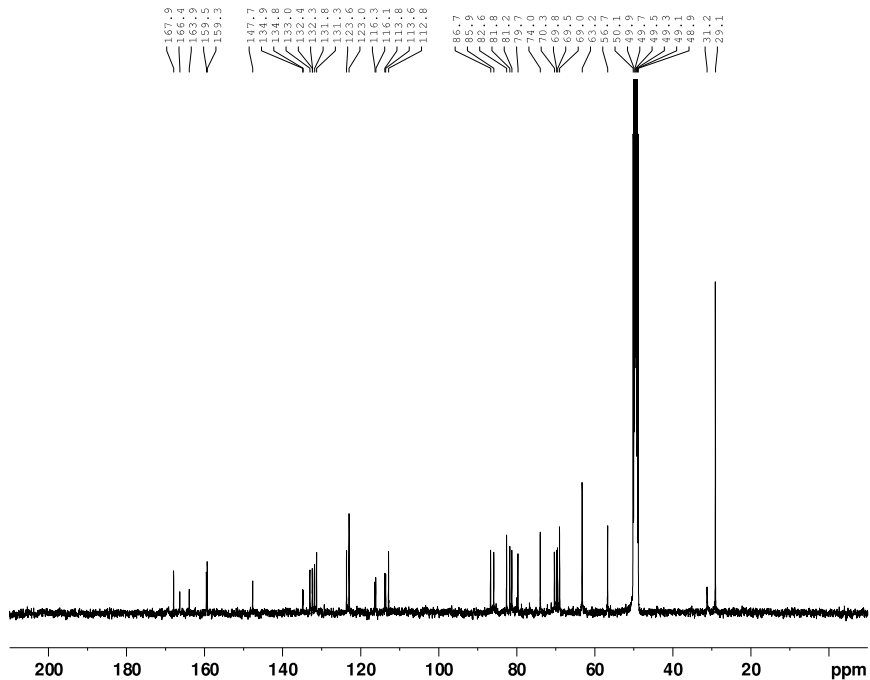


**Appendix B-5**  $^{13}\text{C}$  NMR spectrum of 4-[[tert-butoxycarbonyl)aminoxy]methyl]-3-methoxybenzoic acid

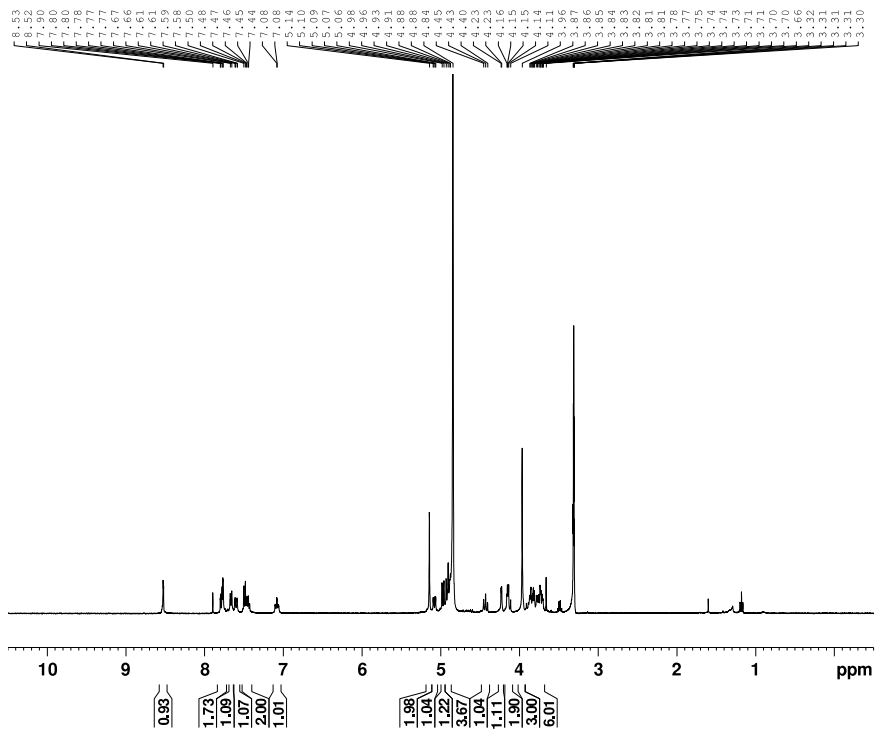


**Appendix B-6** <sup>1</sup>H NMR spectrum of TAZTDG-O3-[[4-(NBoc-aminooxy)methyl]-3-methoxy]-benzoate

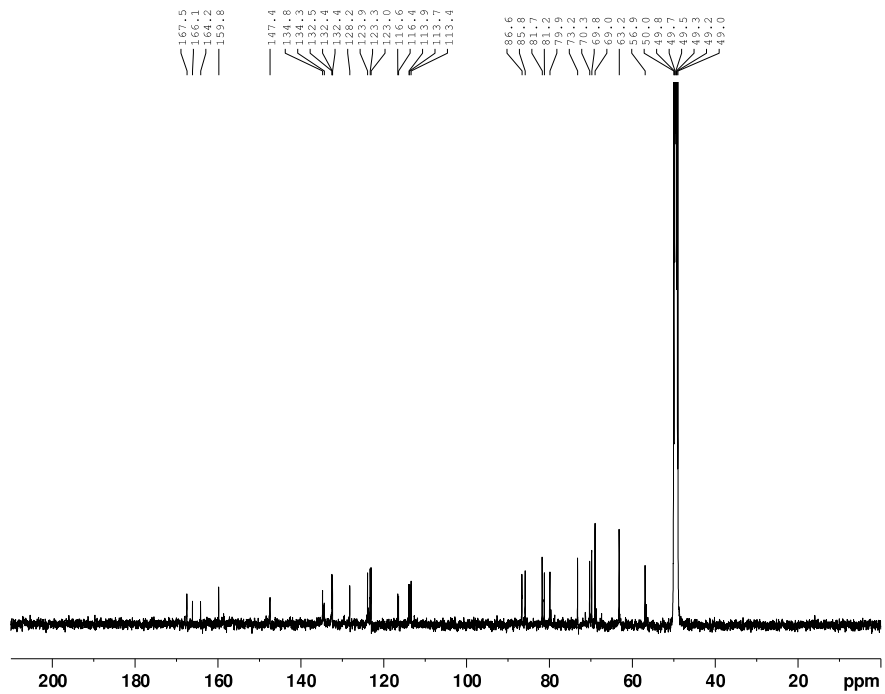




**Appendix B-7**  $^{13}\text{C}$  NMR spectrum of TAZTDG-O3-[[4-(NBoc-aminoxy)methyl]-3-methoxy]-benzoate



Appendix B-8 <sup>1</sup>H NMR spectrum of ZJ



Appendix B-9 <sup>13</sup>C NMR spectrum of ZJ

## Appendix C: Supporting information for Chapter 4

SDB_id <sup>a</sup>	SBD clone sequence <sup>b</sup>	Peptide region <sup>c</sup>	Glycan <sup>d</sup>
1	CTGCTGTTTCGCAATACCACTC	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te286
2	CTTCTATTTCGCAATTCGGCTC	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te288
3	CTGCTTTTCGCAATTCGGCTT	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	BLANK3
6	CTGCTGTTTCGCAATTCCACTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te291
9	CTTCTTTTCGCAATTCCTCTA	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	PZ-5080
10	CTACTGTTTTCGCAATACCGCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Ty176
12	CTTCTGTTTCGCAATACCTCTA	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	X03N1
13	CTACTTTTCGCAATTCCTCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te302
15	CTGCTTTTCGCAATACCCCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te305
17	CTACTCTTCGCAATTCGGCTT	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Tr322
18	CTGCTGTTTTCGCAATTCCTCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	X03L2
20	CTGCTTTTCGCAATTCGGCTT	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	X05N1
21	CTACTCTTTTCGCAATTCCTCTT	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te201
22	CTACTGTTTTCGCAATTCCTCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	X05M1
23	CTGCTTTTCGCAATACCCCTT	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te221
24	CTACTATTTCGCAATTCGGCTC	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te222
26	CTGCTATTTCGCAATTCCTCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te223
29	CTGCTATTTCGCAATTCGGCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te224
34	CTTCTTTTCGCAATTCGGCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	D21
35	CTGCTCTTCGCAATTCCTCTG	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Tr269
40	CTTCTGTTTCGCAATTCGGCTT	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te303
42	CTTCTGTTTCGCAATTCGGCTG	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te271
45	CTGCTGTTTTCGCAATTCCTCTG	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te79
47	TTATTATTTCGCAATTCCTCTA	AGTGTGGAGAAGAATGATCAGAAGACTTATCATGCGGGTGGAGGT	Te100
48	TTATTATTTCGCAATTCCTTTA	AGCGTAGAAAAAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te101
49	CTACTGTTTCGCAATTCGGCTG	AGTGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te102
51	CTGCTCTTTTCGCAATTCCTCTA	AGCGTCGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te119
52	CTACTGTTTTCGCAATTCGGCTG	AGTGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te140
56	CTACTGTTTTCGCAATTCCTCTT	AGCGTTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr57
57	TTATTATTTCGCAATTCCTTTA	AGCGTCGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr59
58	CTACTCTTCGCAATTCCTCTC	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr62
59	TTATTATTTCGCAATTCCTTTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr116
60	CTACTGTTTCGCAATTCCTCTG	AGTGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	BLANK60
66	TTATTATTTCGCAATTCCTTTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	D9
67	CTTCTGTTTCGCAATTCCTCTC	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	D8
68	CTACTATTTCGCAATTCGGCTC	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	X06M1
69	CTTCTGTTTTCGCAATTCCTCTG	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	X06M2
70	CTACTGTTTCGCAATTCGGCTC	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr32
71	CTGCTTTTTCGCAATTCCTCTG	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr33
73	TTATTATTTCGCAATTCCTTTA	AGCGTTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr40
74	CTACTCTTTCGCAATTCCTCTG	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr41
77	CTACTCTTTCGCAATTCGGCTG	AGCGTCGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te135
79	CTACTGTTTCGCAATTCCTCTC	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te176
80	CTTCTGTTTCGCAATTCGGCTA	AGTGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tri-AN3
81	CTACTCTTTCGCAATTCGGCTT	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tri-BN3
82	TTATTATTTCGCAATTCCTTTA	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Di-N3
83	CTGCTCTTTCGCAATTCCTCTG	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	X07M2
87	CTACTCTTTCGCAATTCGGCTC	AGCGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr47
88	CTGCTGTTTCGCAATTCCTCTA	AGTGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te72
99	CTTCTCTTTCGCAATTCGGCTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	X01L1
103	TTATTATTTCGCAATTCCTTTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	BLANK103
105	CTGCTTTTTCGCAATTCCTCTT	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te258
112	CTACTGTTTCGCAATTCGGCTG	AGTGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr260
113	CTACTGTTTTCGCAATTCGGCTG	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te319
115	CTACTGTTTCGCAATTCGGCTG	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te321
118	CTACTCTTTCGCAATTCGGCTC	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te327
120	CTGCTGTTTTCGCAATTCGGCTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	BLANK120
126	CTGCTCTTTCGCAATTCGGCTG	AGTGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Gal1f4-m
127	CTGCTATTTCGCAATTCGGCTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	PZ-8015
149	CTACTCTTTCGCAATTCGGCTT	AGCGTGGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te98
150	CTGCTCTTTCGCAATTCGGCTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Te99
152	CTGCTTTTTCGCAATTCCTCTC	AGCGTAGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	Tr117
mCherry01	CTTCTATTTCGCAATTCCTCTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	X03-red
mNeon04	CTACTGTTTCGCAATTCGGCTA	AGTGTGAAAAGAACGATCAAAAAGACTTATCATGCGGGTGGAGGT	LNTgreen

### Appendix C-1. Silent barcodes used to build LiGA library

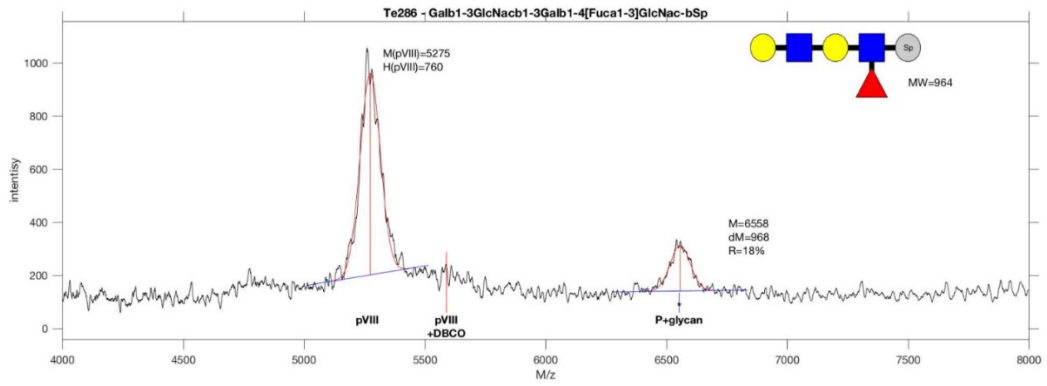
#_list code	symbol nomeclature	#_list code	symbol nomeclature
1 Ty176		21 Te302	
2 PZ-5080		22 Te99	
3 X05L2		23 Te100	
4 D9		24 Di-N3	
5 X05N1		25 Tr120	
6 X05M1		26 Tr117	
7 D8		27 Tr116	
8 D21		28 Tri-AN3	
9 Tr59		29 Tr57	
10 Tr62		30 Tri-BN3	
11 Tr54		31 Te224	
12 Tr55		32 Te221	
13 Tr260		33 Te223	
14 Te98		34 Te258	
15 Te72		35 Te222	
16 X06M1		36 Te259	
17 X06M2		37 Te286	
18 X06NG		38 Te101	
19 Te271		39 Te319	
20 Te327		40 Te135	

Gal   
 GalNAc   
 Galf   
 Glc   
 GlcNAc   
 Man   
 KDN   
 Neu5Ac   
 NeuGAc   
 Fuc

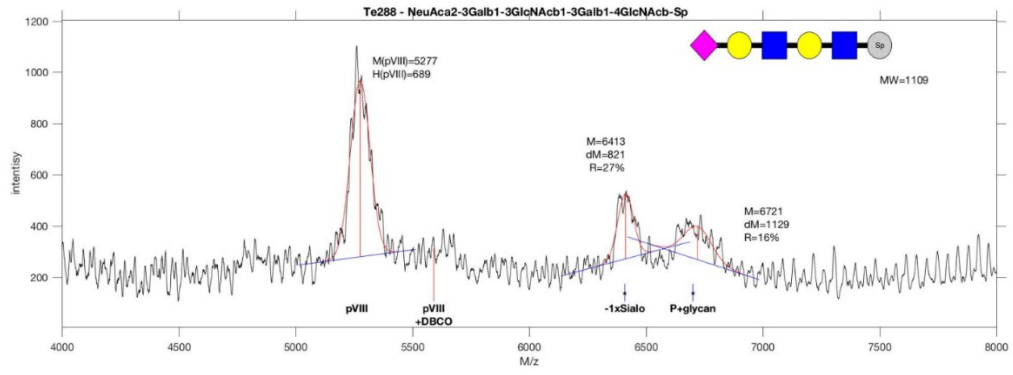
**Appendix C-2.** Glycan structures incorporated into LiGA library(Figure 4-8)

#_list code	symbol nomeclature	#_list code	symbol nomeclature
41 Te102		61 Te303	
42 Tr32		62 Te79	
43 Tr39		63 Te305	
44 3'SLN		64 Te306	
45 Tr33		65 Te119	
46 Tr40		66 Te140	
47 Tr36		67 Te193	
48 6'SLN		68 Te78	
49 Tr43		69 Te97	
50 Tr269		70 X01M1	
51 Tr41		71 X01L1	
52 Tr47		72 X03N1	
53 Tr48		73 X03L2	
54 Tr322		74 PZ-8015	
55 Te201		75 X03MC	
56 Te321		76 X07M2	
57 Te288		77 Galf3	
58 Te176		78 Galf4-low	
59 Te75		79 Galf4-mid	
60 Te291		80 X08M1	

**Appendix C-2(Cont.).** Glycan structures incorporated into LiGA library(Figure 4-8)

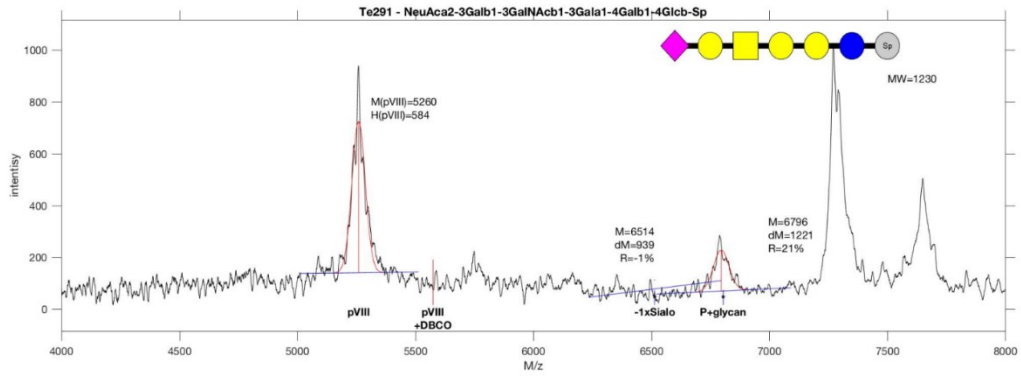


**Appendix C-3.** MALDI-TOF of glycosylated phages, Te286

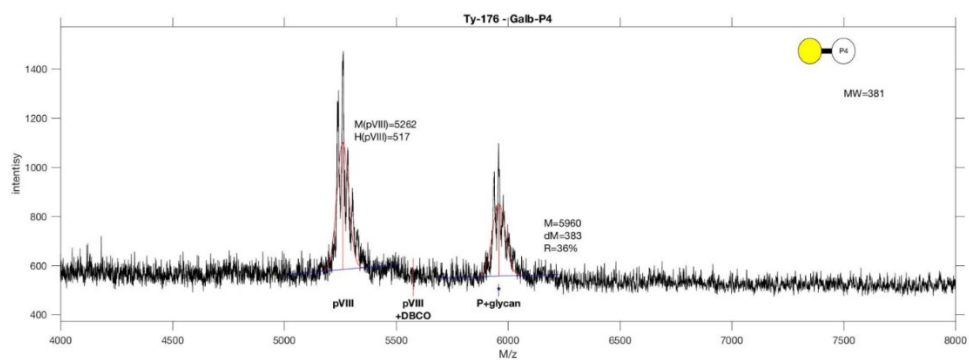


**Appendix C-4. MALDI-TOF of glycosylated phages Te288**

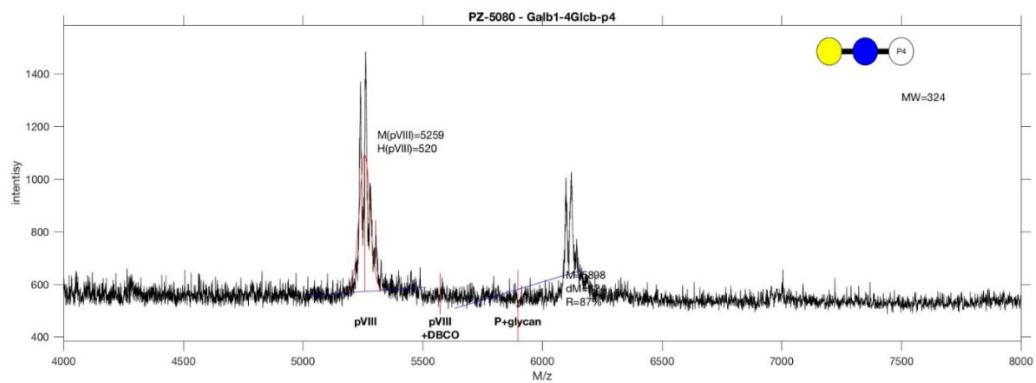




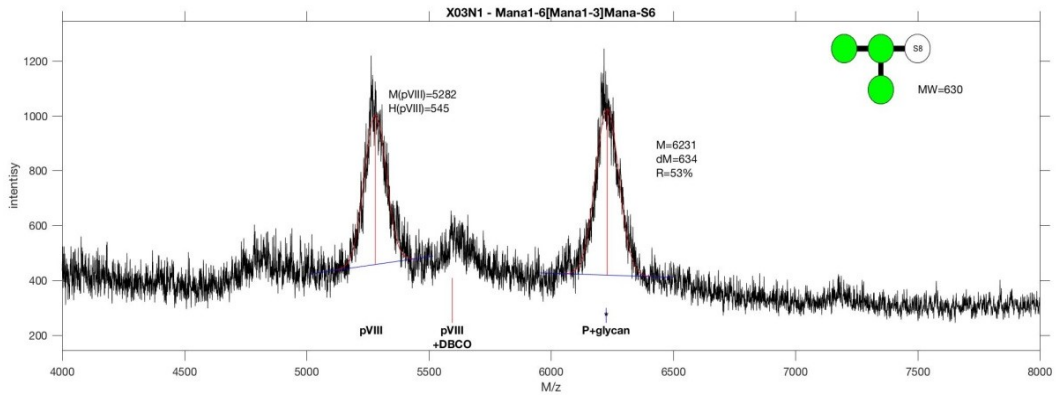
**Appendix C-5. MALDI-TOF of glycosylated phages, Te291**



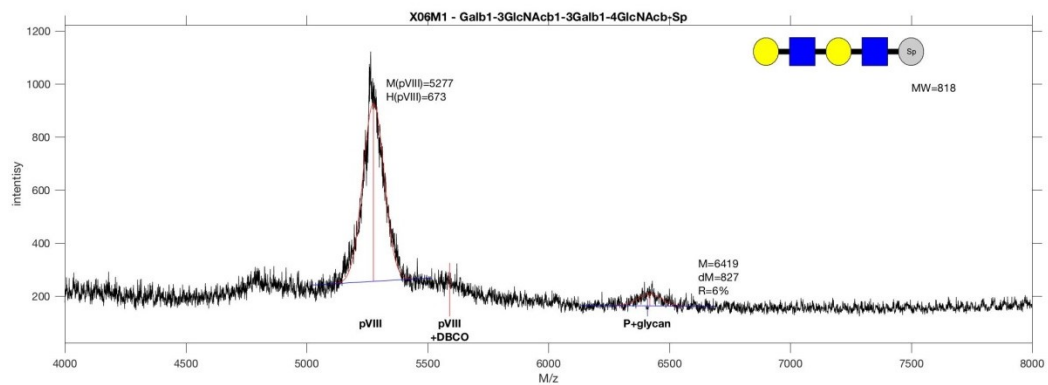
**Appendix C-6.** MALDI-TOF of glycosylated phages, Ty176



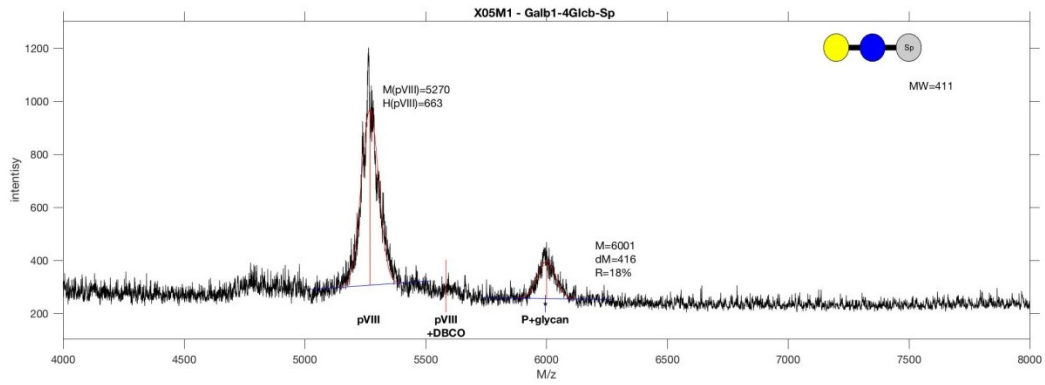
**Appendix C-7.** MALDI-TOF of glycosylated phages, PZ-5080



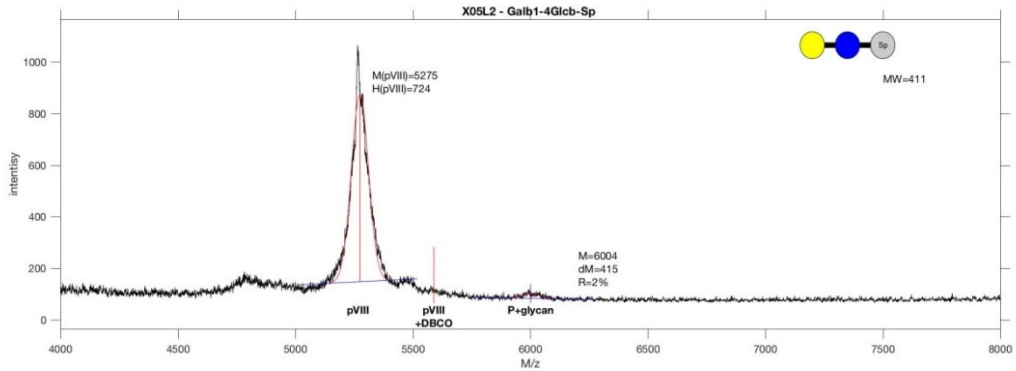
**Appendix C-8. MALDI-TOF of glycosylated phages, X03N1**



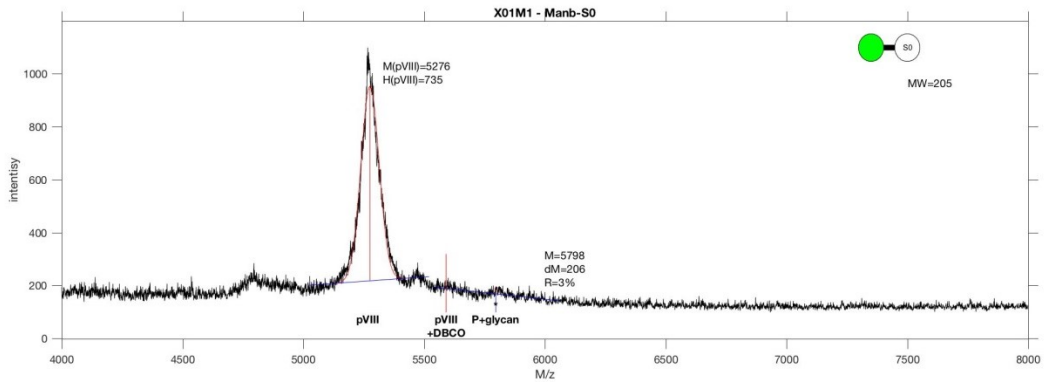
**Appendix C-9.** MALDI-TOF of glycosylated phages, X06M1



**Appendix C-10.** MALDI-TOF of glycosylated phages, X05M1

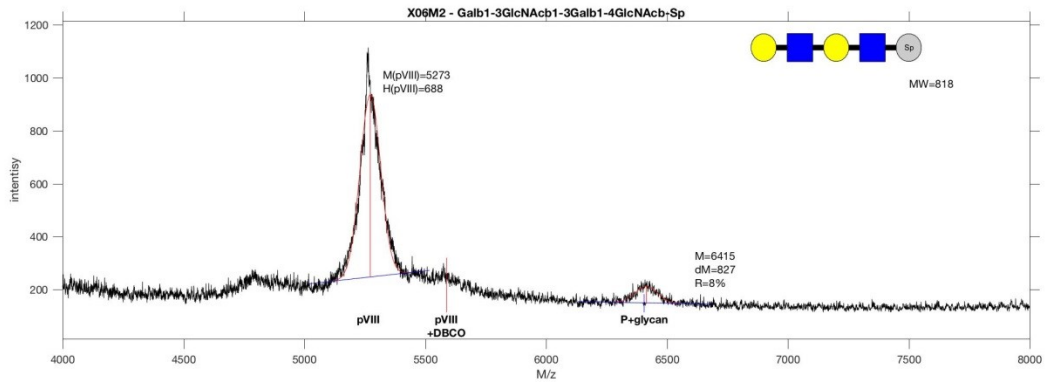


**Appendix C-11.** MALDI-TOF of glycosylated phages, X05L2

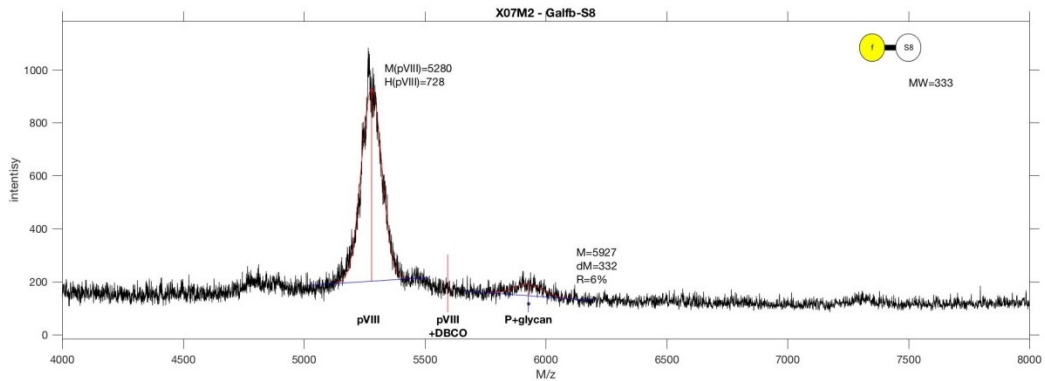


**Appendix C-12.** MALDI-TOF of glycosylated phages, X01M1

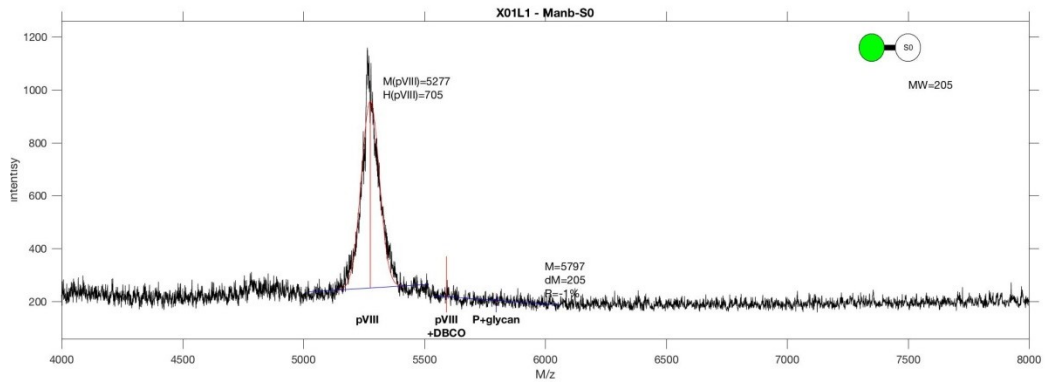




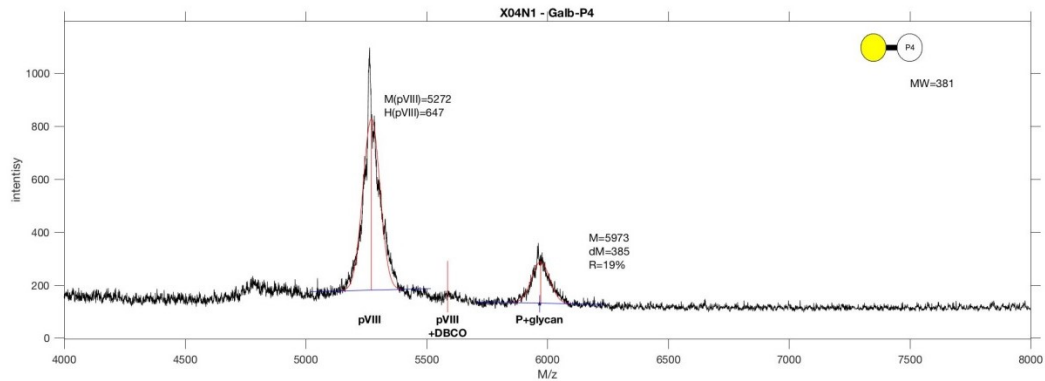
**Appendix C-13.** MALDI-TOF of glycosylated phages, X06M2,



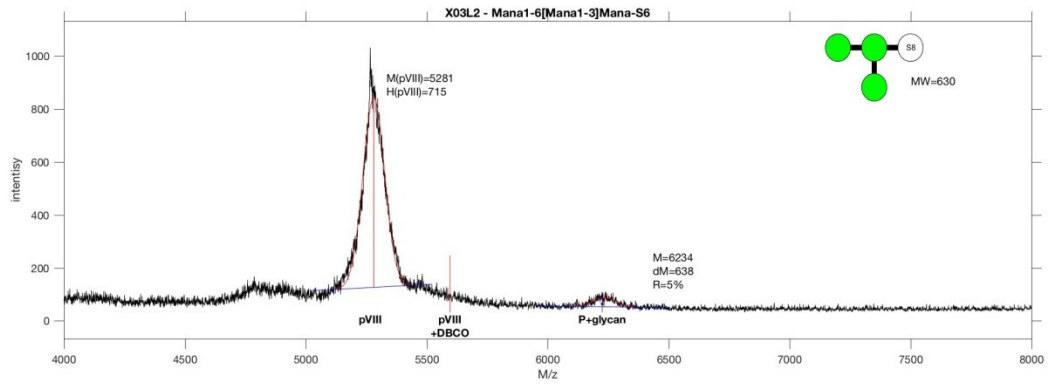
**Appendix C-14.** MALDI-TOF of glycosylated phages, X07M2



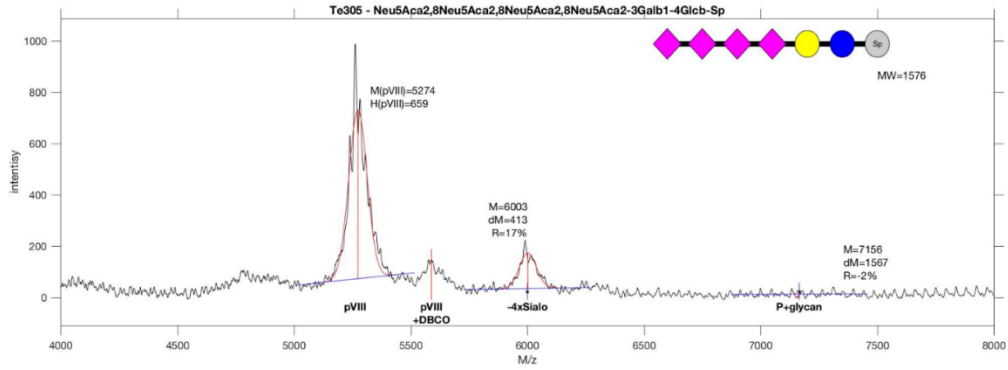
**Appendix C-15.** MALDI-TOF of glycosylated phages, X01L1



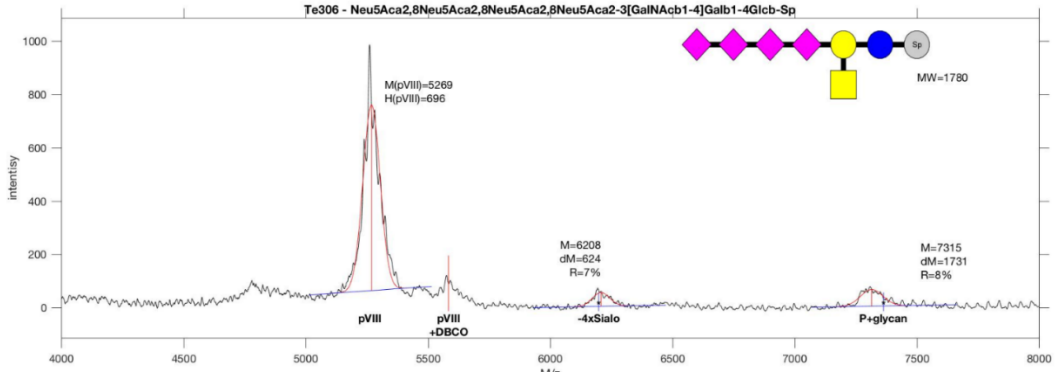
**Appendix C-16.** MALDI-TOF of glycosylated phages, X04N2



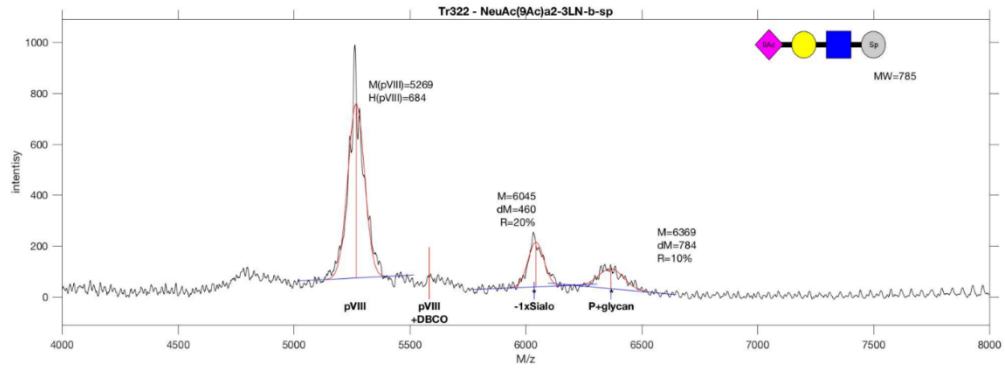
**Appendix C-17.** MALDI-TOF of glycosylated phages, X03L2



**Appendix C-18.** MALDI-TOF of glycosylated phages, Tr305

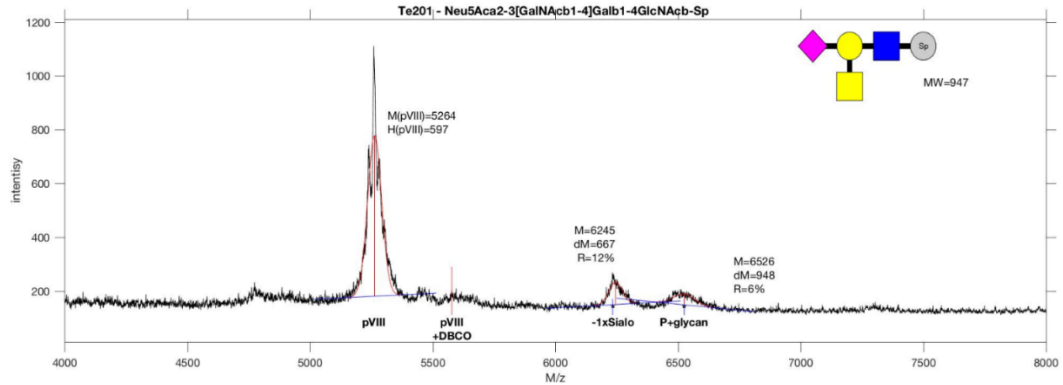


**Appendix C-19.** MALDI-TOF of glycosylated phages, Tr306

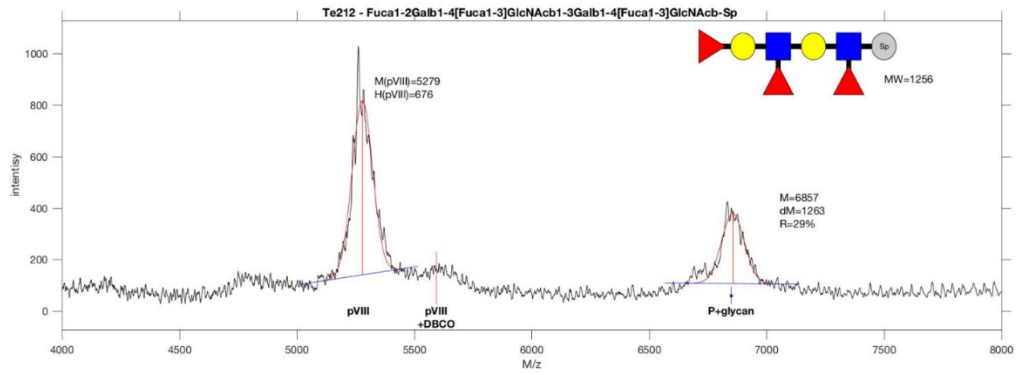


**Appendix C-20.** MALDI-TOF of glycosylated phages, Tr322

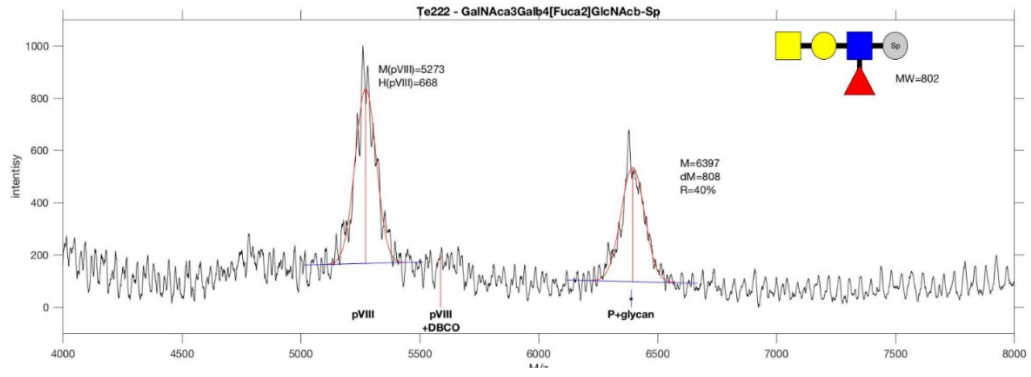




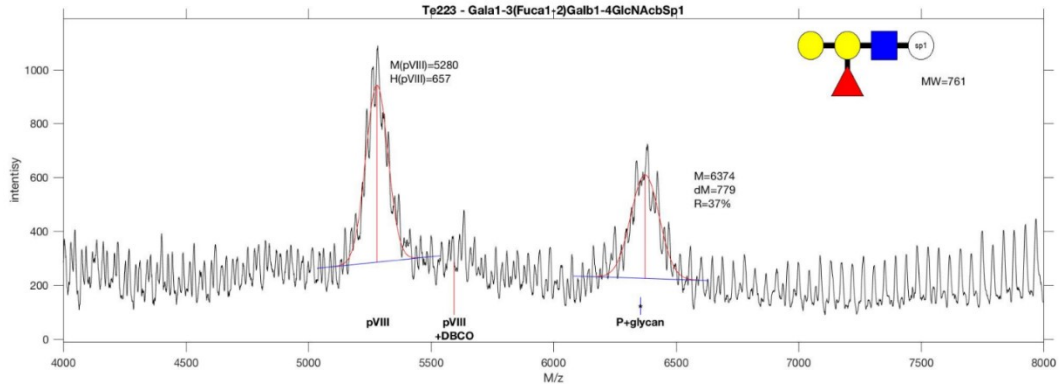
**Appendix C-21.** MALDI-TOF of glycosylated phages, Te201



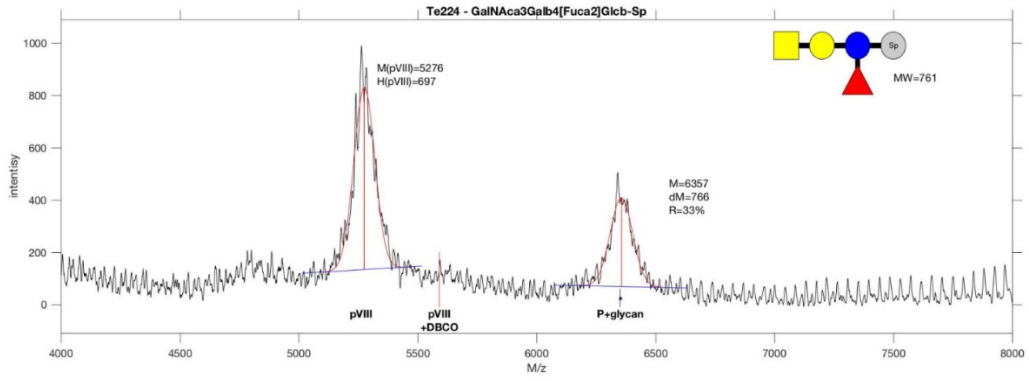
**Appendix C-22.** MALDI-TOF of glycosylated phages Te212



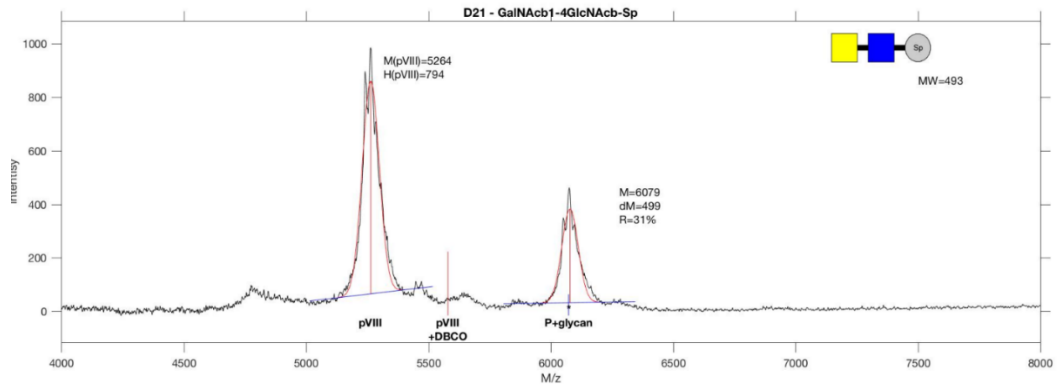
**Appendix C-23.** MALDI-TOF of glycosylated phages, Te222



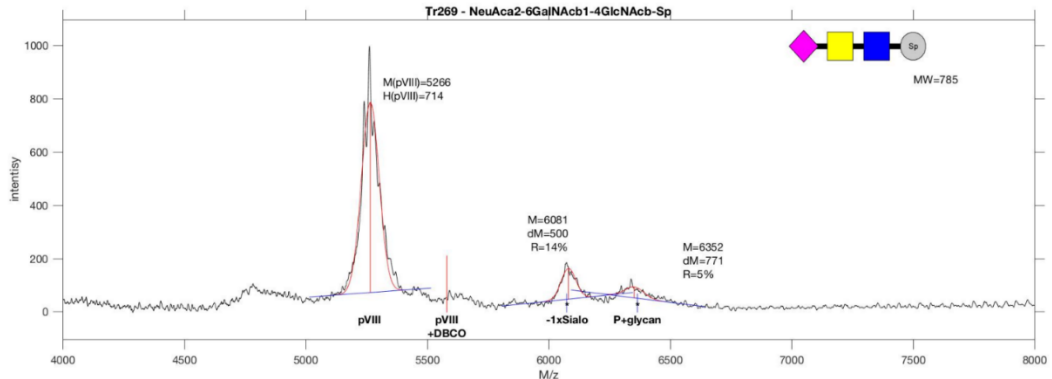
**Appendix C-24.** MALDI-TOF of glycosylated phages, Te201



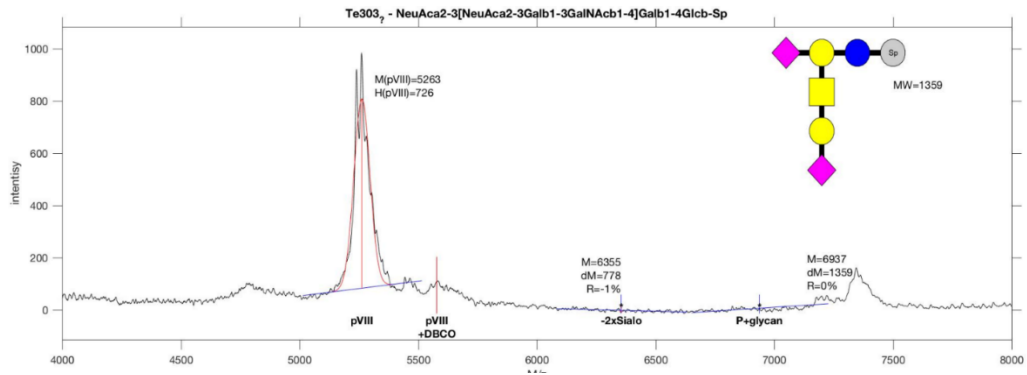
**Appendix C-25.** MALDI-TOF of glycosylated phages, Te212



**Appendix C-26.** MALDI-TOF of glycosylated phages, D21

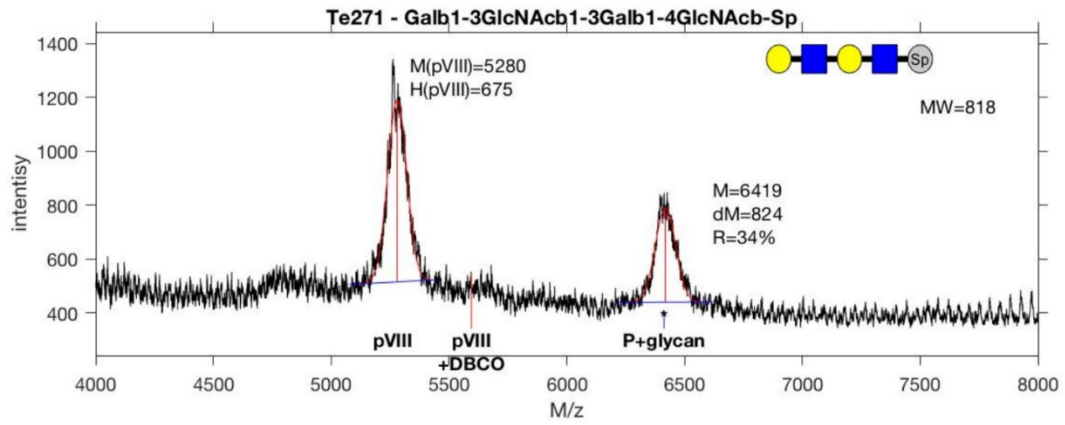


**Appendix C-27.** MALDI-TOF of glycosylated phages, Tr269

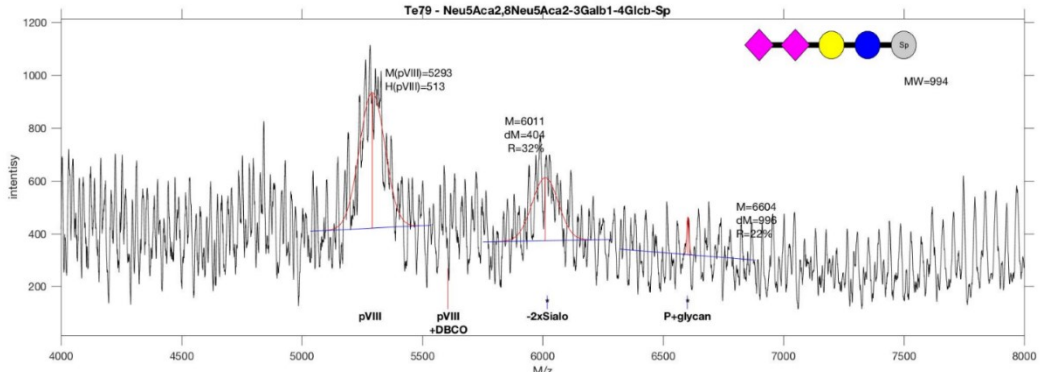


**Appendix C-28.** MALDI-TOF of glycosylated phages, Te303

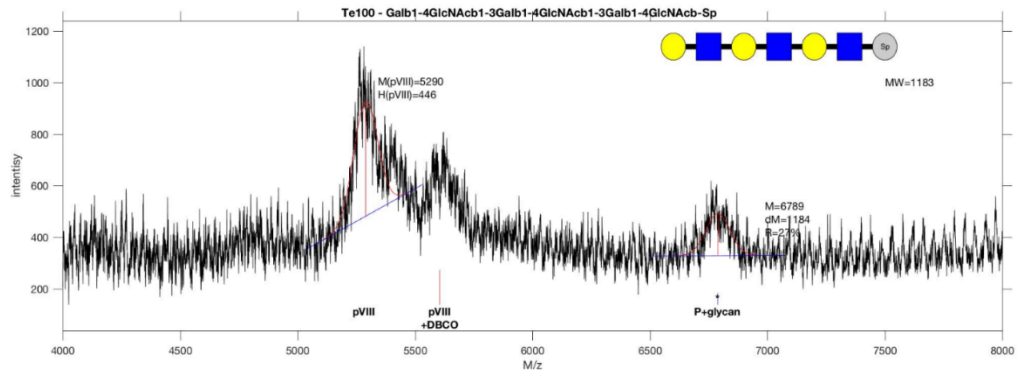




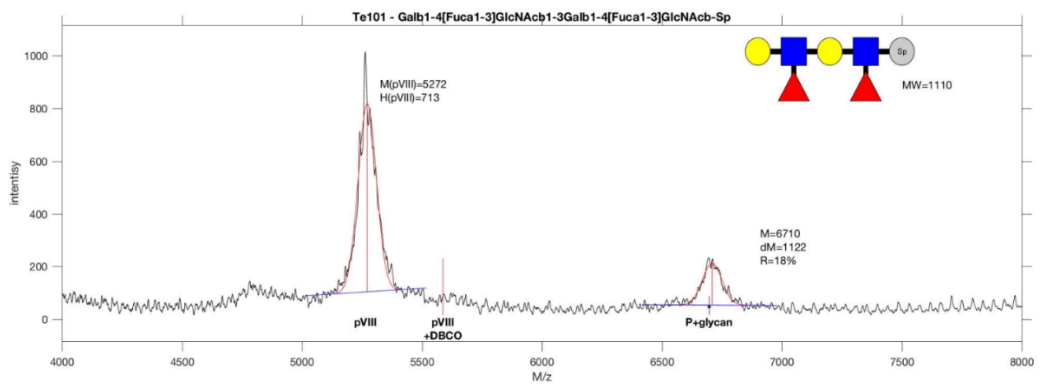
**Appendix C-29.** MALDI-TOF of glycosylated phages, Te271



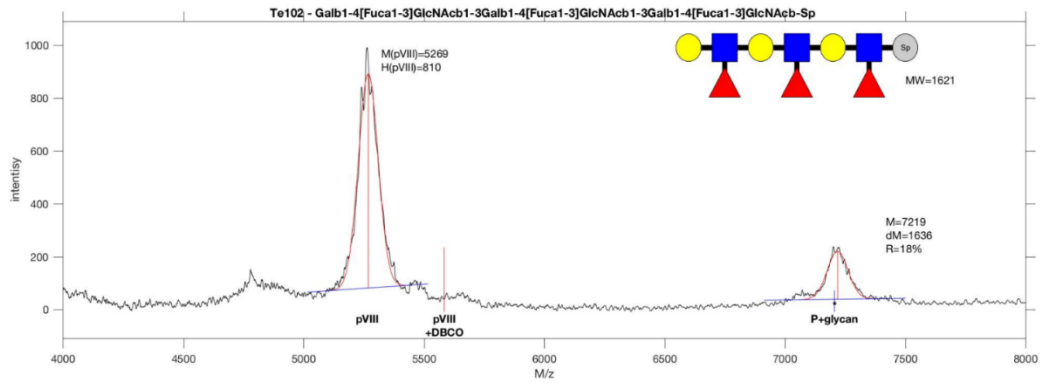
**Appendix C-30.** MALDI-TOF of glycosylated phages, Te79



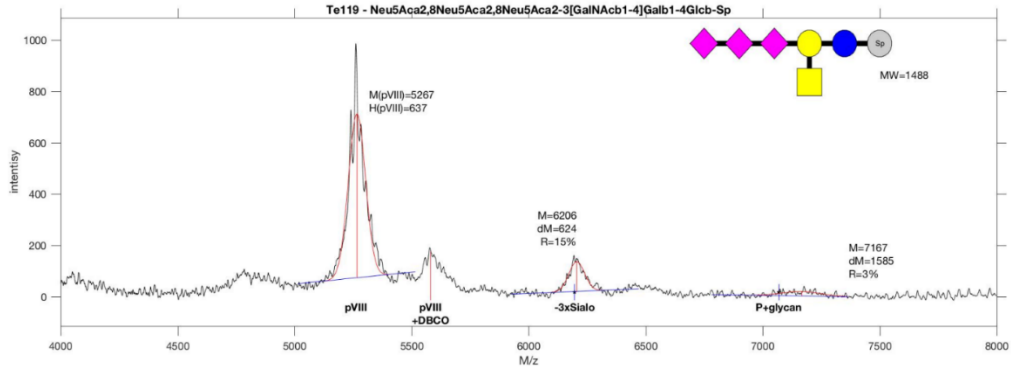
**Appendix C-31.** MALDI-TOF of glycosylated phages, Te100



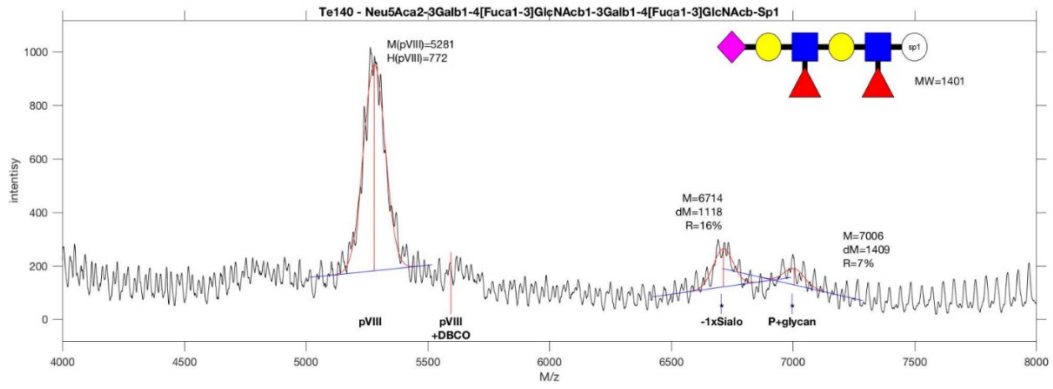
**Appendix C-32.** MALDI-TOF of glycosylated phages, Te101.



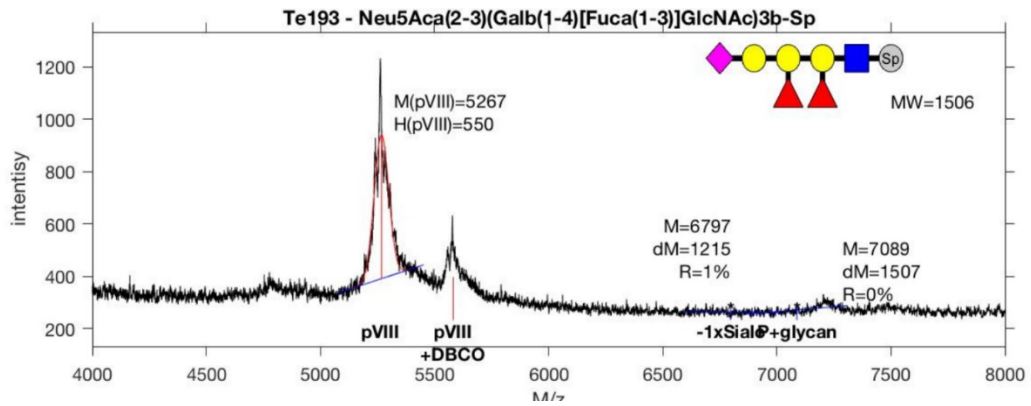
**Appendix C-33.** MALDI-TOF of glycosylated phages, Te102



**Appendix C-34.** MALDI-TOF of glycosylated phages, Te119

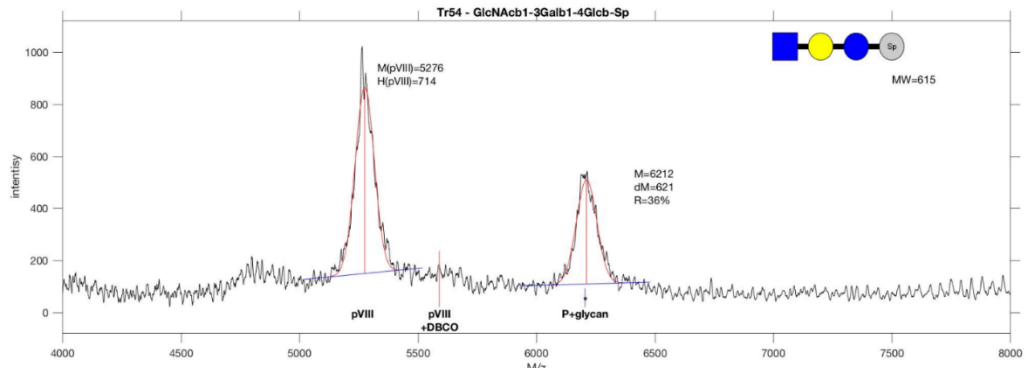


**Appendix C-35.** MALDI-TOF of glycosylated phages, Te140

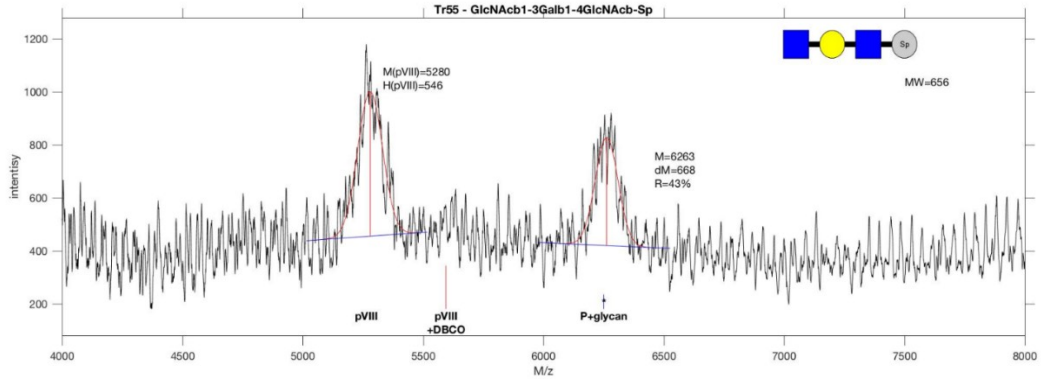


**Appendix C-36.** MALDI-TOF of glycosylated phages, Te193

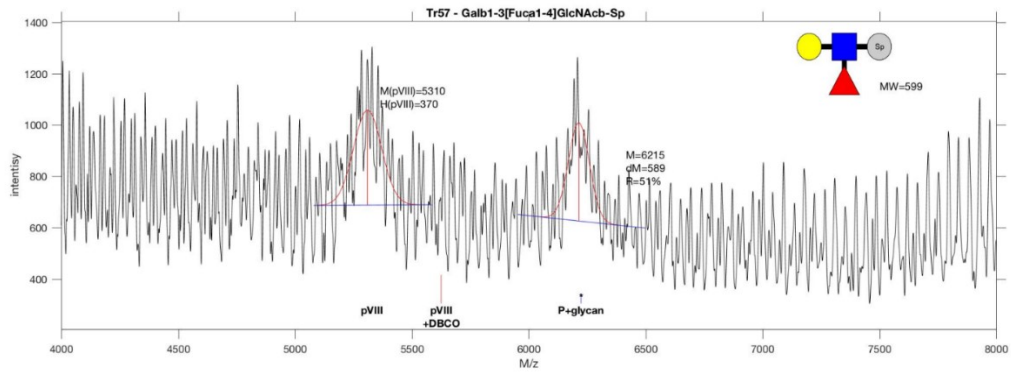




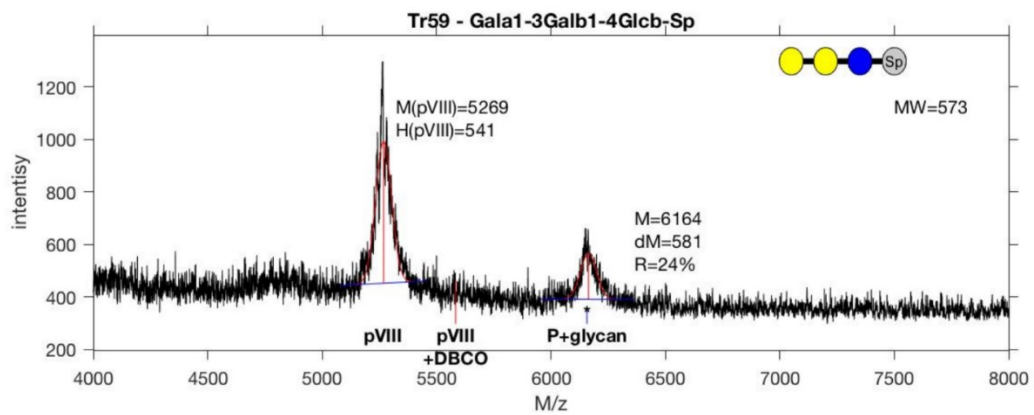
**Appendix C-37.** MALDI-TOF of glycosylated phages Tr54



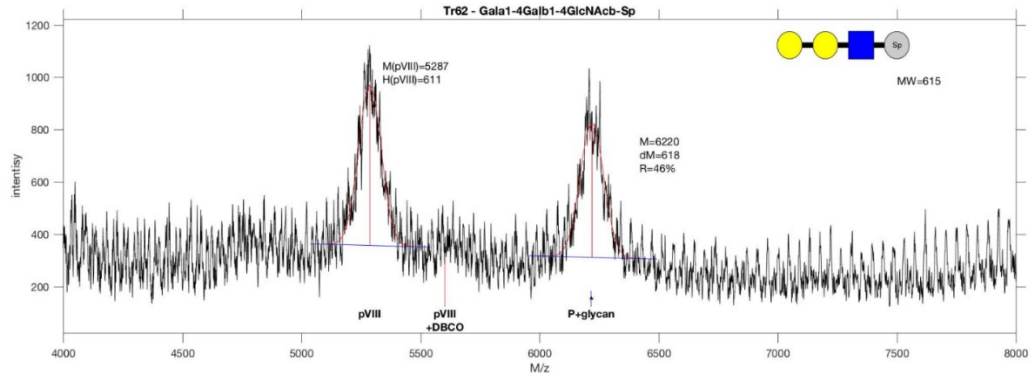
**Appendix C-38.** MALDI-TOF of glycosylated phages, Tr55



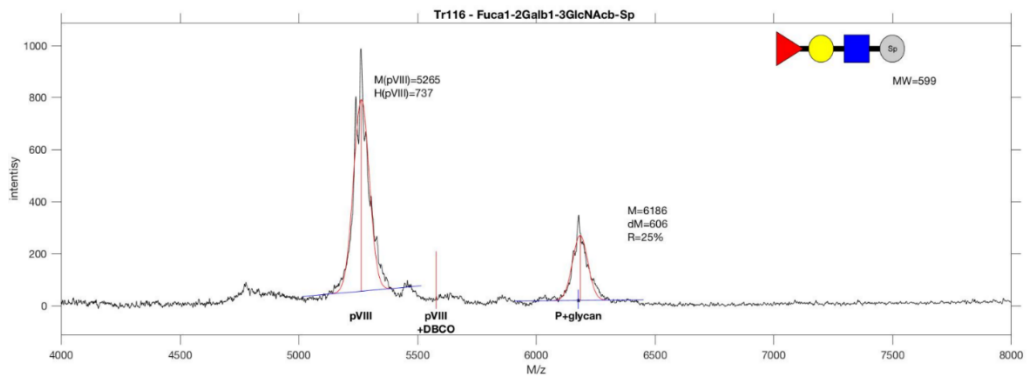
**Appendix C-39.** MALDI-TOF of glycosylated phages, Tr57



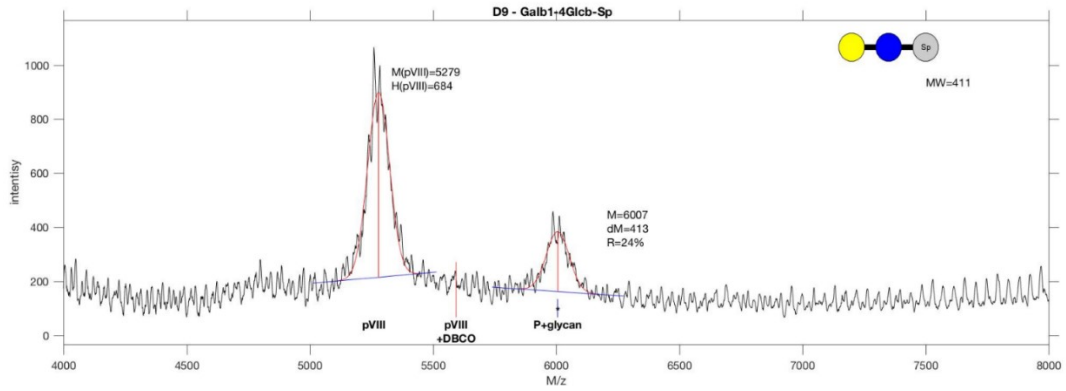
**Appendix C-40.** MALDI-TOF of glycosylated phages Tr59



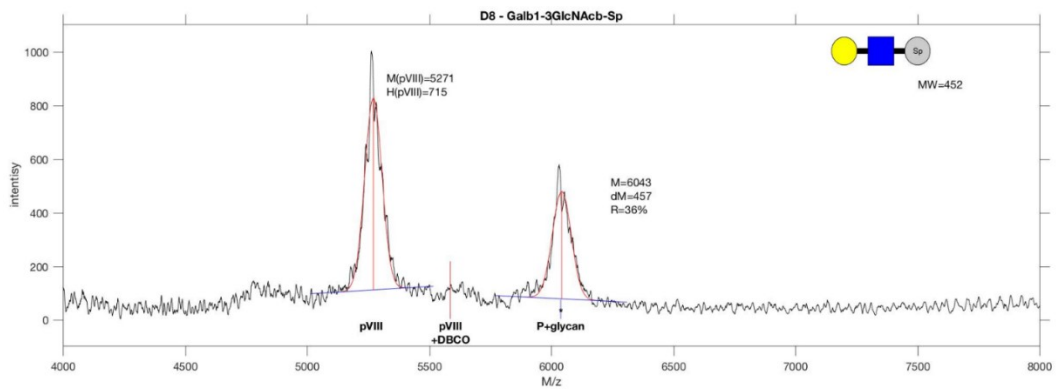
**Appendix C-41.** MALDI-TOF of glycosylated phages, Tr62



**Appendix C-42.** MALDI-TOF of glycosylated phages, Tr116

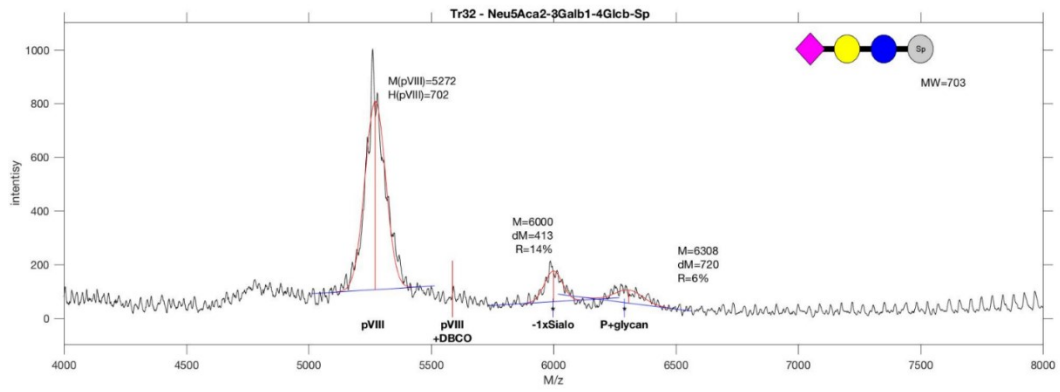


Appendix C-43. MALDI-TOF of glycosylated phages D9

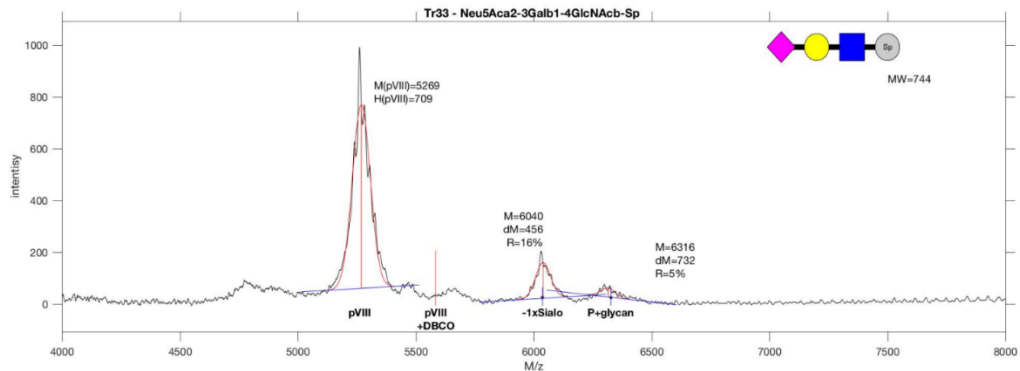


**Appendix C-44.** MALDI-TOF of glycosylated phages, D8

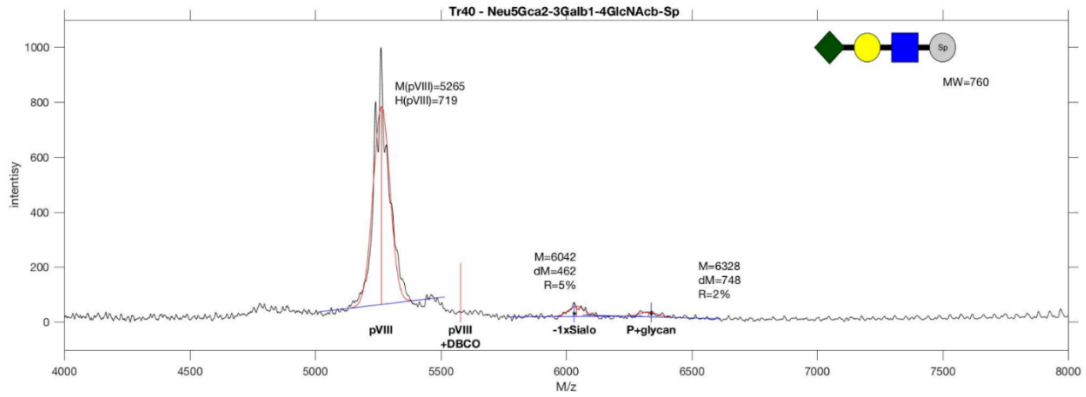




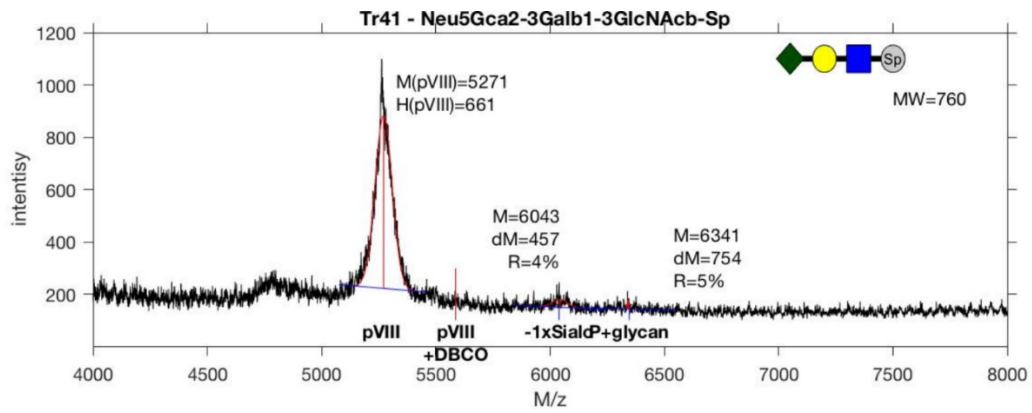
**Appendix C-45.** MALDI-TOF of glycosylated phages, Tr32



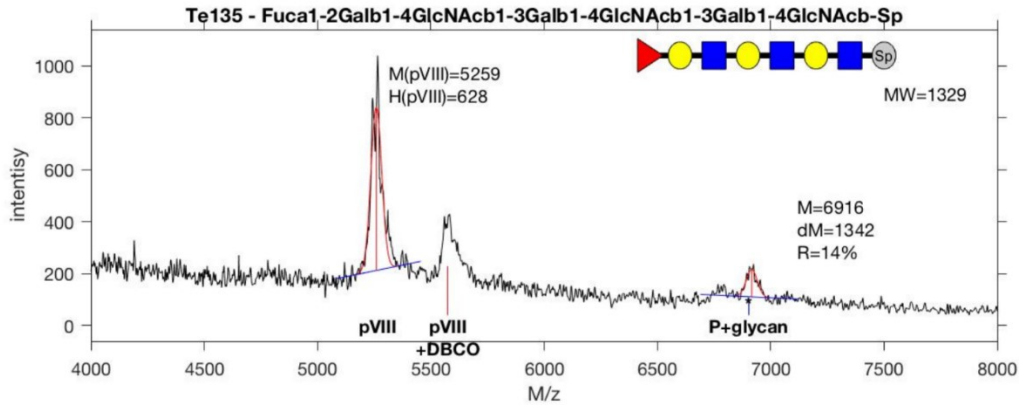
**Appendix C-46. MALDI-TOF of glycosylated phages, Tr33**



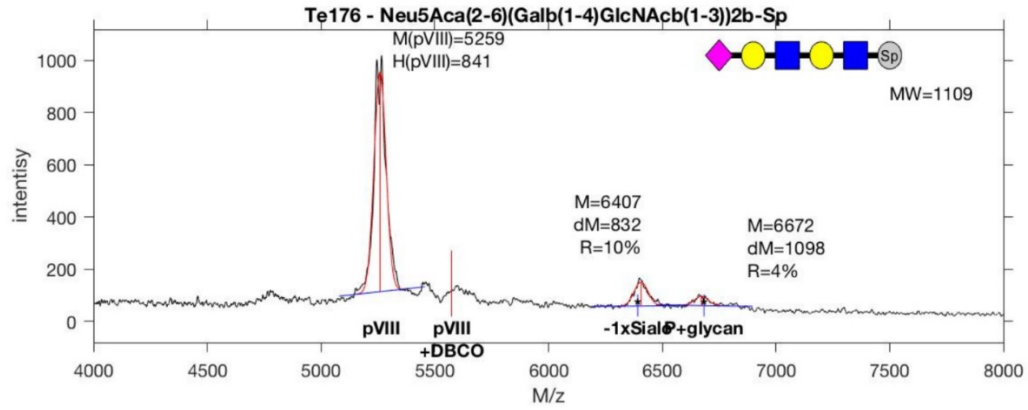
**Appendix C-47.** MALDI-TOF of glycosylated phages, Tr40



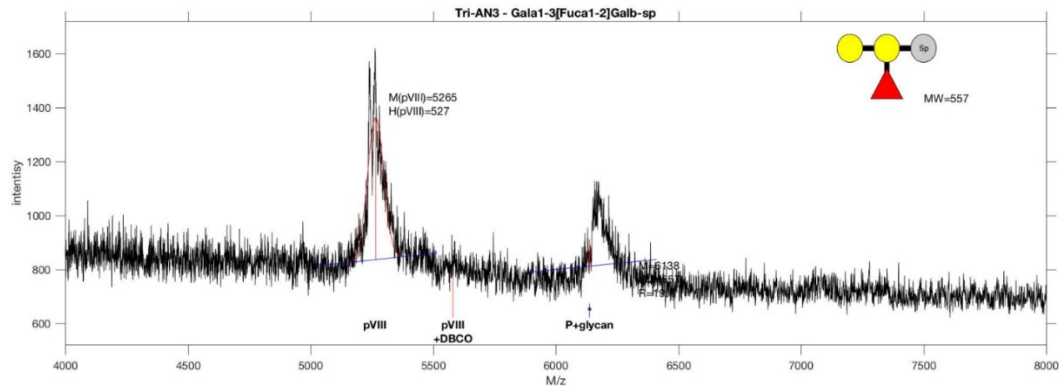
**Appendix C-48.** MALDI-TOF of glycosylated phages, Tr41



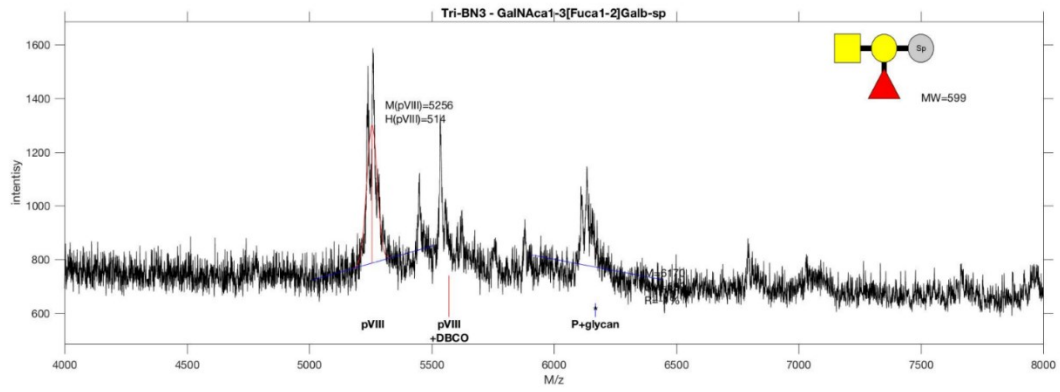
**Appendix C-49** MALDI-TOF of glycosylated phages, Te135



**Appendix C-50.** MALDI-TOF of glycosylated phages, Te176

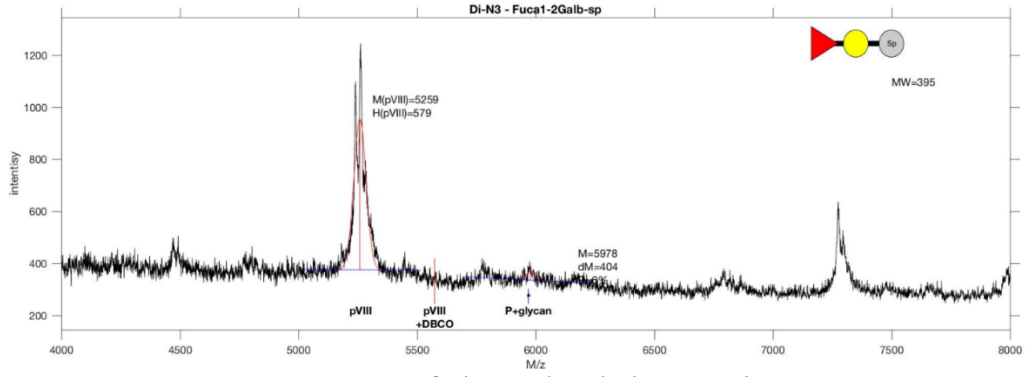


**Appendix C-51.** MALDI-TOF of glycosylated phages, Tri-AN3,

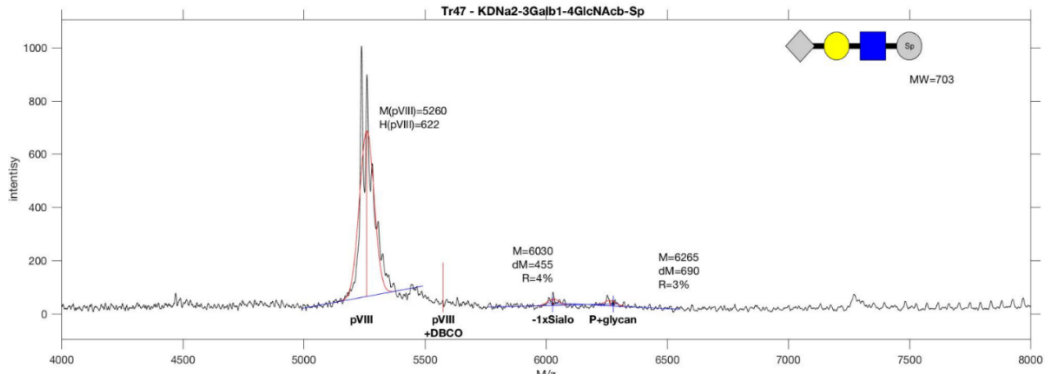


**Appendix C-52.** MALDI-TOF of glycosylated phages, Tri-BN3

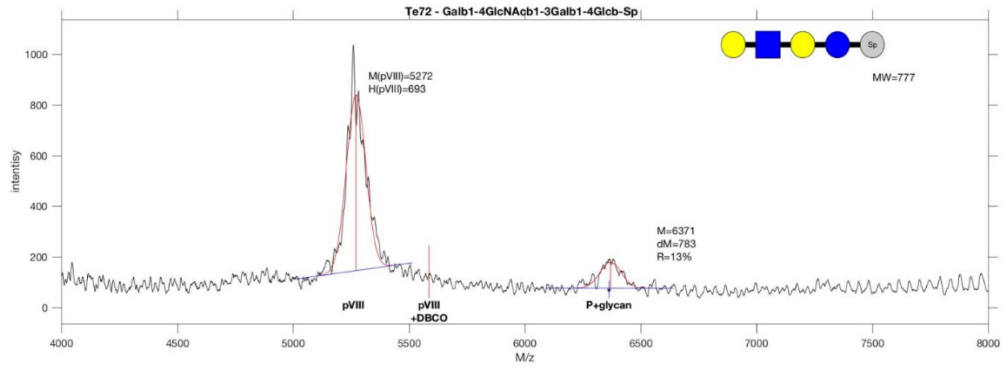




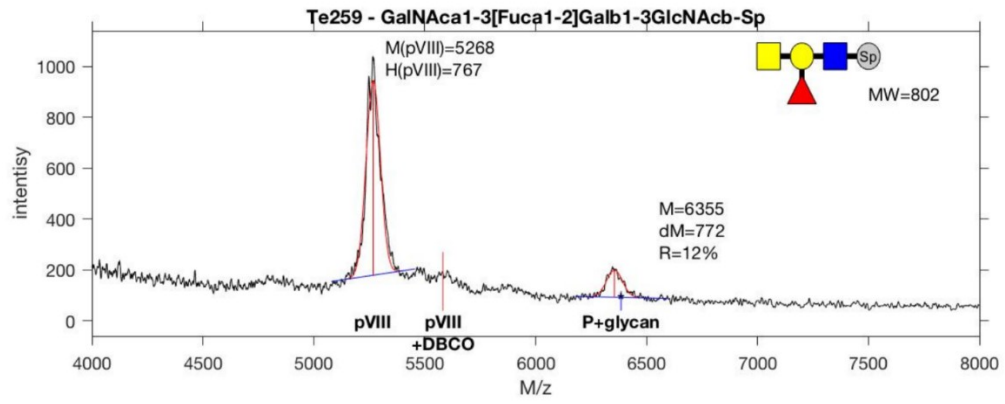
**Appendix C-53.** MALDI-TOF of glycosylated phages, Di-N3



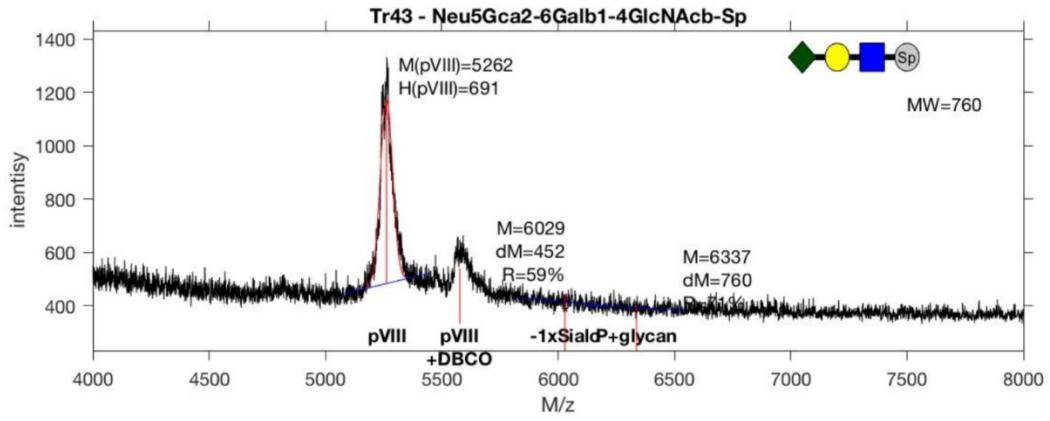
**Appendix C-54.** MALDI-TOF of glycosylated phages, Tr47



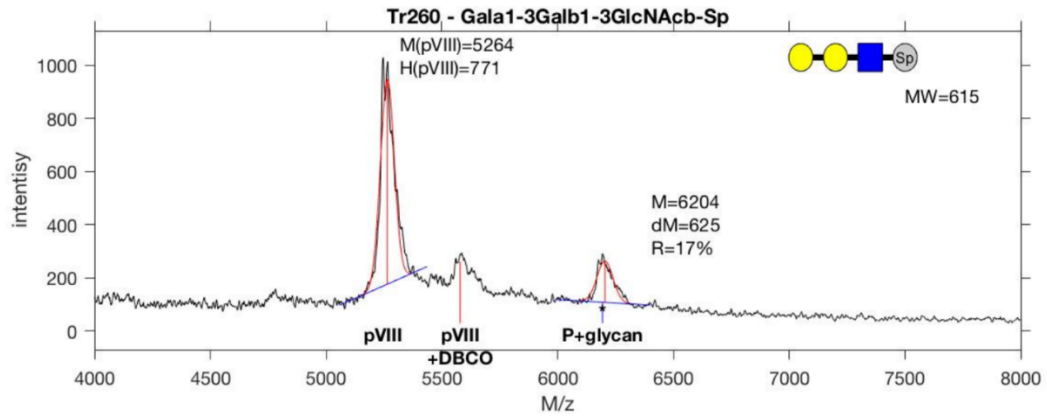
**Appendix C-55.** MALDI-TOF of glycosylated phages, Te72



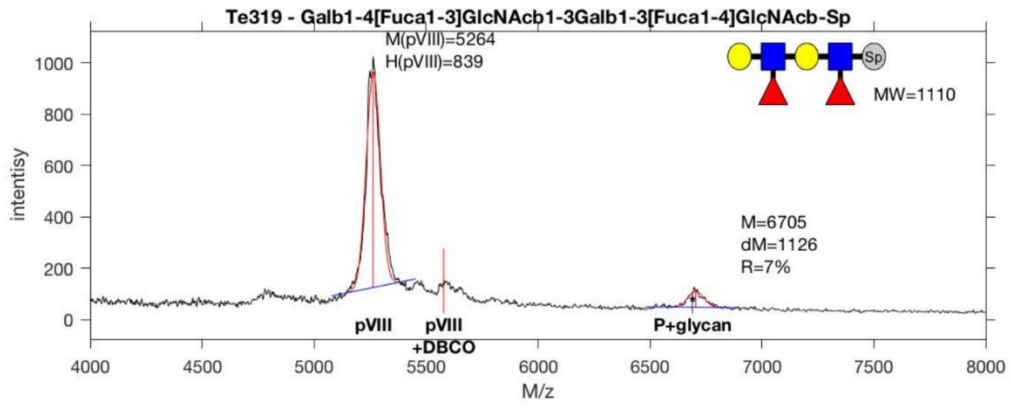
**Appendix C-56.** MALDI-TOF of glycosylated phages Te259



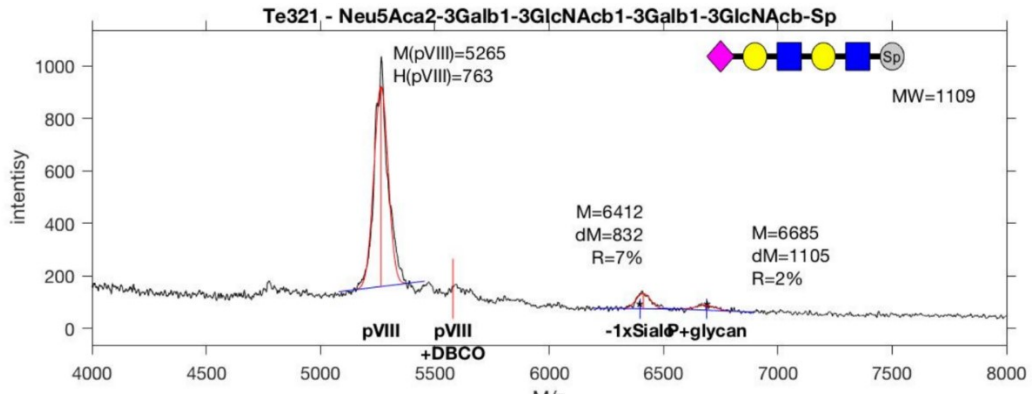
**Appendix C-57.** MALDI-TOF of glycosylated phages, Tr43



**Appendix C-58.** MALDI-TOF of glycosylated phages, Tr260

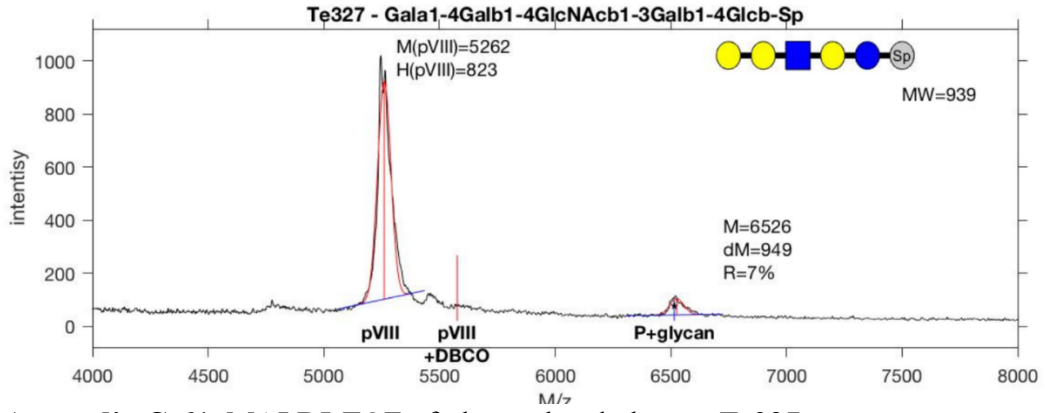


**Appendix C-59.** MALDI-TOF of glycosylated phages, Te319

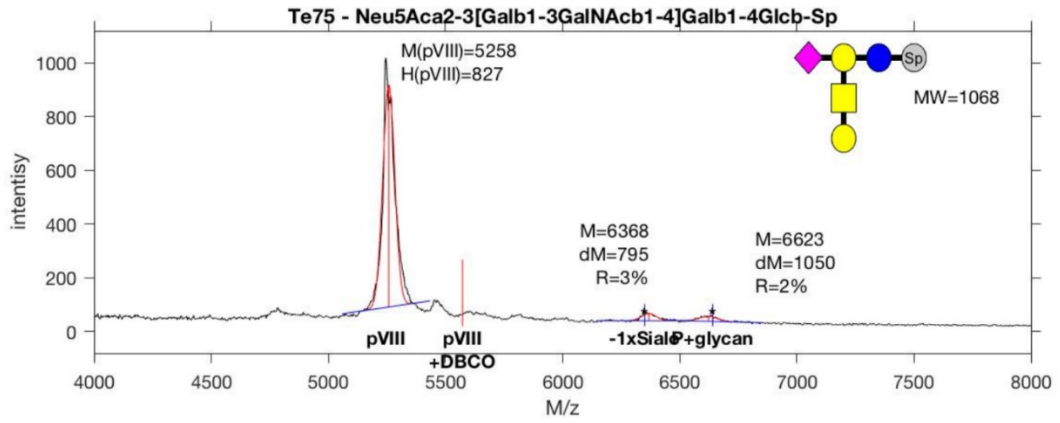


**Appendix C-60.** MALDI-TOF of glycosylated phages, Te321

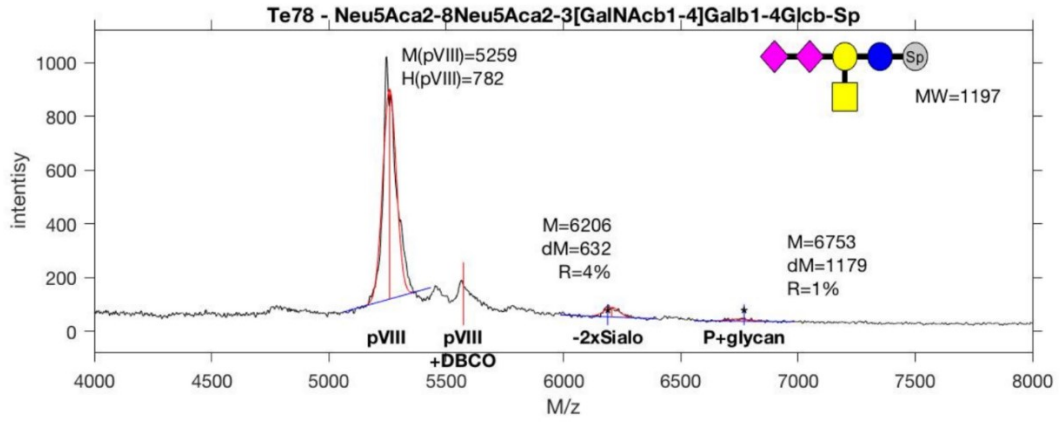




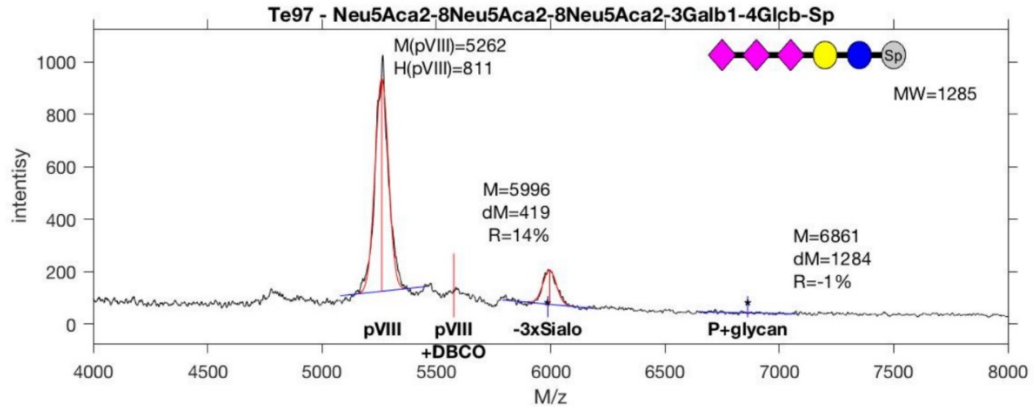
**Appendix C-61.** MALDI-TOF of glycosylated phages, Te327



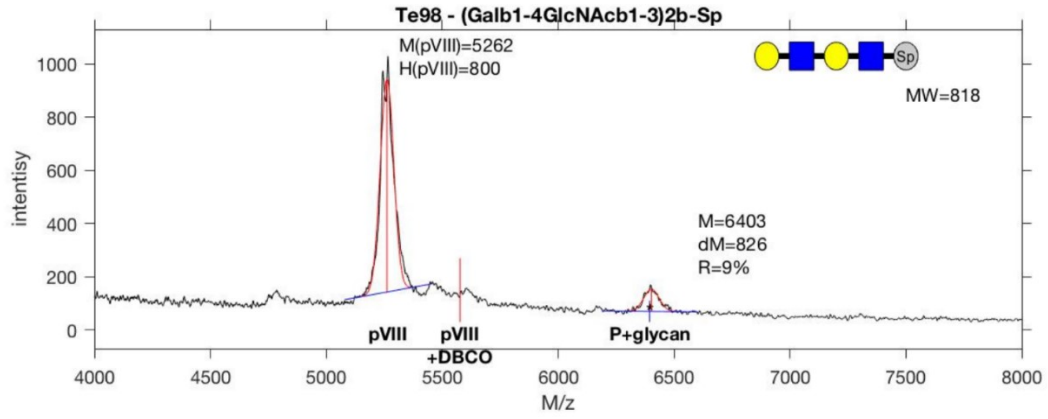
**Appendix C-62.** MALDI-TOF of glycosylated phages, Te75



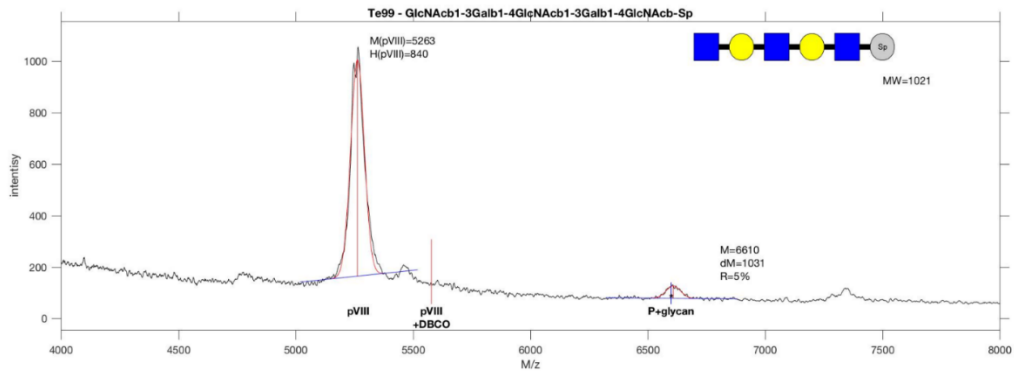
**Appendix C-63.** MALDI-TOF of glycosylated phages, Te78



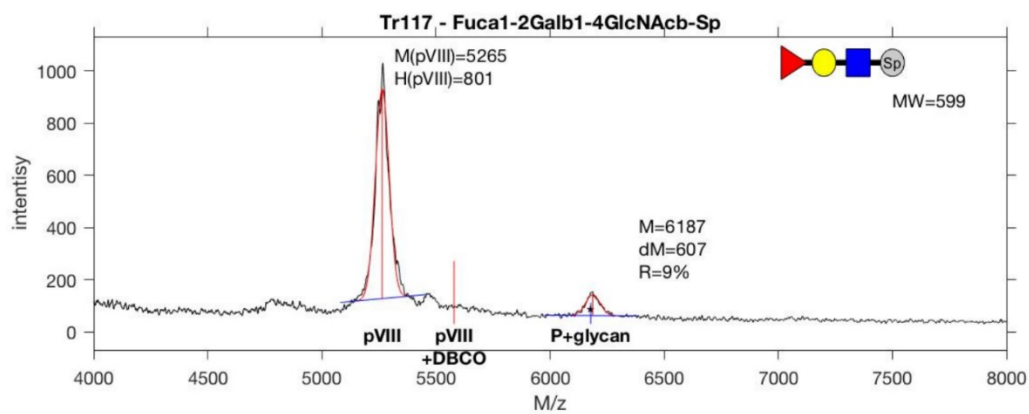
**Appendix C-64.** MALDI-TOF of glycosylated phages, Te97



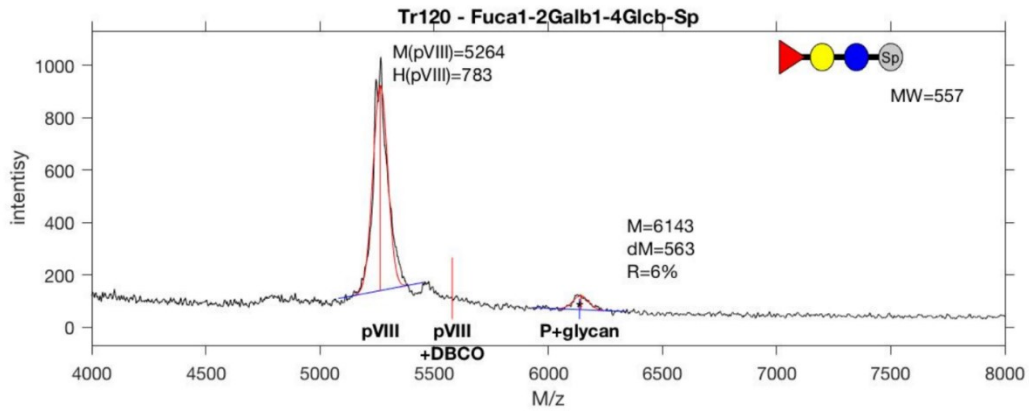
**Appendix C-65.** MALDI-TOF of glycosylated phages, Te98



**Appendix C-67** MALDI-TOF of glycosylated phages, Te99

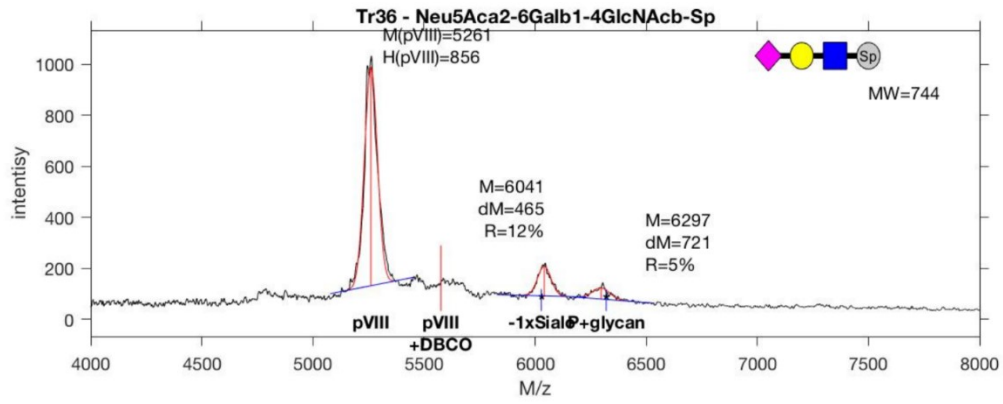


**Appendix C-68.** MALDI-TOF of glycosylated phages Tr117

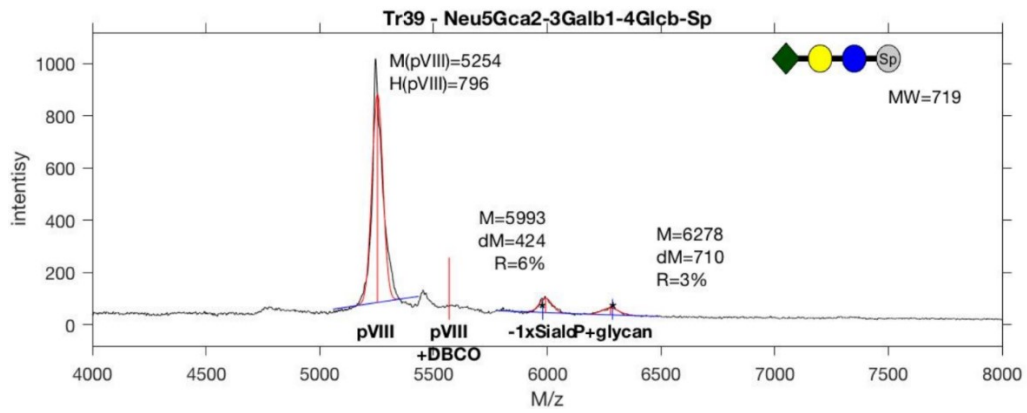


**Appendix C-69** MALDI-TOF of glycosylated phages, Tr120

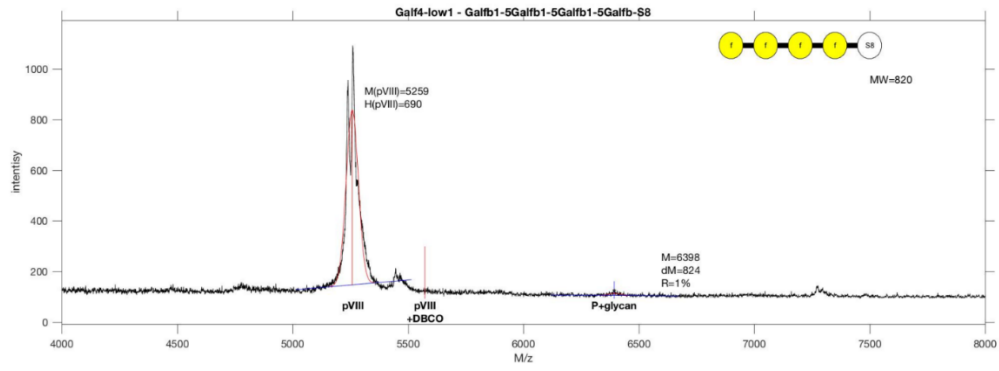




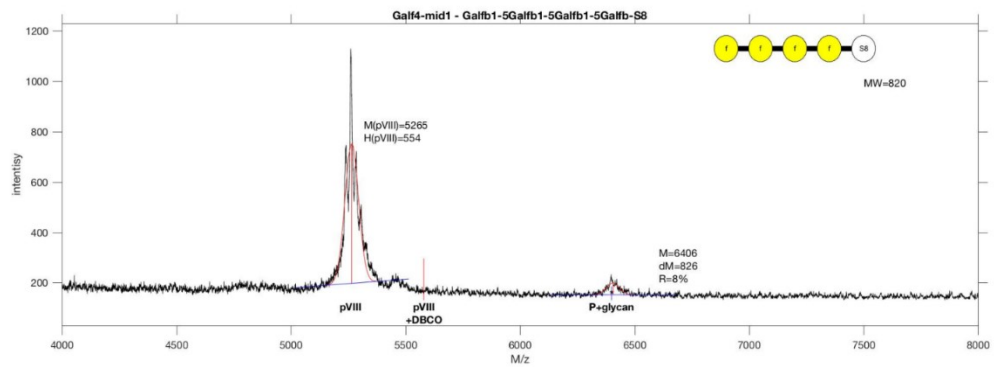
**Appendix C-70.** MALDI-TOF of glycosylated phages, Tr36



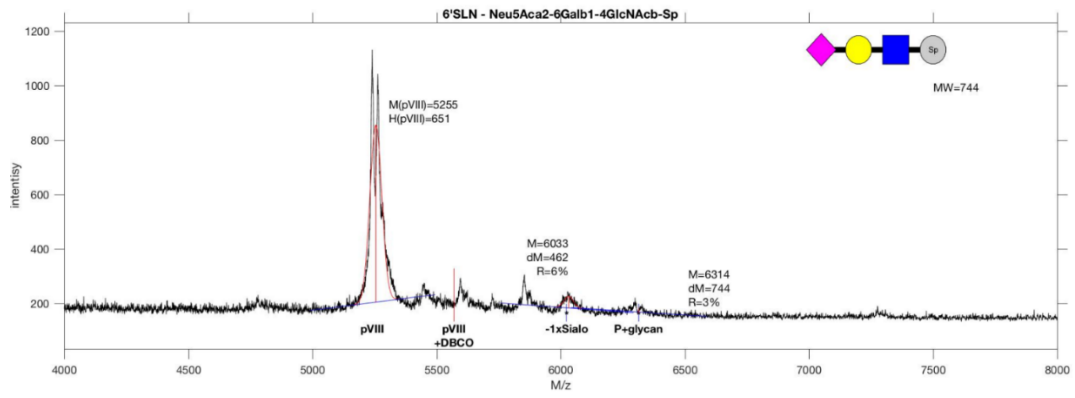
**Appendix C-71.** MALDI-TOF of glycosylated phages Tr39



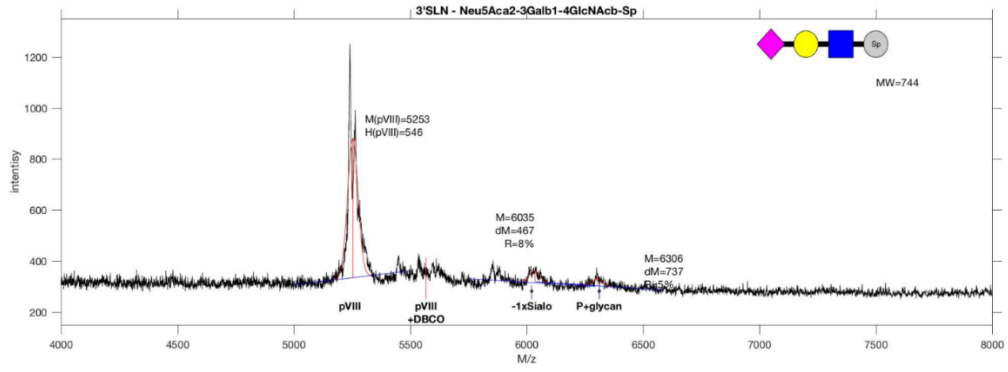
**Appendix C-72.** MALDI-TOF of glycosylated phages, Galf4-1



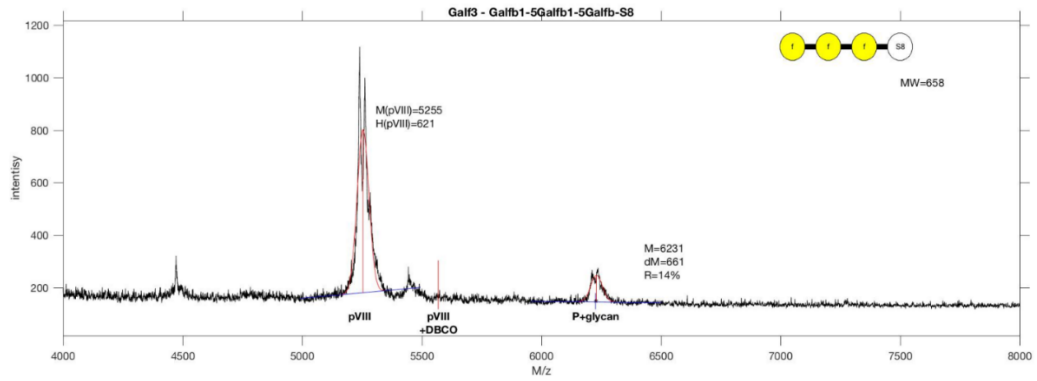
**Appendix C-73.** MALDI-TOF of glycosylated phages, Galf4-m



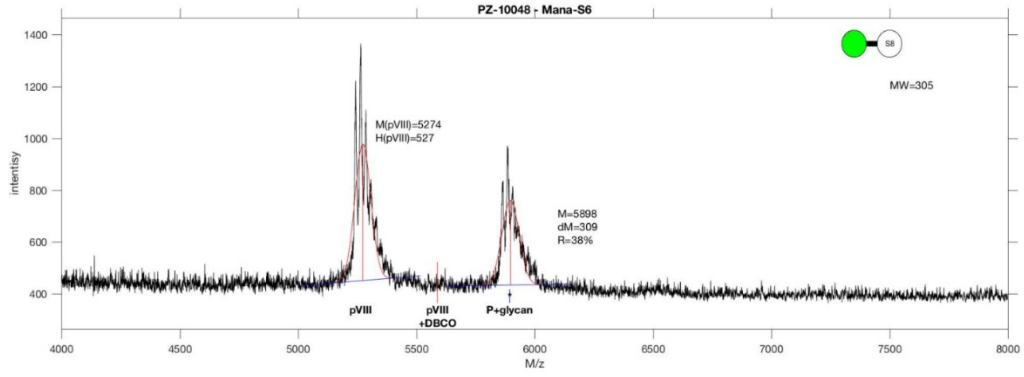
**Appendix C-74.** MALDI-TOF of glycosylated phages, 6'SLN



**Appendix C-75.** MALDI-TOF of glycosylated phages, 3'SLN



**Appendix C-76.** MALDI-TOF of glycosylated phages Galf3,



**Appendix C-78.** MALDI-TOF of glycosylated phages PZ-10048.



## Appendix C-79. Code source for MatLab script **plotONEmaldi.m**

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
addpath (fullfile('/Volumes/Data/!! OUTLINES/Jasmine -
LiGA/MALDI/Mirat/untitledfolder'));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

manual = 0;
filename = 'Phage+Lac_finish reaction.txt';
names = 'Lac';
full_temp = 'Galb1-4G1cb-S2';

XLIM1 = 3000;
XLIM2 = 8000;

%range around the peak for gaussian fitting and
baseline fit
RANGE = 255;

fid = fopen(filename,'r');
    disp([names ' glycan is in ' filename ' '
fgetl(fid)...
        char(10) fgetl(fid)]);

    FORM = '%f %f %*[^\\n\\r]';

    AllVar = textscan(fid,FORM);
    mass = AllVar{1};
    intensity = AllVar{2};

    disp(['Read ' num2str( size(mass, 1)) ' lines'
]);
fclose(fid);

IX = find(mass>XLIM1 & mass<XLIM2);

plot_mass = mass(IX);
plot_intensity = intensity(IX);
YMIN = min(plot_intensity);
YMAX = max(plot_intensity);
YH = YMAX-YMIN;
```

```

SCALE = 1000/YH;
plot_intensity = plot_intensity*SCALE;
YH = YH*SCALE;

YMIN = min(plot_intensity);
YMAX = max(plot_intensity);

margin = 0.1;

h = figure(1);
set(h, 'Units', 'normalized', 'position', [0.2 0.6
0.45 0.25] )

% find the full name based on abbreviated name

%%%%%%%%%% PLOT MALDI TRACE AND MAKE IT PRETTY
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
figure(1);
hold off;
plot(1,1);
hold on;
plot(plot_mass, plot_intensity, '-k');

% add full name of glycan to the title
title([names ' - ' full_temp]);

set(gca, 'yscale', 'lin', 'xscale', 'lin', ...
'TickDir', 'out');
ylabel('intensity');
xlabel('M/z');
xlim([4000 8000]);
ylim([YMIN - margin*YH YMAX + margin*YH]);
drawnow;
hold on;

IX2 = find ( plot_intensity == ...
            max(plot_intensity(plot_mass>5000
& plot_mass<5500)) );
pVIII = plot_mass(IX2(1));

% define the RANGE and make sure it
doesn't exceed the boundaries

```

```

if IX2+RANGE > size(plot_intensity,1)
    H = size(plot_intensity,1);
else
    H = IX2+RANGE;
end

if IX2-RANGE < 1
    L = 1;
else
    L = IX2-RANGE;
end

initial_mass = pVIII;
initial_int = max( plot_intensity(L:H) );

fo =
fitoptions('Method','NonlinearLeastSquares',...
           'StartPoint',[initial_int initial_mass
1 1 1]);
ft = fitype('a1*exp(-((x-
b1)/c1)^2)+e1*x+f1','options',fo);

f =
fit(plot_mass(L:H),plot_intensity(L:H),ft);

X = plot_mass(L:H);
MASS_P8 = f.b1;
MASS_P8_DBCO = MASS_P8 + 315;

HEIGHT_P8 = f.a1;
MAX_P8 = max(f(X));

plot(X, f(X), '-r');
plot(X, f.e1*X + f.f1, '-b');
line(MASS_P8*[1 1], [MAX_P8, MAX_P8-
HEIGHT_P8], 'color','r');

text(MASS_P8*1.01, MAX_P8,...
     ['M(pVIII)=' num2str(round(MASS_P8))
char(10)...
     'H(pVIII)='
num2str(round(HEIGHT_P8))],...
     'VerticalAlignment', 'bottom',...

```

```

        'HorizontalAlignment','left');

text(pVIII, YMIN, 'pVIII', ...
     'FontWeight', 'bold',...
     'VerticalAlignment', 'top',...
     'HorizontalAlignment','center');

line(MASS_P8_DBCO*[1 1], YMIN + [0 0.3*f.a1],
'color', 'r');
text(MASS_P8_DBCO, YMIN, ['pVIII' char(10)
'+DBCO'], ...
     'FontWeight', 'bold',...
     'VerticalAlignment', 'top',...
     'HorizontalAlignment','center');

%%%%%%%%%% EXTRACT THE FULL NAME AND MAKE GLYCAN OBJECT
%%%%%%%%%%

sugar = GlycanLeaf.createObj(full_temp);

hold on; plot(1,1);

drawGlycan('input',sugar.String, 'XY', [7500
YMAX],...
           'spacing', 150, 'angle', pi);

TEXT = ['MW=' num2str(round(sugar.MW)) char(10)];

DEBUG = 0;
[MW(char(10)),b,c] = sugar.MW;
if DEBUG

    for j = 1:numel(sugar.glycans)
        TEXT = [TEXT sugar.glycans{j} ...
                ': ' num2str(round(c(j) ))
char(10)];
    end
end

text(7500, 0.90*YMAX, TEXT, 'HorizontalAlignment',
'left',...
     'VerticalAlignment', 'top');

% set(gca,'xtick',[], 'ytick',[])

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

    MASS_P8_DBCO_GLY = MASS_P8_DBCO +
    MW(char(10));

```

```

    line(MASS_P8_DBCO_GLY*[1 1], YMIN + [0
0.1*HEIGHT_P8], 'color', 'b');

```

```

    text(MASS_P8_DBCO_GLY, YMIN, '*',...
        'FontWeight', 'bold',...
        'VerticalAlignment', 'bottom',...
        'HorizontalAlignment', 'center');
    text(MASS_P8_DBCO_GLY, YMIN, 'P+glycan',...
        'FontWeight', 'bold',...
        'VerticalAlignment', 'top',...
        'HorizontalAlignment', 'center');

```

```

    drawnow;

```

```

    repeating = 1;
    j=0;

```

```

    while repeating>0
        if manual
            choice = questdlg('define a peak?',...
                               'define a peak?',...
                               'define a peak','no mo
peaks','escape','define a peak');
            switch choice
                case 'escape', return;
                case 'no mo peaks', repeating = 0;
                case 'define a peak',

```

```

                    [x]=ginput(1);

```

```

                    j = j+1;
                    lineX(j) = x(1);

```

```

                end
            else
                j = j+1;
                if j==1
                    % find the main peak
                    lineX(j) = MASS_P8_DBCO_GLY;

```

```

IX = (strcmp('Neu5Ac',sugar.glycans)
|...
strcmp('Neu5,9Ac2',sugar.glycans) |...
      strcmp('Neu5Gc',sugar.glycans)
|...
      strcmp('KDN',sugar.glycans) );

repeating = sum(IX);

elseif j==2 && repeating

IX = ~(strcmp('Neu5Ac',sugar.glycans)
|...
strcmp('Neu5,9Ac2',sugar.glycans) |...
      strcmp('Neu5Gc',sugar.glycans)
|...
      strcmp('KDN',sugar.glycans) );
Nsialo = sum( ~IX );

[~,~,c]= sugar.MW;

asialoMW = sum(c(IX)) - 18*(sum(IX)-
1);

MASS_P8_DBCO_GLY = MASS_P8_DBCO +
asialoMW;

lineX(j) = MASS_P8_DBCO_GLY;

line(MASS_P8_DBCO_GLY*[1 1], YMIN + [0
0.1*HEIGHT_P8], 'color', 'b');

text(MASS_P8_DBCO_GLY, YMIN, '*',...
      'FontWeight', 'bold',...
      'VerticalAlignment',
'bottom',...
'HorizontalAlignment','center');
text(MASS_P8_DBCO_GLY, YMIN,...
      ['-' num2str(Nsialo) 'xSialo']
,....
      'FontWeight', 'bold',...
      'VerticalAlignment', 'top',...

```

```

'HorizontalAlignment','center');

        repeating = 0;

    end

    % find the index of data nearest to
the mouse click
    IX2 = max ( find( plot_mass < lineX(j)
));

        % define the RANGE and make
sure it doesn't exceed the boundaries

    if IX2+RANGE > size(plot_intensity,1)
        H = size(plot_intensity,1);
    else
        H = IX2+RANGE;
    end

    if IX2-RANGE < 1
        L = 1;
    else
        L = IX2-RANGE;
    end

    initial_mass = lineX(j);
    initial_int = max(
plot_intensity(L:H) );

    fo =
fitoptions('Method','NonlinearLeastSquares',...
            'StartPoint',[initial_int
initial_mass 1 1 1]);
    ft = fitype('a1*exp(-(x-
b1)/c1)^2)+e1*x+f1','options',fo);

    f = fit(plot_mass(L:H),...
            plot_intensity(L:H),ft);

    X = plot_mass(L:H);
    MASS(j) = f.b1;
    HEIGHT(j) = f.a1;
    MAX(j) = max(f(X));

```

```

        plot(X, f(X), '-r');
        plot(X, f.e1*X + f.f1, '-b');
        line(MASS(j)*[1 1], [MAX(j), MAX(j)-
HEIGHT(j)], 'color','r');

        Delta(j) = round(abs(MASS(j)-
MASS_P8_DBCO));
        Ratio(j) = round(100*HEIGHT(j) /
(HEIGHT(j) + HEIGHT_P8));

        if j==1
            text(MASS(j)+200, MAX(j)*0.7,...
                ['M=' num2str(round(MASS(j)))
char(10)...
                'dM=' num2str(Delta(j))
char(10)...
                'R=' num2str(Ratio(j)) '%'
char(10)],...
                'VerticalAlignment',
'bottom',...
'HorizontalAlignment','left');
        elseif j==2
            text(MASS(j), MAX(j) + (MAX(j)-
HEIGHT(j))*0.1,...
                ['M=' num2str(round(MASS(j)))
char(10)...
                'dM=' num2str(Delta(j))
char(10)...
                'R=' num2str(Ratio(j)) '%'
char(10)],...
                'VerticalAlignment',
'bottom',...
                'HorizontalAlignment','right');
        end
    end
end
end
end

```



## Appendix C-80. Code source for R script DE\_analysis.R

```
# DE analysis of sequencing data
# Daniel Ferrer
# ferrervi@ualberta.ca
# 2019-01-15

library(ggplot2)
library(stats)
library(tidyr)
library(edgeR)

options(scipen=999)

##### EdgeR implementation
#####

x <- read.csv("bg_bc.csv") # compare pair-wise
head(x)

#### 1_Define Groups and replicates:

group <- factor(c(1,1,1,2,2,2,2,2,2)) # 2 groups

#### 2_Create DGEList (EdgeR object)

y <- DGEList(counts=x[,2:10], genes=x[,1]) # specify
your reads are col 2 to 9 and glycans==genes are col 1
y <- calcNormFactors(y) #
Normalization... input data should be no normalized

#### 3_Create design matrix

design <- model.matrix(~group)
y <- estimateDisp(y,design)

#### 4_Data exploration:

plotMDS(y) # PC analysis, the distance between points
on plot

#### 5_DE Analysis

#### 5.1_Method 1: Quasi-likelihood F-test (qlf)
```

```
#### 5.1.1_Compare G3C Vs Input

fit <- glmQLFit(y,design)
qlf.2vs1 <- glmQLFTest(fit,coef=2)

hits <- topTags(qlf.2vs1)
top <- rownames(topTags(qlf.2vs1))
hits

plotMD(qlf.2vs1) #
indicate 2-fold up or down
abline(h=c(-1,1), col="blue")
```

## Appendix C-80. Code source for R script `makeContrast.py`

```
import itertools
import pandas as pd
import numpy as np
import glob as gb

filenames = gb.glob('*.txt') # Gather a list with
filenames

c = 0 # initialize counter to index 1st position on
filename list
for i in range(len(filenames)):
    fo = open((str(filenames[c])[0:20])+'.csv', 'w+')
# nested str() function to create strings.trim name
    with open(filenames[c], 'r+') as f:
        head = f.readline()
# keep 1st line
        fo.write(','.join(head.split(' ')[3:]))
# convert to csv
        for line in itertools.islice(f, 2, None):
# start=2, stop=None
            tokens = line.split(' ')
# make a list of strings
            fo.write(','.join(tokens[3:]))
# convert to csv
            c= c+1
            fo.close()

##### DATA CONSOLIDATION
#####
filenames = gb.glob('*.csv')
d = 0
for f in filenames:
    df = pd.read_csv(filenames[d]).replace(0,1)
    df = df.groupby(['AA'], as_index=
False).agg('sum')
    df.to_csv(filenames[d])
    d = d + 1

##### MAKE CONTRAST TABLES
#####

filenames = gb.glob('*.csv')

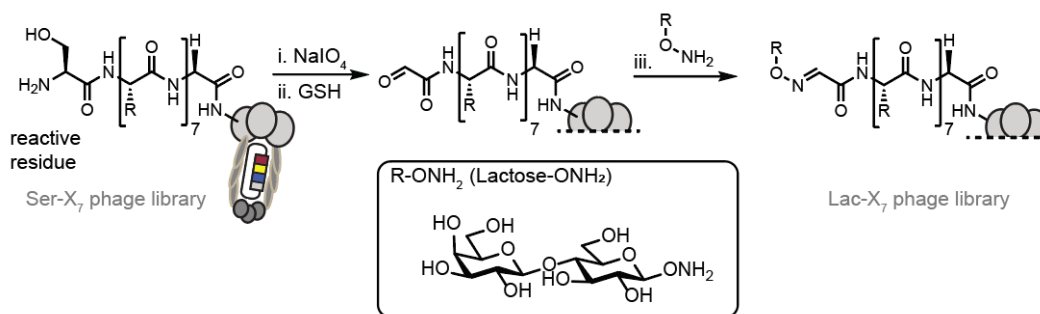
# User needs to define Test and Control experiments
```

```
ctrl = pd.read_csv(filenamees[0]) # Control
test = pd.read_csv(filenamees[1]) # Test

df1 = ctrl.merge(test, on = "glycan", how='outer')
df2 = df1.fillna(value=1.1) # fill missing values
with 1.1
df2.to_csv('bg_sb.csv') # save the data
```

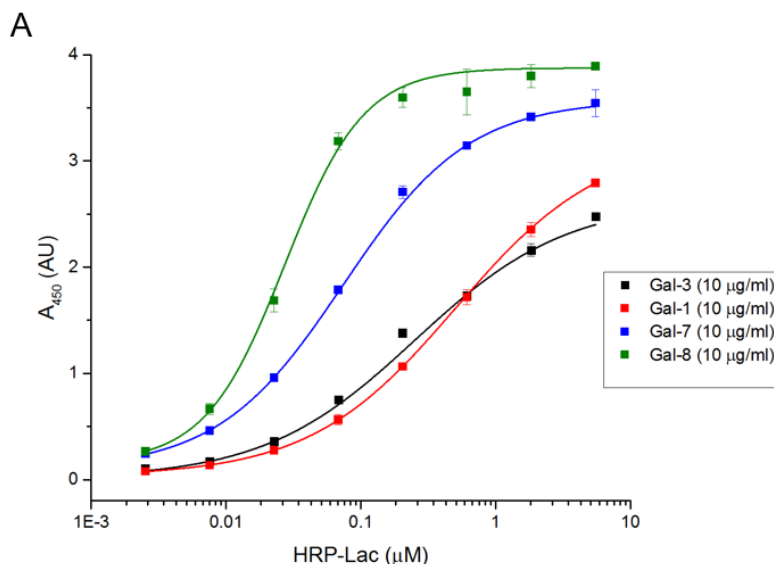
## Appendix D: Selection of lactose-modified phage libraries against galectin family proteins

In this study we decorated random 7mer peptide phage library with anomeric aminoxy lactose substrate to obtain phage displayed peptides bearing lactose residue (Scheme 1). Next, we screened those libraries against plate immobilized galectin-1, galectin-3, galectin-7 or galectin-8.



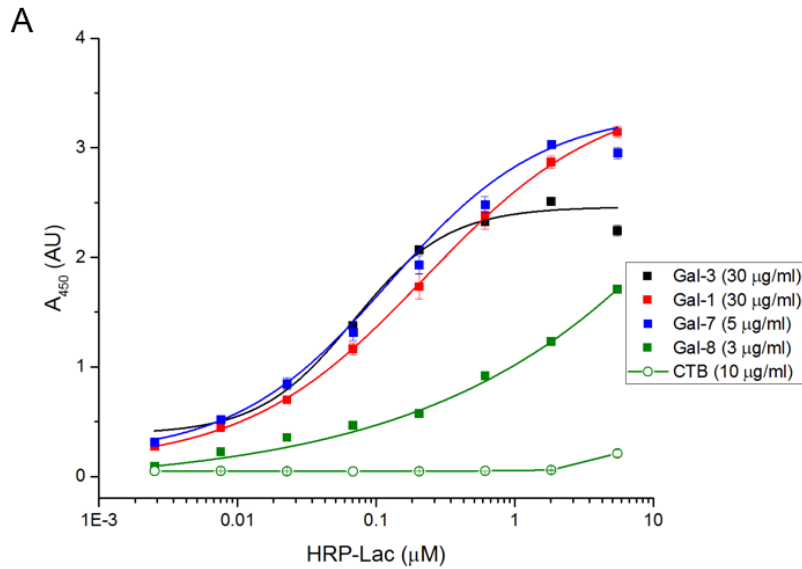
**Scheme 1.** Chemical modification of Ser-X7 phage library with hydroxylamine derivative of lactose (Lactose-O-NH<sub>2</sub>).

Target immobilization, by passive adsorption into plastic 96 well plate resulted in protein-coated plate with non-uniform, binding affinity towards the conjugated lactose-horseradish peroxidase (Appendix D-1).

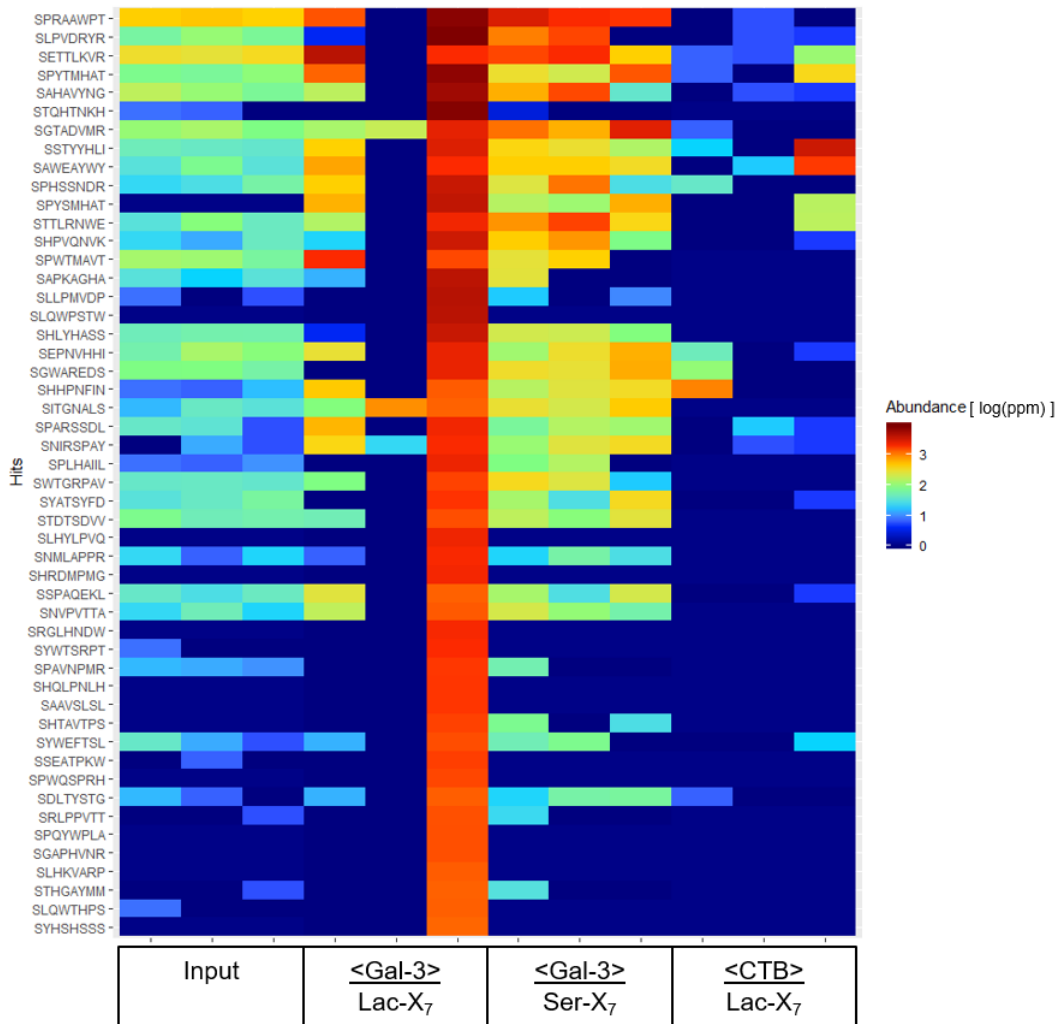


**Appendix D-1** Concentration dependent response curves for the binding of HRP-Lac conjugate to plate-immobilized galectin-1 (Gal-1), galectin-3 (Gal-3), galectin-7 (Gal-7), and galectin-8 (Gal-8).

The binding to HRP-Lac of tandem-dimeric Gal-8 and Gal 7 showed higher  $EC_{50}$  ( $\sim 0.04 \mu\text{M}$ ) as compared with monomeric Gal-3 and Gal-1 ( $EC_{50} \sim 0.4 \mu\text{M}$ ). To achieve uniform activity on coated plate we used different concentrations of proteins and included unrelated carbohydrate binding protein cholera toxin-B (CTB) as control (Appendix D-2).

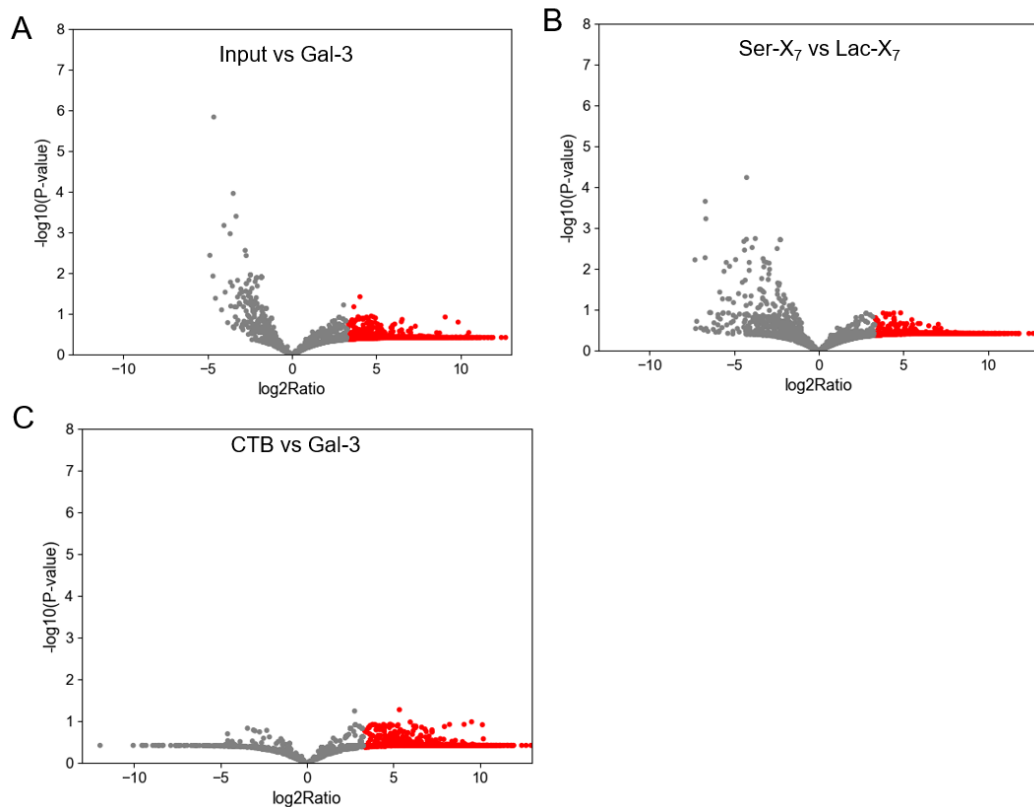


**Appendix D-2.** Curves for the binding of HRP-Lac conjugate to Gal-1, Gal-3, Gal-7, Gal-8 and CTB immobilized on plate. Passive adsorption on plate with different protein concentrations.



**Appendix D-3.** Top 50, most abundant Lac-peptide combinations identified by DNA sequencing from selection experiments on Gal-3 coated wells, compared with input, peptide without lactose (unmodified library = Ser X7) and, selected against unrelated CTB protein target.

These experiment however, are inconclusive. The statistical significance of the magnitude of enrichment for selected sequences against Gal-3 for example was very low. Volcano plots of the differential enrichment analysis for Gal-3 selection compared to input, Ser-X7 and CTB visualized very low statistical significance for enriched sequences (Appendix D-4).

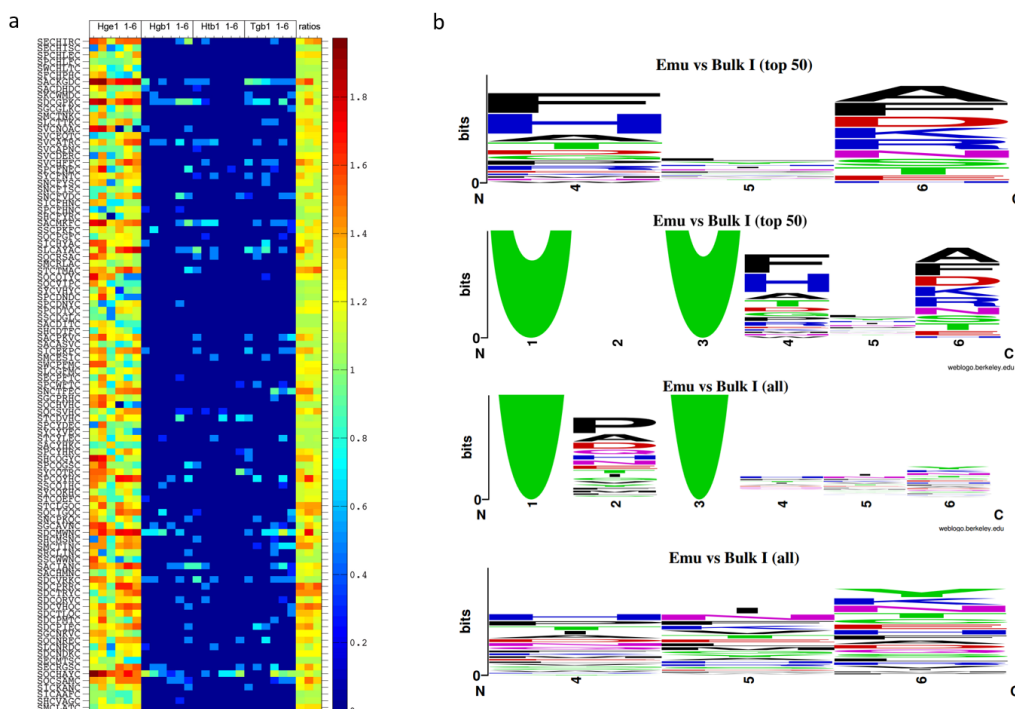


**Appendix D-4.** Volcano plots of DNA-sequencing data from selection of Lac-X<sub>7</sub> phage libraries against Gal-3 compared to Input, Ser-X<sub>7</sub> and CTB.



## Appendix E: Selection of galactose-modified macrocyclic phage libraries against galectin-3.

We modified a SCX<sub>3</sub>C phage library with dichloroacetone derivatives oxime linked to Galactose (DCOGal) or aglycon (DCOMet) and generated macrocyclic glycopeptide libraries displayed on M13 phage. We screened the macrocyclic glycopeptide libraries against Galectin-3 and cholera toxin-B (CTB). From the analysis of deep-sequencing data we identified 176 hits that were 10-fold enriched as compared with results from CTB (Appendix E-1).



**Appendix E-1 A.** Hits present in emulsion-based screening but not present in any other screens. The hits are at least 10 times more abundant compared to all controls, with significance threshold of  $p < 0.05$ . **B.** LOGO plots for the population of sequences enriched on emulsion amplification (Emu) but no on any of the others controls. Only on top 50 hits there could be some consensus Results are based on two independent instances of panning.

Analysis from the hit population by sequence logo identified proline (P) phenylalanine (F), histidine (H), aspartic acid (D), and alanine (A) as conserved reisdues on selected sequences.

