

**Genetically Encoded Fragment-Based Discovery of Inhibitors for
Glycan-Binding Proteins**

by

Simon Ng

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

Doctor of Philosophy

Department of Chemistry
University of Alberta

© Simon Ng, 2016

Abstract

Carbohydrates play vital roles in many disease pathologies, including inflammatory disease, cancer metastasis, autoimmune disease and pathogenic infection. Despite their significance and immense opportunity for pharmaceutical application, carbohydrate-derived drugs only constitute a relatively small portion of therapeutics today. One of the major challenges in drug discovery for glycan-binding proteins is their relatively weak binding affinities for the glycan. The issue is further aggravated by the structural complexity of glycans, which often leads to a bottleneck in generation of various different analogues in sufficient quantity for *in vitro* and *in vivo* studies. Fragment-based lead discovery was used to design better lead compounds for therapeutic targets and demonstrated its potential for challenging targets such as glycan-binding proteins. However, there remains an unmet need for simpler, less expensive, and more efficient routes to identify superior lead compounds.

This thesis approaches these challenges, *i.e.*, the need for simpler and more efficient method to identify lead compounds for glycan-binding proteins, by combining fragment-based lead discovery with phage display. Through the linking of each “displayed peptide” to its encoding DNA, phage libraries are several orders of magnitude ($\sim 10^8$ – 10^9) larger than chemical libraries, but still can be screened against a target in a very efficient way. I pioneered a methodology to selectively and covalently attach a small-

molecule fragments to a phage-displayed peptide libraries (Chapter 2). The resulting bivalent fragment-peptide library demonstrated its application for the selection of potent inhibitors for glycan-binding proteins. By using ConA as a model target of a glycan-binding protein, I discovered novel monosaccharide-peptide conjugates with K_D and IC_{50} values 30–50-fold better than that of the monosaccharide itself (Chapter 3). The conjugates bound competitively to the desired glycan-binding site and functioned as potent inhibitors. The most potent hit, Man-WYDLF, has similar affinity and selectivity as a trimannoside. Interestingly, the peptide bound to a secondary site different from that bound by the disaccharide portion of the trimannoside. The selection identified a novel binding pocket and suggested the potential use of this technology for “hot spot” mapping. I further applied this technology for the selection of mannose-based inhibitors for DC-SIGN, albeit with partial success (Chapter 4). I applied computational solvent mapping (FTMap) to analyze the protein surface of DC-SIGN to identify putative fragment-binding “hot spot”. The results suggested that a “hot spot” is available at ~ 10 Å away from the Ca^{2+} glycan-binding site and could be accessible for library attached to 2-OH position of mannose or anomeric position of fucose. Future work is required to confirm this hypothesis. Finally, I demonstrated a strategy to create the first macrocyclic glycopeptide library displayed on phage (Chapter 5). The disulfide-containing peptide library could be alkylated chemo- and regio-selectively using a dichloro-oxime derivative to form macrocycles stable toward reduction. These

libraries, both linear and cyclic glycopeptides, show potential for use in fragment-based ligand discovery, “hot spot” mapping, and selection of potent ligands for many protein targets. Ligands resulting from the selection will strongly complement and “guide” the rational design of more potent and “drug-like” compounds or *vice versa*.

Preface

Chapter 2 is based on published works and reproduced with permission from S. Ng, M.R. Jafari, W.L. Matochko and R. Derda, “Quantitative synthesis of genetically encoded glycopeptide libraries displayed on M13 phage”, *ACS Chem. Biol.* **2012**, 7, 1482–1487. Copyright © 2012 American Chemical Society. I optimized the reactions on phage, developed the method for characterization of on-phage reactions, as well as manuscript composition and edits. M.R. Jafari helped validating the characterization using other type of reactions. W.L. Matochko aided the transformation of SVEK phage. R. Derda was involved with the concept formation, analysis of reaction kinetics and manuscript composition plus edits.

Chapter 3 is based on published works and reproduced with permission from S. Ng, E. Lin, P.I. Kitov, K.F. Tjhung, O.O. Gerlits, L. Deng, B. Kasper, A. Sood, B.M. Paschal, P. Zhang, C.-C. Ling, J.S. Klassen, C.J. Noren, L.K. Mahal, R.J. Woods, L. Coates and R. Derda, “Genetically encoded fragment-based discovery of glycopeptide ligands for carbohydrate-binding proteins”, *J. Am. Chem. Soc.* **2015**, 137, 5248–5251. Copyright © 2015 American Chemical Society. I was responsible for all the phage selection. E. Lin and I synthesized the hits and validated their binding affinity with ITC. I also validated the hits with SPR and STD NMR. R. Derda and I were involved with the analysis of post-selection data. P.I. Kitov synthesized some of the

mannose derivatives and co-developed the competitive binding assay. B.M. Paschal and C.J. Noren provided the phage library used in the first round of selection, whereas K.F. Tjhung made the phage library used in affinity maturation. O.O. Gerlits and L. Coates crystallized the ligand–protein complexes. L. Deng and J.S. Klassen were involved with the ESI-MS binding assay. B. Kasper and L.K. Mahal were responsible for the lectin microarray and its analysis. A. Sood and R.J. Woods performed the MD simulations and analysis. P. Zhang and C.-C. Ling provided the trimannoside derivative. R. Derda and I were involved with the idea formation, manuscript composition and edits with useful discussion provided by E. Lin, P.I. Kitov, K.F. Tjhung, J.S. Klassen, C.J. Noren, L.K. Mahal, L. Coates, and R.J. Woods.

Chapter 4 is based on non-published work with the ideas composed by R. Derda and refined by me. The extracellular tetrameric DC-SIGN is provided by K. Drickamer. I performed most of the experimental works and FTMap analysis. N. Gao synthesized the hits and helped with part of their validations.

Chapter 5 is based on non-published work initiated by me and refined by R. Derda. I performed most of the experimental works. P.I. Kitov helped to synthesize carbohydrate-hydroxylamine derivatives. K.F. Tjhung made the phage libraries used in this work.

Acknowledgements

First and foremost, I want to express my deepest gratitude to my parents for nurturing me and creating a free environment for my early growth and development. Your hardship and sacrifices that assured my later success is priceless and what I deeply owe you for.

The route to complete Ph.D. is a long journey in life. Without support, help and encouragement, the journey would be bleak and full of despair. Along this journey, the person to whom I would like to express my upmost appreciation is my advisor, Prof. Ratmir Derda. I am thankful to Ratmir for many of the thought-provoking discussions, his trust, his unlimited support and perpetual guidance for me over the past five years. I feel fortunate and proud to have the opportunity to work with Ratmir during his early career. His perseverance to push my limits and to unlock my potential is what I am very grateful about.

I am thankful to Prof. Todd L. Lowary and Prof. John S. Klassen for being my committee members. My first compound ever synthesized during my Ph.D. program was made possible by Prof. Lowary, who offered a space in his lab and numerous chemicals to me during our start-up. I feel grateful for his generosity and valuable advice given to me throughout the years. I especially thank Prof. Klassen for being my long-term collaborator and contributing in many aspects of our projects. I would also like to thank Prof. Frederick G. West for taking his time to be the jury of my candidacy exam and

thesis defense. I appreciate Prof. Pehr A.B. Harbury for coming all the way to Edmonton to be my thesis jury.

Next, I would like to thank my fellow group members, past and present, starting from Wadim and Reza who are such awesome colleagues and friends. They have embraced the up-and-down with me along the journey. Each of you has certain talents that I lack of and have learnt from. I am grateful to Edith who is my best lab partner and has demonstrated passion in whatever tasks she was given to. I especially thanks for her help throughout the ConA project. I also want to thank Katrina for her dedicated contributions in some aspect of my projects. She is always approachable and has demonstrated rigorous scientific practices even at such a young age! I am grateful to Frédérique and Pavel, who have imparted their knowledge and offered much valuable advice and help to me. Many thanks to all my internal and external collaborators who have contributed in every possible way to help me conquer the challenges along the journey.

Finally, I wish to acknowledge the fellowship support from Alberta Innovates Technology Futures, funding support from Alberta Glycomics Centre and awards given by Gilead Alberta and University of Alberta. Without these supports, the journey would definitely be less pleasant.

Last but not least, I would like to thank Chantel for accompanying me along my Ph.D. journey. Her patience with me and endless of support is what I am most grateful for and to be cherished forever.

TABLE OF CONTENTS

Abstract	ii
Preface	v
Acknowledgements	vii
List of tables	xiii
List of figures	xiv
List of abbreviations	xvii
Chapter 1: Introduction	1
1.1 Motivation and objectives	1
1.2 Glycan-binding proteins (GBPs).....	4
1.2.1 General principles of inhibitor design for GBPs.....	8
1.3 Fragment-based lead discovery (FBLD)	11
1.3.1 Why FBLD for glycan-binding proteins?.....	12
1.3.2 Design principles and strategies.....	13
1.3.3 Examples of application for GBPs.....	16
1.4 High-throughput FBLD.....	21
1.4.1 Chemotype evolution	22
1.4.2 DNA-encoded chemical library.....	23
1.4.3 Fragment-based display technology	24
1.5 Thesis overview	26
Chapter 2: Quantitative synthesis of genetically encoded glycopeptide libraries displayed on M13 phage.....	29
2.1 Introduction	29
2.2 Results and discussion.....	31
2.2.1 Challenges for characterization of on-phage reactions	31
2.2.2 Method for characterization of on-phage reactions.....	32
2.2.3 Selectivity of the reactions and their optimizations	33
2.2.4 Relationship of capture yield and reaction yield.....	38
2.2.5 Synthesis of glycopeptide on monoclonal phage.....	40

2.2.6	Synthesis of glycopeptide libraries.....	42
2.3	Conclusion.....	43
2.4	Experimental procedures.....	45
2.4.1	Materials and general information	45
2.4.2	Quantification of on-phage reactions.....	45
2.4.3	Synthesis of 2-(aminooxy)ethyl α -D-mannopyranoside	46
2.4.4	Synthesis of H ₂ N-SVEKY-CONH ₂ (SVEKY).....	52
Chapter 3: Genetically encoded fragment-based discovery of glycopeptide ligands for carbohydrate-binding proteins		54
3.1	Introduction	54
3.2	Hypothesis and objective.....	54
3.3	Results and discussion.....	56
3.3.1	Selection of glycopeptide libraries.....	56
3.3.2	Validation of selected hits	59
3.3.3	Optimization and affinity maturation.....	61
3.3.4	Structural analysis of Man-WYD-ConA complexes	62
3.3.5	Effect of linkers on binding affinity.....	65
3.3.6	Binding selectivity of glycopeptides.....	67
3.4	Conclusion.....	70
3.5	Experimental procedures.....	70
3.5.1	Materials and general information	70
3.5.2	Generation of Man-X ₇ and Me-X ₇ libraries.....	71
3.5.3	Phage selection	72
3.5.4	Phage amplification and PCR	73
3.5.5	Ion Torrent sequencing.....	75
3.5.6	Analysis by volcano plot and generation of LOGO	75
3.5.7	Surface plasmon resonance spectroscopy	76
3.5.8	Isothermal titration calorimetry.....	77
3.5.9	1D STD-NMR experiments	78
3.5.10	Competitive binding assay.....	80
3.5.11	Examples for the synthesis of Man-peptide conjugates.....	81
Chapter 4: Phage selection of DC-SIGN inhibitors		85
4.1	Introduction	85
4.2	Previous works on DC-SIGN inhibitors	86

4.3	Results and discussion.....	88
4.3.1	Selection of glycopeptide libraries.....	88
4.3.2	Validation of selected hits	92
4.3.3	Validation of the reported DC-SIGN inhibitor.....	94
4.3.4	Validation of FTMap for glycan-binding proteins	95
4.3.5	FTMap analysis of DC-SIGN.....	98
4.4	Conclusion.....	100
4.5	Experimental procedures.....	101
4.5.1	Materials and general information	101
4.5.2	Generation of Man-X ₇ and AOB-X ₇ libraries	101
4.5.3	Pre-selection of phage libraries	103
4.5.4	Three rounds of phage selection.....	103
4.5.5	Phage amplification and PCR	104
4.5.6	Validation of hits by an inhibition assay.....	106
4.5.7	FTMap analysis.....	107
Chapter 5: Phage-displayed macrocyclic glycopeptide libraries.....		108
5.1	Introduction	108
5.2	Previous syntheses of macrocyclic glycopeptides	110
5.3	Results and discussion.....	112
5.3.1	Our approach to create macrocyclic glycopeptide libraries .	112
5.3.2	Macrocyclization of a model peptides with DCO derivatives	113
5.3.3	Macrocyclization of phage libraries with DCO-biotin.....	120
5.3.4	Macrocyclization of phage libraries with DCA and oxime ligation.....	123
5.3.5	Validation of “pulse-chase” reaction on model peptide.....	125
5.3.6	Synthesis of macrocyclic glycopeptide libraries.....	127
5.3.7	Preliminary selection of macrocyclic glycopeptide libraries	129
5.4	Conclusion.....	131
5.5	Experimental procedures.....	132
5.5.1	Materials and general information	132
5.5.2	Synthesis of dichloro-oxime-biotin (DCO-biotin)	133
5.5.3	Synthesis of dichloro-oxime-carbohydrate	133
5.5.4	Macrocyclization of SWCSC with DCO derivatives.....	135
5.5.5	Kinetic studies of modification of phage with DCO-biotin.....	135
5.5.6	Pulse-chase reaction of SWCSC with DCO-ManL and BIA.....	136

5.5.7 Modification of phage libraries with DCO-carbohydrate derivatives.....	136
Chapter 6: Conclusion and outlook.....	138
6.1 Conclusion.....	138
6.2 Future directions.....	141
References	143
Appendix A: Supporting information for chapter 2.....	161
Appendix B: Supporting information for chapter 3.....	171
Appendix C: Supporting information for chapter 4.....	189
Appendix D: Supporting information for chapter 5.....	196
Appendix E: Wittig reaction on phage	200
Appendix F: Spontaneous cyclization on phage	211

List of tables

Table 3-1. K_D values of synthetic ligands with ConA.	61
---	----

List of figures

Figure 1-1. Molecular structure of glycan-derived drugs (trade name).....	2
Figure 1-2. The roles of glycan-GBP interactions in human diseases.....	6
Figure 1-3. Design principles of Rivipansel.....	9
Figure 1-4. Examples of FBLD-derived drugs.	11
Figure 1-5. Demonstration of fragment growing.....	15
Figure 1-6. Demonstration of fragment linking.	16
Figure 1-7. Design of MAG inhibitor with fragment linking.	17
Figure 1-8. Design of E-selectin inhibitor with fragment linking.	18
Figure 1-9. Design of Siglec-1 inhibitor with <i>in silico</i> fragment growing.	20
Figure 1-10. Chemotype evolution.....	23
Figure 2-1. Biotin-capture assay.	33
Figure 2-2. Profiling of the optimal conditions for oxidation and oxime ligation.	34
Figure 2-3. Optimization of the conditions for <i>in-situ</i> quenching of the oxidation reaction.....	36
Figure 2-4. Profiling of the kinetics of oxidation.....	37
Figure 2-5. Reaction on multiple copies of pIII protein and the kinetics.	40
Figure 2-6. Quantification of glycopeptide formation on monoclonal phage or phage library.	42
Figure 2-7. Synthesis of 2-(aminoxy)ethyl α -D-mannopyranoside.....	47
Figure 3-1. Illustration of generation of glycopeptide library and its selection.	56

Figure 3-2. Overview of the workflow of phage selection.....	57
Figure 3-3. Post-selection analysis.....	58
Figure 3-4. SPR inhibition profiles of Man-X ₇ conjugates.....	59
Figure 3-5. SPR inhibition profiles of control peptide and conjugates.....	60
Figure 3-6. X-ray crystal structure of Man-WYD–ConA complexes.....	63
Figure 3-7. Binding free energy analysis of ConA bound to Man-WYD.....	64
Figure 3-8. Effect of linkers on binding affinity.....	65
Figure 3-9. STD-NMR analysis of the interaction between ConA and glycopeptides.....	66
Figure 3-10. Lectin microarray analysis.....	67
Figure 3-11. Structural and sequence homology of ConA–LcH and ConA–PSA.	68
Figure 4-1. Illustration of generation of glycopeptide library and its selection.....	89
Figure 4-2. The recovery of phage from biopanning.....	90
Figure 4-3. Inhibition assay and validation of hits.....	93
Figure 4-4. Validation of the inhibitor reported in the literature.....	94
Figure 4-5. Computational solvent mapping (FTMap).....	95
Figure 4-6. Mapping of the locations of potential binding fragments near the glycan-binding site of ConA using FTMap.....	96
Figure 4-7. Mapping of the locations of potential binding fragment near the glycan-binding site of galectin-3 using FTMap.....	97
Figure 4-8. Mapping of the locations of potential binding fragment near glycan-binding site of DC-SIGN using FTMap.....	98

Figure 5-1. Macrocyclization of peptides and genetically encoded peptide libraries with dichloro-oxime derivatives.	112
Figure 5-2. Macrocyclization of linear peptide.	114
Figure 5-3. Reaction of SWCSC with DCO-ManS.	116
Figure 5-4. Reaction of SWCSC with DCO-ManL.	117
Figure 5-5. Reaction of SWCSC with DCO-Fuc.	118
Figure 5-6. Reaction of SWCSC with DCO-Gal.	119
Figure 5-7. Efficiency of macrocyclization of the phage libraries.	120
Figure 5-8. Two-step macrocyclization and functionalization of phage libraries with biotin.	124
Figure 5-9. Macrocyclization of linear peptides under pulse-chase conditions.	125
Figure 5-10. Kinetic profile of reaction between SWCSC and BIA.	127
Figure 5-11. Macrocyclization of phage libraries with four different DCO-carbohydrate derivatives.	128
Figure 5-12. Results of selections of cyclic glycopeptides against DC-SIGN and ConA.	130

List of abbreviations

AOB	Aminooxy-biotin
AOMan	2-(Aminooxy)ethyl α -D-mannopyranoside
BIA	Biotin-PEG ₂ -iodoacetamide
BSA	Bovine serum albumin
ConA	Concanavalin A
DCA	1,3-Dichloroacetone
DCO	1,3-Dichloro-oxime derivatives
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3 -grabbing non-integrin
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMF	Dimethylformamide
ESI	Electrospray ionization
FBLD	Fragment-based lead discovery
Fmoc	9-Fluorenylmethoxycarbonyl
FTSC	Fluorescein-5-thiosemicarbazide
Fuc	L-Fucose
Gal	D-Galactose
GalNAc	<i>N</i> -acetyl-D-galactosamine
GBP	Glycan-binding protein
GE-FBD	Genetically encoded fragment-based discovery
Glc	D-Glucose
GlcA	D-Glucuronic acid
GlcNAc	<i>N</i> -acetyl-D-glucosamine
GSH	Glutathione
h	Hour
HBTU	<i>N,N,N',N'</i> -Tetramethyl- <i>O</i> -(1 <i>H</i> -benzotriazol-1-yl)uronium hexafluorophosphate
HEPES	<i>N</i> -(2-hydroxyethyl)piperazine- <i>N'</i> -ethanesulfonic acid

HRMS	High-resolution mass spectroscopy
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
kDa	Kilo dalton
LcH	<i>Lens culinaris</i>
LE	Ligand efficiency
MALDI	Matrix-assisted laser desorption ionization
Man	D-Mannose
MDa	Mega dalton
MeMan	Methyl α -D-mannopyranoside
min	Minute
MOPS	3-(<i>N</i> -morpholino)propanesulfonic acid
MS	Mass spectroscopy
MWCO	Molecular weight cut-off
NeuAc	sialic acid
NMR	Nuclear magnetic resonance
PBS	Phosphate buffer saline
PDB	Protein data bank
PFUs	Plaque forming units
PSA	<i>Pisum sativum</i>
RMSD	Root-mean-square deviation
RP-HPLC	Reverse phase high-performance liquid chromatography
RT	Room temperature
SA	Streptavidin
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
STD	Saturation-transfer difference

<i>t</i> Bu	Tert-butyl
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
UAA	Unnatural amino acids
WT	Wild-type
Xgal	5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside
Xyl	D-Xylose

Chapter 1: Introduction

1.1 Motivation and objectives

Carbohydrates play vital roles in many disease pathologies, including inflammatory disease, cancer metastasis, autoimmune disease and pathogenic infection. For example, in the case of inflammatory disease, aberrant trafficking of immune cells to inflamed tissues has been attributed to the enhanced expression of E-selectin, which binds to the sialyl Lewis^x glycan structure.¹ In another example, galectin-1, a *N*-acetyl-lactosamine-binding protein, is overexpressed by breast cancer tumors that promote their metastasis to the lung by creating an immunosuppressive microenvironment at the site of tumor growth.² Many pathogens have the capacity to specifically recognize the oligosaccharides expressed on host tissues and exploit these interactions for virulent infection.³ Despite their significance and immense opportunity for pharmaceutical application, carbohydrate-derived drugs only constitute a relatively small portion of therapeutics today.⁴ The most well-known drugs derived from glycans are the viral neuraminidase inhibitors—zanamivir (Relenza) and oseltamivir phosphate (Tamiflu)—for the prevention of infection by the influenza virus (Figure 1-1). Other successful drugs include α -glucosidase inhibitors—miglitol (Glyset) and voglibose (Glustat)—for the treatment of diabetes; and a pentasaccharide heparinoid—fondaparinux (Arixtra)—for the treatment of thrombosis. Despite the success of these drugs, considerable numbers of glycan-binding

proteins still remain untapped as therapeutic targets. This deficiency is mainly due to the challenge in the design of potent inhibitors suitable for pharmaceutical use. Weak intrinsic affinity of glycan-protein interaction and poor pharmacokinetic properties of glycans, such as low bioavailability and short plasma half-life, are the major hurdles.

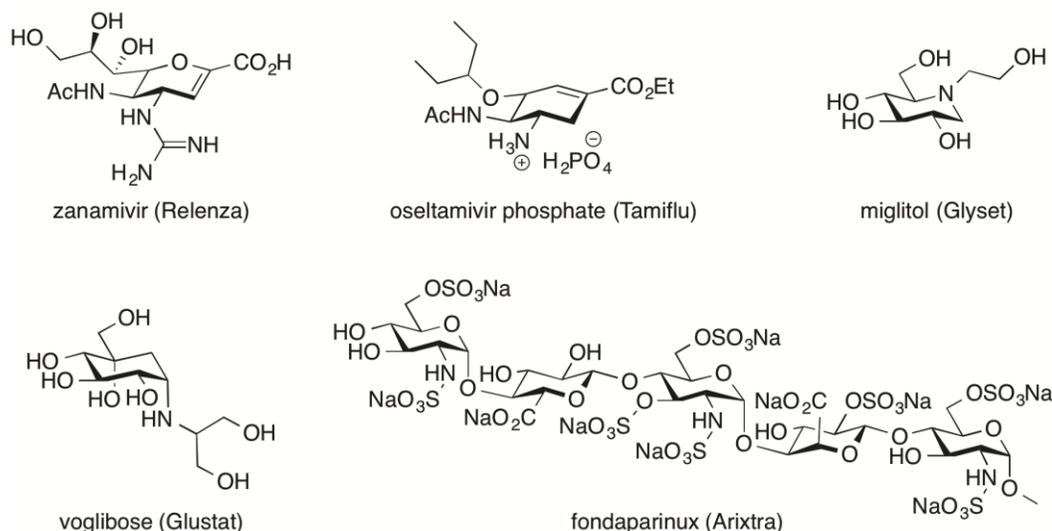


Figure 1-1. Molecular structure of glycan-derived drugs (trade name).

Analysis of the human genome estimates that there are approximately 921 “glycan-related” genes, of which 198 are glycan-binding proteins.⁵ Historically, the lack of tools for studying glycan-protein interactions and the inherent complexity of glycans have been the major barriers for scientists to fully appreciate the importance of glycomics in biology. The breakthrough in technologies, especially the inception of the glycan microarray in 2002 by several independent groups,⁶⁻⁹ have contributed significantly to our understanding of the role of glycomics.^{10,11} Currently, the Consortium for Functional Glycomics (CFG) provides a glycan microarray screening service.

The latest microarray contains 609 different mammalian glycans printed onto *N*-hydroxysuccinimide-activated glass microscope slides (<http://www.functionalglycomics.org>). Protein samples submitted to CFG for screening on the glycan microarray allow structural identification of their binding glycan and provide valuable information about the glycan specificities in a high-throughput fashion. As a result, the binding ligands of GBPs that were previously unknown, or more potent ligands of GBPs other than the ones previously known, are revealed rapidly.¹² Through these efforts, novel glycan-protein interactions and their biological roles are being elucidated at a much faster pace.¹¹⁻¹³ New discoveries and promise emerge rapidly, while at the same time raising new challenges, particularly in the translation of “curiosity-driven” discoveries into clinical applications. We now have accumulated significant amounts of knowledge about GBPs, their various subfamilies and associated binding ligands, yet what the field lacks currently are highly active and selective molecular probes to aid the investigation of the role of GBPs *in vivo*. Furthermore, molecular probes showing promising therapeutic potential need to be further considered and developed into drugs for the treatment of GBPs-associated diseases. To keep up with the need for new therapeutics, new technologies have to be developed to increase the efficiency of drug discovery in glycomics research. As a result, high-throughput and reliable routes to the discovery of carbohydrate-based inhibitors are highly desirable.

The main objective of my Ph.D. research was to develop new technology that has the potential to accelerate the discovery of potent ligands for inhibiting glycan-protein interactions. I combine the power of genetically encoded libraries with fragment-assisted selection. Carbohydrates, due to their well-defined binding modes, are excellent starting points for fragment-based design. In this thesis, I will demonstrate the strategy for attaching covalently any kind of carbohydrate fragment to a phage-encoded peptide library with immense diversity ($>10^8$ unique peptides). The fragment serves as an “anchor” at carbohydrate-binding site and assists the search of promising peptide binding to satellite site(s). One screen, in a high-throughput fashion, can potentially identify peptide binding to the target in synergy with the carbohydrate, resulting in enhanced affinity as well as selectivity. Toward the end, I hope to optimize the technology to become more efficient, more practical, and potentially applicable to a wide range of targets, thereby removing one of the major roadblocks in lead discovery.

1.2 Glycan-binding proteins (GBPs)

Glycosylation of proteins and lipids is the most abundant and structurally diverse posttranslational modification found in nature.¹⁴ It is estimated that more than 50% of all eukaryotic proteins deposited in the SWISS-PROT protein sequence data bank are glycosylated.¹⁵ Although the human glycome⁵ is built on just nine common monosaccharides (Glc, GlcNAc, Gal, GalNAc, Man, Fuc, GlcA, Xyl, NeuAc),^{16,17} the vast possibilities of different

linkages, in term of anomeric and peripheral branching of hydroxyl groups, and secondary modifications such as sulfation and epimerization, rapidly increase the diversity of oligosaccharides. Interestingly, glycan-binding proteins (GBPs) are able to effectively decode the glycan “pattern” and complexity *via* specific and direct binding. These recognitions play key roles in crosstalk within the cell, cell-to-cell adhesion, immune recognition, and metastasis of cancer cells. For example, cell-to-cell adhesion modulated by the interaction between selectins and sialyl Lewis^x facilitated the rolling and initial attachment of leukocytes to the site of inflammation (Figure 1-2).¹⁸ Cancer cells have been showed to exploit the expression of selectin ligands and use rolling-like responses to facilitate their metastasis to other sites.¹⁹ Generation of new blood vessels by endothelial cells, a.k.a angiogenesis, is critical for tumor survival and inhibition of angiogenesis has been a promising anti-cancer treatment. However, aberrant remodeling of the endothelial cell surface glycome was demonstrated as the culprit for the resistance to anti-angiogenetic treatment.²⁰ In anti-VEGF refractory tumors, overexpressed galectin-1 binds to the remodeled complex *N*-linked glycan on vascular endothelial growth factor receptor 2 (VEGFR2), activates “VEGF-like” signaling and eventually leads to compensatory angiogenesis.²⁰ This finding provided a model for the tumors to escape anti-VEGF treatment in mouse models of lung cancer.²⁰

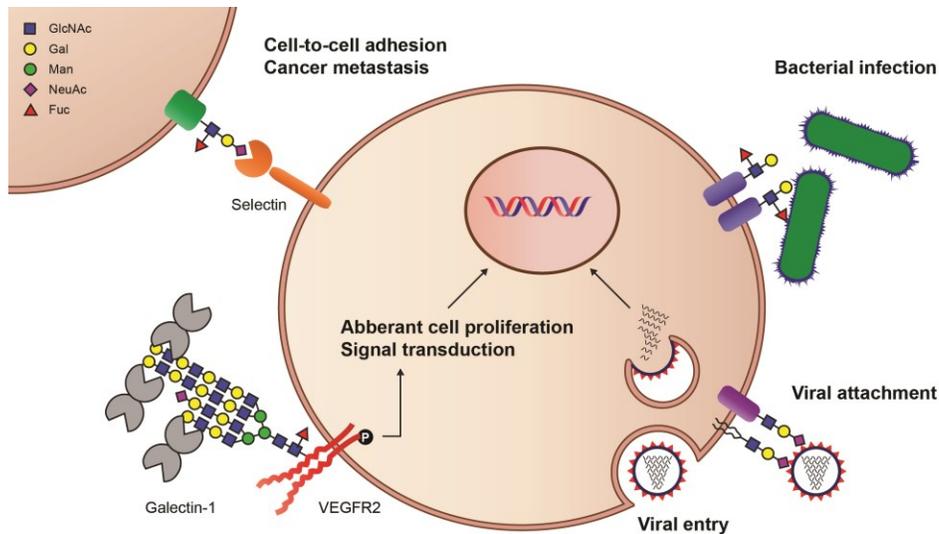


Figure 1-2. The roles of glycan-GBP interactions in human diseases.

Furthermore, almost all bacterial and viral pathogens exploit glycan-GBP interactions as a means for entry into human cells during infection (Figure 1-2). GBPs expressed on bacterial cell surfaces recognize mostly fucosylated histo-blood groups or sialylated epitopes on human cells. These key recognitions play vital roles in the first step of bacterial infection. Viruses use GBP for attachment to their host cells through the binding with diverse carbohydrate conjugates, ranging from gangliosides and glycolipids to glycoproteins. The example shown in Figure 1-2 illustrates the binding of hemagglutinin, capsid protein of influenza virus, to sialylated glycoproteins or glycolipids expressed on human cell, leading to endocytosis of the viral particles.

Although inhibitors of glycan-GBP interactions showed great promise for therapeutic application, they remain relatively underexplored when

compared to other conventional drug targets. Successful examples of small-molecule drugs capable of inhibiting key glycan-GBP interactions are limited (see Figure 1-1).^{4,21} Most of these drugs are designed to target enzymes, such as glycosidases, that use carbohydrates or their derivatives as substrates. Enzymes are generally considered as druggable targets. They represent close to half (47%) of all targets with drugs currently approved on the market.²² The hallmark of a druggable target is the presence of the so-called “hydrophobic pocket” characterized as a deep solvent-accessible surface lined with hydrophobic amino acid side chains.²³ Another feature is the presence of cluster of binding “hot spots” that have the propensity to bind small-molecule compounds with a variety of structures.²⁴ However, GBPs are generally believed to be undruggable.⁴ Two main factors contribute to this belief. First, the physiological ligands of GBP, *i.e.*, glycans, present multiple challenges that include: (i) weak intrinsic affinity to GBPs (K_D in the range of 100 μ M–10 mM);²⁵ (ii) high polarity, which leads to poor oral bioavailability and short plasma half-life; (iii) inefficient hit-to-lead optimization due to the structural complexity of most glycans; and potentially (iv) significant challenges and cost of synthesis during scale-up. The second factor is associated with the proteins themselves. The glycan-binding sites of GBPs often have shallow topologies and are highly solvent-exposed,²⁶ making it difficult to target these sites with drug-like molecules. The shallow binding landscape results in fast off-rates of glycans and eventually constrains the maximal binding affinity a glycan can attain.⁴ Nature has evolved to

compensate for low affinity by employing the multivalent presentation of glycan and clustering of GBPs at the cell surface, as a means to maximize their functional activity *in vivo*.^{27,28} For this reason, display of multiple copies of glycans on a synthetic scaffold is a common strategy to yield inhibitors with nanomolar potency. This approach has been the subject of many reviews²⁹⁻³¹ and is beyond the scope of this chapter.

1.2.1 General principles of inhibitor design for GBPs

Over the last two decades, significant advances have been made in the design of *monovalent* carbohydrate-based inhibitors for GBPs (see reviews^{4,31}), despite the challenges associated with GBPs mentioned in the previous section. Several strategies were deployed to improve the binding affinity. For example, one can reduce the entropic cost due to the immobilization of a ligand upon its binding. A successful example is the recent advance in pan-selectin inhibitor, Rivipansel,^{32,33} which has been licensed to Pfizer and being tested in phase 3 clinical trials for treatment of sickle-cell disease. Although the strategy for its successful design has been reviewed extensively in literature,⁴ I briefly overview it here because it helps to introduce the concepts of fragment-based design. Based on a combination of rational design and empirical medicinal chemistry, the physiological glycans of selectin, sialyl Lewis^x, was minimized to a more rigid molecule with a pre-organized bioactive conformation (Figure 1-3). Specifically, the *N*-acetyl-D-glucosamine moiety in sialyl Lewis^x (Figure 1-3C) does not participate in the binding and only functions as a linker. It can be replaced

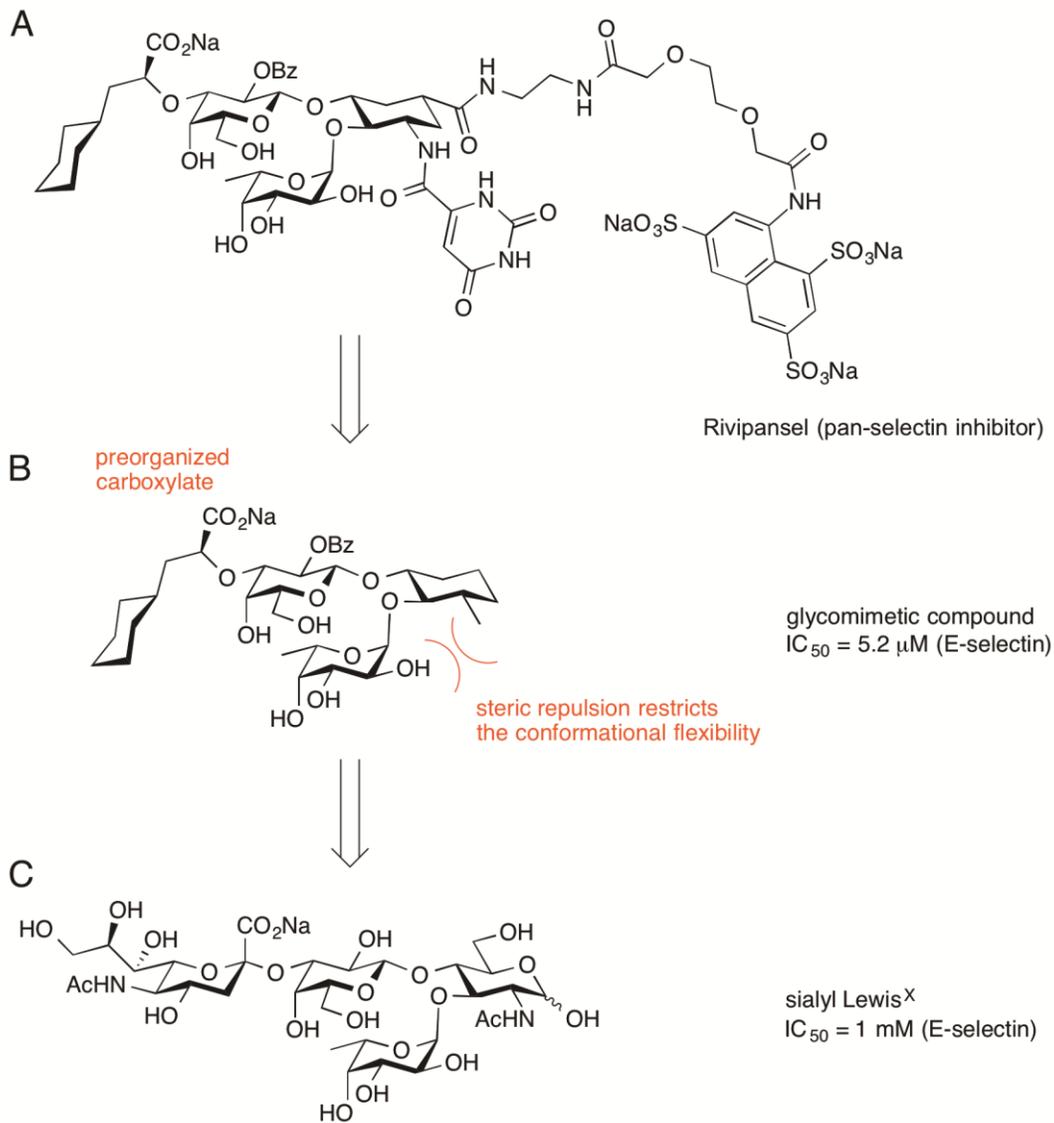


Figure 1-3. Design principles of Rivipansel.

with a cyclohexane ring (Figure 1-3B). A methyl substituent properly placed on the ring further restricts the conformational flexibility and allows the galactose and fucose moieties to be presented in their optimal binding orientations. Furthermore, only the carboxylate moiety of the *N*-acetyl neuraminic acid is critical for recognition by selectins. It is best replaced with (*S*)-cyclohexyl lactic acid. Eventually, these efforts effectively reduced the

entropic cost of ligand binding, thus leading to a more potent inhibitor. Replacement of sugar moieties (Figure 1-3C) with glycomimetic structures (Figure 1-3B) also improves the pharmacokinetics, by reducing the polar surface, leading to a drug with better absorption and cell permeability.

Besides reducing the entropic penalty, another strategy employed in the design of Rivipansel was the increase of enthalpic gain by occupying a potential secondary site located in regions near the glycan-binding site. Fragments that bind to the second site could be linked covalently to a glycan or glycomimetic compound to generate a bivalent molecule (Figure 1-3A). Such molecule thus binds synergistically to both the glycan-binding site (primary site) and the secondary site, leading to substantial improvement of binding affinity. This approach effectively addresses the weak affinity of the glycan due to the shallow landscape of the glycan-binding site. This idea is not new and has been practiced for decades with the first major advance stemming from industrial efforts toward inhibiting the FK506-binding protein.³⁴ The approach is known as fragment-based lead discovery (FBLD), which is the central topic for the technology described in the following chapters. I will first describe this concept in the following section, discuss why it is well-suited for GBPs, and then review several examples of its application for GBPs.

1.3 Fragment-based lead discovery (FBLD)

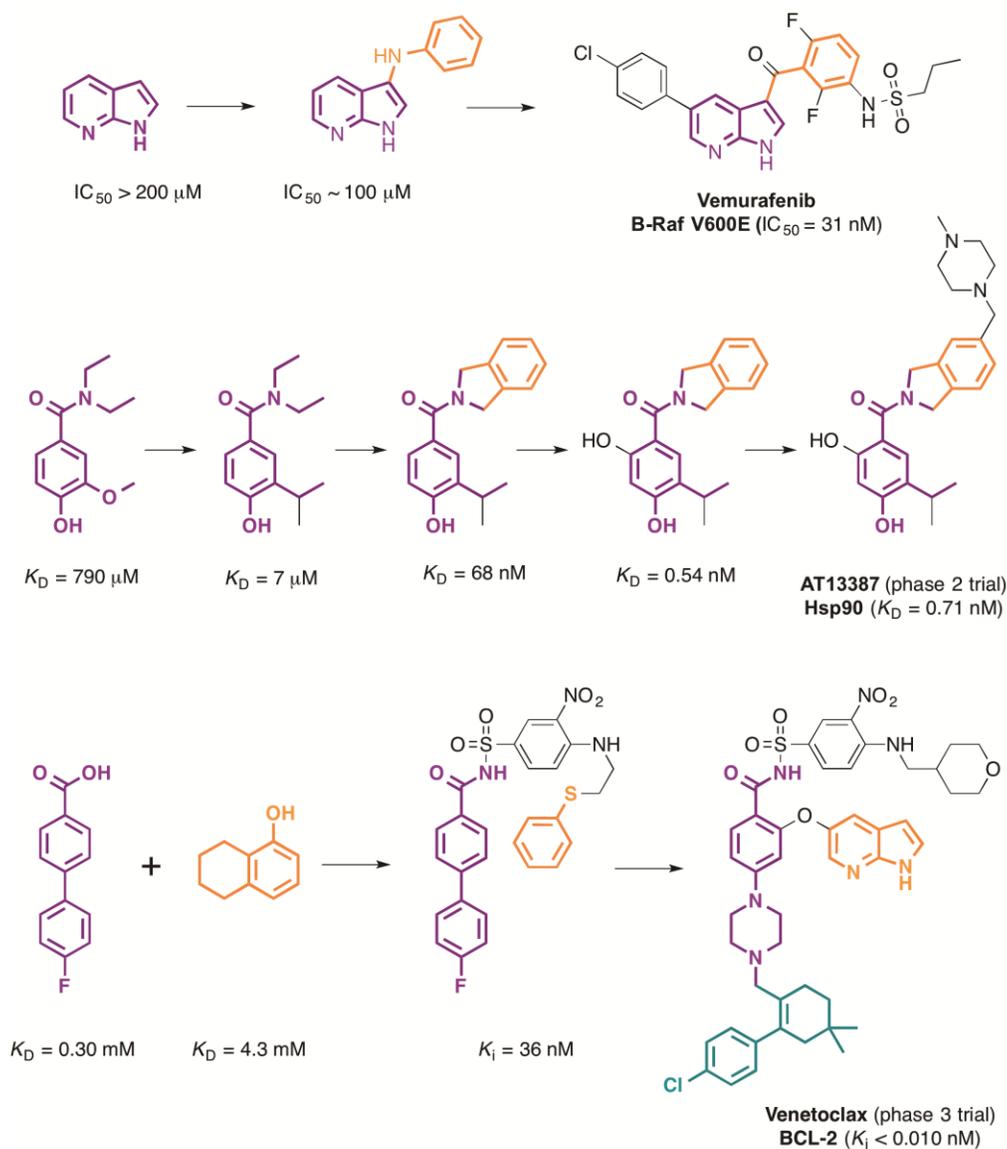


Figure 1-4. Examples of FBLD-derived drugs.

Vemurafenib as an inhibitor of oncogenic B-Raf kinase (V600E mutant).³⁵
AT13387 as an inhibitor of heat shock protein 90 (Hsp90).^{36,37} Venetoclax as an inhibitor of B-cell lymphoma 2 (BCL-2).^{38,39}

Fragment-based lead discovery (FBLD) has revolutionized the way drugs are discovered in pharmaceutical industry. It showed promise to generate potent inhibitors, even for challenging targets such as protein-

protein interactions, when other approaches have failed.⁴⁰ The first drug derived from FBLD, Zelboraf (Vemurafenib), received FDA-approval in 2011; since then, more than 30 potential candidates have been evaluated in various stages of clinical trials.⁴¹ Three examples are shown in Figure 1-4. FBLD has caused a paradigm shift in many different aspects of drug discovery and will continue to impact the field.

1.3.1 Why FBLD for glycan-binding proteins?

One of the key aspects of FBLD is that small fragments with low affinity ($K_D > 1$ mM), though initially believed to be not very useful, actually demonstrate potential to advance into high-affinity leads ($K_D < 100$ nM). Experimental observations^{42,43} validated that, despite their weak affinity, fragments form high-quality interactions, which could be measured by ligand efficiency (LE). LE is a useful and popular parameter for the assessment of the quality of fragments.^{36,44,45} LE is defined as the Gibbs free energy of binding (ΔG) per heavy atom count. LE of 0.30 is usually regarded as the minimal cut-off of good-quality fragment. For example, LE is equal to 0.30 when fragment of 14 non-hydrogen atoms has ΔG of 4.1 kcal mol⁻¹ ($K_D = 1$ mM) at 25 °C (see equations below for the calculations).

$$\text{Equation 1: } LE = \frac{\Delta G}{\text{heavy atom count}} = \frac{4.1}{14} = 0.30$$

$$\text{Equation 2: } \Delta G = -RT(\ln K_D) = -1.987E10^{-3} \times 298.15 \times \ln 10^{-3} = 4.1$$

The higher the LE, the better the binding efficiency of a molecule. However, a practitioner of FBLD should regard this LE cut-off as a guideline

rather than definite “rule” and know when to loosen the “rule” based on his/her own judgment; for example, low LE values are often expected for protein-protein interactions.⁴⁵ In view of the LE cut-off, monosaccharides are in fact excellent candidates for FBLD despite their weak affinity for GBPs. A monosaccharide ($C_6H_{12}O_6$) with K_D of 1 mM has LE value of 0.34 (*i.e.*, $0.41 \text{ kcal mol}^{-1}$ divided by 12) and thus satisfies the criteria. However, more complex glycans, such as a disaccharide with the same K_D value has $LE < 0.17$, which is two-fold lower than that of a monosaccharide since the heavy atom count is doubled for the disaccharide. Another reason why carbohydrates can be suitable fragments for FBLD is the availability of distinct binding modes in glycan-GBPs interactions. These well-defined modes and excellent specificities are due to the formation of specific and directional hydrogen bonds between the hydroxyl groups of carbohydrates and their binding sites. Therefore, the binding modes are maintained at the expected sites during lead expansion and optimization. This widely observed principle in glycan-protein interaction is a key factor for success because fragments with unknown binding mode/site are less useful for FBLD.

1.3.2 Design principles and strategies

In practice, to advance a weakly binding fragment into a high-affinity lead, strategies such as fragment growing and fragment linking are commonly employed. For fragment growing, a library of fragments is screened by common biophysical techniques, *e.g.*, NMR, isothermal titration calorimetry, surface plasmon resonance spectroscopy, thermal shift assays

and X-ray crystallography. Once the hits are identified and validated, their precise binding modes are analyzed with either NMR spectroscopy, or preferably X-ray crystallography. Under the guidance of structural information, the selected fragment is “grown” from the original binding site to enable additional interactions with adjacent sites. This could be done by increasing the complexity and size of the fragment in an iterative and constructive manner (see specific example in Figure 1-5). These steps convert the fragment into a lead compound with higher potency. This strategy works well for targets with concave binding pockets, which are less solvent-exposed and have binding hot spots in close proximity.

However, for targets such as glycan-binding proteins and protein-protein interactions that possess shallow binding sites and involve a large surface area when binding to their endogenous ligands, fragment growing could be quite challenging. In these cases, fragment linking can be employed. The goal is to identify two fragments that bind simultaneously to two distinct sites and devise a strategy to link them together to form a larger molecule of higher potency (see specific example in Figure 1-6). In contrast to fragment growing, fragment linking is much more challenging. Its success depends on the identification of suitable linker that is able to connect two fragments without compromising their original binding modes. Limited choices of linker chemistry could further complicate the design; large entropic penalties can arise when a flexible and long linker is used, whereas a rigid linker might reduce the degrees of freedom and conformational space accessible by both

fragments.^{27,46} Understanding the balance between flexibility and rigidity of the linker remains a challenge. In the case of GBPs, the primary site, *i.e.*, glycan-binding site, is usually known and its geometry is well-understood. Therefore, the effort relies on identifying the “unknown” second site, the binding fragment to occupy this site and the linker joining it.

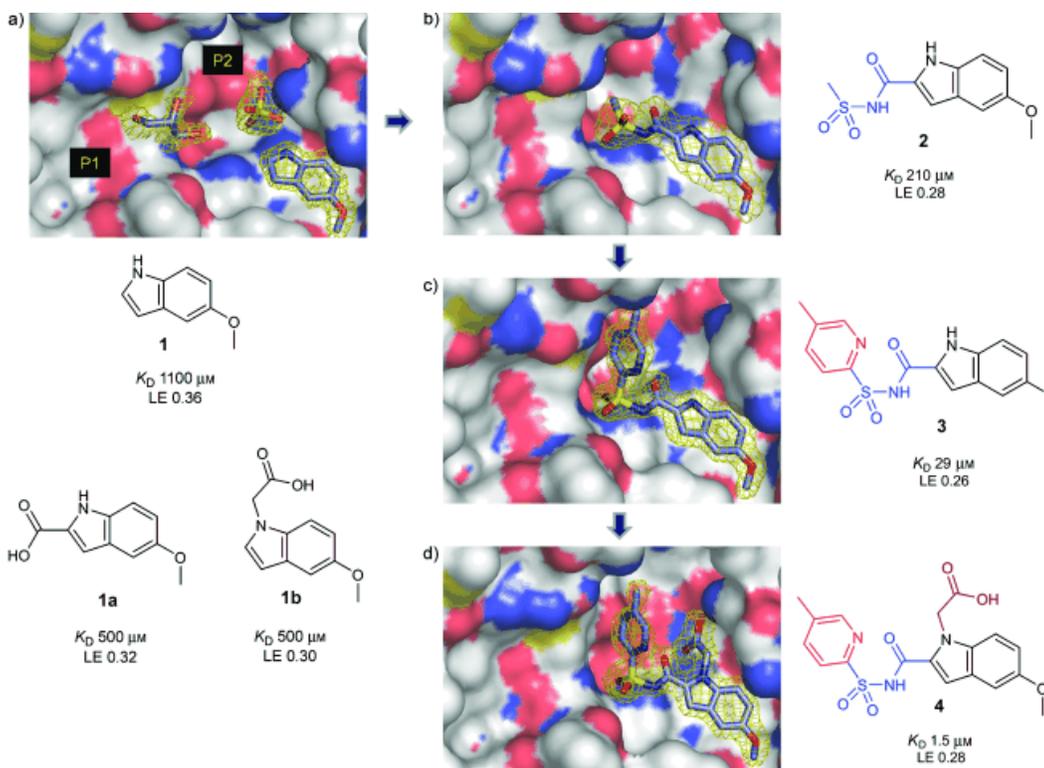


Figure 1-5. Demonstration of fragment growing.

Abell and co-workers have employed fragment growing for the discovery of inhibitors for *Mycobacterium tuberculosis* pantothenate synthetase.⁴⁷ Screening of fragment library identified indole **1** as the potential hit. Two additional sites were found, as shown by P1 site bound with glycerol, and P2 site bound with sulfate. These observations suggested potential sites for the growing of fragment from the C2 and N1 position of indole **1** as exemplified by **1a** and **1b** respectively. A series of synthetic elaborations led to compound **4** which exhibited ~730-fold improvement of binding affinity relative to the starting fragment **1**. The figure is reprinted with permission from the publisher. Copyright © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

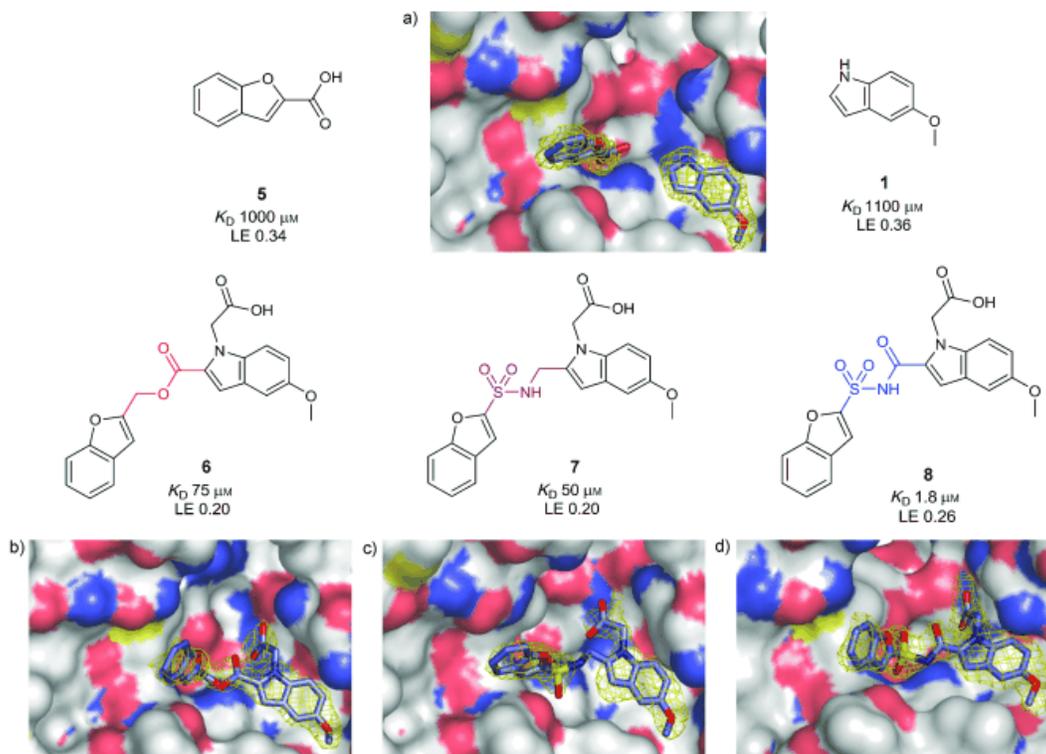


Figure 1-6. Demonstration of fragment linking.

Abell and co-workers further demonstrated the utility of fragment linking based on the same target. Their efforts allow direct comparison of the two distinct strategies for the first time.⁴⁷ Fragment **1** and **5** were identified to occupy two distinct sites simultaneously. Three different linkers were tested and the most rigid linker, acyl sulfonamide, was found to be the best. The resulting compound **8** ($K_D = 1.8 \mu\text{M}$) derived from fragment linking has similar structure and comparable affinity to compound **4** ($K_D = 1.5 \mu\text{M}$) derived from fragment growing. The figure is reprinted with permission from publisher. Copyright © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

1.3.3 Examples of application for GBPs

Both fragment growing and linking have been successfully applied to GBPs to generate glycan-based inhibitors or agonists with nanomolar affinity. The first example is Siglec-4, also known as myelin-associated glycoprotein (MAG), which is one of the protein components that inhibits the growth of neural axon after brain injury.⁴⁸ Schnaar, Paulson and co-workers showed

that glycan-based inhibitors could reverse the action of MAG and enhance the regeneration of axon in cell culture model of rat cerebellar granule neurons.⁴⁹ Based on the concept of fragment linking, Ernst's group has successfully designed novel MAG inhibitor with nanomolar binding affinity.⁵⁰ They started with glycan mimic **1** with low binding affinity ($K_D = 137 \mu\text{M}$, LE

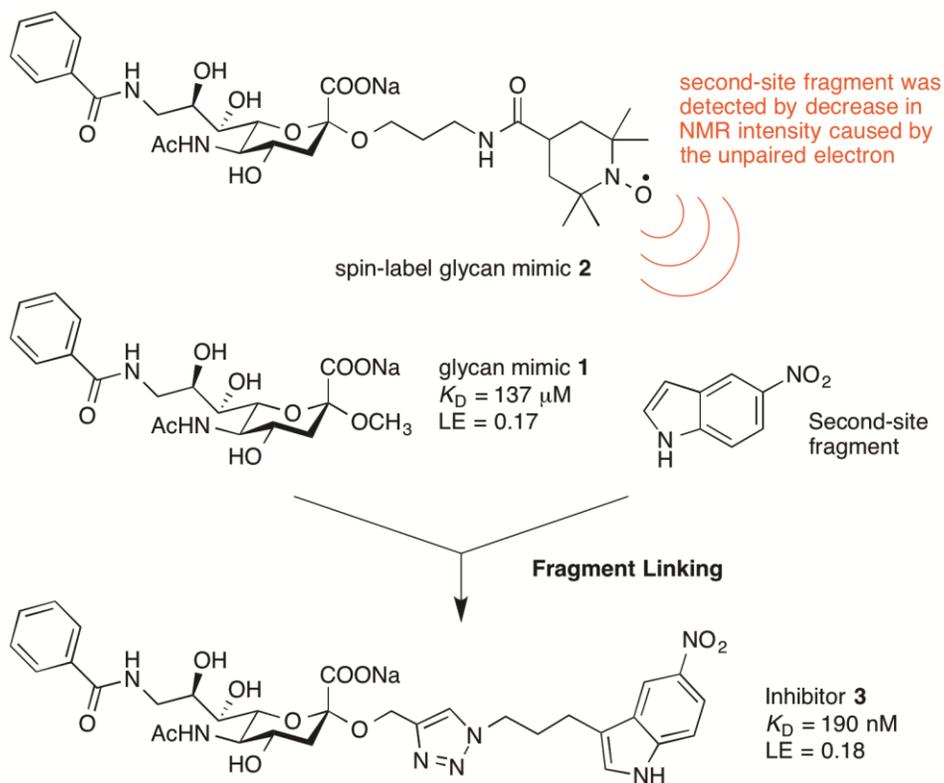


Figure 1-7. Design of MAG inhibitor with fragment linking.

= 0.17) to the glycan-site of MAG (Figure 1-7). Next, they identified the second-site fragments by spin label enhanced NMR screening^{51,52} using a glycan mimic, **2**, that carried a paramagnetic moiety with an unpaired electron (Figure 1-7). The second-site fragment experiences a paramagnetic relaxation effect if, and only if, both fragment and glycan mimic **2** are bound

simultaneously to their protein target. An NMR screen of 80 different fragments led to the identification of 5-nitroindole as the most promising second-site fragment. To join glycan mimic **1** and 5-nitroindole together, a productive linker was identified through *in-situ* click chemistry,⁵³ using MAG as the template for catalyzing the click reaction between an alkynyl derivative of **1** and an azide derivative of 5-nitroindole. The resulting conjugate, **3**, exhibited potent binding to MAG with a K_D value of 190 nM (LE = 0.18), which represents ~720-fold increase in affinity relative to glycan mimic **1** (K_D = 137 μ M, LE = 0.17).

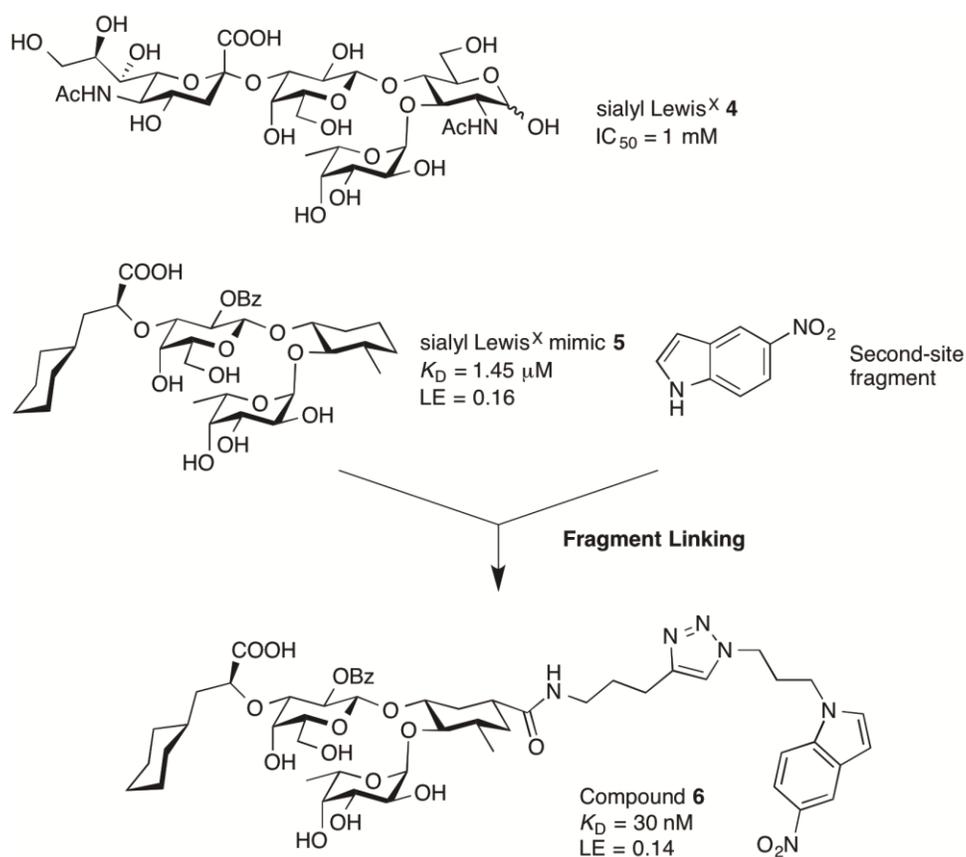


Figure 1-8. Design of E-selectin inhibitor with fragment linking.

Two different groups showed that blocking E-selectin holds promise for the treatment of inflammatory-related disease³³ using Rivipansel or the inhibition of cancer metastasis using anti-E-selectin antibody.⁵⁴ The physiological glycan ligand for E-selectin, sialyl Lewis^x (**4**) is too polar and not potent enough to be a useful oral drug ($IC_{50} \approx 1$ mM, Figure 1-8). The Ernst group has deployed FBLD to design E-selectin inhibitors with nanomolar potency.⁵⁵ Sialyl Lewis^x mimic **5** ($K_D = 1.45$ μ M, LE = 0.16) of similar size to **4**, was chosen as the starting point (Figure 1-8). Using an NMR screen similar to that used in the design of MAG inhibitors, the team identified the same 5-nitroindole moiety as the optimal second-site fragment. However, attempts to design a suitable linker by *in situ* click chemistry was not successful.⁵⁵ The reasons could be the flat-binding site of E-selectin, which did not drive the alkyne and azide groups sufficiently into their optimal orientations for acceleration of the click chemistry.⁵⁵ Furthermore, the issue is aggravated by the low concentration of the ternary complex, which must be formed between the low micro- and milli-molar affinities of both fragments and E-selectin.⁵⁵ Consequently, they resorted to the synthesis and activity screening of a library of 20 compounds containing linkers of various length. The screen identified compound **6** as the most potent antagonist for E-selectin with a K_D value of 30 nM (LE = 0.14). This value represents an ~48-fold improved affinity relative to the starting glycomimetic fragment **5** ($K_D = 1.45$ μ M, Figure 1-8). The prolonged association and dissociation phases observed for compound **6** using SPR

spectroscopy suggested that an induced fit and conformational change of the protein might occur during the binding event. This successful example of designing nanomolar inhibitors for E-selectin demonstrated the potential use of FBLD for other glycan-binding proteins, although the identification of an optimal linker is still a challenging process.

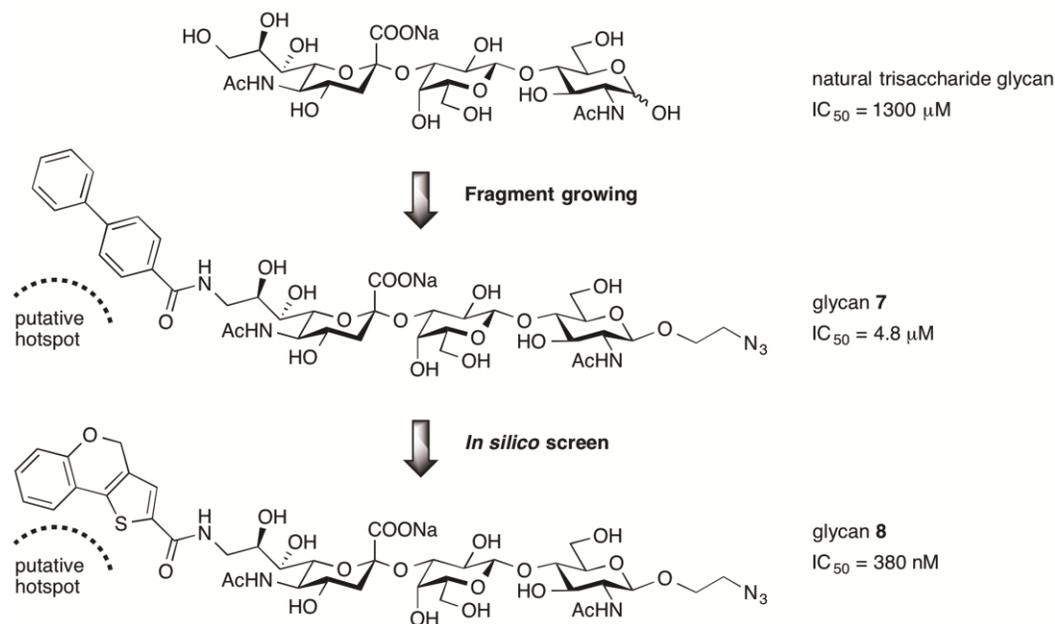


Figure 1-9. Design of Siglec-1 inhibitor with *in silico* fragment growing.

Previously, liposomes decorated with Siglec-1-targeting glycan **7** (IC_{50} = $4.8 \mu\text{M}$, Figure 1-9) was employed for antigen delivery to macrophages *in vitro*.⁵⁶ However, these liposomes were not selective enough for *in vivo* targeting possibly due to the cross-reactivity with Siglec-G, which is expressed on other cells, *e.g.*, dendritic cells and B cells.⁵⁶ Recently, Paulson's group has successfully developed a more potent and selective Siglec-1 agonist by adopting the fragment-growing approach.⁵⁷ Based on the co-

crystal structure of glycan **7** and Siglec-1, they conducted a virtual screening of a library (~8400 compounds) to identify promising fragments that might bind to the putative hotspot occupied by the biphenyl substituent of **7** (Figure 1-9). A *in silico* docking experiment identified six compounds. The synthesis and subsequent validation by a competitive inhibition assay identified bivalent ligand **8** as the most potent lead with an IC₅₀ value of 380 nM. This value represents a 13-fold increase in potency compared to the previous glycan **7** (IC₅₀ = 4.8 μM) and an impressive 3,400-fold improvement relative to the natural trisaccharide glycan (IC₅₀ = 1300 μM, Figure 1-9). The result could be explained by the improved shape complementarity between the ligand and Siglec-1, and additional contacts provided by the newly identified fragment. Using liposomes decorated with agonist **8**, the Paulson's group demonstrated selective targeting and delivery of antigen to Siglec-1⁺ macrophages in a mouse model.⁵⁸ The liposomes triggered the activation of invariant natural killer T cells, which recognized the antigens presented on the macrophages.⁵⁸

1.4 High-throughput FBLD

FBLD revolutionized lead discovery and provided an alternative to conventional high-throughput screening. However, there remains an unmet need for simpler, less expensive, and more efficient routes to identify superior lead compounds. A new method that can overcome the limitations of FBLD should include: (i) a sensitive system for the discrimination of

weakly binding fragments from false positive due to artifacts;⁵⁹ (ii) a higher-throughput method for screening of large libraries and the identification of hits; (iii) a ranking system that can prioritize the hits and determine a short-list of hits for downstream validation; and (iv) a powerful screen and a tailored library for identification of both fragments and its optimal linker in one screen. The last demand is particularly challenging, yet important. Stivers and co-workers has showed that even linkers of similar length can exhibit differences in conformational strain and flexibility, which can bring up to 3.2 kcal mol⁻¹ in entropic penalty during binding. At room temperature, this energetic penalty corresponds to ~2 orders of magnitude increase in the dissociation constant (K_D).⁶⁰

1.4.1 Chemotype evolution

Several emerging technologies have the potential to address the issues mentioned above, yet still adhere to the principles of FBLD. The first approach is chemotype evolution pioneered by Erlanson and Hansen at Carmot Therapeutics.⁶¹ This technology allows rapid, iterative screening of a target-oriented library using a multiwell-plate-based assay to identify more potent hits. The technology is akin to high-throughput screening but the library is generated in a unique way (Figure 1-10). An anchor fragment, or “bait”, serves as a starting point and is covalently linked to 7500 fragments to generate a two-component library. The complete library of 7500 members is proprietary but the chemistry to covalently link the “bait” with fragment is shown in Figure 1-10. Once the library was assembled, it can be screened

directly without any purification. The “bait” can contain any known inhibitor, substrate, co-factor or peptide binder for the target of interest. During the screen, the “bait” assists the identification of novel fragments that bind to the site adjacent to the binding site of the “bait”. The resulting hits can serve as new bait for evolution through another round of screening. Up to 14,000 novel compounds, containing two different baits, can be generated in parallel and screened per target per month. Furthermore, hits emerging from screening the two-component library in the presence of linker partially addresses the “linking challenge” of FBLD. The team has employed this technology to identify nanomolar inhibitors for kinase and protein-protein interactions after minor optimization of the linker.

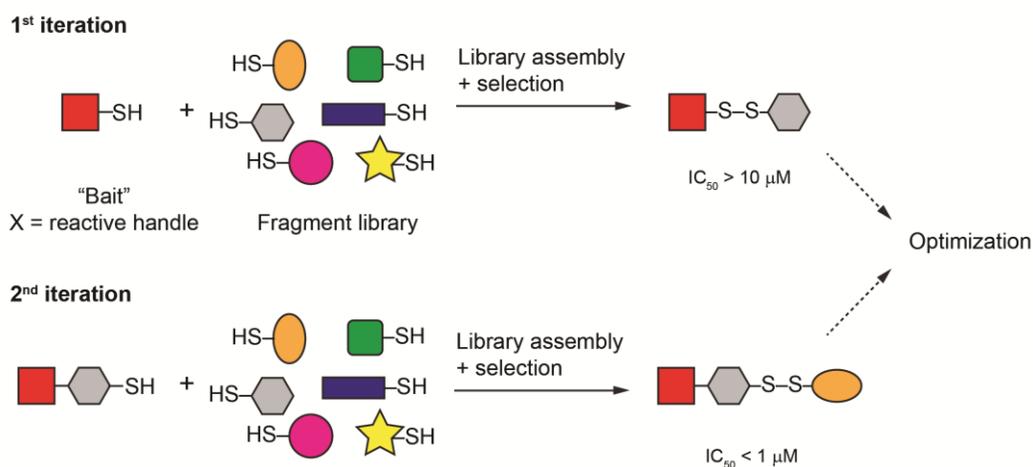


Figure 1-10. Chemotype evolution. Each library member is covalently linked to the “bait” through disulfide bridging.

1.4.2 DNA-encoded chemical library

Another innovative technology that could be potentially applied for FBLD is DNA-encoded chemical library developed by the research groups of

Harbury,⁶² Liu,⁶³ and Neri⁶⁴ independently in the early 2000's. Today, each technology has been validated in multiple peer-reviewed publications; furthermore, several different variations have emerged from biotech companies,^{65,66} such as Ensemble Therapeutics, Philochem, and Viperger. For more comprehensive coverage, I recommend a recent review.⁶⁷ The technologies differ in library synthesis,⁶⁷ but the subsequent screen uses the same approach: affinity-based selection of DNA-tagged ligands followed by DNA sequencing for the identification of the structure of the hits. The major advantage of this technology is the diversity of the library ($>10^5$), which is several orders of magnitude larger than conventional fragment library ($\sim 10^2$ – 10^4). Another advantage of DNA-encoded screening is the potential to identify the hits even from minute quantities of ligands (*e.g.*, femtomolar of each library member) using small amount of proteins (*e.g.*, few micrograms). In theory, 50% of library members with $K_D = 1 \mu\text{M}$ would be captured by the protein present at $1 \mu\text{M}$ concentration, which translates to $10 \mu\text{g/mL}$ of protein with 10 kDa in size. Therefore, the technology is highly sensitive and high-throughput. Recent advances in the encoding strategy demonstrated by Neri and Scheuermann has allowed fragment pairs to be paired up and selected for synergistic binding,⁶⁸ suggesting its potential use in FBLD.⁶⁹

1.4.3 Fragment-based display technology

In recent years, the union of display technologies with the concepts of FBLD has emerged.⁷⁰⁻⁷⁷ Historically, libraries displayed on phage or mRNA are limited to natural amino acids. Advances in bio-conjugation⁷⁸ and site-

specific incorporation of unnatural amino acids⁷⁹ make it possible to introduce small molecules or fragments onto these peptide libraries.^{80,81} As a result, *fragment-based display technologies* merge the idea of FBLD with phage- or mRNA-displayed libraries. For example, the pioneering work of Roberts and co-workers in 2003 demonstrated that mRNA-displayed peptide library can be covalently modified with penicillin.⁷⁰ Screening of this library against penicillin-binding protein has identified penicillin-peptide conjugates with inhibitory activity ~100-fold better than that of the penicillin itself.⁷⁰

Another example is the selection of bivalent kinase inhibitors from a phage-displayed peptide library displaying a known kinase inhibitor (staurosporine).⁷² Interestingly, even non-covalent tethering mediated by coiled-coil interaction allowed for the identification of cyclic peptide fragments that enhanced the binding of staurosporine. The authors observed that the peptide fragment binds to an unknown second site. To validate the hit, the authors designed a 30 Å linker to join the cyclic peptide and staurosporine together. The resulting bivalent inhibitor ($IC_{50} = 2.6$ nM) exhibited 90-fold improved affinity relative to staurosporine ($IC_{50} = 243$ nM) and at least 5-fold selectivity across a panel of 90 kinases.⁸²

Peptide-based fragments pose several issues such as poor cell membrane permeability, limited routes for administration, low enzymatic stability, and short plasma half-life. However, I anticipate that fragment-based display technology could be a highly efficient and empirical tool for “hot spot” mapping. The hits emanating from such screen could be further

optimized to improve their pharmacokinetic properties. Innovation and new strategies for improving the pharmacokinetic properties of peptide leads are emerging continuously.⁸³⁻⁸⁵ These strategies include stapling of peptides,⁸⁶ macrocyclization of peptides,⁸⁷ the use of unnatural amino acids and *N*-methylation of amide nitrogens.⁸⁸⁻⁹⁰ For the past three decades, the numbers of peptide drugs entering clinical studies have grown exponentially.⁸³ This trend capitalizes on several favorable properties of peptides, such as their higher selectivity for the target, predictable metabolism and lower toxicity, shorter time to market, and much higher success rates (14%) of transitioning from clinical stages to final drugs as opposed to small molecules (7%).⁸⁴

1.5 Thesis overview

In Chapter 1, I have described the motivation of my research and the main objectives of the thesis. I discussed the biological significance of targeting glycan-protein interactions and identified the challenges. Then, I briefly described the general strategies to tackle these problems using fragment-based lead discovery for targets like glycan-binding proteins (GBPs). The design principles behind and recent examples of GBPs inhibitors derived from this approach were discussed. Finally, I reviewed several emerging technologies that hold promise for improving the fragment-based approach, which make it simpler, more efficient and higher-throughput.

In Chapter 2, I present my work that describes a strategy for conjugation of carbohydrates to linear peptide libraries displayed on M13

phage. I describe the challenges in characterizing the reaction yield of on-phage reaction and demonstrated, for the first time, a quantitative characterization method. The facile characterization allowed me to optimize the on-phage reaction and, for the first time, generate a library of glycopeptides with 10^8 diversity encoded by phage.

In Chapter 3, I demonstrate the first application of fragment-based display technology for glycan-binding protein. The library generated in Chapter 2 made it possible to discover highly potent and selective hits for a model glycan-binding protein. My co-workers and I validated that the carbohydrate fragment binds at the expected site. Interestingly, peptide fragments bind to a previously unknown satellite site and this binding contributes substantial enthalpic gains. This observation demonstrates the potential of the technology for “hot spot” mapping. I observed that the changes in the linker connecting the carbohydrate and peptide fragment after the screen have a pronounced effect on their binding affinity, confirming a critical role of linker during the selection process.

Chapter 4 describes the second application of the technology for a therapeutically relevant target (DC-SIGN). Starting from a mannose fragment, I found inhibitors of moderately better potency than the mannose itself. Although the leads were not attractive for further development, I proposed structural reasons for the unsuccessful screen and outlined future directions for its improvement.

Chapter 5 focuses on the design of macrocyclic glycopeptide library displayed on M13 phage. The library consists of macrocyclic peptides with a ketone functionality that can be used to attach any small-molecule fragment through oxime ligation. I will describe the modification on model peptide and phage libraries. Then, I will present specific examples of making a macrocyclic glycopeptide library displayed on phage. Finally, I will describe an initial screening of these libraries with GBPs.

Lastly, in Chapter 6, I will summarize my works. I will discuss the major limitations of the technology and outline future directions and proposals for the improvement of this technology.

Chapter 2: Quantitative synthesis of genetically encoded glycopeptide libraries displayed on M13 phage

2.1 Introduction

Biological encoding of information allows handling libraries of large diversity. Unlike chemically synthesized libraries, libraries of peptides encoded by DNA or RNA can be amplified from a single copy and optimized using directed evolution. Phage display is the most common strategy for the discovery of functional peptides from genetically encoded libraries.^{91,92} The technology has been the source of numerous FDA-approved drugs and drug-candidates in clinical trials.⁹³ Expanding the use of phage to display and encode molecules other than natural polypeptides makes it possible to select and evolve molecules with properties not found in peptides.^{70,94-96} Chemical modification of phage-displayed peptide libraries is one of the simplest approaches to encode non-natural moieties.^{96,97} The identity of the products could be deciphered from phage DNA that encodes the starting material (peptide), only if the modification is regioselective and quantitative.^{80,98} Such an approach to building diversity is ubiquitous in nature: organisms from all kingdoms of life use post-translational modifications to diversify the structure of genetically encoded polypeptide libraries.⁹⁹ Several recent reports have highlighted the power of chemically modified phage-displayed libraries.^{96,97}

The broad utility of this approach, however, is plagued by a lack of techniques that can be used to characterize chemical modifications on mixtures of 10^9 phage-displayed peptides.⁸⁰ In this chapter, I developed a characterization technique that addresses this deficiency. To demonstrate its utility, I perform quantitative and regioselective modification of commercially available phage-display libraries, in as short as 1.5 h, to generate a library of 2×10^8 glycosylated peptides; each of them is attached to an information carrier. Furthermore, chemical modification had minimal interference on phage infectivity; modified libraries, thus, could be amplified and applied to affinity selection just like any other phage-displayed library.

For the generation of chemically modified libraries, M13 phage display has several advantages over the display of peptides on yeast cells¹⁰⁰ or “naked” DNA/RNA.^{101,102} The chemical composition of M13 proteins is simpler than that of a cell surface, and the growth of M13 phage libraries in *E. coli* culture is simpler than the generation of DNA/RNA-displayed libraries. Additionally, several M13 phage-displayed libraries are available from commercial vendors. Previous attempts to generate chemically modified libraries used engineered M13 phage that contained unnatural amino acids (UAA) such as selenocysteine^{103,104} and *p*-azidophenylalanine.¹⁰⁵ Thiol handles were also used to introduce modification in Cys-free M13 phage.⁹⁶ Unfortunately, neither UAA-containing libraries nor libraries on Cys-free M13 are commercially available. These libraries require significant genetic engineering of phage, which compromises their growth rates; the decrease in

amplification rate, in turn, can be detrimental for the selection process.¹⁰⁶ We bypassed the complexity associated with “unnatural” M13 phage and built libraries starting from commercially available libraries, such as Ph.D.-7, a library of $<1 \times 10^9$ random heptapeptides displayed on M13KE phage. Our strategy employs *N*-terminal Ser and Thr residues, which upon oxidative cleavage by NaIO_4 yield bio-orthogonal aldehyde handles.¹⁰⁷ These *N*-terminal amino acids are absent from native M13 proteins but can be found at 20–30% abundance in random peptide libraries.¹⁰⁸

2.2 Results and discussion

2.2.1 Challenges for characterization of on-phage reactions

Modification of phage-displayed peptide libraries is a reaction that is performed on 10^9 molecules simultaneously. Because the pIII protein is displayed on phage at a low copy number, each molecule is present in small amounts. MALDI and other MS techniques, which are often used to characterize chemical modification on viral coat proteins present in high copy number,¹⁰⁹ fail to detect pIII, which constitutes less than 1% (w/w) of M13 virion. Heinis and co-workers used MALDI to monitor reactions on recombinant purified pIII subunit,⁹⁶ but the efficiency of the reaction cannot be validated in the context of intact M13-virion or an entire library of 10^9 peptides. To date, the only characterization methods that could assess reactions on phage, or libraries of phage, was reported in a patent by Winter and co-workers. The authors used Western blot and fluorescent

densitometry analysis of pIII, isolated from M13 virion using SDS-PAGE.¹¹⁰ I found this technique difficult to adapt because it consumed over 10^{11} particles of phage, required complete removal of labeling reagents, and in our hands was not reproducible (Appendix A-1). Most importantly, neither MS- nor SDS-PAGE-based methods can allow determination of yield with sufficient accuracy.¹¹⁰

2.2.2 Method for characterization of on-phage reactions

To quantify reactions on clonal phage and libraries of phage, I developed a straightforward and sensitive assay, which capitalized on the ability of individual phage to generate plaques in agar overlays.^{111,112} I distinguished individually modified and non-modified phage particles using a capture reagent, here aminoxy-biotin (AOB), which underwent covalent ligation with aldehydes displayed on phage particles. Upon incubation with streptavidin (SA)-coated magnetic beads, biotinylated clones were captured. The remaining non-biotinylated clones were quantified as plaque forming units (PFUs). The ratio of PFUs before and after capture thus quantified the yield of the modification (Figure 2-1A). I have validated the assay and demonstrated that non-specific binding of phage to the beads was negligible (Figure 2-1B). I routinely captured and quantified phage particles present at 10^5 PFU mL⁻¹ (<200 aM). This assay was reliable for detection of phage with concentration as low as 10^3 PFU mL⁻¹ (Figure 2-1C).

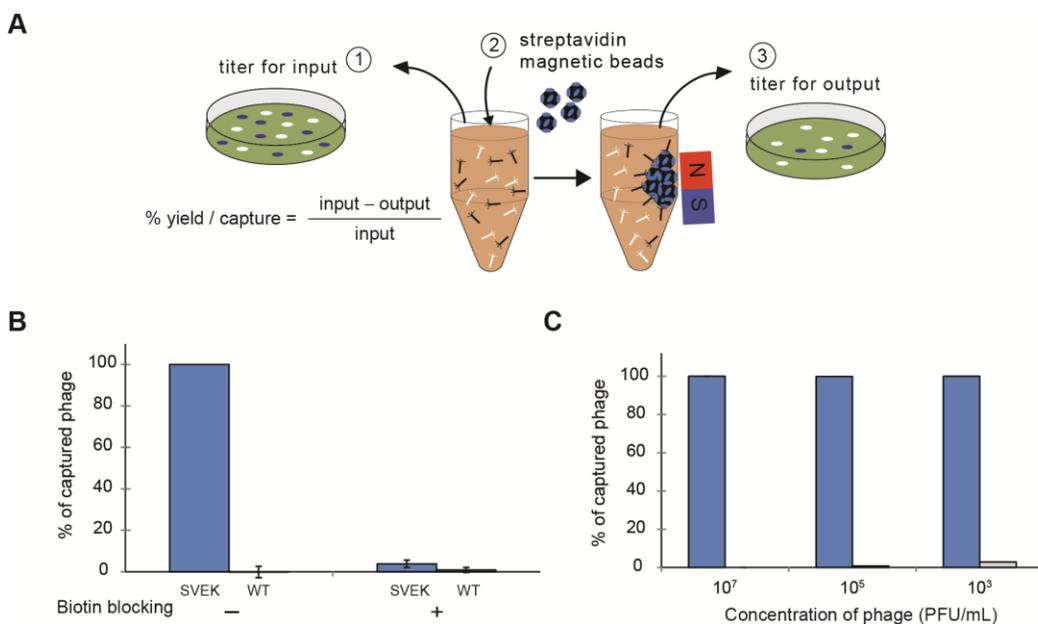


Figure 2-1. Biotin-capture assay.

(A) To quantify the biotinylated phage, the titer of clones in the supernatant was determined before and after capture. The titers were used to derive the percentage of capture or modification yield. (B) To validate the pull-down assay, a positive control was prepared by reacting SVEK phage with biotin NHS ester. Capture of the biotinylated phage was quantitative (99.99%) but the non-biotinylated WT phage, present in the same solution, was not captured. Blocking of the binding site on streptavidin with excess of biotin (1 mM) inhibited the capture. Both controls demonstrated the specificity of the capture. The data is an average from three independent experiments. The error bar is equal to one standard deviation. (C) The assay allowed capture and quantification of phage as low as 10^3 PFU mL^{-1} (~ 2 attomolar).

2.2.3 Selectivity of the reactions and their optimizations

To prove that reactions occurred only at the desired location on the phage, I performed all reactions in a mixture of two phage clones: (i) phage displaying a 16-residue peptide on pIII protein with Ser at its *N*-terminus (SVEK phage) and (ii) wild-type phage (WT phage) displaying an *N*-terminal Ala on pIII that lacks the extra 16-residue peptide but is otherwise identical in amino acid composition to SVEK phage. The SVEK phage carries a *lacZ* reporter gene and forms blue plaque in a bacterial lawn on an IPTG/Xgal

plate, whereas WT phage forms clear plaque under these conditions. A simple “blue/white” screen¹¹³ thus quantified the regioselectivity of the reaction.

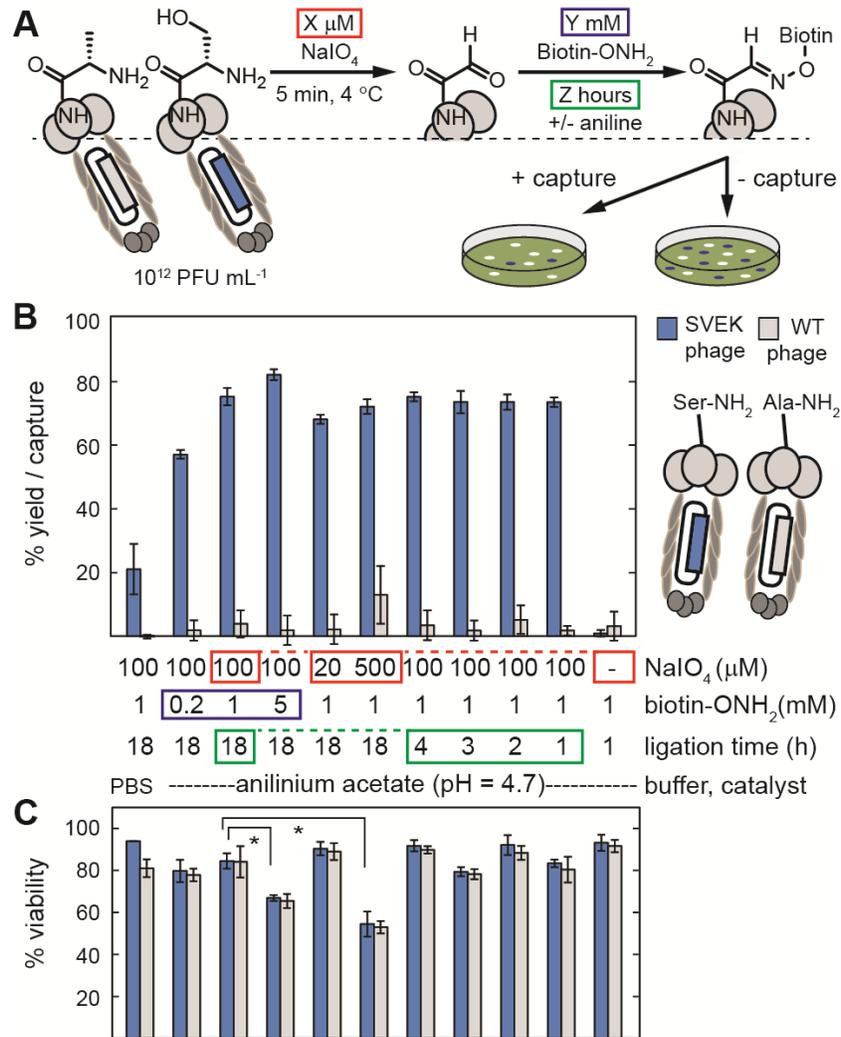


Figure 2-2. Profiling of the optimal conditions for oxidation and oxime ligation.

(A) Scheme of reactions and quantification of biotinylated phage with plaque-forming assay. Oxidation was quenched by 1 mM *N*-acetyl-DL-methionine. (B, C) Effects of NaIO₄ and AOB concentration, duration of ligation, pH, and catalyst have on the yields of capture (B) and phage viability (C). Data is an average of the three biotin-capture assays; error bars are one SD (*denotes $p < 0.05$, as determined by two-tail unequal variance Student test).

Reaction of SVEK phage with AOB was undetectable in the absence of oxidation but occurred after NaIO₄ treatment (Figure 2-2B). When a mixture of SVEK and WT phage were both exposed to NaIO₄ and AOB, WT phage was not modified to any appreciable level (Figures 2-2B). These results unambiguously showed that only phage bearing *N*-terminal Ser can react with NaIO₄ and participate in the two-step reaction. The availability of rapid and simple quantification made it possible to screen seven orthogonal parameters to maximize the yield of the reaction and the viability of phage. These parameters included the concentration of NaIO₄ and aminoxy-biotin (AOB), the duration of oxidation and ligation, and the role of pH, catalyst, and concentration of phage (Figures 2-2). The use of aniline as a catalyst^{114,115} and performing the reaction in acidic buffer (pH 4.7) greatly enhanced the rate of oxime ligation and led to 74% capture in 1 h. In contrast, less than 30% of the phage was modified even after 18 h of incubation in neutral buffer (pH 7.4). Increasing the concentration of NaIO₄ beyond 100 μM was detrimental to phage viability (Figure 2-2C). AOB concentration of 1 mM was optimal; higher concentrations resulted in a marginal increase in reaction yield and a significant decrease in viability (Figure 2-2C). The use of an appropriate quencher was critical to ensure that unreacted NaIO₄ did not interfere with oxime formation. Thiol-based quenchers improved the capture by >25% when compared to a sulfide quencher, such as *N*-acetyl-DL-methionine (Figure 2-3). In most reactions, I used glutathione (GSH) as the quencher due to its minimal interference with phage viability.

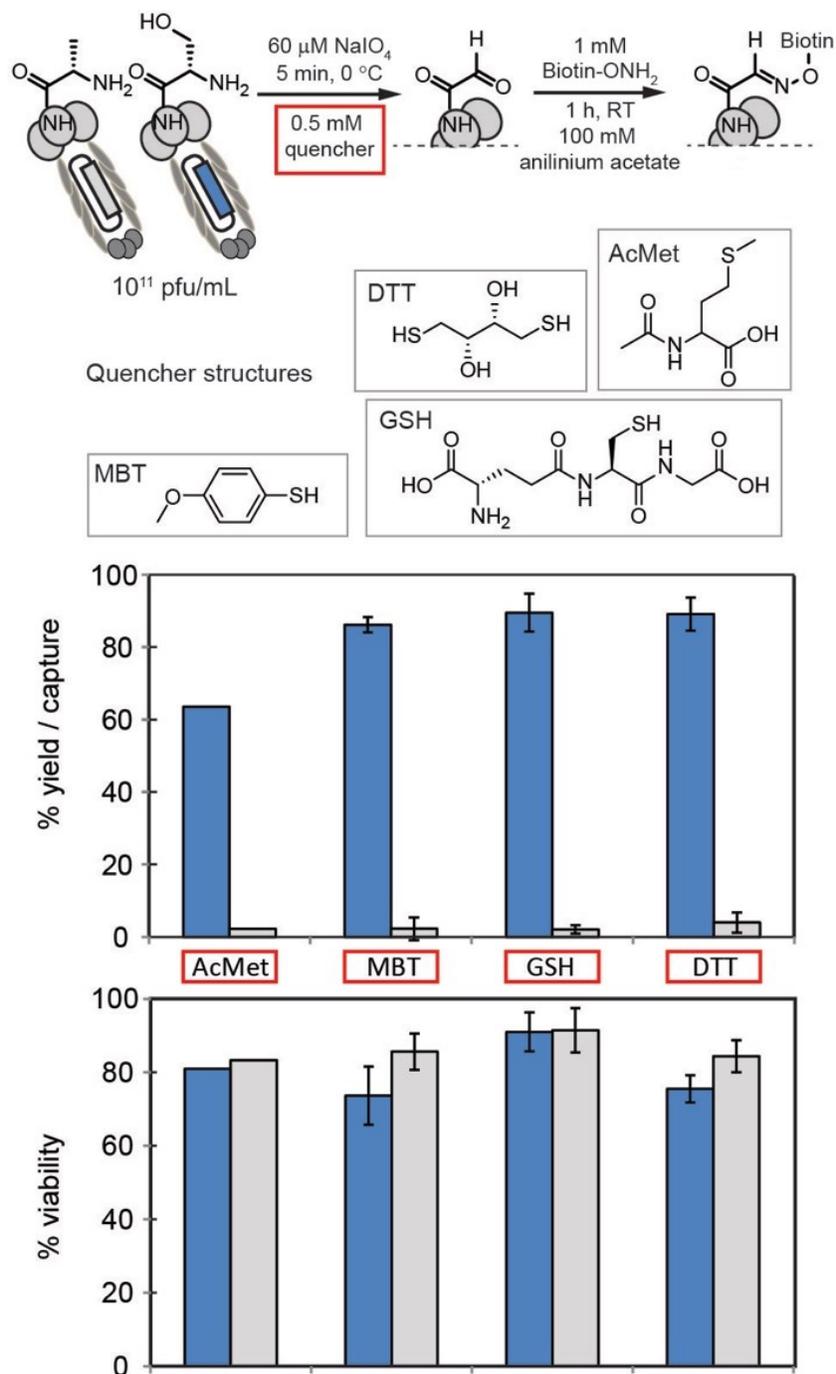


Figure 2-3. Optimization of the conditions for *in-situ* quenching of the oxidation reaction.

Following oxidation, quench and AOB ligation, the number of modified phage was quantified using biotin-capture assay. The value is an average of at least two independent experiments performed on different days. Error bars are equal to one standard deviation. Abbreviation: AcMet = *N*-acetyl-DL-methionine; MBT = 4-methoxybenzenethiol; GSH = glutathione; DTT = dithiothreitol.

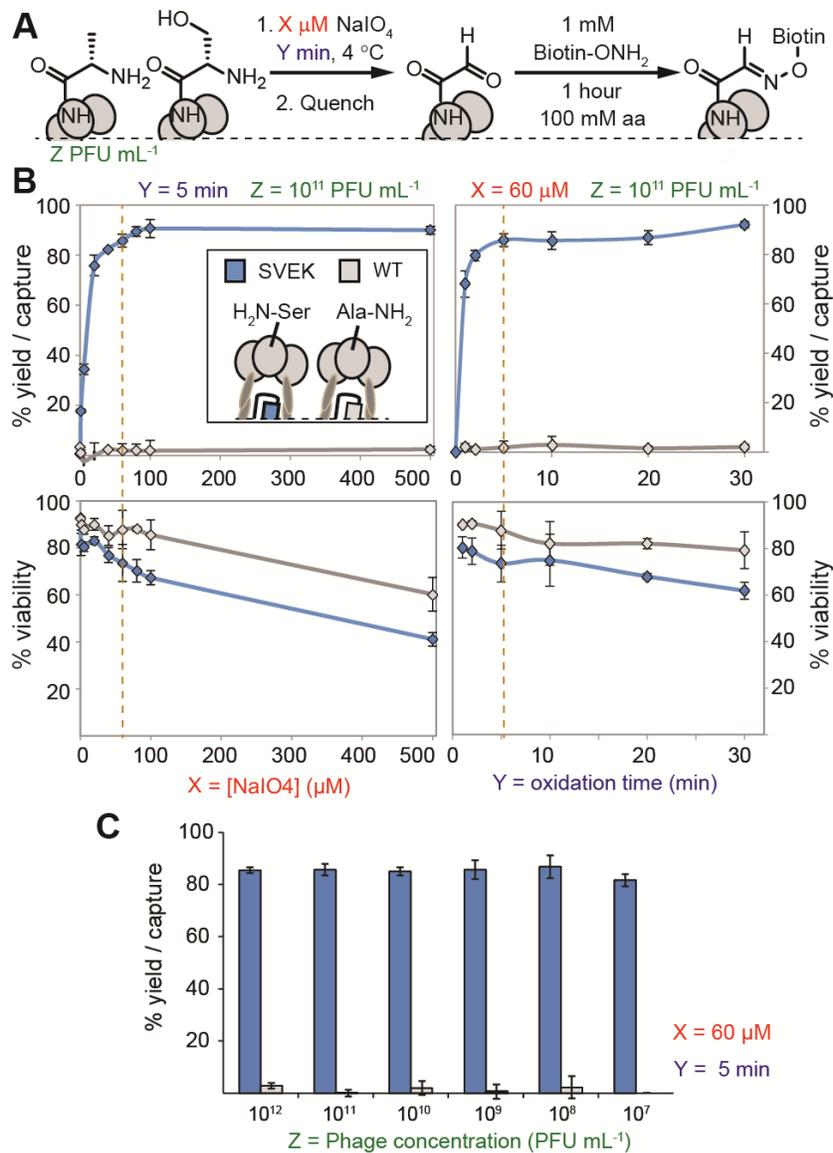


Figure 2-4. Profiling of the kinetics of oxidation.

(A) Oxidation was performed on a mixture of SVEK and WT phage and quenched by the addition of 500 μM 4-methoxybenzenethiol (MBT); aa = anilinium acetate buffer (pH = 4.7). (B) The optimal $[\text{NaIO}_4]$ and reaction time that maximize the yield and viability are 60 μM and 5 min. (C) Varying phage concentrations did not affect the reaction kinetics. The results are an average of at least two independent experiments; error bars are one SD.

The yield of capture increased with increasing concentration of oxidant or reaction time (Figure 2-4B); however, increasing both parameters had a negative effect on phage viability. Taking all the above factors into

consideration, I arrived at an optimal condition that maximized both capture and phage viability, and reproducibly yielded modification of $86 \pm 3\%$ of phage: (i) 5 min exposure to $60 \mu\text{M NaIO}_4$ at 4°C ; (ii) 10 min quenching with $500 \mu\text{M GSH}$; followed by (iii) 1 h incubation with 1 mM AOB . Modifications of phage with reagents present in large excess compared to phage should exhibit pseudo-first-order kinetics. The efficiency of such reaction should be independent of the concentration of phage. Indeed, the optimal condition gave similar capture across 10^7 – 10^{12} PFU mL^{-1} of phage (Figure 2-4C). We note that, at 10^7 PFU mL^{-1} , the concentration of phage is $<20 \text{ fM}$ and quantification of reaction efficiency using any other method would be impossible.

2.2.4 Relationship of capture yield and reaction yield

The reactions on phage that has 3–5 copies of the pIII protein could yield phage with 0, 1, 2, 3, 4, or 5 biotins (B_0 – B_5). If the reactions on pIII are independent, binomial statistics links reaction yields on individual pIII with B_0 (*i.e.*, non-captured phage with 0 biotin) (Figure 2-5A) and yields an equation that could be used to derive the rate constants (k) from the biotin-capture (Figure 2-5B). We observed agreement of k among reactions performed on phage (Figure 2-5C), on model peptides (Figure 2-5D), and on similar reactants reported in the literature.¹¹⁵ For example, k of aniline-catalyzed oxime ligations were $3.3 \pm 0.5 \text{ M}^{-1} \text{ s}^{-1}$ (phage), $7.8 \pm 2.3 \text{ M}^{-1} \text{ s}^{-1}$ (peptide), and $8.6 \pm 2.0 \text{ M}^{-1} \text{ s}^{-1}$ (literature), whereas k of non-catalyzed ligations were $0.018 \pm 0.005 \text{ M}^{-1} \text{ s}^{-1}$ (phage) and $0.020 \pm 0.001 \text{ M}^{-1} \text{ s}^{-1}$

(literature).¹¹⁵ In all studies, the reactions on phage saturated at <90% capture ($B_0 = 10\%$). From B_0 , binomial statistics estimates that the phage contained 1–3 copies of modified peptide (Figure 2-5E). The calculation of the percent yield of the modified peptides depends on the number of pIII proteins (37% assuming $N = 5$ and 54% for $N = 3$). Furthermore, pIII-displayed peptides could be subjected to proteolysis.¹¹⁶ As the oxidation and oxime ligations proceed to >90% conversion on SVEKY peptide (Figure 2-5D), the low apparent yields on phage could indicate that phage contains only 1–3 reactive sites (peptides).

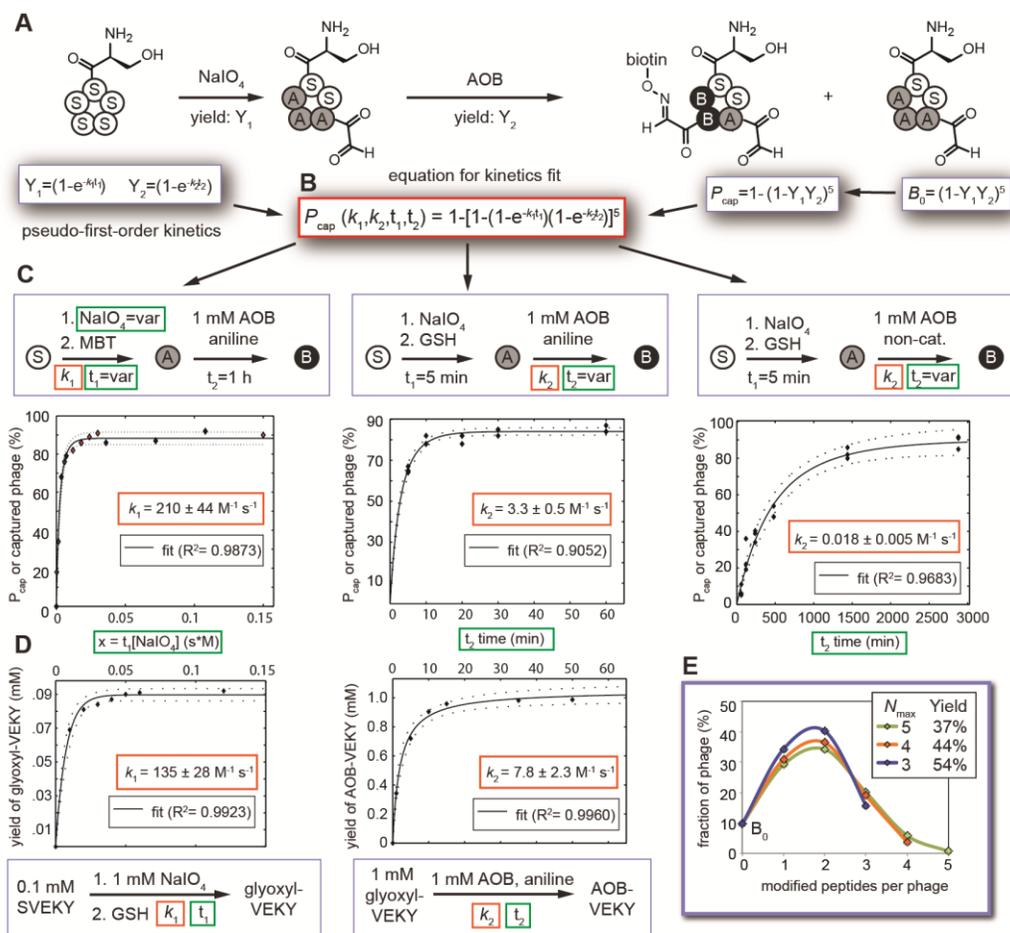


Figure 2-5. Reaction on multiple copies of pIII protein and the kinetics. (A) Binomial distribution of biotin-free phage (B_0) and phage that can be captured (P_{cap}). (B) Derivation of the rate of formation of P_{cap} from binomial distribution and pseudo-first-order kinetics. (C) Fit of equation B to capture data yielded the rate constants (k_1 , k_2) for oxidation, aniline-catalyzed ligation (pH 4.7), and non-catalyzed ligation (NaOAc buffer, pH 4.5) on phage. (D) Rate constants of reactions on peptides, as monitored by HPLC, were similar to those in panel C. Kinetics contains data from 2–3 independent experiments, k are reported as average and 95% confidence range of the fit. (E) In the population where $B_0 = 10\%$, phage contains 1, 2, or 3 copies of AOB-modified peptides. The yield of modification could be estimated only if the maximum number of reactive sites (N_{max}) is known.

2.2.5 Synthesis of glycopeptide on monoclonal phage

The pull-down assay can quantify both the appearance of aldehydes as a result of oxidation and their disappearance due to oxime ligation (Figure

2-6A). We observed that the capture yield did not change when the AOB-labeled phage was incubated with methoxylamine (Figure 2-6B). This result suggested that the oxime linkage was thermodynamically stable under the reaction conditions, and a phage reacted with AOB does not react further with any other aminoxy reagent. This observation was further confirmed by analogous reactions performed on a tetrapeptide ($\text{H}_2\text{N-SVEK-CO}_2\text{H}$) and verification of the products by ESI-MS (Appendix A-2). Building on this observation, we could quantify modification of phage with any molecule that carries a functional group that reacts irreversibly with an aldehyde. For example, if oxidized phage is first exposed to methoxylamine, subsequent exposure to AOB quantifies the amount of unreacted aldehydes. The ligation yield of methoxylamine (<85%) was calculated as the difference of yields between reaction 1 and 2 (Figure 2-6B). Similarly, I have quantified the ligation yield of oxidized phage with 2-(aminoxy)ethyl α -D-mannopyranoside (AOMan) to be <80% (Figure 2-6B). We note that, because a single modification with biotin is sufficient for phage capture, the follow-up reaction using AOB thus quantified the number of phage that contained no unreacted aldehyde in any of the five copies of pIII. To ensure that phages were stable for the duration of the selection process, I demonstrated that the oxime linkage was stable for at least six days when phages were stored in neutral solution at 4 °C (Appendix A-3).

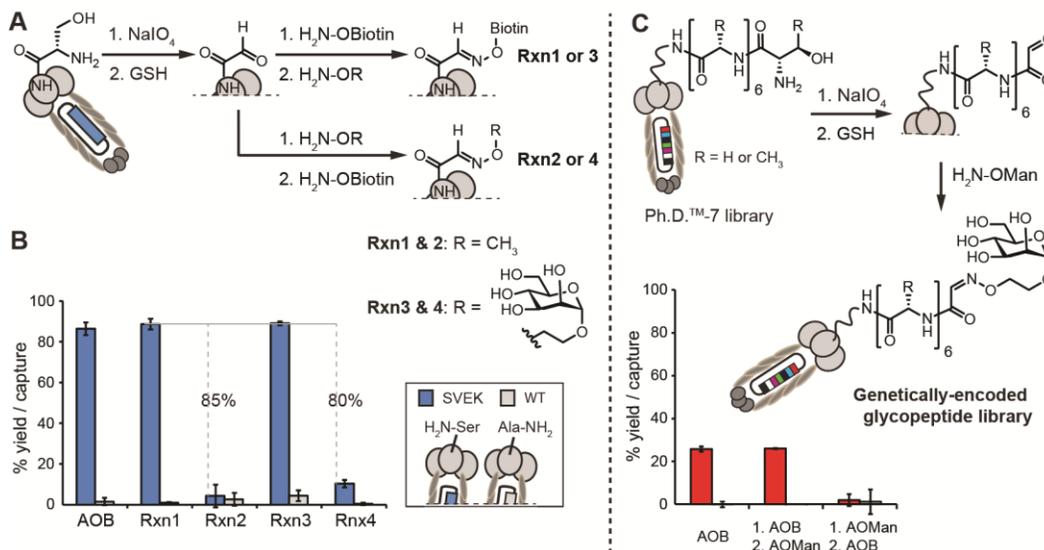


Figure 2-6. Quantification of glycopeptide formation on monoclonal phage or phage library.

(A) Changing the order of addition of reagents quantified the coupling yield of methoxylamine or AOMan with oxidized phage. (B) Coupling yields were determined by subtracting the yield of reaction 2 or 4 from the yield of reaction 1 or 3. (C) A similar strategy was used to quantify the efficiency of reaction on Ph.D.-7 library. In panels B and C, the value is an average of at least two independent experiments run on different days; error bars are one SD.

2.2.6 Synthesis of glycopeptide libraries

The biotin-capture assay allowed, for the first time, accurate yield determination of reactions on 10⁹ diverse peptides present at subnanomolar concentration. I observed that 26% of the clones in the phage library (Ph.D.-7) could be oxidized and reacted with AOB (Figure 2-6C). The observed value is similar to the fraction of *N*-terminal Ser and Thr present in the library as determined by deep sequencing.¹⁰⁸ Eventually, I demonstrated the synthesis of a library of glycopeptides with a simple carbohydrate epitope (*e.g.*, mannose, Figure 2-6C). We envision that a phage-displayed library of glycopeptides could accelerate the discovery of high-affinity ligands for

carbohydrate-binding proteins (*e.g.*, lectins). These proteins regulate pathogen invasion, immune response, and cancer development, but potent inhibitors for lectins are scarce due to the low affinity of lectin–carbohydrate interaction. Hybrid ligands (*i.e.*, glycopeptides) could potentially bind to the carbohydrate-binding site and adjacent sites in a synergistic fashion.^{117,118} We believe that the availability of a large, genetically traceable library of glycopeptides will aid in the selection of optimal side chains of the secondary binding ligand as well as the optimal spacing and flexibility of the linker.

2.3 Conclusion

Characterization of reaction efficiencies is a cornerstone of chemical synthesis. Generation of genetically encoded libraries of peptide derivatives through modification of phage libraries hinges on quantitative conversion of all reactive peptides in the library. Although strategies for modifying phage libraries were proposed many years ago,^{95,96} lack of simple and quantitative characterization tools remains one of the roadblocks to the development of this technology. Reactions applied to an entire phage library have typically been optimized using one sequence with the assumption that all peptides with reactive groups react similarly. In practice, variations in structural, electrostatic, and steric factors in the diverse peptides inevitably influence the efficiency of any reaction. For example, generating a library of peptides with *N*-terminal Ser¹¹⁹ does not guarantee 100% modification because Ser-Pro sequences within this library could undergo rapid cyclization after

oxidation to glyxoylyl-Pro.¹²⁰ This self-reactivity precludes the ligation with *O*-alkyl hydroxylamines. On the other hand, we demonstrated that AOB tagged a subpopulation of 2.5×10^8 reactive peptides within the random library of 10^9 peptides (Figure 2-6C). Biotin tagging thus could complement genetic engineering by marking the reactive clones within libraries. If necessary, the AOB-tagged phage clones could be isolated and reamplified to yield a sublibrary that form an oxime quantitatively. The library of 10^8 peptides containing an aldehyde functionality, which was produced within 30 min in >90% yield, could be readily derivatized using any glycan containing an aminoxy functional group.¹²¹ Moreover, an aldehyde-containing library is a rich starting point for a wide variety of reactions, such as Wittig reactions, that proceed with high efficiency in water and tolerate a range of functionalities.¹⁰² As a result, one commercially available phage library could provide multiple opportunities for the generation of novel genetically encoded libraries of peptide derivatives. We believe that the availability of this “starting material” coupled with the simplicity of the methodology described herein will render the synthesis of custom chemically modified phage libraries accessible to a broad range of chemical biology researchers in academic and industrial laboratories.

2.4 Experimental procedures

2.4.1 Materials and general information

The M13KE vector encoding SVEK phage with the sequence of SVEKNDQKTYHA-GGGS-pIII was a generous gift from Beth Paschal at New England Biolabs (NEB). The vector was transformed into chemically-competent *E. coli* to produce the corresponding phage; for transformation, amplification, PFU assays and other general phage methods see protocols and manuals at the phage display section at www.neb.com. Insert-free WT phage was isolated and sub-cloned from clear plaques, which appear in a population of “blue” plaques (*lacZ* (+) library) after several rounds of bulk amplification of phage library. BigDye® DNA sequencing was performed to confirm that WT phage has an alanine on *N*-terminus of pIII protein. Ph.D.™-7 library (lot # 0061101) was purchased from NEB. All other chemicals for synthesis were purchased from Sigma-Aldrich, TCI, and Acros and used without purification. Fluorescein-5-thiosemicarbazide (FTSC) was purchased from GenoLite Biotek. Aminoxy-biotin was purchased from Invitrogen and Cayman Chemical. PBS contains 10 mM sodium phosphate dibasic, 2 mM potassium phosphate monobasic, 137 mM sodium chloride and 2.7 mM potassium chloride with pH of 7.4 after preparation.

2.4.2 Quantification of on-phage reactions

The following is a representative procedure (other conditions can be found in the figure legends). A mixed solution of SVEK and WT phage (99 μ L,

10^{11} PFU mL⁻¹ in PBS, pH 7.4) was oxidized with NaIO₄ (add 1 μL of 6 mM in ddH₂O) for 5 min on ice in the dark and quenched with glutathione (add 1 μL of 50 mM in ddH₂O) for 10 min at room temperature (RT). Treatment with aminoxy reagent (add 101 μL of 2 mM in sterile 200 mM anilinium acetate, pH 4.7) for 1 h at RT yielded the corresponding biotinylated peptide or glycopeptide. For reaction involved AOB, reaction mixture was diluted by 10⁵-fold with binding buffer (PBS, 0.1% BSA) to quench the reaction and to ensure that traces of the biotin reagent do not saturate the binding site of SA-coated magnetic beads. The diluted phage solution (100 μL) was agitated with SA-coated magnetic beads (5 μL, Bioclone Inc., binding capacity: 2 pmol μL⁻¹) for 15 min at RT and captured on a magnetic separator. The phages in the supernatant were quantified by plaque-forming assay. The yield of reaction was determined as $(A - B)/A \times 100\%$, where *A* and *B* are the titers of phage before and after capture. Note: Ph.D.-7 library should be dialyzed (24 h, 4 °C, 10 K MWCO) against PBS (5 L, changed twice after 4 h interval) prior to the reaction to remove the glycerol (storage buffer).

2.4.3 Synthesis of 2-(aminoxy)ethyl α-D-mannopyranoside

The aminoxy mannopyranoside **5** (AOMan) used to generate the glycopeptide library was made according to Figure 2-7. Details of the procedures are described below.

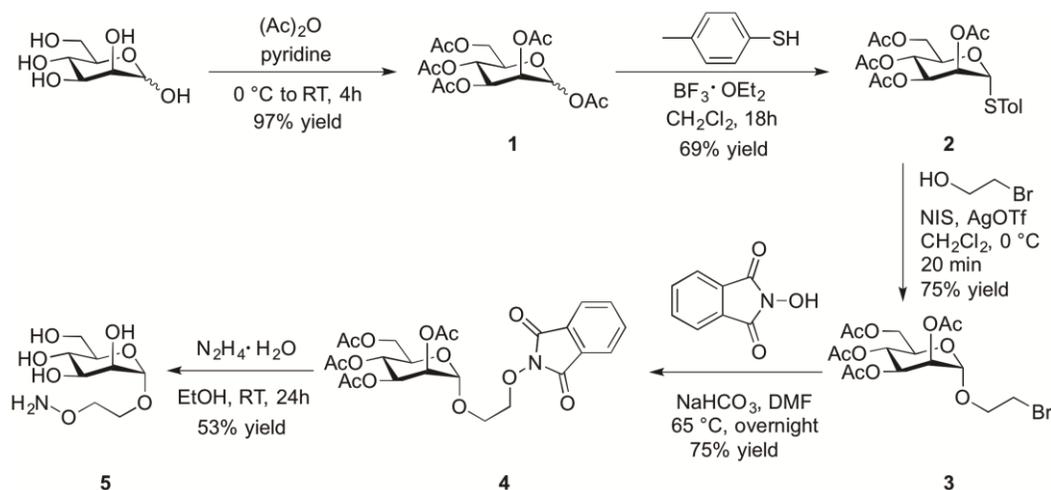
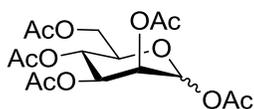


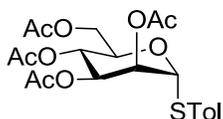
Figure 2-7. Synthesis of 2-(aminooxy)ethyl α -D-mannopyranoside.



1,2,3,4,6-Penta-*O*-acetyl- α/β -D-mannopyranose (1)

Acetic anhydride (7.40 mL, 78.3 mmol) was added to a stirred solution of D-mannose (2.00 g, 11.1 mmol) in pyridine (20 mL) at 0 °C. The mixture was allowed to warm to room temperature and was stirred for an additional 4 h. The mixture was then poured into ice water (ca. 100 mL) and extracted with ethyl acetate (3 \times 80 mL). The combined organic layers were washed with saturated NaHCO₃ solution (3 \times 150 mL), 1 N HCl (3 \times 80 mL), dried over Na₂SO₄, filtered, and concentrated *in vacuo* to yield compound **1** (4.21 g, 97% yield) as a mixture of anomers (7:1 α/β): ¹H NMR (500 MHz, CDCl₃) δ = 6.08 (d, *J* = 1.8 Hz, 1 H), 5.35 – 5.33 (m, 2 H), 5.26 – 5.25 (m, 1 H), 4.27 (dd, *J* = 5.0, 12.5 Hz, 1 H), 4.09 (dd, *J* = 2.5, 12.4 Hz, 1 H), 4.06 – 4.03 (m, 1 H), 2.17 (s, 3

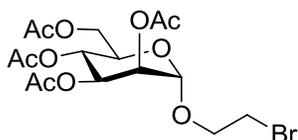
H), 2.16 (s, 3 H), 2.08 (s, 3 H), 2.04 (s, 3 H), 2.00 (s, 3 H); ^{13}C NMR (126 MHz, CDCl_3) δ = 170.7, 170.1, 169.8, 169.6, 168.2, 90.7, 70.7, 68.9, 68.5, 65.7, 62.2, 21.0, 20.9, 20.8, 20.8, 20.8; HRMS (ESI) calcd for $\text{C}_{16}\text{H}_{22}\text{O}_{11}\text{Na}$ $[\text{M}+\text{Na}]^+ m/z$ = 413.1054, found 413.1048. (see Appendix A-4 for NMR spectra)



4-Methylphenyl 2,3,4,6-tetra-*O*-acetyl-1-thiol- α -D-mannopyranoside (**2**)

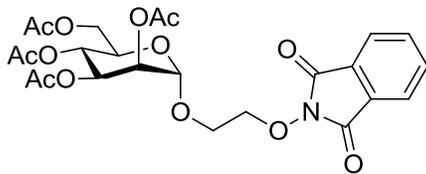
To a cold solution (4 °C) of compound **1** (1.78 g, 4.6 mmol) and 4-methylbenzenethiol (0.68 g, 5.5 mmol) in anhydrous CH_2Cl_2 (30 mL), $\text{BF}_3\cdot\text{OEt}_2$ (0.67 mL, 5.5 mmol) was added dropwise. After stirring at room temperature for overnight, the mixture was diluted with CH_2Cl_2 (60 mL), washed with saturated NaHCO_3 solution (2×100 mL) and brine (100 mL), and dried over Na_2SO_4 . After removal of drying agent and solvent, the crude product was further purified on silica gel (120 g) with a gradient of 0-30% ethyl acetate in hexane using CombiFlash® Rf. Removal of solvent and drying *in vacuo* yielded glycosyl donor **2** as white solid (1.43 g, 69% yield): ^1H NMR (500 MHz, CDCl_3) δ = 7.38 – 7.35 (m, 2 H), 7.11 (dd, J = 0.6, 8.4 Hz, 2 H), 5.48 (t, J = 2.3 Hz, 1 H), 5.40 (d, J = 1.1 Hz, 1 H), 5.32 – 5.30 (m, 2 H), 4.56 – 4.52 (m, 1 H), 4.28 (dd, J = 5.9, 12.3 Hz, 1 H), 4.09 (dd, J = 2.4, 12.3 Hz, 1 H), 2.31 (s, 3 H), 2.13 (s, 3 H), 2.06 (s, 3 H), 2.05 (s, 3 H), 2.00 (s, 3 H); ^{13}C NMR (126 MHz, CDCl_3) δ = 170.6, 170.0, 169.9, 169.8, 138.5, 132.7, 130.0, 128.9, 86.1, 71.0, 69.5, 69.5, 66.5, 62.6, 21.2, 20.9, 20.8, 20.8, 20.7; HRMS (ESI) calcd for

C₂₁H₂₆O₉Na [M+Na]⁺ *m/z* = 477.1190, found 477.1180. (see Appendix A-5 for NMR spectra)



2-Bromoethyl 2,3,4,6-tetra-*O*-acetyl- α -D-mannopyranoside (3)

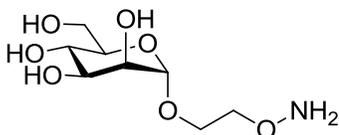
To a mixture of glycosyl donor **2** (1.29 g, 2.8 mmol) and 2-bromoethanol (0.3 mL, 4.3 mmol) in anhydrous CH₂Cl₂ (28 mL) containing 4 Å molecular sieves, I added *N*-iodosuccinimide (0.83 g, 3.7 mmol) followed by silver triflate (0.07 g, 0.3 mmol). After stirring at 0 °C for 20 min, TLC indicated complete conversion of the reaction. The reaction was quenched with a few drops of triethylamine. The mixture was filtered through Celite and concentrated. The residue was purified on silica gel (40 g) with a gradient of 0–50% ethyl acetate in hexane using CombiFlash® Rf. Removal of solvent and drying *in vacuo* yielded compound **3** as white crystals (0.97 g, 75% yield): ¹H NMR (500 MHz, CDCl₃) δ = 5.34 (dd, *J* = 3.5, 10.1 Hz, 1 H), 5.30 – 5.26 (m, 2 H), 4.86 (d, *J* = 1.8 Hz, 1 H), 4.26 (dd, *J* = 5.9, 12.5 Hz, 1 H), 4.14 – 4.11 (m, 2 H), 3.97 (td, *J* = 6.3, 11.2 Hz, 1 H), 3.88 (td, *J* = 5.7, 11.2 Hz, 1 H), 3.51 (t, *J* = 6.1 Hz, 2 H), 2.15 (s, 3 H), 2.10 (s, 3 H), 2.04 (s, 3 H), 1.99 (s, 3 H); ¹³C NMR (126 MHz, CDCl₃) δ = 170.7, 170.1, 170.0, 169.9, 97.9, 69.6, 69.2, 69.1, 68.6, 66.2, 62.6, 29.7, 21.0, 20.9, 20.8, 20.8; HRMS (ESI) calcd for C₁₆H₂₃BrO₁₀Na [M+Na]⁺ *m/z* = 477.0367, found 477.0358, 479.0340. (see Appendix A-6 for NMR spectra)



***N*-2-[2-[(2,3,4,6-tetra-*O*-acetyl- α -D-mannopyranosyl)oxy]ethoxy]-
phthalimide (4)**

N-Hydroxyphthalimide (0.46 g, 2.8 mmol) was deprotonated with NaHCO₃ (0.31 g, 3.7 mmol) in DMF (10 mL) at 65 °C for 1 h. Compound **3** (0.85 g, 1.9 mmol) was then added into the mixture. After stirring at 65 °C for overnight, the mixture was diluted with CH₂Cl₂ (70 mL), washed with saturated NaHCO₃ solution (3 × 50 mL) and brine (2 × 50 mL), and dried over Na₂SO₄. The mixture was filtered and the filtrate was concentrated. The residue was purified on silica gel (80 g) with a gradient of 0-50% ethyl acetate in hexane using CombiFlash® Rf. Removal of solvent and drying *in vacuo* yielded compound **4** as white crystals (0.76 g, 75% yield): ¹H NMR (500 MHz, CDCl₃) δ = 7.86 – 7.83 (m, 2 H), 7.78 – 7.73 (m, 2 H), 5.30 – 5.27 (m, 2 H), 5.22 (dd, *J* = 1.8, 2.9 Hz, 1 H), 4.92 (d, *J* = 1.7 Hz, 1 H), 4.41 (t, *J* = 4.4 Hz, 2 H), 4.31 (dd, *J* = 5.0, 12.3 Hz, 1 H), 4.19 (tdd, *J* = 2.5, 4.7, 7.1 Hz, 1 H), 4.14 (dd, *J* = 2.1, 12.0 Hz, 1 H), 4.07 – 4.03 (m, 1 H), 3.92 – 3.87 (m, 1 H), 2.15 (s, 3 H), 2.10 (s, 3 H), 2.04 (s, 3 H), 1.95 (s, 3 H); ¹³C NMR (126 MHz, CDCl₃) δ = 170.9, 170.1, 169.9, 169.7, 163.5 (×2), 134.7 (×2), 128.9 (×2), 123.9 (×2), 98.3, 77.0, 69.5, 69.1, 68.8, 66.7, 66.2, 62.6, 21.0, 20.9, 20.9, 20.8; HRMS (ESI) calcd for

$C_{24}H_{27}NO_{13}Na$ $[M+Na]^+$ $m/z = 560.1375$, found 560.1366. (see Appendix A-7 for NMR spectra)

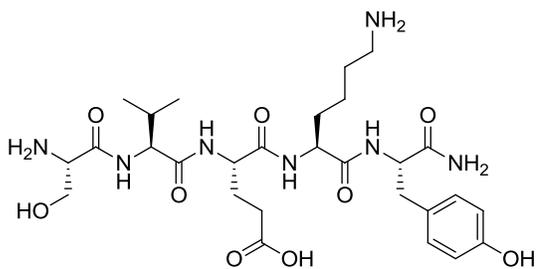


2-(Aminoxy)ethyl α -D-mannopyranoside (5, AOMan)

To a solution of compound **4** (0.25 g, 0.47 mmol) in ethanol (10 mL), hydrazine hydrate (0.41 mL of 64% solution, 8.4 mmol) was added and reaction was allowed to stir at room temperature overnight. The mixture was then filtered through Celite, rinsed with cold ethanol, and concentrated. The residue was purified on silica gel (12 g) with a gradient of 0–25% methanol in CH_2Cl_2 using CombiFlash® Rf. Removal of solvent and drying *in vacuo* yielded aminoxy-mannose **5** as pale-yellow oil (0.06 g, 53% yield): 1H NMR (400 MHz, CD_3OD) $\delta = 4.79$ (d, $J = 1.8$ Hz, 1 H), 3.90 – 3.84 (m, 2 H), 3.84 – 3.80 (m, 4 H), 3.73 – 3.68 (m, 2 H), 3.67 – 3.62 (m, 1 H), 3.60 – 3.58 (m, 2 H); ^{13}C NMR (101 MHz, CD_3OD) $\delta = 101.7, 75.7, 74.6, 72.5, 72.1, 68.6, 66.6, 62.9$; HRMS (ESI) calcd for $C_8H_{18}NO_7$ $[M+H]^+$ $m/z = 240.1078$, found 240.1073. (see Appendix A-8 for NMR spectra)

It is important to note that acetone should be avoided during the synthesis and work-up of AOMan because the compound rapidly forms an oxime with acetone.

2.4.4 Synthesis of H₂N-SVEKY-CONH₂ (SVEKY)



Rink amide resin (0.30 g, 0.71 mmol/g, 0.21 mmol) was weighed into a polypropylene column (10 mL) equipped with a frit. The resin was swelled with CH₂Cl₂ (4 mL) for 15 min and subsequently washed with DMF (4 mL). The resin was deprotected with 20% (v/v) piperidine/DMF (4 mL) for 1 min, and the deprotecting step was repeated again for 10 min. The resin was washed with DMF (5 × 4 mL) to remove excess of reagents. Fmoc-Tyr(*t*Bu)-OH (0.39 g, 0.84 mmol, 4 eq.) and HBTU (0.32 g, 0.84 mmol, 4 eq.) was dissolved in DMF (3 mL), and subsequently added into the resin. After adding DIPEA (0.29 mL, 1.68 mmol, 8 eq.), the column was rocked for 30 min. After removing the excess of reagents, the resin was washed with DMF (4 × 4 mL). The deprotection and coupling cycle were repeated with the procedures just described above using the following amino acid derivatives: Fmoc-Lys(Boc)-OH (0.39 g), Fmoc-Glu(*O**t*Bu)-OH (0.36 g), Fmoc-Val-OH (0.29 g), and Fmoc-Ser(*t*Bu)-OH (0.32 g). After the last coupling step, *N*-terminal Fmoc-group was removed by treatment of the resin with 20% (v/v) piperidine/DMF (1 mL) for 1 min, and the deprotecting step was repeated again for 10 min. The resin was washed with DMF (5 × 4 mL), followed by CH₂Cl₂ (5 × 4 mL). The peptide was cleaved by subjecting the resin to TFA/H₂O/TIPS/phenol (4 mL,

85/5/5/5 (v/v/v/w)) for 2 h. After washing with TFA (1 ml), the combined TFA solutions were concentrated to 1 mL under a stream of nitrogen and added dropwise to cold diethyl ether (10 mL). The mixture was incubated on ice for 30 min, centrifuged (2000 g, 5 min), and washed with cold diethyl ether (3 × 10 mL) to yield the resulting peptide as white solid (64 mg, 49% yield): ¹H NMR (400 MHz, D₂O) δ = 7.18 (d, *J* = 8.5 Hz, 2 H), 6.86 (d, *J* = 8.5 Hz, 2 H), 4.60 (dd, *J* = 6.0, 9.2 Hz, 1 H), 4.37 (dd, *J* = 6.0, 8.6 Hz, 1 H), 4.29 – 4.18 (m, 3 H), 4.06 – 3.93 (m, 2 H), 3.13 (dd, *J* = 6.0, 14.1 Hz, 1 H), 3.02 – 2.90 (m, 3 H), 2.51 – 2.33 (m, 2 H), 2.15 – 2.04 (m, 1 H), 2.04 – 1.87 (m, 2 H), 1.76 – 1.59 (m, 4 H), 1.42 – 1.22 (m, 2 H), 0.96 (t, *J* = 6.6 Hz, 6 H); ¹³C NMR (101 MHz, D₂O) δ = 176.9, 175.6, 173.1, 172.7, 167.8, 154.4, 130.6, 128.2, 117.8, 115.4, 60.3, 59.6, 54.5, 54.4, 53.5, 52.7, 39.1, 36.2, 30.4, 30.1, 29.8, 26.3, 21.9, 18.3, 17.7; HRMS (ESI) calcd for C₂₈H₄₄N₇O₉ [M–H]– *m/z* = 622.3206, found 622.3208. (see Appendix A-9 for NMR spectra)

Chapter 3: Genetically encoded fragment-based discovery of glycopeptide ligands for carbohydrate-binding proteins

3.1 Introduction

Carbohydrate–protein interactions are central to a variety of normal and pathological processes, including inflammation, cell-to-cell adhesion, metastasis, and recognition of pathogens by the immune system. Although inhibition of some of these interactions would be therapeutically valuable, the development of effective inhibitors for lectins is challenging due to the synthetic complexity and usually low intrinsic affinity of the native oligosaccharides.^{122,123} The generally shallow landscape of many carbohydrate-binding sites and the lack of strong hydrophobic interactions with the ligand results in fast off-rates (k_{off}).⁴ The interaction can be reinforced by using a non-carbohydrate synergistic motif that can fill the remaining space in the carbohydrate-binding site or occupy area on the protein surface adjacent to the principal binding site.¹²⁴⁻¹²⁶

3.2 Hypothesis and objective

Partial¹²⁴⁻¹²⁶ or complete^{127,128} replacement of glycan with simple structural blocks has been successful in the discovery of lectin ligands for therapeutic targets, such as E-Selectin⁵⁵ and MAG.⁵⁰ This replacement often yields ligands with improved metabolic stability and pharmacokinetic

properties for applications such as probing lectin function,¹²⁹ design of therapeutics,^{4,130} and vaccines.^{130,131} Generally, these ligands are developed through rational design, requiring multistep synthesis and complex chemical manipulations.¹²³ Genetic encoding of peptide libraries offers a 10⁵-fold increase in the throughput and selection of active peptides that bind lectins.¹³²⁻¹³⁶ However, binding of peptides lacking a carbohydrate fragment to lectins is often non-specific, and these peptides often target non-carbohydrate-binding sites.¹³⁷ Selections that use the peptide solely as a linker, rather than as a recognition element, can yield peptides as a multivalent scaffold for Man9 and yield glyco-oligomers with avidity in the sub-nanomolar range.¹³⁸ We pursued an approach in which we aim to identify peptides that can serve as a moiety that synergizes with glycans in binding to lectins (Figure 3-1), rather than acting as a neutral linker or standalone recognition element. Indeed, synthetic conjugates of peptides and carbohydrates are known to yield effective inhibitors of carbohydrate-protein interactions,^{139,140} but the throughput of synthesis of these ligands is limited. Here, we employ genetically encoded fragment-based discovery (GE-FBD) to identify peptide fragments (Figure 3-1) capable of forming a synergistic interaction together with a glycan moiety by taking advantage of the immense diversity of a genetically encoded peptide library.^{73,141}

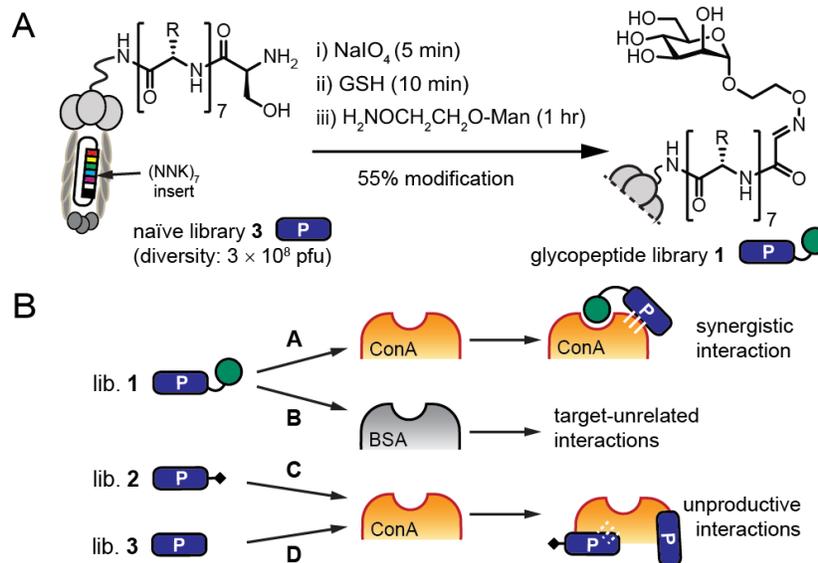


Figure 3-1. Illustration of generation of glycopeptide library and its selection.

(A) Modification of phage-displayed peptide library (**3**) yielded Man-X₇ (**1**). (B) Libraries **1** (Man-X₇), **2** (methyl-X₇), and **3** (Ser-X₇) were panned against two targets (ConA or BSA) in parallel.

3.3 Results and discussion

3.3.1 Selection of glycopeptide libraries

The GE-FBD procedure starts with the synthesis of a glycopeptide-presenting phage library with diversity of $<10^8$, via periodate oxidation of a peptide library with fixed *N*-terminal serine residues, and subsequent chemical modification of the resulting bioorthogonal aldehyde functionality (Figure 3-1A).¹⁴¹ Next, I employed deep-sequencing¹⁰⁸ and multiple panning controls to identify functional ligands, even after a single round of panning. In the proof-of-concept experiment with the model lectin, concanavalin A (ConA), I panned an α -mannopyranoside-oxime-terminated library (Man-X₇,

1) against ConA and a control target (bovine serum albumin, BSA) (screens **A** and **B**, Figure 3-1B). In parallel, I panned control libraries terminated with *O*-methyl-oxime (methyl-X₇, **2**) and an unmodified naïve library (**3**) against ConA (screens **C** and **D**, Figure 3-1B). The four combinations of the target and the library—**A–D**—were incubated, rinsed, and processed on a 96-well plate, each in 3–6 replicates. After panning, recovery of phage DNA, and ion torrent sequencing, I acquired 10^5 sequences per replicate (Figure 3-2).

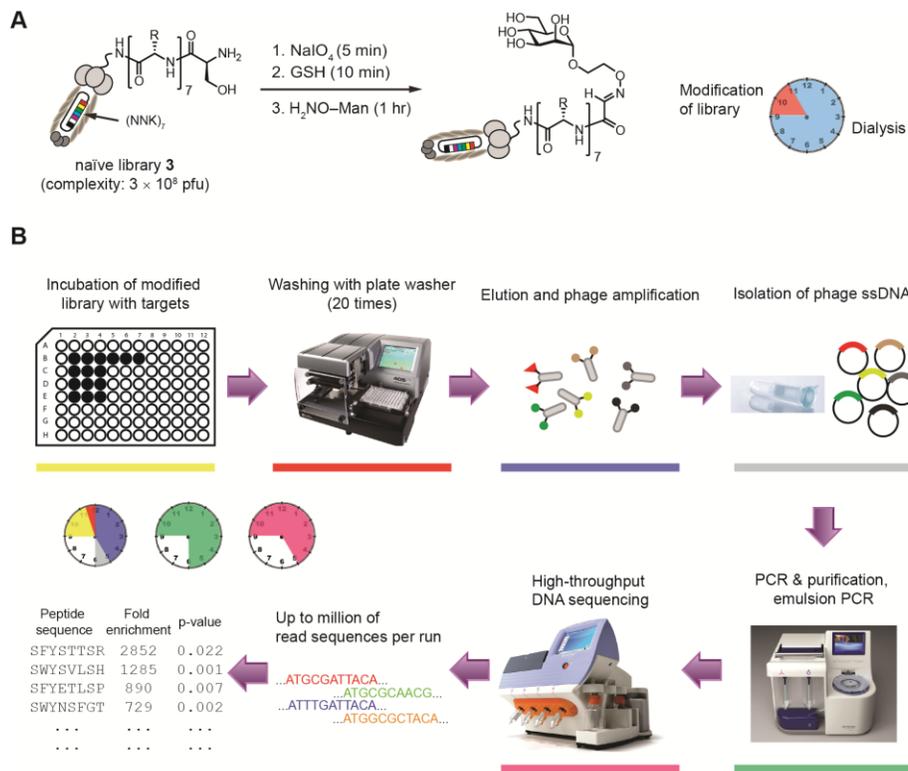


Figure 3-2. Overview of the workflow of phage selection.

The entire ligand search could be accomplished within a week. (A) Modification of phage library and subsequent purification by dialysis required a day. (B) Incubation of the modified library with targets on 96-well plate (yellow, 2 h); washing of plate (red, 0.5 h); elution and phage amplification (blue, 5 h); isolation of phage ssDNA (grey, 1 h); PCR of library DNA with 15-barcoded primers, E-Gel purification and template preparation on Ion OneTouch™ 2 System (green, one day); DNA sequencing with Ion PGM™ System (magenta, one day).

A student's *t*-test identified a set of 231 sequences that were significantly ($p \leq 0.05$) enriched by >5-fold in the screen against ConA but not against BSA (designated as set **A/B** in Figure 3-3A). In the **A/B** set, we identified a weak "consensus motif" (Figure 3-3C). By adding data from the control sets **A/C** and **A/D**, and calculating the intersection of $(\mathbf{A/B}) \cap (\mathbf{A/C}) \cap (\mathbf{A/D})$, we reduced the number of hits from 231 to 86 (Figure 3-3B); these 86 hits shared a pronounced consensus motif: Man-[WYF]Y[SDEA] (Figure 3-3C). Refer to Appendix B-1 for sequences of the 86 hits.

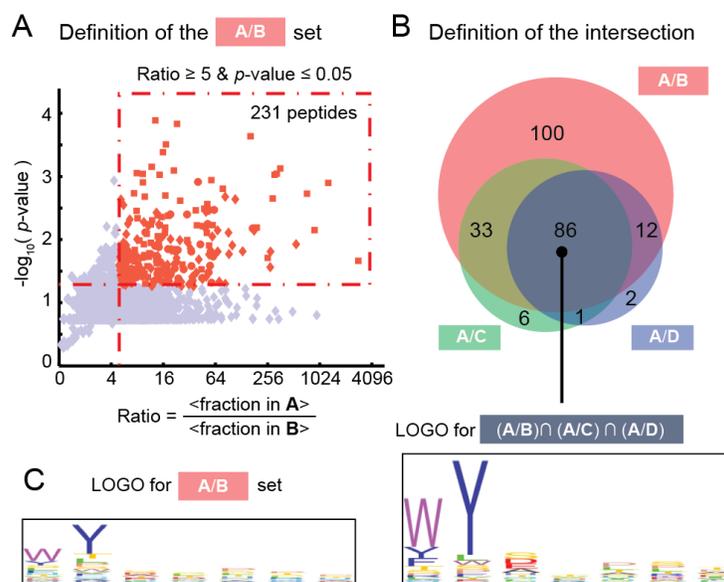


Figure 3-3. Post-selection analysis.

(A) A volcano plot defined sequences from library **1** that were significantly enriched in the ConA screen more than in the BSA screen. We abbreviate the set of these sequences as **A/B** and defined the sets **A/C** and **A/D** analogously. (B) 86 sequences belong to all three sets ("intersection"). (C) Consensus motifs defined by LOGO in set **A/B** and in the intersections with sets **A/C** and **A/D**.

3.3.2 Validation of selected hits

Several ligands chosen from the pool of 86 hits, based on their correlation to the consensus motif (Figure 3-3C) and their rank of abundance in screen **A**, were chemically synthesized and their activities evaluated by several methods. All ligands were able to inhibit the binding of ConA to the dextran-coated surface in a surface plasmon resonance (SPR) spectroscopic assay and exhibited lower IC_{50} values (11–34 μM) than methyl α -D-mannopyranoside (MeMan) (144 \pm 7 μM , Figure 3-4). The monosaccharide was essential to the activity; the corresponding unmodified peptides did not inhibit the binding of ConA to dextran (Figure 3-5A). Specific peptide sequences were required for synergistic binding, as the affinities of glycopeptides with randomly chosen sequences were indistinguishable from those of MeMan (Figure 3-5B).

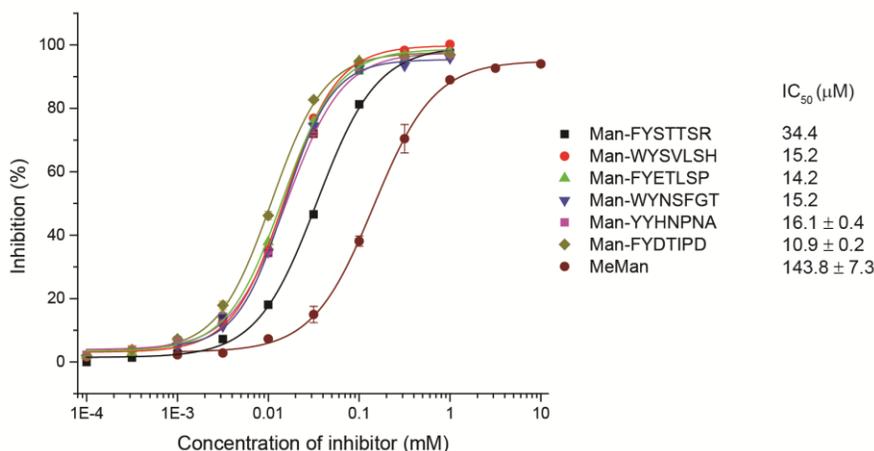


Figure 3-4. SPR inhibition profiles of Man- X_7 conjugates.

Inhibitions of ConA (200 $\mu\text{g}/\text{mL}$) binding to the dextran-coated surface (CM5 chip) were measured with SPR using 6 Man- X_7 conjugates or MeMan as the inhibitor. IC_{50} values with standard error were measured in three independent experiments.

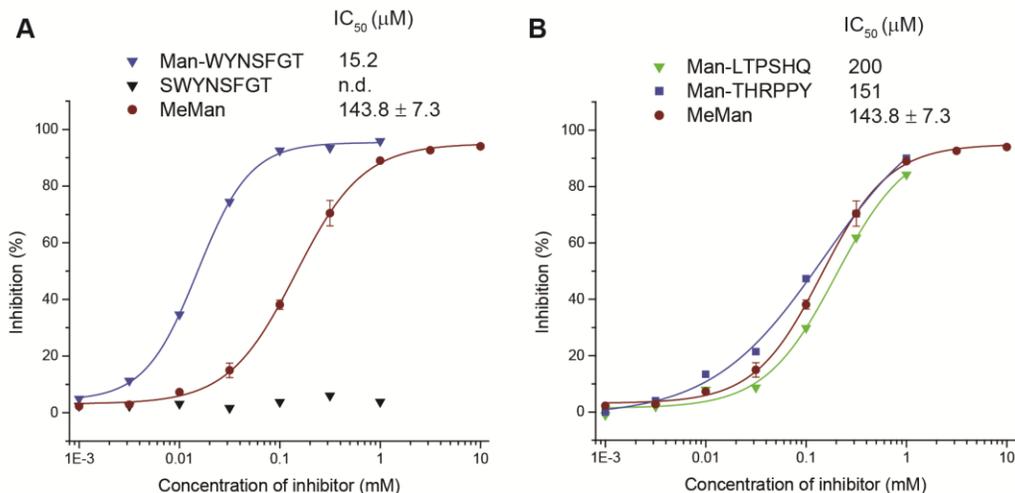


Figure 3-5. SPR inhibition profiles of control peptide and conjugates. (A) Inhibitions of ConA (200 μg/mL) binding to the dextran-coated surface (CM5 chip) were measured with SPR using Man-X₇ conjugate, Ser-X₇ or MeMan as the inhibitor. IC₅₀ of SWYNSFGT was not determined (n.d.) due to insignificant inhibition even with inhibitor concentration as high as 1 mM. (B) Inhibition profile of Man-X₆ conjugates and MeMan. These conjugates have IC₅₀ values similar to that of MeMan suggested that the inhibition mainly stem from the mannopyranosyl moiety and is independent of the peptide moiety. IC₅₀ values with standard error were measured in three independent experiments.

I validated the SPR findings using isothermal titration calorimetry (ITC) and found that all 10 selected Man-X₇ conjugates (**L2–L11**), including the six SPR-validated hits, could bind to ConA with higher affinity than MeMan (Table 3-1). ITC also confirmed the lack of activity in false positive hits. For example, **L12** ligand present in the **A/B** set but not the intersect (Figure 3-3B) showed no improvement in affinity over MeMan. By including **C** and **D** controls, we eliminate these false positives. The dissociation constants (K_D) of the corresponding peptides lacking Man fragment were found to be >1 mM (**L8.1**, **L9.1**, **L9.2**, **L11.1**; Table 3-1 and Appendix B-2).

Therefore, the synergistic sequences discovered in our screen would not be identified by screening unmodified peptide libraries.

Table 3-1. K_D values of synthetic ligands with ConA.

Ligand ^a	K_D (μ M) ^b	Ligand ^a	K_D (μ M) ^b
L1 MeMan	137.4 \pm 6.1	L19 Man- WY	12.0
L2 Man-YWEFTSL	55.2	L20 Man- WYD	10.0 \pm 0.9
L3 Man-F Y STTSR	42.6	L20.1 SWYD	>1000
L4 Man-AWEAYWY	29.2	L20.2 Me-WYD	>1000
L5 Man-F Y LGSDI	23.9	L21 Man- WYE	9.6
L6 Man-Y H NPNNA	20.9	L22 ^d Man- WYG-OH	9.8 \pm 1.1
L7 Man-F Y DTIPD	17.2	L23 Man- WYD ANHSKPL	6.0
L8 Man-F Y ETLSP	15.9	L24 Man- WYDR QETRFER	4.6
L8.1 SFYETLSP	>1000	L25 Man- WYDL HHSRTR	4.5
L9 Man- WY SVLSH	14.1	L26 Man- WYDL YHPVQH	4.3
L9.1 SWYSVLSH	>1000	L27 Man- WYEL LDDDDIT	5.3
L9.2 Me-WYSVLSH	>1000	L28 Man- WYDQ FPPLHQ	5.1
L10 ^c Man-Y Y DLM ^{ox} QT	12.0	L29 Man- WYDN FDTIFA	5.0
L11 Man-Y Y DLMQT	11.1	L30 Man- WYDLF DNINS	4.3 \pm 0.4
L11.1 SYDLMQT	>1000	L31 Man- WYDRF PPHES	3.7
L12 Man-HTHDSVE	151.5	L32 Man- WYDR	7.5
L13 Man-Y YD	20.1	L33 Man- WYDL	6.1
L14 Man-F YD	18.9	L34 Man- WYDFF	5.8
L15 ^d Man- WY -OH	17.1 \pm 0.4	L35 Man- WYEIF	5.3
L16 Man- WYS	14.3	L36 Man- WYDRF	4.9
L17 Man- WYA	13.5	L37 Man- WYDLF	4.6 \pm 0.4
L18 Man- WYH	12.7	ref. 17 Man3	2.6

^aIn general, red residues are essential for the strongest enhancement of affinity, while blue residues provide additional benefit.

^b K_D was derived from ITC data (Appendix B-3). The errors are one standard deviation of the mean value measured by 2–5 independent experiments.

^cM^{ox} is methionine sulfoxide.

^dWith the exception of **L15** and **L22**, all peptides were C-terminal amides (CONH₂).

3.3.3 Optimization and affinity maturation

LOGO analysis¹⁴² suggested that residues 4–7 contribute minimally to the binding. I therefore truncated the heptapeptide to a tripeptide and,

indeed, observed similar K_D values for the corresponding Man-peptides (compare **L7** to **L14**, **L9** to **L16**, and **L11** to **L13** in Table 3-1). I then performed two more rounds of panning using a focused library of Man-WY[D/E]X₇ (where X is any 20 amino acids, except Cys) and a panning–sequencing–analysis routine as described above. All of the hits bind to ConA at single-digit μ M affinity (**L23–L31**, Table 3-1) and showed on average 5-fold lower K_D than ligands from round one (**L2–L11**, Table 3-1). Truncation of the hits revealed that again the residues proximal to the glycan dominate the interactions. Man-WYDLFDNINS (**L30**) exhibited 33-fold lower K_D over MeMan (4.3 vs 140 μ M), and retained most of its potency when truncated to Man-WYDLF (**L37**). The K_D of **L37** (4.6 μ M) was close to that of α -Man-(1 \rightarrow 3)-[α -Man-(1 \rightarrow 6)]- α -Man (Man3, K_D = 2.6 μ M).¹⁴³ The origin of enhanced binding affinities of all Man-peptide conjugates are driven by enthalpic gain (see Appendix B-4 for thermodynamic data).

3.3.4 Structural analysis of Man-WYD–ConA complexes

To determine whether Man-WYD is mimicking Man3, we obtained the crystal structure of its complex with ConA at 1.73 Å resolution (Figure 3-6A). The mannopyranosyl moiety of the synthetic ligand occupies the same binding site as the (1 \rightarrow 6)-linked Man residue of the Man3 (Figure 3-6B) and displays the same H-bonding pattern (Figure 3-6C). However, the peptide moiety, WYD, does not occupy the same shallow area as the remaining disaccharide portion of the Man3 but instead resides in a somewhat deeper

cavity located next to Tyr12, which forms a wall between the two cavities on the protein surface at the tip of each arm of the ConA tetramer.

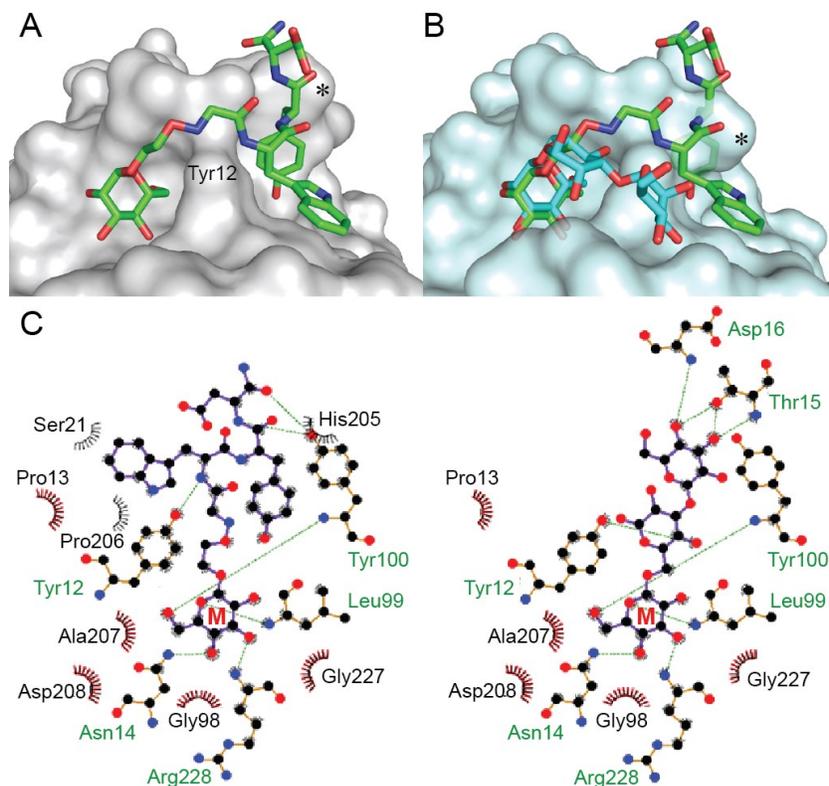


Figure 3-6. X-ray crystal structure of Man-WYD–ConA complexes. (A) Structure of Man-WYD (green) co-crystallized with ConA (PDB: [4CZS](#)). (B) Superimposition of Man-WYD with the Man3-ConA complex (cyan, PDB: [1CVN](#)), generated by aligning the protein chain A backbone atoms (RMSD = 0.32 Å). * indicates His205 residue. (C) Contact analysis of ConA bound to Man-WYD (left) and to the Man3 (right). The equivalent mannopyranoside residues are labeled with “M”.

Bound Man-WYD has more extensive van der Waals contact with the protein (contact area 662 Å², Figure 3-6C) when compared to the Man3 (contact area 204 Å²). Additionally, the complex revealed a ligand-induced fit: the conformation of the His205 residue has been changed to open a latent hydrophobic site, which is masked by this residue in both native ConA and

Man3-ConA structures (Figure 3-6A–B, asterisks). This site is now occupied by the Tyr side chain of Man-WYD. Hence, our screening identified a novel class of ligands that would be difficult to discover by structure-based design, which commonly employs docking of proposed ligands into a rigid protein.

Molecular dynamics (MD) simulations employing the GLYCAM/AMBER force field¹⁴⁴ evaluated the stability of the complex (see MD analysis in Appendix B-5) and permitted estimation of the contributions to affinity made by each residue in the protein and the ligand (Figure 3-7 and Appendix B-5).¹⁴⁵ Nearly 60% of the binding free energy is provided by H-bond/electrostatic interactions between the protein and the glycan, while hydrophobic interactions with the Trp and Tyr residues contribute the remainder. This observation is consistent with the view that H-bonds to the sugar provide specificity, whereas hydrophobic interactions enhance affinity.¹⁴⁶

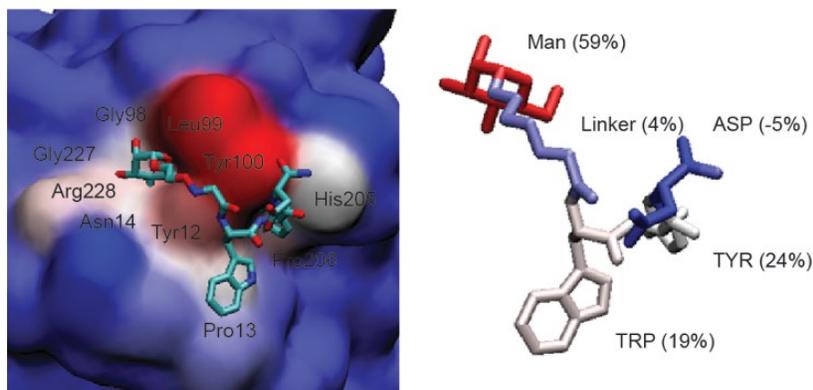


Figure 3-7. Binding free energy analysis of ConA bound to Man-WYD. Strength of interaction energies scaled from red (strongest) to blue (weakest) in the protein (left) and in the ligand, with percent contribution shown in parentheses (right).

3.3.5 Effect of linkers on binding affinity

A linker of some kind is essential in the GE-FBD approach. Minor changes, such as replacing the oxime fragment in **L20** with a hydrolytically stable linker of similar length (**L20.5**, Figure 3-8), results in only a 3-fold loss in binding affinity, possibly due to loss of the H-bond between the nitrogen atom of the oxime and the hydroxyl group of Tyr12, while shortening of the linker (**L20.3**) or incorporation of an allyl group (**L20.4**) completely abolishes the synergistic effect and returns the affinities to the level of the monosaccharide itself.

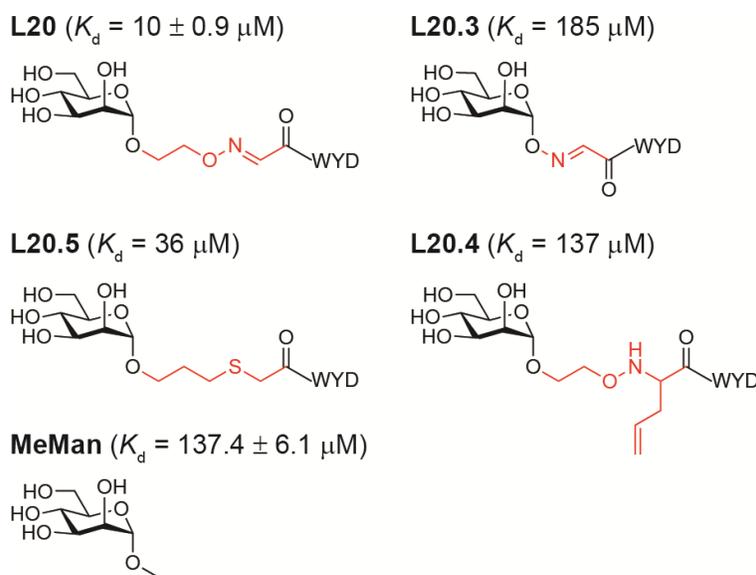


Figure 3-8. Effect of linkers on binding affinity.

Saturation transfer difference (STD) NMR spectroscopy (Figure 3-9) clearly demonstrated a substantial decrease of the interactions between WYD fragment and ConA in the ligand with a shorter linker, **L20.3**. We hypothesized that oxime functionality might be involved in molecular recognition. STD-NMR detected significant contacts between ConA and the

proton of the oxime (ON=CH), as well as Tyr and Trp of **L20** and **L15**. Both oxime and the aromatic rings play significant roles in the interaction of the ligand with the protein (Figure 3-6). These same protons in the control ligand (**L20.3**) exhibited much weaker signal in the STD-NMR, and, thus, significantly less contact with the protein. The results were in-line with the ITC measurements and crystallographic data. A shortened linker ablates the geometry necessary for synergistic binding.

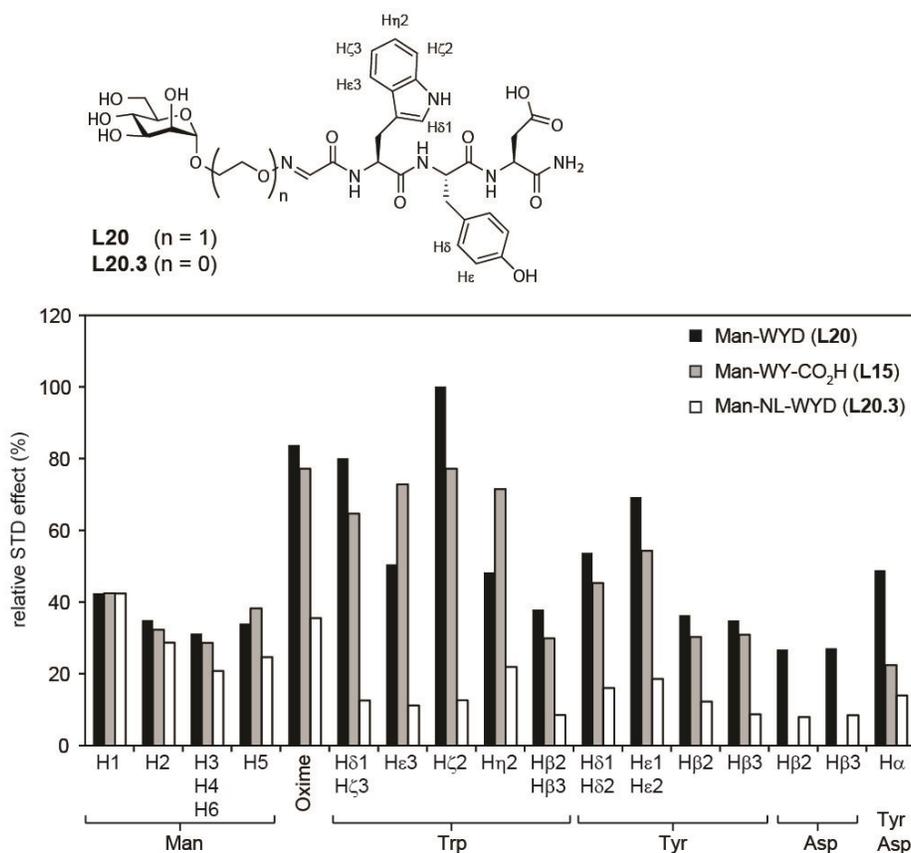


Figure 3-9. STD-NMR analysis of the interaction between ConA and glycopeptides.

To facilitate the ligand comparison, Man H1 of **L20.3** was set as the reference and normalized to 42% (same for **L20** and **L15**). The relative STD effects for other protons were then estimated based on this reference proton. The STD effect of Trp H α was not determined due to the complete attenuation of its signals by WATERGATE W5 sequence. Refer to Appendix B-6 for raw NMR spectra.

3.3.6 Binding selectivity of glycopeptides

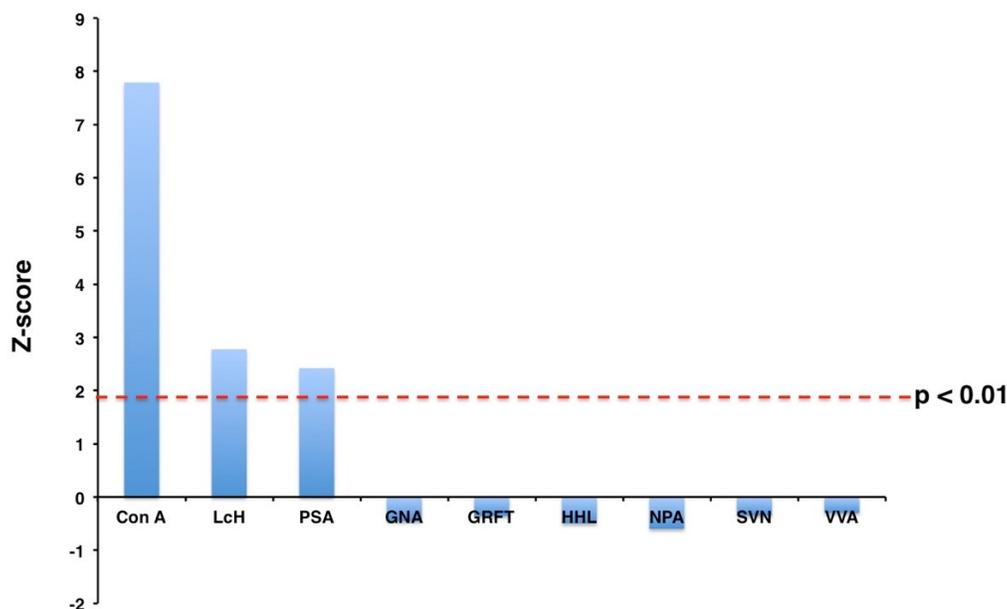
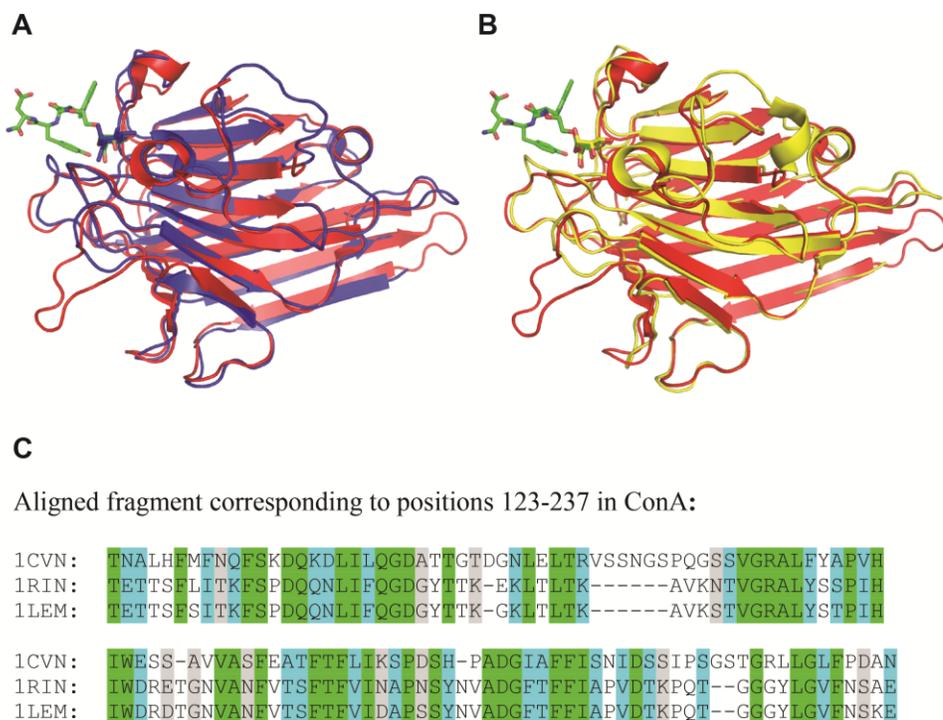


Figure 3-10. Lectin microarray analysis.

Z-scores for mannose-binding lectins from a lectin microarray incubated with 1.85 μM Cy3-labeled Man-WYK-OH are shown. The $p < 0.01$ cutoff is indicated (red dashed line, $Z = 1.95$). Of the 85 lectins tested, only ConA, LcH, and PSA met this significance threshold, while other mannose-binding proteins (GNA, GRFT, HHL, NPA, SVN, and VVA, see Appendix B-6 for details) did not.

To assess the selectivity of peptide binding to ConA over other lectins, we used a lectin microarray containing 85 lectins of varied specificities (Appendix B-7).¹⁴⁷ I synthesized a Man-WYK-OH probe (**L38**), which contained a fluorescent probe Cy3 attached to the side chain of lysine. Of the 85 lectins, 17 lectins exhibit selectivity for Man, yet only three show statistically significant binding to **L38**. Specifically, at 1.85 μM concentration, **L38** bound strongly to ConA but only weakly to two other Man-binding lectins (Figure 3-10), *Lens culinaris* (LcH) and *Pisum sativum* (PSA), both of which have a high degree of sequence similarity in their binding site and a

folded structure similar to that of ConA (Figure 3-11). ITC confirmed that the peptide fragment (WYDLF) provides a mere 2-fold enhancement of the binding of monosaccharide moiety to LcH and PSA, but a 35-fold benefit for ConA (Figure 3-12, Appendix B-8).



Where green marks exact match, blue – highly conservative substitution, grey – conservative substitution.

Figure 3-11. Structural and sequence homology of ConA-LcH and ConA-PSA. (A) Superimposition of Man-WYD-ConA complex (red, PDB: 4CZS) with the Glc-LcH complex (blue, PDB: 1LEM), generated by aligning the protein backbone atoms (RMSD = 0.81 Å). (B) Superimposition of Man-WYD-ConA complex (red, PDB: 4CZS) with the Man-PSA complex (yellow, PDB: 1RIN), generated by aligning the protein backbone atoms (RMSD = 0.76 Å). (C) Multiple sequence alignment of ConA (PDB: 1CVN), PSA (PDB: 1RIN), and LcH (PDB: 1LEM) using Clustal Omega from EMBL-EBI.

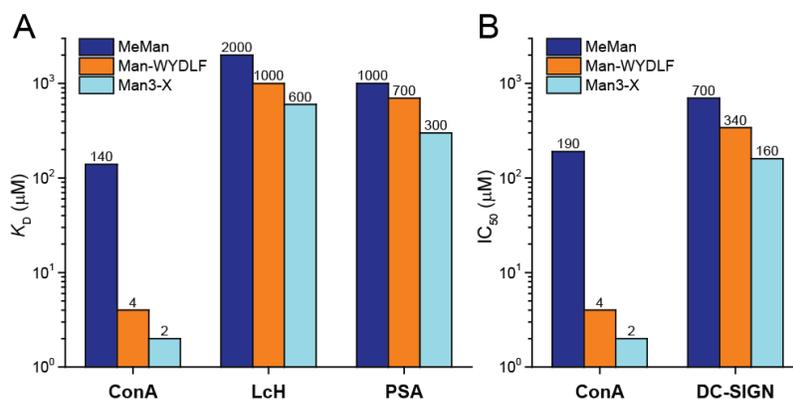


Figure 3-12. Binding selectivity of glycopeptide.

Binding of Man-WYDLF is specific to ConA, providing significantly less affinity for related Man-binding proteins as measured by (A) ITC and (B) inhibition studies.

We also tested the binding of the Man-peptides to DC-SIGN, an immune system lectin that recognizes the same Man3 motif as ConA and mediates the transmission of HIV during viral infection.¹⁴⁸ An ESI-MS assay¹⁴⁹ measured an enhanced affinity for binding of DC-SIGN to **L20** but not to the control ligand **L20.3** (Appendix B-9). The measured K_D of **L20** ($600 \pm 7 \mu\text{M}$) closely resembles the K_D of Man3 ($590 \mu\text{M}$) binding to DC-SIGN.¹⁵⁰ I demonstrated that Man-WYDLF is able to inhibit the binding of ConA to Man3-containing glycoprotein (horseradish peroxidase) with $IC_{50} = 4 \mu\text{M}$ (Figure 3-12B, Appendix B-10). The inhibitory effect of Man-WYDLF is comparable to that of Man3-X ($IC_{50} = 2 \mu\text{M}$, X = 6-azidohexyl). Affinities measured by inhibition are consistent with the ITC experiments (Figure 3-12A). The binding specificity of Man-WYDLF to ConA, LcH, PSA, and DC-SIGN closely mimics that of Man3. The affinity increases in the order LcH < PSA < DC-SIGN < ConA (Figure 3-12). While the hydrophobic interactions added by

WYDLF in the the glycopeptide are structurally distinct from the interactions provided by the dimannoside unit in Man3, both interactions add shape complementarity of ligand with the receptor and yield similar affinity and specificity for Man3 and Man-WYDLF.

3.4 Conclusion

In conclusion, I describe the discovery of ligands for challenging targets (lectins) from chemically modified phage libraries. This report expands the use of GE-FBD, which is known to work effectively with anchor fragments that have nanomolar affinity.⁷² Here we show that, despite their weak affinity, glycans can also serve as anchors in fragment-based discovery and yield synthetically accessible and potent glycopeptide ligands. These ligands bind competitively to the carbohydrate-recognition domain (CRD) with specificity akin to that of complex oligosaccharides and exhibit novel interactions with the CRD pocket. Therefore, GE-FBD may provide an alternative route to rapid enhancement of specific binding ligands for many targets starting from its known ligand or its essential binding fragment.

3.5 Experimental procedures

3.5.1 Materials and general information

PBS contains 10 mM sodium phosphate dibasic, 2 mM potassium phosphate monobasic, 137 mM sodium chloride and 2.7 mM potassium chloride with pH of 7.4 after preparation. ConA from *Canavalia ensiformis*

(Jack bean) was purchased from Sigma-Aldrich. ConA (2.5–10 mg, monomeric MW = 26500 Da) was dissolved in HEPES buffer (1.0 mL, 50 mM HEPES, 150 mM NaCl, 1 mM CaCl₂, pH 7.2). After incubating for overnight at 4 °C, the ConA mixture was centrifuged for 2 min at 14000 rpm and syringe-filtered (0.22 µm). The final subunit concentration of ConA solution was determined by UV absorbance at 280 nm ($A_{280} = 1.37 \times [\text{mg/mL ConA}]$).¹⁵¹ LcH from *Lens culinaris* (lentil) and PSA from *Pisum sativum* (pea) were purchased from Medicago. The monomeric MW of LcH and PSA are 24500 and 23500 Da, respectively.¹⁵² The final subunit concentrations of LcH or PSA solution were determined by UV absorbance at 280 nm ($A_{280} = 1.26 \times [\text{mg/mL LcH}]$ or $A_{280} = 1.50 \times [\text{mg/mL PSA}]$).¹⁵² Horseradish peroxidase (HRP) was purchased from Sigma-Aldrich. All solutions used for phage work are sterilized either by autoclave or by filter sterilization (0.22 µm)

3.5.2 Generation of Man-X₇ and Me-X₇ libraries

N-SerX₇ phage-displayed peptide library **3** (complexity: 3×10^8 pfu) was generated according to the referred protocol.¹⁵³ Prior to the chemical modification, the phage library was dialyzed extensively (4 °C, 10K MWCO) against two changes of PBS (5 L) to remove the storage buffer, which contains 50% (v/v) glycerol. The phage library (1 mL, $\sim 4 \times 10^{12}$ PFU mL⁻¹) was oxidized with 0.06 mM sodium periodate (by adding 10 µL of 6 mM solution in MQ water) at 4 °C for 5 min. The oxidation was quenched with 0.5 mM glutathione (by adding 10 µL of 50 mM solution in MQ water) at RT for

10 min. To monitor the oxidation, a small portion of the oxidized library was treated with aminoxy-biotin and captured with biotin-capture assay as described in a previously published method.¹⁴¹ Typically, 60% of the fractions of phage library were successfully oxidized. The oxidized library was distributed into two separate portions of 0.75 mL and 0.25 mL, and they were treated with 1 mM 2-(aminoxy)ethyl α -D-mannopyranoside (by adding 0.75 mL of 2 mM solution in 200 mM anilinium acetate buffer, pH 4.7) and 1 mM methoxylamine (by adding 0.25 mL of 2 mM solution in 200 mM anilinium acetate buffer, pH 4.7) respectively. The reaction mixtures were incubated for 1 h at RT, after which, the excess of reagents was removed by dialysis (4 °C, 10K MWCO) against two changes of PBS (5 L) to yield the Man-X₇ library **1** and Me-X₇ library **2**. To quantify the reaction efficiency, right after the oxime ligation, a small portion of the library was treated with aminoxy-biotin and captured with a biotin-capture assay as described in a previously published method.¹⁴¹ Typically, 55% of the fractions of phage library were successfully modified with the reagents.

3.5.3 Phage selection

12 wells of a 96-well polystyrene plate were coated with a solution of ConA (100 μ L, 100 μ g/mL) in PBS overnight at 4 °C. These wells, plus an additional three empty wells, were blocked with a solution of 2% (w/v) BSA in PBS (300 μ L) for 1 h at RT. The Man-X₇ library **1** (0.9 mL), Me-X₇ library **2** (0.3 mL), and N-Ser-X₇ library **3** (0.3 mL) were blocked with an equal volume

of 2× blocking solution (4% (w/v) BSA in PBS) for 1 h at RT. After blocking, the plate was rinsed with washing solution (3 × 300 μL, 0.1% (v/v) Tween-20 in PBS) using 405™ Touch Microplate Washer (BioTek). The selection of Man-X₇ library **1** against ConA (denoted as screen **A**) was performed in six replicates. The control selections, *i.e.*, Man-X₇ against BSA (**B**), Me-X₇ against ConA (**C**), and *N*-Ser-X₇ against ConA (**D**), were performed in triplicate in parallel with screen **A**. Specifically, the solutions of library **1**, **2**, or **3** were added into the corresponding wells (200 μL/well, ~1 × 10¹¹ pfu/well). After incubating for 1 h at RT, the unbound phage was rinsed with the washing solution (20 × 300 μL) using the plate washer. Phage remaining on the well was eluted for 9 min at RT by adding 200 μL of glycine elution buffer (0.2 M glycine-HCl, pH 2.2, 1 mg/mL BSA). The elution buffer was transferred into a 1.5-mL microcentrifuge tube and immediately neutralized with 50 μL of 1 M Tris-HCl (pH 9.1). An additional washing solution (200 μL) was added to the well to recover the remaining phage and subsequently combined with the eluate. Up to this point, the selections yielded 15 different eluates (0.45 mL per sample).

3.5.4 Phage amplification and PCR

The eluted phage was amplified separately by adding the eluates (15 × 0.45 mL) into 3 mL of ER2738 culture (1:100 dilution of overnight culture). The phage and bacterial mixtures were incubated for 4.5 h at 37 °C with vigorous shaking. The cultures were centrifuged (15 min, 4700 rpm) at 4 °C

to pellet the bacterial cells. The supernatants (~3.4 mL) containing the amplified phage were poured into a fresh tube. The ssDNA of the amplified phage was extracted using QIAprep spin M13 kit (Qiagen, #27704) according to manufacturer's instructions. To allow multiplexing, 15 reverse barcoded primers were designed with adapters compatible with Ion Torrent sequencing.¹⁵⁴ The library DNA was subjected to PCR amplification with the barcoded primers flanking the variable region. Briefly, the library DNA (15 samples, 50 ng each) was amplified in a total volume of 50 μ L with 1 \times Phusion[®] buffer, 50 μ M each dNTPs, 500 μ M MgCl₂, 1 μ M forward primer, 1 μ M reverse barcoded primer, and one unit Phusion[®] High-Fidelity DNA Polymerase. PCR was performed using the following thermo cycler program: a) 98 $^{\circ}$ C 30 s, b) 98 $^{\circ}$ C 10 s, c) 60 $^{\circ}$ C 20 s, d) 72 $^{\circ}$ C 30 s, e) repeat b)–d) for 34 cycles (total 35 cycles), f) 72 $^{\circ}$ C 5 min, g) 4 $^{\circ}$ C hold. The dsDNA fragments from the PCR were quantified by running at 2% (w/v) agarose gel in Tris-Borate-EDTA buffer at 100 volts for ~45 min using a low molecular weight DNA ladder as a standard (NEB, #N3233S). The dsDNA fragments (15 samples, 40 ng per sample) were pooled together and purified on E-Gel[®] SizeSelect[™] 2% agarose gel (Invitrogen, #G6610-02). The desired band corresponding to 121 bp with reference to the ladder was collected with RNase-free water and the concentration was determined by Qubit[®] Fluorimeter (Invitrogen, #Q32851) using the manufacturer's protocol.

3.5.5 Ion Torrent sequencing

Ion PGM™ Template OT2 200 Kit (Life Technologies) was used to prepare the DNA template for sequencing. Briefly, the pooled and purified dsDNA fragments were hybridized onto Ion Sphere Particles (ISPs) and amplified by emulsion PCR using Ion OneTouch™ 2 System according to manufacturer's protocol. The fraction of ISPs loaded with the DNA template was determined with Qubit® Fluorimeter (Invitrogen) according to the manufacturer's protocol. The ISPs loaded with the DNA template were enriched and deposited in Ion 316™ chip. The DNA sequencing was performed on Ion PGM™ System using Ion PGM™ Sequencing 200 Kit v2. The FASTQ file generated from the sequencing data was processed by in-house MATLAB script that identified the barcodes and constant flanking regions, and extracted the reads of the correct length (24 bp only) corresponding to the TCT(NNK)₇ structure.

3.5.6 Analysis by volcano plot and generation of LOGO

This plot identified sequences isolated from the screen **A** that increased significantly in abundance against sequences isolated from the control selection. The copy number of each sequence is normalized by dividing the copy number by the total number of reads in each replicate. Sequences not observed in a specific replicate were assigned a copy number of zero. For volcano analysis, the ratio of each sequence was calculated by dividing the mean fraction of the particular sequence in the screen **A** by that in the control screen (e.g., **B**, **C**, or **D**). Because the denominator must not be a

zero when taking the ratio, sequences with zero copy number found in all three replicates are assigned with 0.3 copy number before taking the normalization. The significance of the ratio was assessed using one-tailed, unequal variance Student *t*-test. The ratio is considered to be statistically significant if the calculated *p*-value ≤ 0.05 . Only sequences with ratio ≥ 5 and *p*-value ≤ 0.05 were included in the set **A/B**, **A/C**, and **A/D**. The volcano plot was generated using in-house MATLAB script. Sequence LOGO was generated using the MATLAB function *seqlogo* with *StartatValue* set as 2 and *EndatValue* set as 8 (define the range of position to be considered in the sequence).

3.5.7 Surface plasmon resonance spectroscopy

The buffer was sterile-filtered before use. All ligands were pre-dissolved in a 5% volume of DMF followed by the addition of HEPES buffer (50 mM HEPES, 150 mM NaCl, 1 mM CaCl₂, pH 7.2). A solution of ConA tetramer (0.4 mg/mL) was prepared using the same buffer. The measurements were recorded on a BIAcore 2000 instrument using a CM5 chip (carboxylated dextran) as the binding target for ConA. To perform the competitive inhibition, ConA (0.4 mg/mL) was mixed with an equal volume of inhibitor and this mixture (50 μ L) was injected (10 μ L/min). Inhibitor concentrations of 0, 0.0001, 0.0003, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, and 1 mM were tested. Two more data points (3 and 10 mM) were included for inhibition with MeMan. The solution of inhibitor in the absence of ConA was injected each time before the injection of the ConA/inhibitor mixture of the

same concentration. The response values were used for subtraction to account for the bulk effect caused by the inhibitor itself. The chip was regenerated after each injection with the regeneration buffer (6 M guanidinium chloride). Bound ConA response values were assessed during the equilibrium binding portion of the curve (280 s after injection). The corrected response value (R_{\max}) of bound ConA in the absence of inhibitor was set as 0% inhibition. The degree of inhibition by the inhibitor was calculated with the equation $R_{\max} - R_{\text{inh}}/R_{\max}$, where R_{inh} is the corrected response value given by the bound ConA in the presence of certain concentration of inhibitor.

3.5.8 Isothermal titration calorimetry

ITC experiments were performed using a Microcal VP-ITC instrument. Ligands were pre-dissolved in a small amount of DMF followed by the addition of HEPES buffer (50 mM HEPES, 150 mM NaCl, 1 mM CaCl_2 , pH 7.2). The final solvent was HEPES buffer containing 2% (v/v) DMF. In the cases where ligands have poor solubility in the prepared buffer, up to 5% (v/v) DMF was used. The ConA solution was prepared with the same buffer as the ligand. All solutions were degassed with MicroCal ThermoVac unit prior to use. All titrations were carried out at 30 °C. An initial injection of 2 μL followed by a total of 41 injections of ligand solution (7 μL) were added at intervals of 4 min into the solution of ConA (cell volume = 1.44 mL) while stirring at 300 rpm. Typically, the initial concentrations of ConA and ligands were 0.05–0.30 mM and 2 mM respectively, unless otherwise specified. The

quantity $c = K_a M$, where M is the initial macromolecule concentration, is of importance in ITC. All experiments were performed with c values in the range of $1 < c < 100$, except for the ligands with $K_a \leq 1000 \text{ M}^{-1}$, where preparation of ConA with concentration $> 1 \text{ mM}$ is difficult due to solubility issue. In the cases where heat of dilution is significant, titration data obtained by making an identical injection of ligand into the buffer without ConA was subtracted from the titration data obtained in the presence of ConA. The data point produced by the first injection was discarded prior to curve fitting to account for the diffusion effect during the equilibration process. The experimental data were fitted to a non-interacting one-site binding model using Origin software supplied by Microcal, with ΔH (enthalpy change), K_a (association constant) and n (number of binding sites per monomer) as adjustable parameters. The free energy change (ΔG) and entropy contributions ($T\Delta S$) were determined from the standard equation: $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$, where T is the absolute temperature and $R = 1.987 \text{ cal mol}^{-1} \text{ K}^{-1}$. For ITC measurements of ligand binding to LcH or PSA, exactly the same protocol was used.

3.5.9 1D STD-NMR experiments

The experiments were conducted on Agilent/Varian VNMRS 600 MHz spectrometer at a probe temperature of 300 K. The NMR samples were prepared by pre-dissolving the ligand in a small amount of DMSO- d_6 (14 μL), followed by the addition of ConA in deuterated PBS (686 μL). Deuterated PBS

was prepared from 1× PBS (pH 7.4) by two cycles of lyophilization and redissolution in D₂O. The final samples contained ligand/ConA in a ratio of 40:1 (2 mM ligand, 0.05 mM ConA). The NMR experiments involved the selective saturation of protein resonances at 0.2 ppm (30 ppm for reference spectra) using a cascade of 20 Gaussian-shaped pulses (50-ms duration, 1-ms delay between each pulse) resulting in a total saturation time of 1.02 s. The WATERGATE W5 sequence¹⁵⁵ was used to suppress the residual HDO signal. A 30-ms $T_{1\rho}$ filter was applied to suppress the protein background. STD-NMR spectra were obtained by subtracting the saturated spectra from the reference spectra *via* phase cycling. The integral regions of the reference spectra were copied to the STD-NMR spectra to guarantee identical boundaries and unbiased ratio of the particular integrals. Relative STD effects were calculated according to the equation $E_{STD} = (I_0 - I_{sat})/I_0 = I_{STD}/I_0$ by comparing the intensity of the signals in the STD-NMR spectrum (I_{STD}) with intensity of the signals in a reference spectrum (I_0). A control STD experiment using the free ligand (Man-WYD) was performed under the same experimental conditions to verify true ligand binding. No signal was present in the STD-NMR spectra, indicating that the effects observed in the presence of the protein were due to true saturation transfer with negligible artifacts. The reference spectra and STD-NMR spectra of Man-WYD and Man-NL-WYD are shown in Appendix B-6.

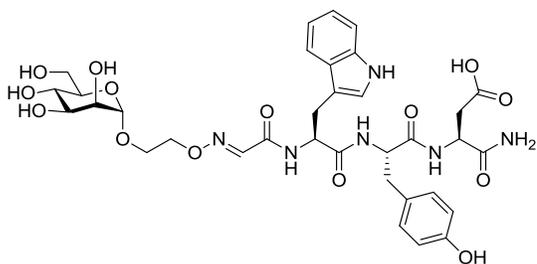
3.5.10 Competitive binding assay

The tetrameric extracellular domain of dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) was a gift from Professor Kurt Drickamer (Imperial College, London). The competitive binding assay involves the binding of immobilized lectins, i.e., ConA or DC-SIGN, to horseradish peroxidase (HRP), a glycoprotein containing trimannoside epitope, under the competition of the studied inhibitors, i.e., MeMan, Man-WYDLF, or Man3-X. The buffer used in the experiment was a solution of 50 mM MOPS, 150 mM NaCl, and 2mM CaCl₂ (pH 7.4). A solution of ConA or DC-SIGN (10 µg/mL) dissolved in the buffer was used to coat a polystyrene plate (Costar #3369) to have a final volume of 50 µL/well. The plate was sealed with a membrane and kept in the fridge overnight. In a separate non-binding surface plate (Corning #3641), a 3-fold serial dilution was performed for the solutions of the inhibitor dissolved in the buffer. The diluted solutions were then mixed with an equal volume of a solution of HRP (2 µg/mL) dissolved in the same buffer. The coating solution of lectin was aspirated using 405™ Touch Microplate Washer (BioTek) and subsequently washed with the washing solution (10 × 300 µL, the same buffer containing 0.1% (v/v) Tween-20). The mixture of the inhibitor and HRP probe was transferred accordingly onto the plate coated with the lectin to have a final volume of 50 µL/well. The plate was incubated at RT for 1 h. Then, the plate was washed with the washing solution (10 × 300 µL) and the TMB substrate (50 µL) was added to each well. After 5 min incubation for

ConA-coated well or 10 min incubation for DC-SIGN-coated well, 1 M phosphoric acid (50 μ L) was added to quench the colorimetric assay. The color developed was read at 450 nm with a 96-well plate reader. The data was fitted with a logistic function using Origin software to determine the half maximal inhibitory concentration (IC_{50}) of the studied inhibitor.

3.5.11 Examples for the synthesis of Man-peptide conjugates

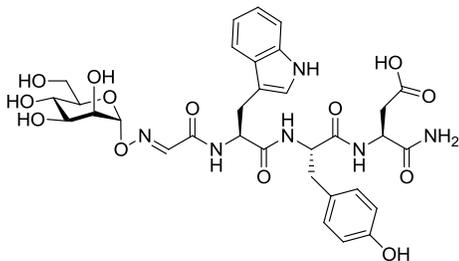
Man-WYD (L20)



SYWD (5.7 mg, 10 μ 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 sodium periodate (2.6 mg, 12 μ mol, 1.2 eq.). The reaction mixture was incubated for 10 min at RT. To quench the oxidation, the solution was added to glutathione (37 mg, 120 μ mol, 12 eq.) and mixed rapidly to ensure the dissolution of glutathione. After incubation for 10 min at RT, 2-(aminooxy)ethyl α -D-mannopyranoside (2.6 mg, 11 μ mol, 1.1 eq.) dissolved in 200 mM anilinium acetate (0.25 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 purified on 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^{-1} to yield the

product as a white fluffy powder (5.0 mg, 66% isolated yield) after lyophilization. The purity of the product was determined with an analytical RP-HPLC system (flow rate: 1 mL min⁻¹) using a gradient of solvent A and solvent B. ¹H NMR (600 MHz, deuterated PBS + 0.1% (v/v) DMSO-*d*₆) δ = 7.61 (d, *J* = 7.9 Hz, 1 H), 7.56 (s, 1 H), 7.50 (d, *J* = 7.9 Hz, 1 H), 7.24 (t, *J* = 7.9 Hz, 1 H), 7.20 – 7.13 (m, 2 H), 6.96 (d, *J* = 8.1 Hz, 2 H), 6.74 (d, *J* = 8.1 Hz, 2 H), 4.88 (br. d, 1 H), 4.69 (t, *J* = 6.9 Hz, 1 H), 4.50 – 4.42 (m, 2 H), 4.41 – 4.33 (m, 2 H), 3.98 – 3.91 (m, 2 H), 3.83 - 3.75 (m, 3 H), 3.75 – 3.70 (m, 1 H), 3.66 (t, *J* = 9.6 Hz, 1 H), 3.62 – 3.56 (m, 1 H), 3.29 – 3.17 (m, 2 H), 2.92 (dd, *J* = 7.4, 13.9 Hz, 1 H), 2.79 (dd, *J* = 7.4, 13.9 Hz, 1 H), 2.60 (dd, *J* = 6.7, 16.0 Hz, 1 H), 2.50 (dd, *J* = 6.7, 16.0 Hz, 1 H); HRMS (ESI) calcd for C₃₄H₄₁N₆O₁₄ [M-H]⁻ *m/z* = 757.2686, found 757.2691. (see Appendix B-11 for NMR spectrum)

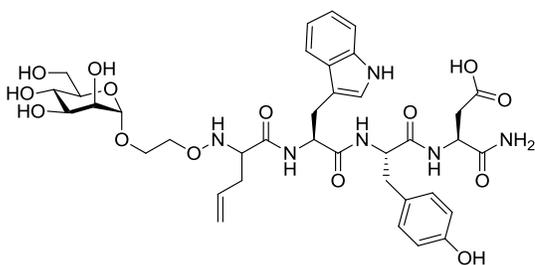
Man-NL-WYD (L20.3)



The synthesis and purification are identical to that of Man-peptides, other than the use of *O*-α-D-mannopyranosyl oxyamine instead of 2-(aminoxy)ethyl α-D-mannopyranoside during the oxime ligation. The product was obtained as a white fluffy powder (4.3 mg, 60% isolated yield): ¹H NMR (500 MHz, D₂O + 0.1% (v/v) DMSO-*d*₆) δ = 7.65 (s, 1 H), 7.60 (d, *J* = 7.9 Hz, 1 H), 7.49 (d, *J* = 7.9 Hz, 1 H), 7.24 (t, *J* = 7.9 Hz, 1 H), 7.21 (s, 1 H), 7.16

(t, $J = 7.9$ Hz, 1 H), 6.93 (d, $J = 8.4$ Hz, 2 H), 6.74 (d, $J = 8.4$ Hz, 2 H), 5.53 (br. d, 1 H), 4.68 (t, $J = 7.1$ Hz, 1 H), 4.50 (t, $J = 7.2$ Hz, 1 H), 4.42 (t, $J = 7.2$ Hz, 1 H), 4.14 (dd, $J = 1.9, 3.2$ Hz, 1 H), 3.90 – 3.72 (m, 4 H), 3.68 – 3.60 (m, 1 H), 3.29 – 3.17 (m, 2 H), 2.86 – 2.72 (m, 3 H), 2.61 (dd, $J = 7.2, 17.0$ Hz, 1 H); HRMS (ESI) calcd for $C_{32}H_{37}N_6O_{13}$ $[M-H]^-$ $m/z = 713.2424$, found 713.2430. (see Appendix B-12 for NMR spectrum)

Man-allyl-WYD (L20.4)



To a mixture of allyl bromide (12 μ L, 132 μ mol, 20 eq.) and indium (3 mg, 26 μ mol, 4 eq.), a solution of Man-WYD **L20** (5 mg, 6.6 μ mol, 1 eq.) dissolved in DMF/H₂O/MeOH (200 μ L, 1:1:2) was added. The mixture was agitated for 2 h at RT, after which, the solid was filtered. The filtrate was injected into a semi-preparative RP-HPLC system. The reaction mixture was purified on 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^{-1} to yield the product as a white fluffy powder (2 mg, 38% isolated yield): ¹H NMR (500 MHz, D₂O + 0.1% (v/v) DMSO-*d*₆) $\delta = 7.62$ (d, $J = 7.7$ Hz, 1 H), 7.48 (d, $J = 8.3$ Hz, 1 H), 7.26 – 7.20 (m, 2 H), 7.15 (t, $J = 7.7$ Hz, 1 H), 7.01 (d, $J = 8.4$ Hz, 2 H), 6.79 (d, $J = 8.4$ Hz, 2 H), 5.49 – 5.37 (m, 1 H), 5.03 – 4.92 (m, 2 H), 4.77 (br. d, 1 H), 4.71 – 4.65 (m, 1 H), 4.56 (t, $J = 6.4$ Hz, 1 H), 4.46 (t,

J = 7.2 Hz, 1 H), 3.89 – 3.79 (m, 2 H), 3.75 – 3.69 (m, 2 H), 3.69 – 3.51 (m, 6 H), 3.51 – 3.44 (m, 1 H), 3.24 (dd, J = 7.2, 14.9 Hz, 1 H), 3.13 (dd, J = 7.2, 14.9 Hz, 1 H), 2.93 – 2.77 (m, 3 H), 2.71 – 2.65 (m, 1 H), 2.12 (t, J = 6.9 Hz, 2 H); HRMS (ESI) calcd for C₃₇H₄₇N₆O₁₄ [M–Na][–] m/z = 799.3156, found 799.3163. (see Appendix B-13 for NMR spectrum)

Chapter 4: Phage selection of DC-SIGN inhibitors

4.1 Introduction

Dendritic cells (DCs) play a major role in both innate and adaptive immune responses. DCs efficiently capture and process antigens. Upon maturation and migration to lymph nodes, they present the antigen fragments to naïve T-cells, leading to their clonal selection and differentiation. The induced cytotoxic T-cell can recognize and eliminate cancer cells carrying the antigen. Owing to these properties, DCs are the first-in-class natural targets for antigen delivery.^{156,157} Studies have shown that presentation of antigens by DC increases substantially, by at least 100-fold, when a specific receptor is engaged, as opposed to non-targeted delivery.¹⁵⁸⁻¹⁶⁰ Upon binding to the antigen, the uptake receptors on DC internalize and translocate the antigen to the endosome for downstream processing and presentation. Most of these uptake receptors recognize carbohydrate-based ligands.

Among them, dendritic cell-specific ICAM3 grabbing non-integrin (DC-SIGN) is exceptional, because it is predominantly expressed on DCs,¹⁶¹ and involved in antigen uptake.¹⁶² DC-SIGN recognizes both high-mannose oligosaccharides on viruses and Lewis^x-containing self glycans. It belongs to a superfamily of C-type lectins that require calcium ion for binding to glycans.¹⁶³ Delivery of antigens to DCs *via* an endocytic pathway mediated by DC-SIGN represents a promising strategy for immunotherapy. Multiple

studies have shown that antigens modified with a DC-SIGN-binding glycan^{164,165} or antigens encapsulated in Lewis^X-modified liposomes¹⁶⁶ lead to efficient antigen uptake and increased T-cell responses. However, these glycans share a common motif recognizable across various glycan-binding receptors, which might pose a selectivity issue. On the other hand, HIV viruses have been shown to exploit the binding to DC-SIGN to promote efficient infection of T-cells.¹⁶⁷ Blocking the interaction between DC-SIGN and high-mannose oligosaccharides decorating on the capsid proteins of HIV viruses could be considered as a novel anti-viral strategy. For example, Penades and co-workers have shown that oligomannoside-functionalized gold nanoparticles are able to inhibit the *trans*-infection of T-cells by HIV at nanomolar concentrations.¹⁶⁸ The Bernardi group demonstrated that dendrimers decorated with trimannoside mimics can inhibit the same process, albeit at micromolar concentrations.¹⁶⁹

4.2 Previous works on DC-SIGN inhibitors

Enormous efforts have been made in the past decade to design potent and selective DC-SIGN agonists/antagonists. Novel designs based on multivalent scaffolding has successfully achieved DC-SIGN antagonists with activity at nanomolar concentrations (see review^{31,170}). However, the design of *monovalent* ligands for DC-SIGN has been extremely challenging. None of the monovalent ligands discovered so far have dissociation constants (K_D) or half-maximal inhibitory concentrations (IC_{50}) crossing the nanomolar barrier

(*i.e.*, less than 1 μM). Most of these compounds only exhibit moderate enhancement of affinity relative to the monosaccharide.^{124,171-176} The most potent monovalent inhibitors, to our knowledge, are quinoxalinone-based small molecules discovered by the Kiessling lab using high-throughput screening.^{128,177} These compounds do not bear any carbohydrate functionality, yet they exhibit >1000-fold enhancement of inhibitory activity relative to *N*-acetyl-mannosamine. However, the mechanism of their inhibition remains unclear. In my recent personal communication with Laura Kiessling, she suggested that these inhibitors bind to an allosteric site and on-going investigations are in progress in her lab to elucidate the binding site. Binding to non-glycans binding site is a common outcome for screening of GBPs with non-glycan ligands.¹³⁷

The design of more potent monovalent ligand that target glycan-binding sites is important. Every increment of affinity in a *monovalent* ligand has a measurable impact on the affinity of a *multivalent* ligand,¹⁷⁸ and might permit the use of constructs with lower valency. Most importantly, *monovalent* ligands of higher affinity could impart greater selectivity toward the target.¹⁷⁸ Here, I sought to discover *monovalent* glycopeptide-based inhibitors for DC-SIGN because peptides can be easily expressed and incorporated into protein antigens to target them to DC.

4.3 Results and discussion

Previously, we have developed genetically encoded fragment-based discovery (GE-FBD) for the identification of low micromolar inhibitors for a model target (ConA) using glycopeptide libraries displayed on M13 phage (Chapter 3). The library contains a monosaccharide fragment that serves as an anchor at a glycan-binding site and assists the search for synergistic peptide sequences binding to previously unknown secondary site(s). The GE-FBD demonstrated the potential use of the technology for mapping of “hot spots” that reside in the vicinity of the glycan site. In this chapter, I describe application of this technology to DC-SIGN. I aimed to select for inhibitors of higher potency than the monosaccharide ligands of DC-SIGN and identify whether there are any potential secondary sites located nearby its mannose-binding site.

4.3.1 Selection of glycopeptide libraries

GE-FBD begins with the synthesis of a phage-displayed glycopeptide library. First, a serine-terminated naïve library (SX₇, where X = any 20 natural amino acid) with a diversity of $\sim 10^8$ was oxidized with sodium periodate as previously described.^{141,179} The resulting aldehyde provides a unique handle to which a fragment, such as mannose-hydroxylamine derivative (Man) or aminoxy-biotin (AOB), could be ligated (Figure 4-1A). The latter fragment is used to monitor the efficiency of the ligation in a phage library ($\sim 40\text{--}60\%$ of the library is chemically modified)⁷⁷ and serves as a

control in the selection. Biopanning of modified libraries, or enrichment of a specific binder, involved parallel incubation of Man-X₇ or AOB-X₇ libraries with plate-immobilized DC-SIGN, or a blank well, all in triplicate (Figure 4-1B). The last two experiments serve as negative controls.

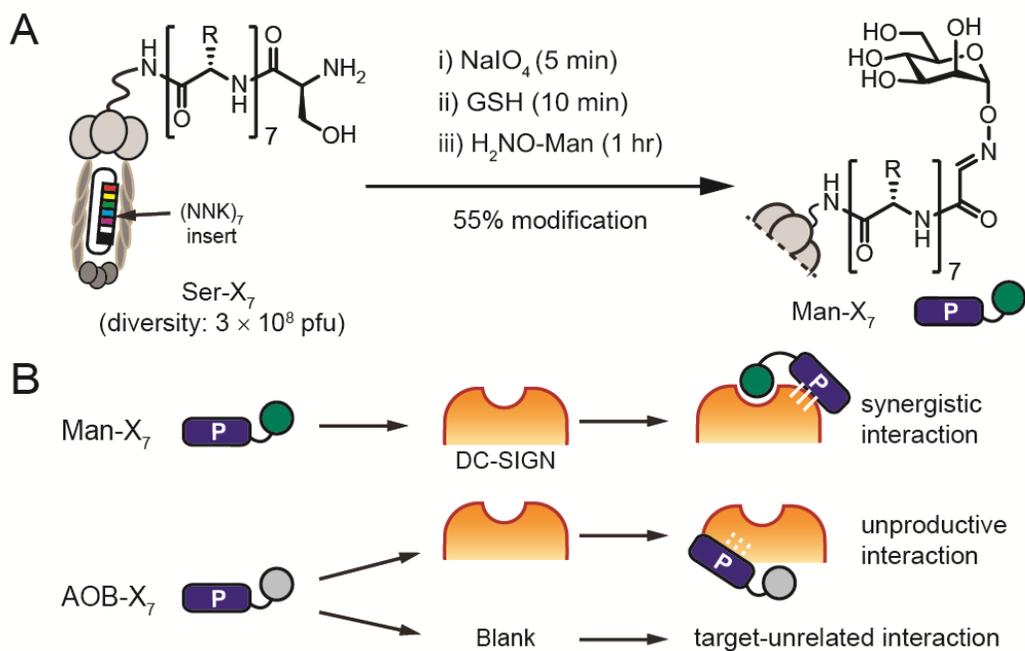


Figure 4-1. Illustration of generation of glycopeptide library and its selection.

(A) Modification of serine-terminated peptide library displayed on M13 phage yield glycopeptide library (Man-X₇). (B) Biopanning of Man-X₇ or AOB-X₇ on plate-immobilized DC-SIGN, and AOB-X₇ on blank well were performed in parallel during each round of selection.

I was able to recover about 50-fold more phage titers from DC-SIGN-coated wells than from the blank wells (Figure 4-2A). The result implied significant binding of phage libraries to the target and yet background binding to the polystyrene surface is negligible. The intrinsic binding of the peptide libraries to DC-SIGN regardless of the fragment attached to the N-terminus of the peptides indicated that peptides could bind to DC-SIGN at

multiple sites. This finding can be explained by a tetrameric structure of the extracellular domain (ECD) containing a repeating hydrophobic neck region. Such binding of peptides outside of the carbohydrate-recognition domain (CRD) have plagued many previous genetically encoded selections leading to identification of peptides that bind outside of CRD.^{132,133,136,137}

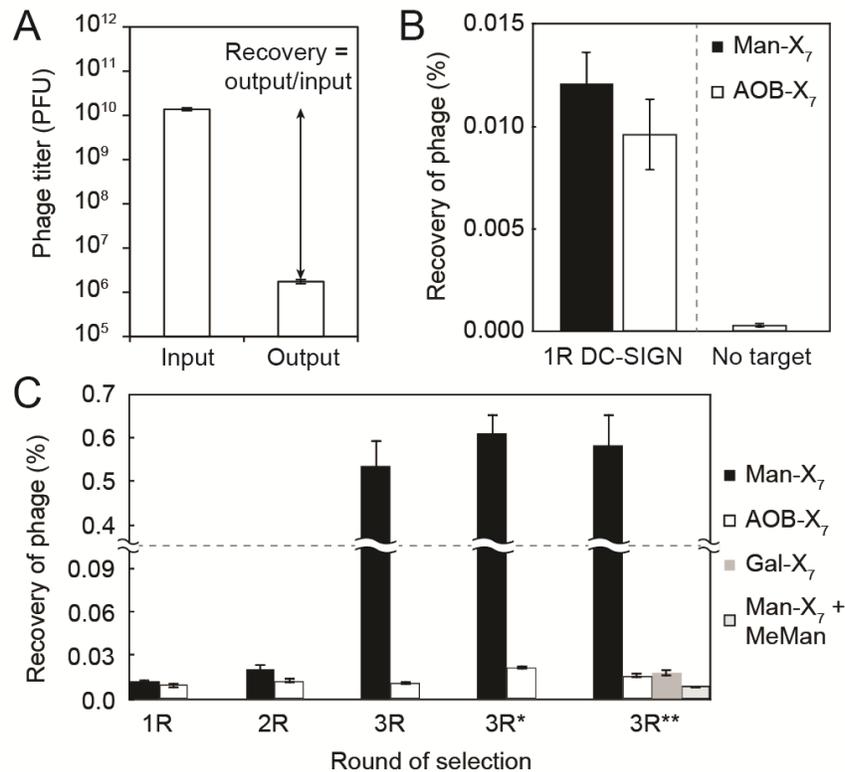


Figure 4-2. The recovery of phage from biopanning.

(A) Recovery of phage is the mean of eluted phage titers divided by the mean of input phage titers for biopanning ran in triplicate. Error bars represent the propagation of error for calculating the recovery. (B) Recovery of phage after the first round of biopanning. (C) Recovery of phage after three rounds of biopanning on DC-SIGN. At the third round of selection (3R), a library modified with Man (Man-X₇) was significantly enriched, whereas the same library modified with biotin or Gal was not. Panning of Man-X₇ in the presence of soluble MeMan also abrogated the enrichment. 3R, 3R* and 3R** are independent biopanning performed on separate days.

To deplete the peptides that have the potential to bind outside of CRD, I incubated the naïve library with the target prior to the first round of selection. The unbound phage library remaining in the supernatant was then modified with the Man fragment to enrich CRD-binding glycopeptides through subsequent rounds of selection. Indeed, at the third round of selection (see 3R in Figure 4-2B), I observed significant amounts of phage recovered from the mannose-modified library as opposed to the biotin-modified library. The percentage of recovery increased by ~50-fold from 0.011 ± 0.001 % (AOB-X₇) to 0.53 ± 0.06 % (Man-X₇). Such enrichment was reproducible, at least three times, when the biopanning was repeated on separate days (see 3R* and 3R** in Figure 4-2B). I further demonstrated that the enrichment was the result of the specific interaction of mannose with DC-SIGN. For example, modification of the enriched library with an irrelevant glycan, *e.g.*, galactose-hydroxylamine derivative (Gal) and its selection exhibited a factor of 50-fold decrease in recovery. Although Man-X₇ and Gal-X₇ differ by their anomeric positions and C2/C4 epimeric positions, mannose is essential for the recognition. Furthermore, the enrichment for Man-X₇ at 3R** was completely abrogated in the presence of 500 mM methyl α -mannopyranoside (MeMan) added during the biopanning (Figure 4-2B). These results suggested that the observed enrichment of the phage library is achievable if, and only if, the library is modified with an appropriate fragment anchor, *i.e.*, mannose, and the binding of this library to DC-SIGN could be interrupted by inhibition. Next, I recovered phage clones from all

three rounds of biopanning, including the controls, and subjected the phage to amplification, DNA extraction, and PCR of DNA with barcoded primers as described in section 4.5.5. Pooled amplicons were sequenced using Illumina NextSeq platform and analyzed using the volcano plot analysis employed in section 3.5.6 to identify the potential hits.

4.3.2 Validation of selected hits

To validate the inhibitory activity of the potential hits, I randomly selected six glycopeptides from the list of potential hits, synthesized them, and tested for their ability to compete with trimannoside-containing horseradish peroxidase (Sigma #77332) for binding to the plate-immobilized DC-SIGN ECD (Figure 4-3A). All synthetic glycopeptides have relative IC_{50} (rIC_{50}) in the range of 0.27–0.48 when compared to that of MeMan ($rIC_{50} = 1$, Figure 4-3B). In other words, their IC_{50} values are 2- to 4-fold lower than that of MeMan ($IC_{50} = 0.7 \pm 0.1$ mM). Some of these synthetic inhibitors have comparable activity to 6 azidohexyl α -Man-(1→3)-[α -Man-(1→6)]- α -Man (Man3-X, $rIC_{50} = 0.31$). To this end, the phage selection has resulted in a new class of inhibitors with comparable potency to Man3 derivative but only with moderate enhancement relative to MeMan (2–4 fold). These results are similar to the experience of other groups working on DC-SIGN,^{124,171-176} where the efforts to design highly potent *monovalent* glycan-based inhibitors for DC-SIGN were met with some challenges.

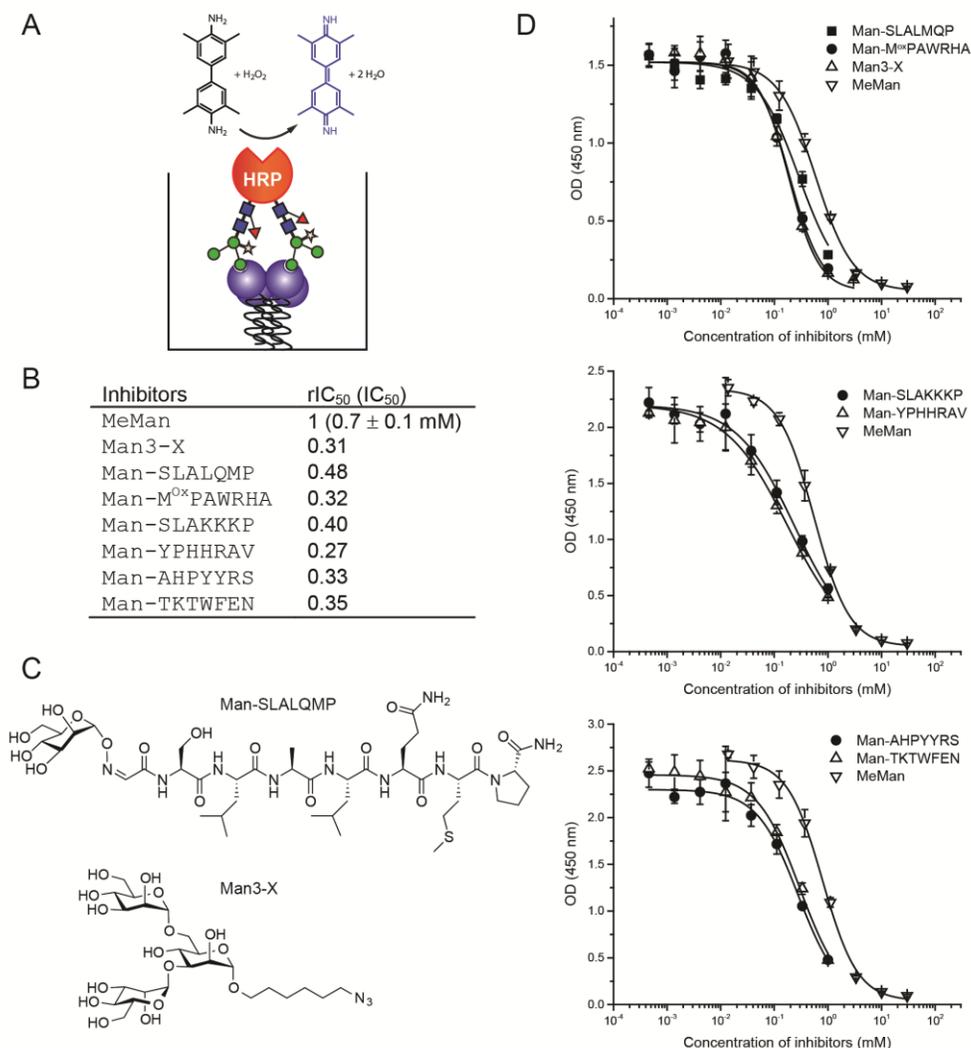


Figure 4-3. Inhibition assay and validation of hits.

(A) Scheme of the assay that measures the ability of soluble inhibitors to disrupt the binding of trimannoside-decorated HRP to plate immobilized DC-SIGN ECD. (B) Summarized data of relative IC₅₀ (rIC₅₀) for glycopeptides (Man-X₇) or 6 azidoheptyl α -Man-(1 \rightarrow 3)-[α -Man-(1 \rightarrow 6)]- α -Man (Man3-X). In each assay, methyl α -mannopyranoside (MeMan) was used as an internal control. rIC₅₀ was calculated as the ratio of IC₅₀ of test compound to the IC₅₀ of MeMan measured on the same plate. (C) Structures of glycopeptide and Man3-X. (D) Raw data of the inhibition profiles. The absorbance at 450 nm (OD = optical density) is an average of three independent wells. Errors are one standard deviation. The data was fitted with logistic function using Origin software.

4.3.3 Validation of the reported DC-SIGN inhibitor

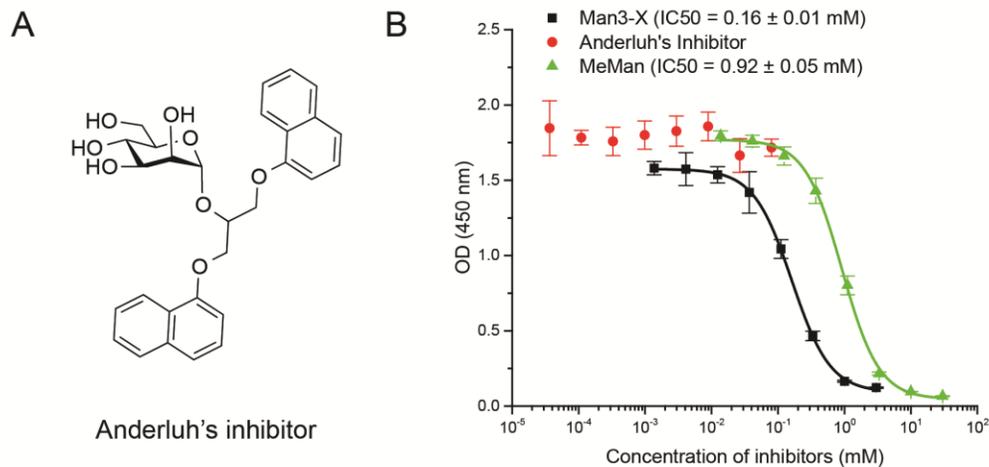


Figure 4-4. Validation of the inhibitor reported in the literature.

(A) Chemical structure of the inhibitor with IC_{50} value $40 \mu\text{M}$ reported in by the Anderluh group.¹⁸⁰ (B) In our assay, I did not observe any inhibitory effect of the compound at concentrations up to $80 \mu\text{M}$. Positive controls, Man3-X and MeMan, studied along with the inhibitor exhibited reproducible results.

To compare our inhibitors to other reported DC-SIGN inhibitors, I synthesized a mannose-based inhibitor designed by Anderluh and co-workers using the computational docking tool FlexX (Figure 4-4).¹⁸⁰ In their report, the inhibitor was shown to block the binding of HIV gp120 to immobilized DC-SIGN ECD with an IC_{50} value of $40 \mu\text{M}$.¹⁸⁰ As a result, this inhibitor exhibited 75-fold better potency than L-fucose ($IC_{50} = 3 \text{ mM}$ on the same assay) and the authors claimed that it is the most potent mannose-based monovalent inhibitor available to date. However, in our hands, we could not reproduce the activity of the inhibitor (Figure 4-4). The inhibitor has poor solubility in aqueous buffer. No inhibitory effect was observed for concentrations tested up to $80 \mu\text{M}$ (attempts to dissolve more compound was

unsuccessful even in the presence of 2% DMSO). This discrepancy with the reported results could be due to artifact of insoluble aggregates or the use of different assay. To confirm our assay is valid and suitable for the studies, we will be collaborating with groups of John Klassen (Alberta) and/or Christoph Rademacher (Max Planck Institute, Postdam) to evaluate these inhibitors using orthogonal assays such as an ESI-MS binding assay^{149,181} or a cell-based inhibition assay. Furthermore, I will be contacting Kiessling to request her non-carbohydrate inhibitors of DC-SIGN to complement our study and serve as a positive control.

4.3.4 Validation of FTMap for glycan-binding proteins

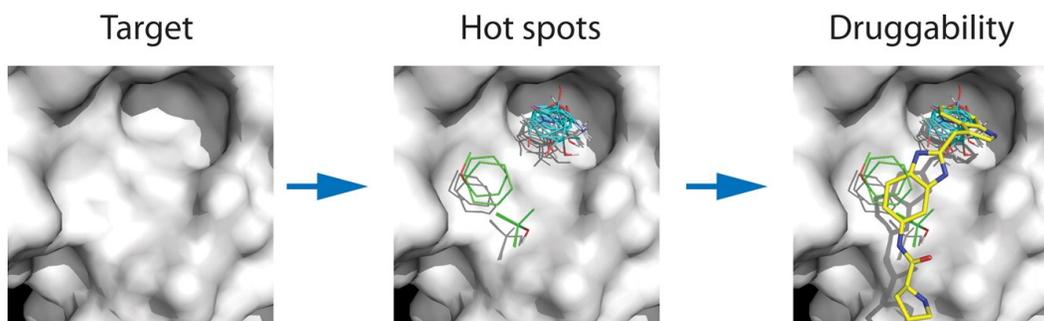


Figure 4-5. Computational solvent mapping (FTMap).

The figure¹⁸² is reprinted with permission from the publisher. Copyright © 2015, American Chemical Society.

We hypothesized that the unsuccessful selection could originate from the lack of fruitful “hot spots” in close proximity to the glycan site. To test this hypothesis, I employed FTMap, a validated computational solvent mapping server developed by Kozakov *et. al.*,¹⁸²⁻¹⁸⁴ to identify binding energy hot spots of DC-SIGN. FTMap has previously been used to identify regions of proteins, so-called “consensus sites”, that bind to multiple probe molecules *in*

silico. These consensus sites have been shown to correlate with regions of known sites for binding of small-molecule fragments (Figure 4-5).^{182,183}

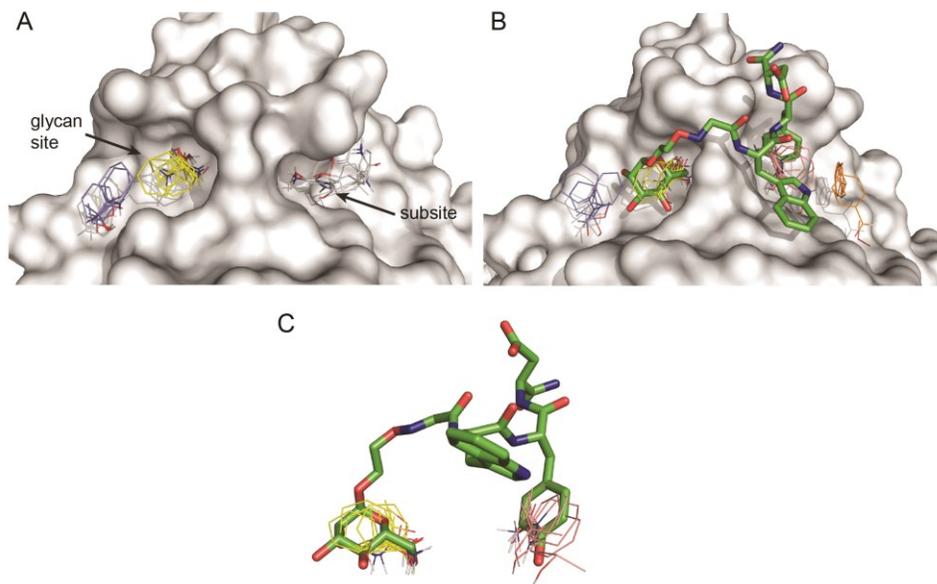


Figure 4-6. Mapping of the locations of potential binding fragments near the glycan-binding site of ConA using FTMap.

Probe clusters binding at consensus sites identified by FTMap are shown as thin sticks. Oxygen and nitrogen atoms are in red and blue, respectively. (A) Mapping of apo ConA (PDB: 1NLS). (B) Mapping of ConA co-crystallized with Man-WYD (green, PDB: 4CZS). FTMap accurately identified the glycan site and subsites for both ligand-free and ligand-bound ConA. (C) Mannose and Tyr side chain of Man-WYD inhibitor previously identified by GE-FBD overlap well with 16 yellow and 16 salmon probes identified by FTMap.

As FTMap has not been validated for glycan-binding proteins, I first used ConA and galectin-3 to determine the effectiveness of FTMap for the identification of hot spots in GBPs. In ConA, the approach accurately identified the glycan-binding site as the main hot spot. It also identified a secondary hot spot located in close vicinity to the glycan site (Figure 4-6A). We have previously validated this hot spot by co-crystallization of ConA with Man-WYD—a low micromolar inhibitor discovered with GE-FBD approach.

FTMap correctly predicted the overlap of probe clusters with Tyr side chain of the inhibitor (Figure 4-6B–C). In galectin-3, FTMap successfully outlined the glycan-binding site with high-ranking probe clusters and revealed several subsites as potential hot spots (Figure 4-7). This prediction coincides with the location of the binding sites for several known galectin-3 inhibitors.¹⁸⁵

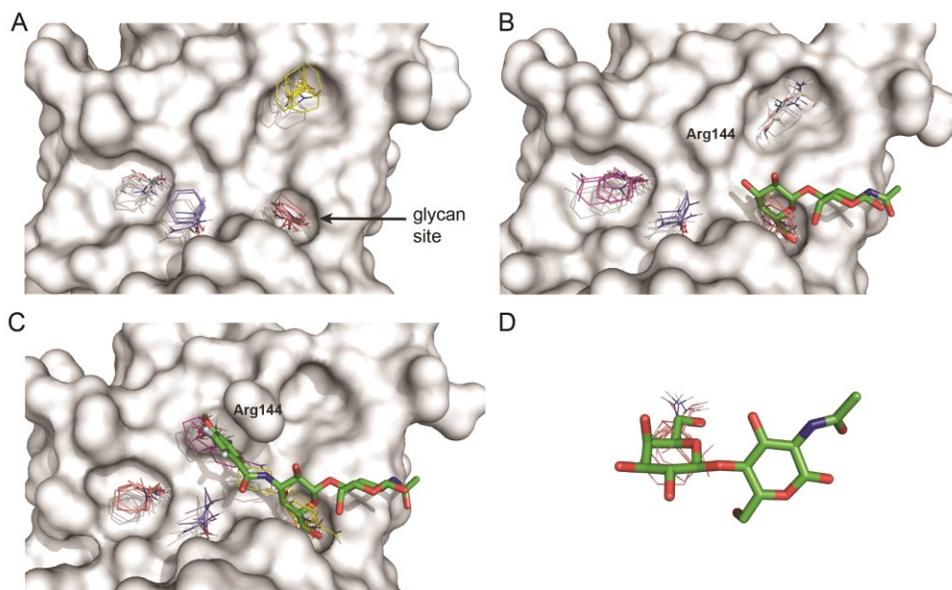


Figure 4-7. Mapping of the locations of potential binding fragment near the glycan-binding site of galectin-3 using FTMap. Probe clusters binding at consensus sites identified by FTMap are shown as thin sticks. Oxygen and nitrogen atoms are in red and blue, respectively. (A) Mapping of apo galectin-3 (PDB: 3ZSL). (B) Mapping of LacNAc–galectin-3 complexes (PDB: 1KJL). FTMap successfully identified the glycan-binding site and subsites in both ligand-free and ligand-bound galectin-3. (C) Mapping of galectin-3 co-crystallized with a LacNAc-based inhibitor (PDB: 1KJR). Though the glycan site was accurately pinpointed, a new subsite was revealed due to a significant conformational change of Arg144. The change was induced by the favorable cation- π interaction between the guanidino group of Arg144 and the aromatic group of inhibitor. These results suggest that a conformational change might expose new hot spots but this hot spot is missed by FTMap when mapping an apo protein. (D) The galactose of LacNAc shown in panel B overlaps well with 15 cluster probes in salmon color.

4.3.5 FTMap analysis of DC-SIGN

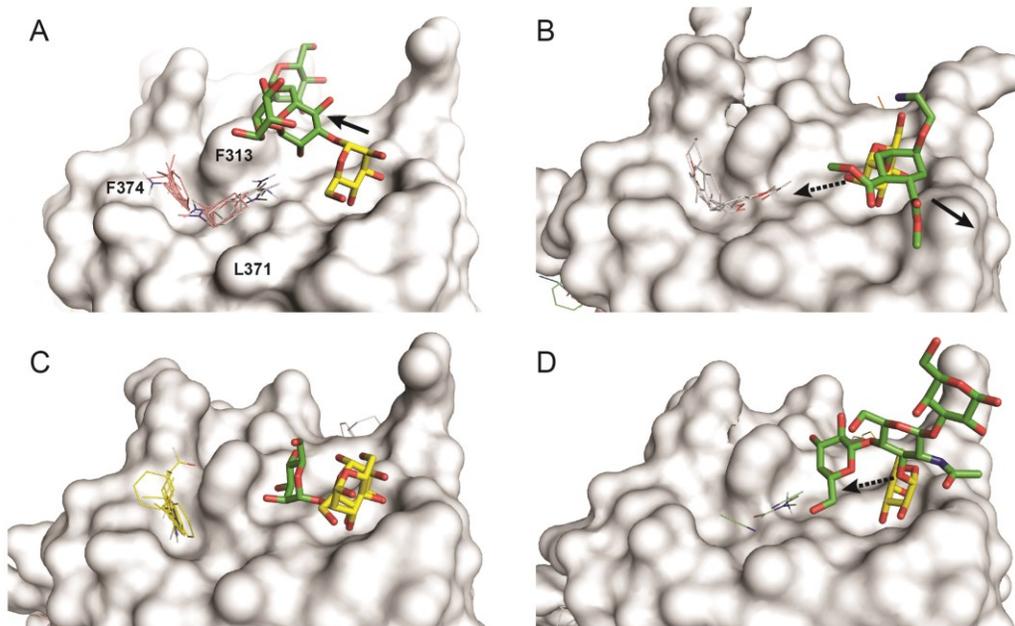


Figure 4-8. Mapping of the locations of potential binding fragment near glycan-binding site of DC-SIGN using FTMap.

(A) Mapping around the Man4-bound structure (PDB: 1SL4). (B) Mapping around the pseudo-1,2-mannobioside-bound structure (PDB: 2XR5). (C) Mapping around the Man2-bound structure (PDB: 2IT6). (D) Mapping around the Lewis^X derivative-bound structure (PDB: 1SL5). In (A–D) the glycan site of DC-SIGN contains the mannose (A–C) or fucose (D) residue (yellow stick) chelating the calcium ion. Probe clusters binding at consensus sites identified by FTMap are shown as thin sticks. The solid arrows illustrate the direction into which the peptide aglycones can be pointing in Man-X₇ conjugates with α -anomeric linkage between Man and *N*-terminus of the peptide. Dashed arrows represent potential attachment points for the access of library to the groove identified with FTMap. Oxygen and nitrogen atoms are in red and blue respectively.

After the validation of FTMap, I used it to analyze several known DC-SIGN structures co-crystallized with four different ligands— α -Man-(1 \rightarrow 3)-[α -Man-(1 \rightarrow 6)]- α -Man-(1 \rightarrow 6)- α -Man (Man4),¹⁸⁶ pseudo-1,2-mannobioside,¹⁸⁷ α -Man-(1 \rightarrow 2)- α -Man (Man2),¹⁸⁸ and a Lewis^X derivative¹⁸⁶ (Figure 4-8). The calcium ion, which is essential for chelating the vicinal diol of

mannose or fucose, was removed by the server before the mapping (retention of the ion was tried, but an error emerged during the mapping). It is possible that removal of calcium was the reason why no probe cluster was found at the glycan site. I found that the top-ranking consensus sites were >15 Å away from the glycan-binding site, and thus are not useful for the design of inhibitors of a reasonable size. Large entropic penalties might arise due to the long linker necessary to bridge the sites. The landscape around the glycan-binding site of DC-SIGN in the putative glycopeptide-binding region is topologically flat and featureless. This analysis could explain the modest affinities of these ligands (*i.e.*, K_D or IC_{50} values of >100 μ M). However, a groove surrounded by Phe313, Leu371, and Phe374 was identified by FTMap in three out of four structures as a prospective hot spot (Figure 4-8A-C), although the ranking of this site was relatively low compared to other “consensus sites” of DC-SIGN identified by FTMap. Coincidentally, part of this groove has been previously identified by the Bernardi lab using the GRID program¹⁸⁹ and used in the design of fucosylated inhibitors.¹²⁴ FTMap analysis suggested that the mannose fragment used in this work might not be an optimal anchor for GE-FBD. Structural data showed that mannose has two different modes for binding to DC-SIGN (compare Figure 4-8A and 4-8B). One of the binding modes, which is also the major binding mode of single mannose,¹⁸⁷ clearly steers the aglycone away from the groove in the opposite direction (see solid arrow in Figure 4-8B). Mannose in the other configuration (Figure 4-8A) has better chance to position the peptide

fragment toward the groove but a steric clash with the bulky side chain of Phe313 is inevitable (see solid arrow in Figure 4-8A). I believe that peptide fragments attached to the α -anomeric position of fucose (see dashed arrow in Figure 4-8D) or the 2-OH group of mannose (see dashed arrow in Figure 4-8B) might have a better chance of reaching and engaging with the groove.

4.4 Conclusion

In summary, through the use of GE-FBD technology from 10^8 mannose-peptide conjugates, we have discovered a novel class of inhibitors for DC-SIGN, albeit with moderate enhancement in activity relative to a monosaccharide. The modest affinity correlates well with reports of other groups^{171,174,176} in the design of mannose-based inhibitors for DC-SIGN and suggests that mannose might not be an optimal fragment for FBD. Given the 10^8 heptapeptide fragments, the screen can find only modest combinations of mannose and heptapeptides. Computational mapping with FTMap paralleled the results of our GE-FBD screen and suggested that secondary hot spots in close proximity to the glycan-binding site are simply not available. The only potential hot spot identified by FTMap within 10 Å of glycan binding-site might be accessed by peptide fragments originating from an α -anomeric position of fucose. We do not see these observations as a limitation of GE-FBD technology. Instead, we suggest that the observed synergy between GE-FBD and FTMap can be used for productive design of fragments and libraries

for GE-FBD and similar approaches. This work offers valuable insights for the future design of DC-SIGN inhibitors.

4.5 Experimental procedures

4.5.1 Materials and general information

MOPS buffer contains 50 mM MOPS, 150 mM NaCl, 2 mM CaCl₂ with pH of 7.4 after preparation. Tetrameric DC-SIGN ECD was expressed and purified by Prof. Drickamer's lab. Horseradish peroxidase (HRP) was purchased from Sigma-Aldrich (Sigma #77332). All solutions used for phage work were sterilized by filtration through 0.22 μm filters. *O*-α-D-mannopyranosyl oxyamine was synthesized using a method adapted from the literature.¹⁴¹ Man3-X was synthesized by C.C. Ling and co-workers according to the literature.⁷⁷ Aminoxy-biotin (#10009350) was purchased from Cayman Chemical. The mannose-peptide conjugates were synthesized according to the methods described in section 3.5.10 and were characterized by UPLC-MS (see Appendix C for the spectra). The inhibitor reported by Anderluh was synthesized according to published methods¹⁸⁰ (see Appendix C for the ¹H and ¹³C NMR spectra). HRMS (ESI) spectra were recorded on Agilent 6220 oaTOF mass spectrometer using either positive or negative ionization mode.

4.5.2 Generation of Man-X₇ and AOB-X₇ libraries

An *N*-SerX₇ phage-displayed peptide library (complexity: 3 × 10⁸ pfu) was generated according to the reported protocol.^{153,179} Prior to chemical

modification, the phage library was dialyzed extensively against 5 L PBS (4 °C, pH 7.4, two buffer changes over 24 h, 10K MWCO) to remove the storage buffer, which contains 50% (v/v) glycerol. For chemical modification, the phage library (100 μ L, $\sim 2 \times 10^{12}$ pfu/mL) was oxidized with 0.06 μ M sodium periodate (by adding 1 μ L of 6 mM solution in MQ water) at RT for 5 min. The oxidation was quenched by the addition of 0.5 mM glutathione (by adding 1 μ L of 50 mM solution in MQ water) at RT for 10 min. The oxidized library was distributed into two separate portions. One was treated with an equal volume of *O*- α -D-mannopyranosyl oxyamine (2 mM Man-ONH₂ in 200 mM anilinium acetate buffer, pH 4.7). The other was treated with aminoxybiotin (2 mM AOB in 200 mM anilinium acetate buffer, pH 4.7). The reaction mixtures were incubated for 1 h at RT, after which the modified libraries were diluted 10 times with MOPS buffer for panning experiments. To monitor the oxidation, a small portion of the AOB-X₇ library was diluted and captured with the biotin-capture assay previously described in Chapter 2 (section 2.4.2). Typically, 60% of the fractions of the phage library were successfully oxidized. To quantify the modification with Man-ONH₂, a small portion of the Man-X₇ library was treated with AOB and captured with the biotin-capture assay previously described in Chapter 2 (section 2.4.2). Typically, 40–60% of the fractions of phage library were successfully modified with the Man-ONH₂.

4.5.3 Pre-selection of phage libraries

A 96-well polystyrene plate (Corning #3369) was coated with a solution of DC-SIGN ECD (100 μ L, 10 μ g/mL) in MOPS buffer overnight at 4 $^{\circ}$ C. After coating, the plate was rinsed with washing solution (10 \times 300 μ L, 0.1% (v/v) Tween-20 in MOPS buffer) using 405™ Touch Microplate Washer (BioTek). Pre-selection was performed by incubating the dialyzed SerX₇ library (100 μ L, $\sim 2 \times 10^{12}$ pfu/mL) with a single DC-SIGN-coated well for 1 h at RT. The library was then transferred into a fresh microcentrifuge tube and modified with Man-ONH₂ or AOB as described in section 4.5.2. The modified libraries were diluted 10 times with MOPS buffer.

4.5.4 Three rounds of phage selection

The selection of Man-X₇ library against DC-SIGN was performed in three separate wells in parallel with the control selections, *i.e.*, AOB-X₇ against DC-SIGN, AOB-X₇ against blank wells. Specifically, the modified libraries were added into the corresponding wells (100 μ L/well, $\sim 1 \times 10^{10}$ pfu/well). After incubating for 1 h at RT, the unbound phage was rinsed with the washing solution (5 \times 300 μ L) using the plate washer. Phage remaining on the well were eluted by adding 100 μ L of glycine elution buffer (0.2 M glycine-HCl, pH 2.2, 1 mg/ml BSA). After 9 min of incubation at RT, the elution buffer was transferred into a 1.5-mL microcentrifuge tube containing the neutralization buffer of 1 M Tris-HCl (25 μ L, pH 9.1). The recovered phage libraries, 9 samples in total, were processed further as described in

section 4.5.5. After phage amplification, the amplified phage recovered from the selection of Man-X₇ library against DC-SIGN were modified with Man-ONH₂ or AOB as described in section 4.5.2. The modified phage then underwent a second round of selection using the same procedure as described above. Similarly, a third round of selection followed the same procedure of modification and panning, except for the following modifications: (i) a lower concentration of DC-SIGN ECD (1 µg/mL) was used during the plate coating; (ii) the input for panning was lowered to $\sim 1 \times 10^9$ pfu/well; and (iii) the stringency of washing was increased by washing 10 times.

4.5.5 Phage amplification and PCR

The eluted phage was amplified separately by adding the eluates (125 µL) into 3 mL of ER2738 culture (1:100 dilution of overnight culture). The phage and bacterial mixtures were incubated for 4.5 h at 37 °C with vigorous shaking. The cultures were centrifuged (15 min, 4700 rpm) at 4 °C to pellet the bacterial cells. Half (~ 1.5 mL) of the amplified phage from the selection of the Man-X₇ library against DC-SIGN was transferred to a fresh centrifuge tube and precipitated with PEG/NaCl. The precipitated library will serve as the input for a subsequent round of selection. The remaining supernatants containing the amplified phage were transferred into fresh tubes and the phage ssDNA was isolated using QIAprep spin M13 kit (Qiagen, #27704) according to manufacturer's instructions. The ssDNA was then converted to

Illumina-compatible dsDNA amplicon by PCR. Briefly, the ssDNA (~50–100 ng) was combined with 1× Phusion® buffer, 200 μM dNTPs (each), 0.5 μM forward and reverse primers, and one unit Phusion® High-Fidelity DNA Polymerase in a total volume of 50 μL. Forward (F) and reverse (R) primer sequences, 5' → 3':

F:

5'-CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTC
TTCCGATCT**XXXX**CCTTTCTATTCTCACTCT-3'

R:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCT**XXXX**ACAGTTTCGGCCGA-3'

The **XXXX** in the primer sequence denotes the four-nucleotide-long barcodes used to trace multiple samples in one Illumina sequencing experiment. The temperature cycling protocol for PCR was as follows: 95 °C for 30 s, followed by 25 cycles of 95 °C for 10 s, 60.5 °C for 15 s and 72 °C for 30 s, and then a final extension at 72 °C for 5 min before holding at 4 °C. The resulting amplicons were pooled (20 ng per sample) together and purified on E-Gel® SizeSelect™ 2% agarose gel (Invitrogen, #G6610-02). Sequencing was performed using the Illumina NextSeq platform (The Donnelly Sequencing

Centre at The Donnelley Centre for Cellular and Biomolecular Research, University of Toronto) and analyzed as described in section 3.5.6.

4.5.6 Validation of hits by an inhibition assay

The competitive binding assay involves the binding of immobilized DC-SIGN ECD to horseradish peroxidase (HRP), a naturally glycosylated protein that displays a trimannoside epitope. Inhibitors, *i.e.*, MeMan, Man3-X, or Man-peptide conjugates, were tested for their ability to compete with HRP for binding to DC-SIGN. A solution of DC-SIGN ECD (10 $\mu\text{g}/\text{mL}$) in MOPS buffer was used to coat a polystyrene plate (Costar #3369) to have a final volume of 50 $\mu\text{L}/\text{well}$. The plate was sealed with a membrane and kept in the fridge overnight. On a separate non-binding surface plate (Corning #3641), a 3-fold serial dilution was performed for the solutions of the inhibitor dissolved in MOPS buffer. The diluted solutions were then mixed with an equal volume of the solution of HRP (2 $\mu\text{g}/\text{mL}$) in MOPS buffer. The coating solutions of DC-SIGN were then discarded and the wells were subsequently washed with the washing solution ($3 \times 200 \mu\text{L}$, MOPS buffer containing 0.1% (v/v) Tween-20). The mixture of the inhibitor and HRP probe was transferred accordingly to the plate coated with protein to obtain a final volume of 50 $\mu\text{L}/\text{well}$. The plate was incubated at RT for 1 h, then, washed with the washing solution ($3 \times 200 \mu\text{L}$). TMB substrate (50 μL) was added to each well. After 10 min incubation, 1 M phosphoric acid (50 μL) was added to quench the colorimetric assay. The color developed was read at 450 nm with

a 96-well plate reader. The data was fitted with logistic function using Origin software to determine the half maximal inhibitory concentration (IC₅₀) of the studied inhibitor.

4.5.7 FTMap analysis

The jobs of computational solvent mapping was submitted to FTMap server (<http://ftmap.bu.edu/login.php>).¹⁸⁴ The required input variable is the PDB ID, which corresponds to the structure of interest available on protein data bank (PDB, <http://www.rcsb.org/pdb/home/home.do>). When more than one subunit is available for a single structure, chain A was chosen for analysis during the job submission. The results and the structural figures were analyzed and rendered by PyMOL.

Chapter 5: Phage-displayed macrocyclic glycopeptide libraries

5.1 Introduction

Synthetic macrocycles hold great promise for the development of drugs targeting proteins that are hard to target by conventional small-molecule drugs.^{190,191} The pharmacodynamic and pharmacokinetic properties of the macrocyclic peptides are generally more favorable as compared to their linear analogues.^{192,193} One reason is the increased proteolytic stability. The flexibility in linear peptides allows different shapes to be adopted, thus increases their susceptibility to enzymatic degradation. In contrast, restricting the conformation of peptides by macrocyclization has been shown to improve their proteolytic stability.¹⁹⁴⁻¹⁹⁶ Macrocyclization of peptides can also increase their binding affinity by preorganizing the ligand into a biologically active conformation and minimizing the contribution of unproductive conformations.¹⁹⁷⁻²⁰⁰ This preorganization allows the macrocycle to interact with a larger surface area of the target protein while maintaining productive conformation. Although linear peptide ligands of similar size can span similar areas, maintenance of one productive conformation and elimination of all other conformations results in large entropic penalty and often decreases the binding affinity.

In addition, some studies have suggested that cyclization of a peptide could have a favorable impact on its pharmacokinetic properties by

improving oral absorption. A notable example is cyclosporine A, an oral macrocyclic peptide drug that achieves its unexpected high bioavailability by shielding its four amide hydrogen bond donors while passively diffusing through cell membrane.²⁰¹ Lokey and co-workers hypothesized that the formation of intramolecular hydrogen bonds within the macrocyclic peptide could enhance cell membrane permeability by reducing the energetic cost of desolvating hydrogen bond donors during membrane penetration.^{202,203} They further designed a cyclic hexapeptide with three *N*-methyl groups that showed 28% bioavailability in rats.⁸⁹ Due to the potentially favorable pharmacodynamic/kinetic properties of macrocyclic peptides, there is a need for continuous development of approaches that permit the synthesis of macrocycles²⁰⁴ or macrocyclic peptides in a simple, robust and straightforward fashion.^{87,96,205,206}

The discovery and development of new macrocyclic ligands for proteins usually mandates the synthesis of libraries of macrocycles. Amongst many library syntheses available, the technologies that permit genetic encoding of macrocyclic libraries are the most attractive because they allow identification of the target-binding compounds from a mixture of 10^8 – 10^{15} derivatives present in the same solution. This selection approach is technologically simpler than the screening on multiwell-plate and higher throughput than one-bead-one-compound²⁰⁷ approaches. Although several synthetic approaches are available for constructing genetically encoded libraries, we envision that the simplest approach is constraining existing,

biologically produced, and genetically encoded, *linear* libraries by chemical cross-linkers. Such an approach by-passes the need for complex multi-step synthesis of DNA-conjugated molecules,^{63,64} proximity-driven ligations,^{63,208} and bottom-up synthesis of peptide oligomers assisted by genetically encoded reaction routing.⁶²

Since its inception in 1985,²⁰⁹ phage display has been one of the dominant technology for the discovery of drugs. Antibodies and peptides derived from this technology have been one of the fastest growing contributors to FDA-approved drugs. In Chapter 3, I described a novel approach, termed “genetically encoded fragment-based discovery” (GE-FBD), for the identification of potent ligands for glycan-binding proteins from the *linear* glycopeptide libraries displayed on M13 phage. The chemistry we used to generate these libraries—*N*-terminal serine-mediated oxime ligation—is suitable for the production of *linear* glycopeptide libraries (see Chapter 2). In this chapter, I aim to generate *cyclic* glycopeptide libraries displayed on a M13 phage, using only one-step and rapid chemical modification of readily available phage-displayed libraries.

5.2 Previous syntheses of macrocyclic glycopeptides

Previously, several synthetic approaches for making the cyclic glycopeptides have been reported. For example, head-to-tail macrocyclization of a glycosylated peptide thioester by thioesterase has been demonstrated by Wong and Walsh to afford cyclic glycopeptides.²¹⁰

Alternatively, Wu and Guo achieved macrocyclization through amide coupling between the *N*- and *C*-termini after the cleavage of a linear glycopeptide from beads.²¹¹ Although these methods could be useful for the generation of a cyclic glycopeptide library, it is not suitable for the modification of a phage library because the peptides are attached on phage coat protein, and thus lack the *C*-terminus required for cyclization using these methods. Recently, Mihara and co-workers reported pseudo-cyclic glycopeptide libraries on phage based on a β -loop stabilized by the non-covalent interactions between antiparallel β -strands.⁷³ The authors decorated the library with mannose *via* a disulfide bond with a constant Cys in the peptide library. Such disulfide linkage, however, is susceptible to reduction and disulfide exchange. Furthermore, the expression of peptides with single or odd number of Cys on phage is known to be challenging²¹²⁻²¹⁴ due to (i) trapping of phage in the periplasm of *E. coli*, (ii) mis-folding of pIII proteins by forming disulfides with the structural Cys in the pIII and (iii) enforced enrichment of library clones with even number of cysteine.²¹⁵ The authors bypassed the problem (ii) by expressing the library in a phage with Cys-free pIII protein, but the drawback of this variant is >100-fold diminished infectivity of phage.²¹⁶ Multiple strategies for cyclization of peptides on phage have been reported by our group²¹⁷ and others,^{96,218} but introducing a carbohydrate fragment using these approaches is not obvious.

5.3 Results and discussion

5.3.1 Our approach to create macrocyclic glycopeptide libraries

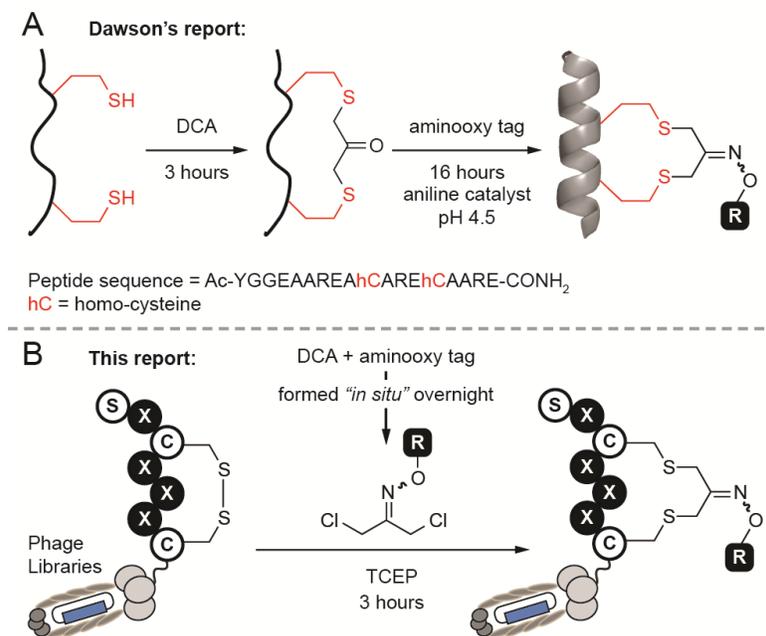


Figure 5-1. Macrocyclization of peptides and genetically encoded peptide libraries with dichloro-oxime derivatives.

(A) Approach reported by Dawson and co-workers. DCA = 1,3-dichloroacetone. (B) We propose to use dichloro-oxime derivatives (DCO) which were formed "in situ" to functionalize phage-displayed peptide libraries in one step.

I was inspired by a recent strategy reported by Dawson and co-workers,²¹⁹ in which dithiol-containing peptides could be cyclized using commercially available 1,3-dichloroacetone (DCA). The macrocyclization introduces a 3-carbon bridge that can stabilize the α -helices of linear peptides flanked by two homocysteines placed at the i and $i + 4$ positions, resulting in an increase of helicity from 22% (linear peptide) to 53% (DCA-constrained peptide).²¹⁹ The most attractive aspect of this chemistry is that the ketone handle embedded in the 3-carbon bridge can be further

derivatized with diverse hydroxylamine derivatives through oxime ligation. In Dawson's report, macrocyclization by DCA, followed by oxime ligation, afforded the macrocyclic peptide with a tag (Figure 5-1A). Unfortunately, we found the original approach to be sub-optimal for the modification of phage-displayed libraries because it requires exposure of a peptide (i.e., phage) to two consecutive reactions. One of them—a 17-hour long oxime ligation of ketone in anilinium buffer (pH 4.5)—is detrimental to the infectivity of phage particles when the reaction is performed in one pot (see Figure 5-8). In this chapter, I modified this methodology to make functionalized phage libraries in one, rapid, biocompatible step (Figure 5-1B).

5.3.2 Macrocyclization of a model peptides with DCO derivatives

We found that dichloro-oxime (DCO) derivatives, such as DCO-biotin can be pre-formed in >99% conversion by overnight incubation of 1,3-dichloroacetone with 1.1 equivalent of aminoxy-biotin in DMF in the absence of any catalysts (see section 5.5.2 for the synthesis and characterization). I was pleased to find that DCO-biotin formed “*in situ*” and used without any purification, afforded a clean macrocyclization in the model synthetic peptide H₂N-Ser-Trp-Cys-Ser-Cys-CONH₂ (SWCSC). In 20 min, 0.1 mM of SWCSC peptide reacted quantitatively with 1 mM of DCO-biotin in Tris buffer (pH 8.5) in the presence of tris(2-carboxyethyl)phosphine (TCEP) to form the macrocyclic peptide **1** as the sole product (Figure 5-2A). The reaction reached >99% conversion in 20 min, according to UPLC-MS analysis, and exhibited a second-order rate constant of $k = 3.8 \text{ M}^{-1} \text{ s}^{-1}$ (Figure 5-2B). I

detected no by-products after the completion of the reaction (20 min). Although both the starting peptide and the macrocyclic product **1** contain nucleophilic *N*-terminal amine, even after incubation for 72 hours in the presence of 10 equivalents of crude DCO-biotin, macrocyclic peptide **1** remained as the major product (Figure 5-2C). Accumulated minor by-products constituted less than 50%. The by-products could not be identified unambiguously based on the *m/z* signals.

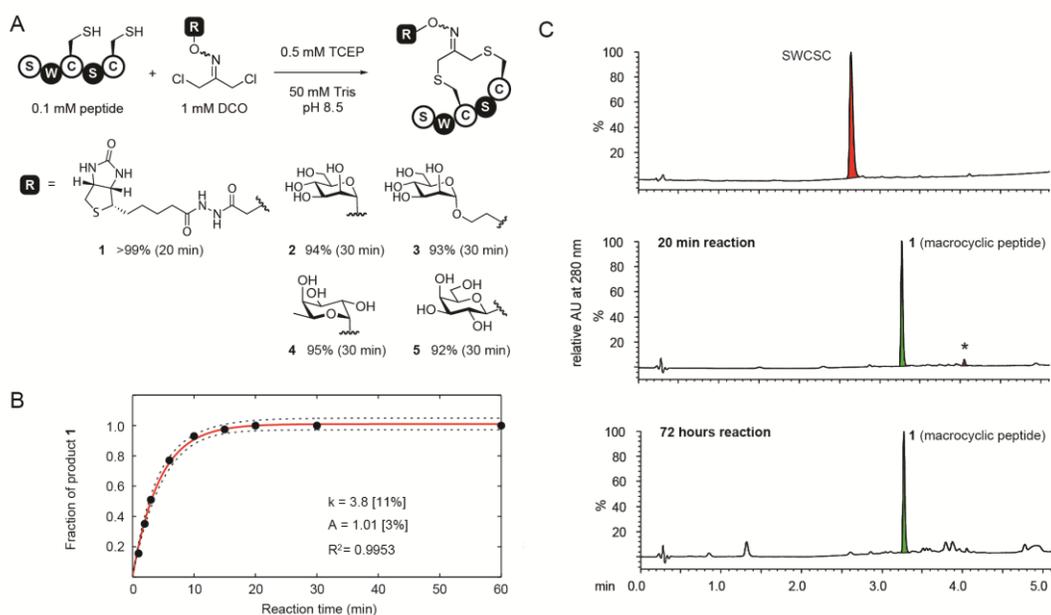


Figure 5-2. Macrocyclization of linear peptide.

(A) The reaction of SWCSC with dichloro-oxime (DCO) derivatives formed the macrocyclic products (**1–5**). Percentages are the conversion yields determined by UV absorbance measured by UPLC. Reaction time is showed in parentheses. (B) The reaction of SWCSC with DCO-biotin under pseudo-first-order conditions gave a second-order rate constant of $k = 3.8 \text{ M}^{-1} \text{ s}^{-1}$ (percentage error in brackets). Reaction proceeded to near completion (>99%) in 20 min as monitored by UPLC-MS. The red line represents the best-fit curve and dashed lines encompass the 95% confidence interval of the fit. (C) UPLC traces of the starting material (SWCSC, top spectrum), and UPLC traces of the reaction between SWCSC and DCO-biotin after 20 min (middle spectrum) or 72 h (bottom spectrum). *: DCO-biotin (weak absorbance at 280 nm).

These observations suggest that any side reaction of the peptide, such as the undesired alkylation of *N*-terminal amine with DCO-biotin, has half-life ($t_{1/2}$) $>2.6 \times 10^5$ s \approx 72 h. It is over 1000 times slower than the alkylation of thiols ($t_{1/2}$ = 182 s) under the same conditions. Unlike the original report,²¹⁹ macrocyclization of the peptide with DCO derivatives was completed in one step and bypassed the need for purification of the peptide-ketone intermediate and subsequent prolonged oxime ligation (16 hours) of the ketone intermediate at pH 4.5.

Encouraged by the success of the model study with DCO-biotin, we extended this approach to four different DCO-carbohydrate derivatives—DCO-ManS, DCO-ManL, DCO-Fuc and DCO-Gal—to form the desired macrocyclic peptide-carbohydrate **2**, **3**, **4** and **5**, respectively (Figure 5-2A). The peptide SWCSC was incubated with each DCO-derivative at room temperature for 5, 15, or 30 min, and then quenched by the addition of acetic acid, followed by analysis using UPLC-MS. The reactions with DCO-carbohydrate derivatives were \sim 2.5 times slower than that with DCO-biotin ($k = 3.8 \text{ M}^{-1} \text{ s}^{-1}$, Figure 5-2B). Still, these reactions proceeded to near completion ($>90\%$) in 30 min with a second-order rate constant of $k = 1.5\text{--}1.8 \text{ M}^{-1} \text{ s}^{-1}$ and formed no by-products (Figure 5-3 to 5-6).

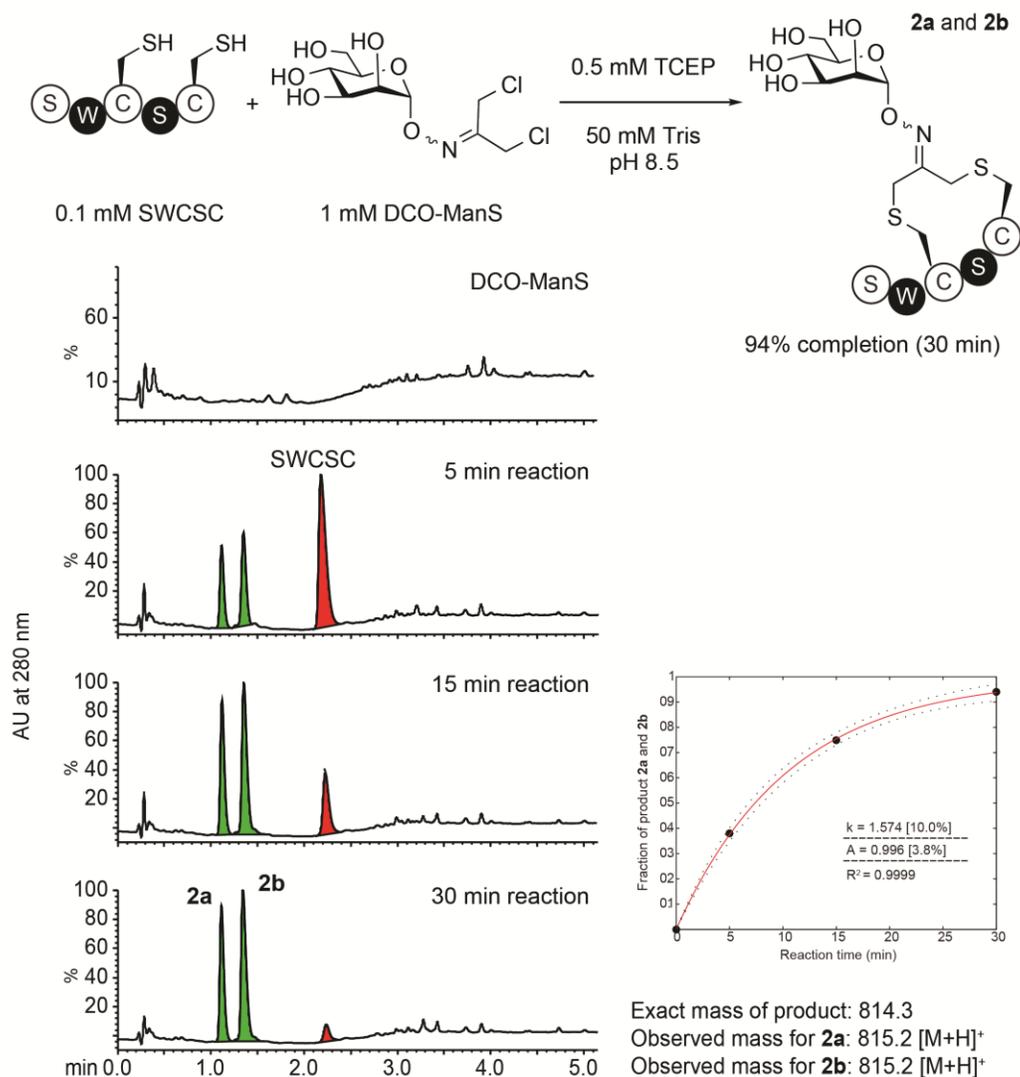


Figure 5-3. Reaction of SWCSC with DCO-ManS.

The reaction proceeded to near completion (94%) in 30 min as monitored by UPLC-MS. UV absorbance was detected at 280 nm. Red and green peaks represent SWCSC and the macrocyclic product (**2a** and **2b**), respectively. **2a** and **2b** are the oxime E/Z isomers and have the same m/z. The reaction was performed under pseudo-first-order conditions. Fitting of the kinetic curve gave a second-order rate constant of $k = 1.6 \text{ M}^{-1} \text{ s}^{-1}$.

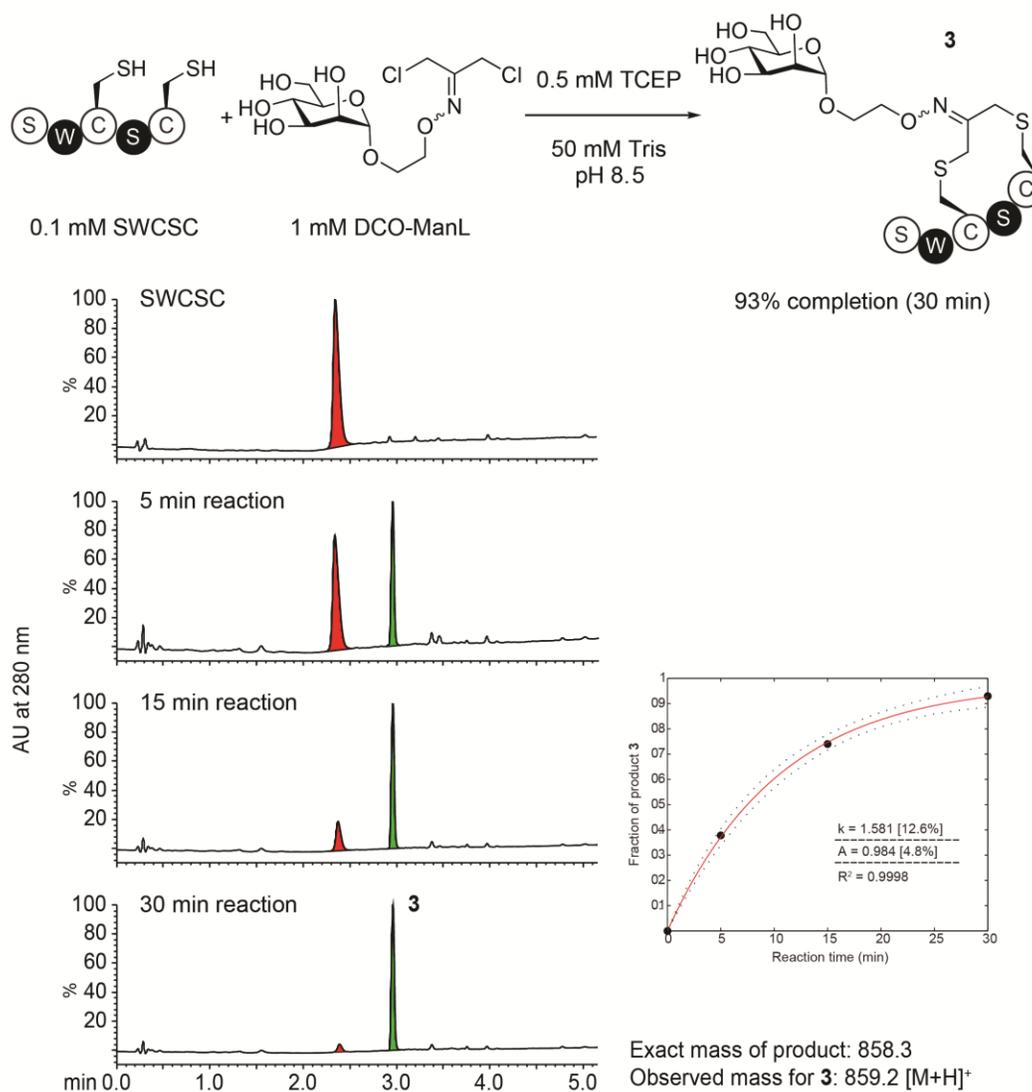


Figure 5-4. Reaction of SWCSC with DCO-ManL.

The reaction proceeded to near completion (93%) in 30 min as monitored by UPLC-MS. UV absorbance was detected at 280 nm. Red and green peaks represent SWCSC and the macrocyclic product (**3**), respectively. The reaction was performed under pseudo-first-order condition. Fitting of the kinetic curve gave a second-order rate constant of $k = 1.6 \text{ M}^{-1} \text{ s}^{-1}$.

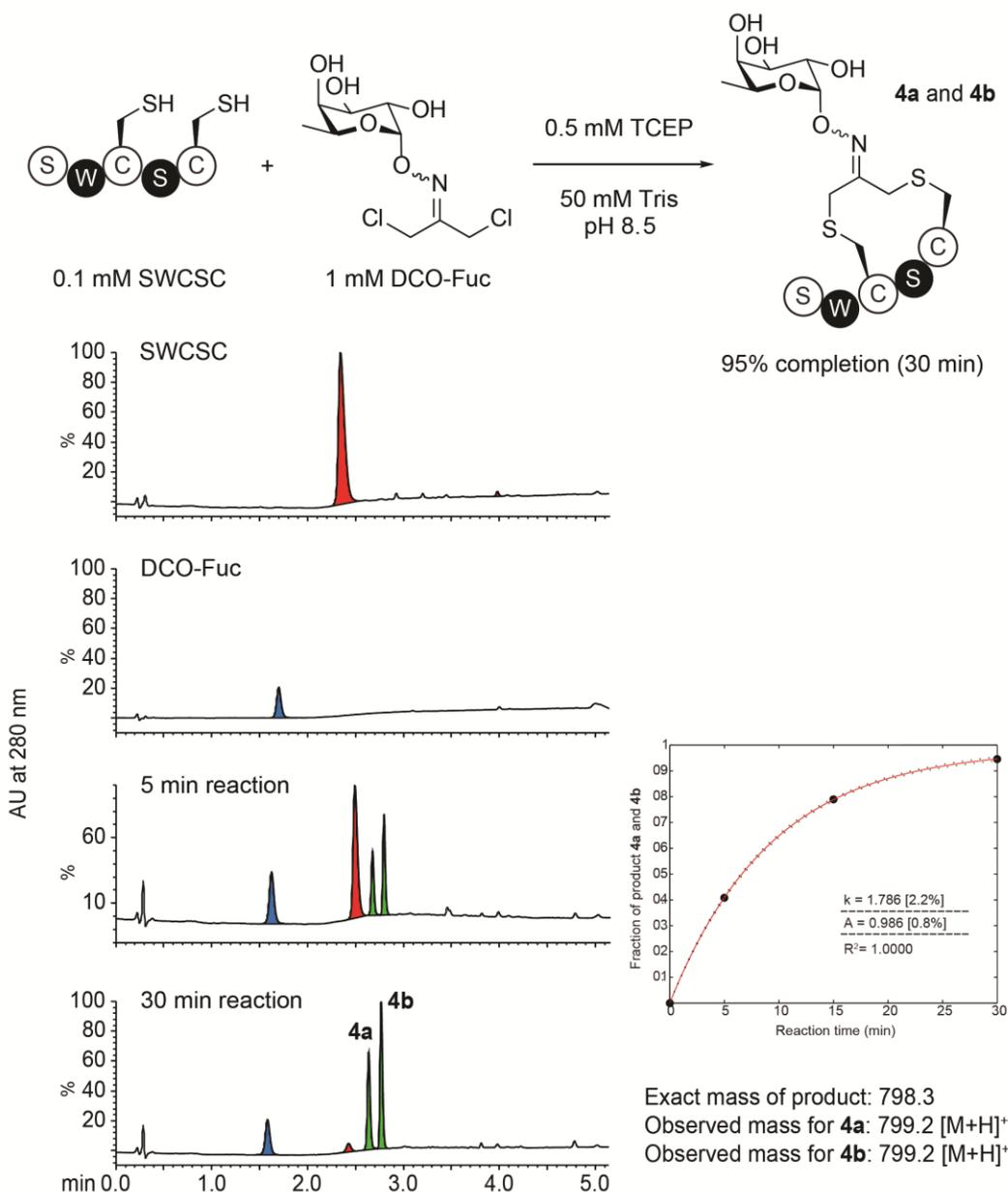


Figure 5-5. Reaction of SWCSC with DCO-Fuc.

The reaction proceeded to near completion (95%) in 30 min as monitored by UPLC-MS. UV absorbance was detected at 280 nm. Red and green peaks represent SWCSC and the macrocyclic product (**4a** and **4b**), respectively. **4a** and **4b** are the oxime E/Z isomers and have the same m/z. Blue peak was observed in the solution of DCO-Fuc, but could not be identified unambiguously by MS, most likely due to poor ionization in electrospray. The reaction was performed under pseudo-first-order condition. Fitting of the kinetic curve gave a second-order rate constant of $k = 1.8 \text{ M}^{-1} \text{ s}^{-1}$.

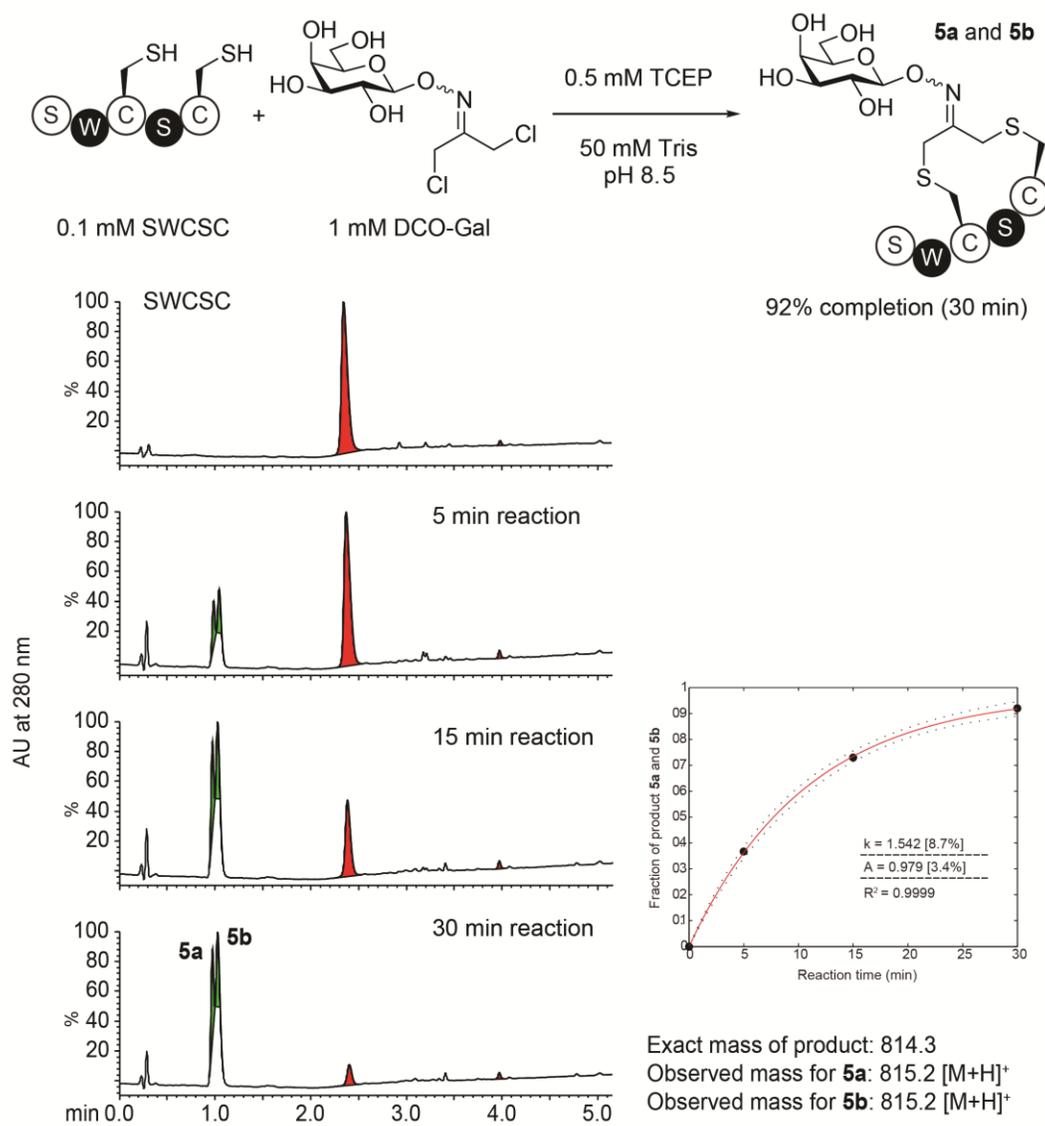


Figure 5-6. Reaction of SWCSC with DCO-Gal.

The reaction proceeded to near completion (92%) in 30 min as monitored by UPLC-MS. UV absorbance was detected at 280 nm. Red and green peaks represent SWCSC and the macrocyclic product (**5a** and **5b**), respectively. **5a** and **5b** are the oxime E/Z isomers and have the same m/z. The reaction was performed under pseudo-first-order conditions. Fitting of the kinetic curve gave a second-order rate constant of $k = 1.5 \text{ M}^{-1} \text{ s}^{-1}$.

5.3.3 Macrocyclization of phage libraries with DCO-biotin

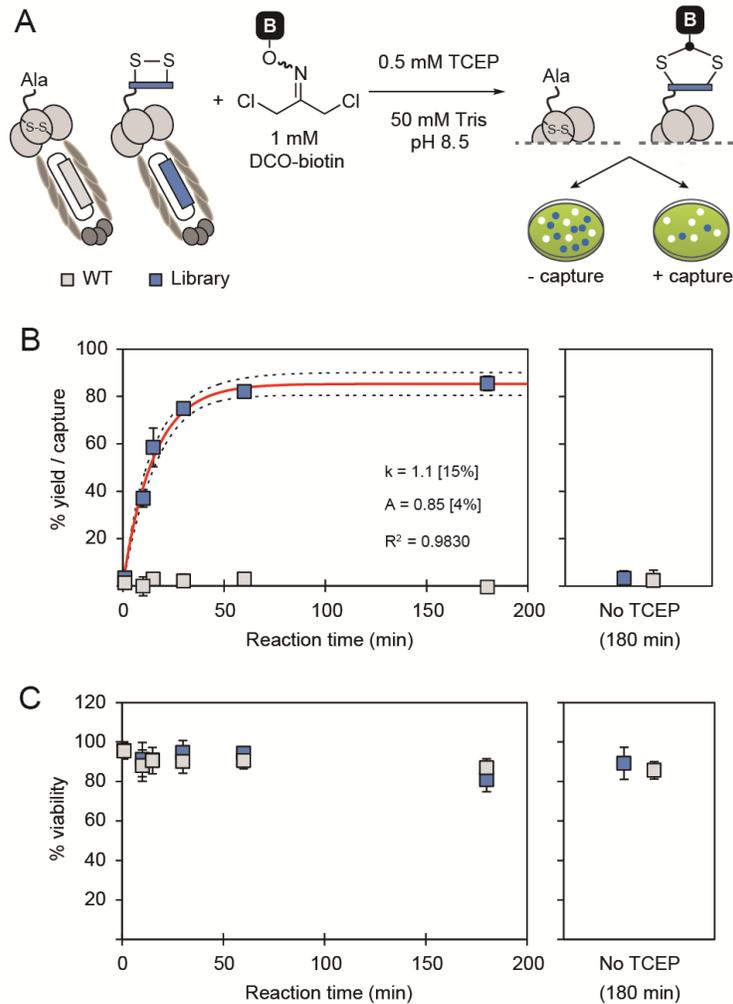


Figure 5-7. Efficiency of macrocyclization of the phage libraries. (A) The reaction of 1 mM DCO-biotin with disulfide-containing phage library was monitored by biotin-capture assay in the presence of wild-type (WT) phage as control. (B) The reaction performed under pseudo-first-order condition gave second-order rate constant $k = 1.1 \text{ M}^{-1} \text{ s}^{-1}$ (percentage error in bracket). Red line represents the best-fit curve and dashed lines encompass the 95% confidence interval of the fit. Reaction with WT phage is insignificant even after 3 hours of incubation. In the absence of TCEP, reaction of neither library nor WT phage proceeded to any detectable extent (>5%) even after 3 hours of incubation. (C) After 3 hours of modification, more than 80% of phage population retained the ability to amplify in bacterial host. Data in panel B and C are mean value from three independent experiments; error bars represent one standard of deviation.

Next, we investigated whether the reaction could be applied to disulfide-containing peptide libraries SXCX₃C and SXCX₄C displayed on M13 phage (X represents 19 any natural amino acids and no Cys, Figure 5-7A). I quantified the efficiency of the on-phage reactions using a biotin-capture assay as described in section 2.4.2. Three independent on-phage reactions were performed on separate days to assure the reproducibility. All on-phage reactions described in this report were performed with a 1:1 mixture of phage that displays a library of peptides (“Lib”), and wild-type (WT) phage, which lacks the displayed disulfide-peptide, but otherwise it is identical to Lib. Two phage populations can be evaluated together because Lib and WT phage form blue and white plaques respectively in agar overlay supplemented by X-Gal. I used the biotin-capture assay developed in our lab to quantify the efficiency of the reaction and perform the kinetic studies.¹⁴¹ Phage libraries were biotinylated and reached saturation at ~85% of capture within 3 hours (Figure 5-7B). Fitting the kinetic curve afforded the rate constant of $k = 1.1 \text{ M}^{-1} \text{ s}^{-1}$, which is ~3.5-fold slower as compared to the reaction on the peptide ($k = 3.8 \text{ M}^{-1} \text{ s}^{-1}$). The difference in rate is not surprising because the phage libraries contain millions of different peptides with varying reactivity. This observation highlights the importance of monitoring the reaction progress in the context of a phage library even if the reaction has been well optimized on the model peptides or the proteins.

Importantly, DCO-biotin did not react with WT phage present in the same solution; biotinylation of WT was negligible (<5%). DCO-biotin reacted

specifically with peptide sequences displayed on Lib phage and it did not have any detectable reactivity with other nucleophilic functionalities common to Lib and WT phage. For example, both WT and Lib phage contain four pairs of internal disulfide bonds within the pIII coat proteins⁸⁰ and a vast number of other nucleophilic functionalities on the pVIII coat proteins (*e.g.*, ~2700 copies of *N*-terminal amines, ϵ -amine of Lys residue, phenolic side chain of Tyr, etc). Lack of biotinylation of WT proved that none of these functionalities reacted with DCO; it alkylated only externally displayed dithiols. Moreover, in the absence of TCEP, neither Lib nor WT phage were biotinylated. This experiment highlights the excellent regioselectivity of the reduction of external disulfides by TCEP. To summarize, modification with DCO takes place only when two conditions are satisfied: (1) a disulfide displayed on the *N*-terminus of the pIII protein is solvent-accessible;^{80,217} and (2) free thiol is available through the TCEP reduction.

The post-translational chemical modification by DCO-biotin is benign to phage. After 3 hours of treatment with DCO-biotin, more than 80% of the phage population still retained its ability to infect the bacterial hosts (Figure 5-7C). This result is in contrast to the report of Heinis and Winter,⁹⁶ in which the modification of phage with 10 μ M of 1,3,5-tris(bromomethyl)benzene (TBMB) for 1 hour resulted in ~80% loss of phage infectivity and much greater losses at higher concentration of TBMB. Heinis and co-workers stated in their report that the loss of infectivity is due to the cross-linking of phage

coat proteins.⁹⁶ In this study, we observed excellent viability of phage after modification with DCO, suggesting that the cross-linking is negligible.

5.3.4 Macrocyclization of phage libraries with DCA and oxime ligation

When I exposed the mixture of phage libraries and WT phage to DCA instead of DCO for 3 hours and then followed by the treatment with aminoxy-biotin (AOB) in anilinium acetate buffer (pH 4.7) for 17 hours, I could capture ~72% of the phage libraries but not WT phage (Figure 5-8). As noted in the introduction, this one-pot two-step manipulation was detrimental to phage viability. Only ~12% of the Lib and WT population survived the 17 hour-long reaction (Figure 5-8B). The reason for the decrease in viability is not clear but we hypothesize that the imine formed by DCA and aniline could be susceptible to attack by the side-chain nucleophiles,^{220,221} leading to cross-linking of phage coat proteins.

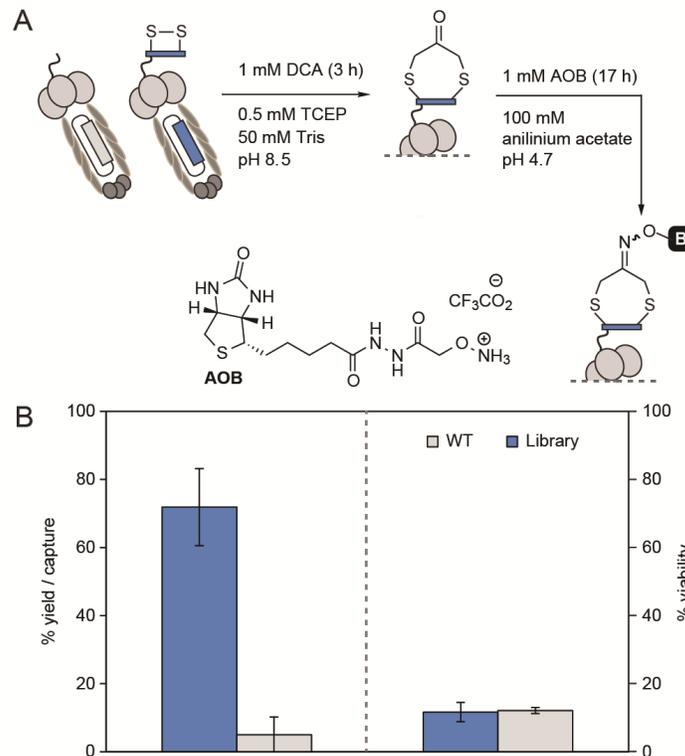


Figure 5-8. Two-step macrocyclization and functionalization of phage libraries with biotin.

(A) After the mixture of phage libraries and WT phage was exposed to DCA for 3 h, an equal volume of 2 mM AOB in anilinium acetate buffer was added to the mixture and incubated at RT for 17 h. (B) I observed ~72% of phage libraries, but not WT phage, were biotinylated under the one-pot two-step reaction. However, this two-step approach might generate side products that are detrimental to phage infectivity as demonstrated by the significant drop of viability of both library and WT phage down to ~12%.

5.3.5 Validation of “pulse-chase” reaction on model peptide

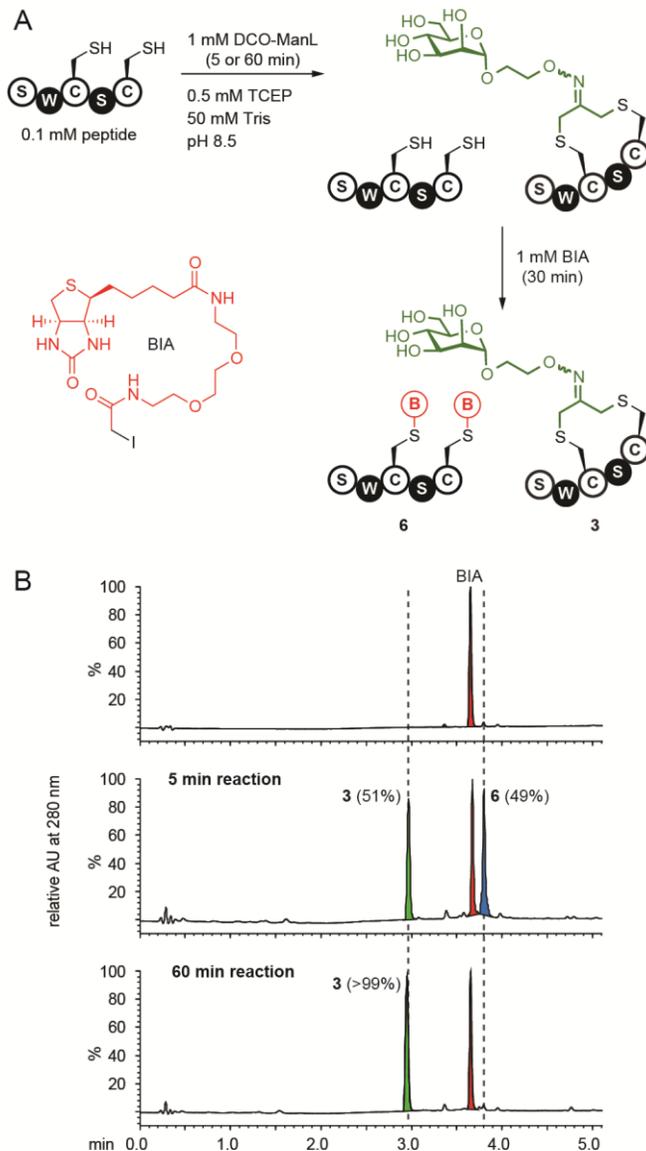


Figure 5-9. Macrocyclization of linear peptides under pulse-chase conditions.

(A) SWCSC is pulsed with 10 eq. of DCO-ManL for 5 or 60 min; and then chased immediately with 10 eq. of BIA for 30 min. (B) The pulse-chase reactions were monitored by UPLC-MS. After pulsing with DCO-ManL for 5 min, 51% of the peptide formed the desired macrocyclic product **3** (green peak, middle spectrum) and 49% of the unreacted peptide was biotinylated twice to give the peptide-biotin conjugate **6** (blue peak, middle spectrum). In contrast, when pulsing with DCO-ManL for 60 min, no peptide-biotin conjugate **6** was observed and nearly all SWCSC formed the desired product **3** (green peak, bottom spectrum). Red peaks: BIA.

To assess the efficiency of the on-phage reaction with DCO derivatives lacking the biotin tag, I employed a “pulse-chase” reaction developed in our group.^{141,217} I first validated the use of the pulse-chase reaction by monitoring the reaction on synthetic peptide by UPLS-MS (Figure 5-9A). I pulsed the peptide with DCO-ManL and chased it with excess of biotin-PEG₂-iodoacetamide (BIA), a thiol-reactive probe with high reactivity ($k = 10.8 \text{ M}^{-1} \text{ s}^{-1}$, Figure 5-10). Peptide pulsed with DCO-ManL for 5 min and chased with BIA formed only two products (Figure 5-9A): 51% of the desired macrocyclic product **3** and 49% of a peptide with two biotin moieties (Figure 5-9B, middle spectrum). UPLC-MS did not detect any peptide that contained DCO-ManL and biotin simultaneously, suggesting that the intramolecular cyclization was rapid and the open-chain peptide intermediates with mono-reacted Cys did not accumulate. Reaction proceeded to near 99% conversion when pulsing with DCO-ManL for 60 min and BIA did not react with the product **3** after 30 min of exposure (Figure 5-9B, bottom spectrum).

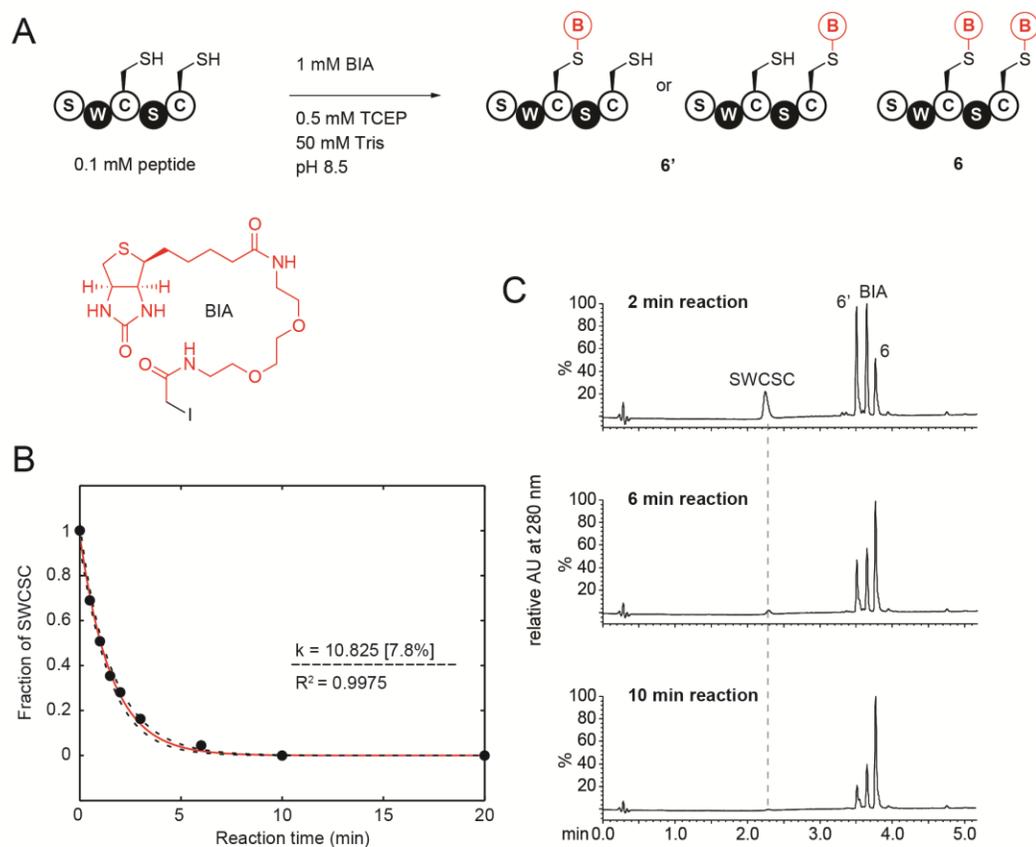


Figure 5-10. Kinetic profile of reaction between SWCSC and BIA. (A) The reaction of SWCSC (0.1 mM) with 10 eq. of BIA (1 mM) proceeded to near completion in 10 min as monitored by UPLC-MS. (B) Fitting of the kinetic curve gave a second-order rate constant of $k = 10.8 \text{ M}^{-1} \text{ s}^{-1}$. (C) UPLC traces of the reaction mixture at selected time point.

5.3.6 Synthesis of macrocyclic glycopeptide libraries

I then applied the validated “pulse-chase” method to monitor the modification of phage libraries with different DCO-carbohydrate derivatives (Figure 5-11). First, the mixture of phage libraries (SXCX₃C and SXCX₄C) and WT phage was incubated separately with four different DCO-carbohydrate derivatives (1 mM) in Tris buffer (pH 8.5) in the presence of 0.5 mM TCEP (Figure 5-11A). After pulsing with DCO for 3 hours, the solutions of phage were immediately diluted by 100-fold to halt the modification. To quantify

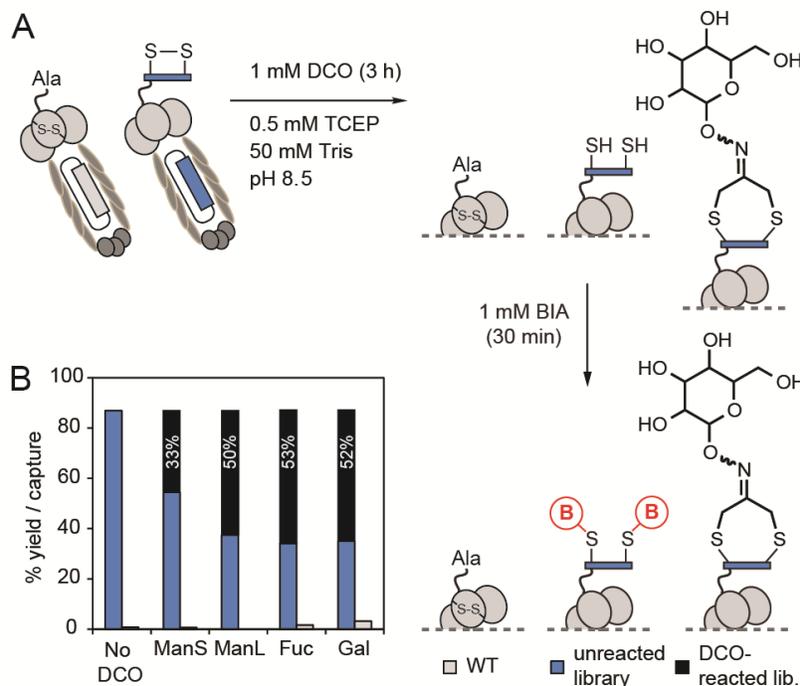


Figure 5-11. Macrocyclization of phage libraries with four different DCO-carbohydrate derivatives.

(A) Mixture of phage libraries and WT phage was pulsed with DCO derivatives for 3 hours, diluted by 100-fold, and then chased immediately with BIA for 30 min. Under these conditions, any unreacted thiolate would be biotinylated, resulting in the capture of phage. (B) Bar graph represents the efficiency of modification. About 87% of the phage population was captured when no DCO was added. The black bars quantified the population of phage libraries (white labels) that were fully modified with DCO-carbohydrate derivatives.

the unreacted phage, I incubated the phage with 1 mM BIA for 30 min. Phage library that was not exposed to any DCO for 3 hours had unreacted thiols, thereby leading to ~87% of capture after exposure to BIA (Figure 5-11B). This observation is consistent with ~85% capture observed in alkylation of phage libraries by DCO-biotin (Figure 5-7B). In contrast, “pulsing” the phage library with various DCO-carbohydrate derivatives for 3 hours, followed by BIA “chasing”, resulting in a decrease of the phage capture (blue bars, Figure

5-11B). These results suggested that 33%, 50%, 53% and 52% of the phage libraries were modified with DCO-ManS, DCO-ManL, DCO-Fuc and DCO-Gal, respectively, after 3 hours (black bars, Figure 5-11B). The incomplete modifications are consistent with our observation that DCO-carbohydrate derivatives ($k = 1.5\text{--}1.8 \text{ M}^{-1} \text{ s}^{-1}$) have slower reactivity than DCO-biotin ($k = 3.8 \text{ M}^{-1} \text{ s}^{-1}$) in the reactions with a model peptide (see Figure 5-2B). We note that WT phage presents in the same solution was not captured in any experiments (Figure 5-11B).

5.3.7 Preliminary selection of macrocyclic glycopeptide libraries

With the phage libraries modified with ManS- or ManL-DCO in hand, I used them for selection against DC-SIGN and ConA. The biopanning was performed as previously described and post-selection analysis was performed by volcano plot analysis as described in section 3.5.6. However, I could not find high-confidence hits from the analysis. Looking at the top 50 most abundant sequences (Figure 5-12A), I realized that they all have similar copy numbers across the panning experiments, regardless of the target used (DC-SIGN or ConA) or the fragment attached (ManS-, ManL, or methyl group). These results are in contrast to our previous selection (see chapter 3 for detail), in which we successfully enriched potent ligands for ConA from a linear glycopeptide library (ManL-X₇). In the case of ConA, true hits already emerged from the top 50 most abundant sequences even with a single round of selection (see black arrow in Figure 5-12B). These hits all share a similar

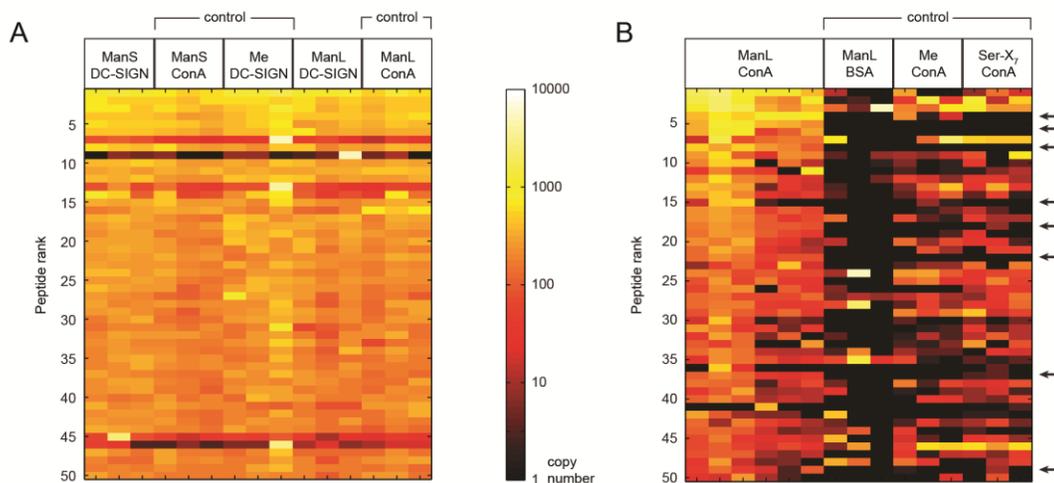


Figure 5-12. Results of selections of cyclic glycopeptides against DC-SIGN and ConA.

(A) The top 50 most abundant peptide sequences were obtained from biopanning on DC-SIGN or ConA using phage libraries modified with ManS, ManL, or methyl (Me) DCO derivatives. For the DC-SIGN selection, no sequence was found to be enriched significantly against the control selections. (B) Previous results of the selection on ConA or BSA using ManL-X₇, Me-X₇, or Ser-X₇ phage libraries (see Chapter 3 for details). In these experiments, we identified peptide sequences, as indicated by the arrows that were statistically and significantly enriched against all controls. These sequences contain a consensus motif [WY] and bind to ConA in synergy with ManL to yield glycopeptides with enhanced affinities relative to the mannose itself.

pattern—high copy numbers in ConA–ManL-X₇ panning and zero to single-digit copy numbers in the control panning. However, this did not occur in the current selection with DC-SIGN and ConA. The observations suggested that the current library format might not be optimal for DC-SIGN and ConA. From our previous selection with DC-SIGN (see Chapter 4 for detail), we concluded that mannose is not the optimal fragment for the selection. In the case of ConA, the conformation of the library might be too rigid and/or the linker length might not be exactly right for optimal access of the “hot spot”

previously identified by our selection of the linear glycopeptide library (see Chapter 3 for detail). In future work, we will assess the utility of these macrocyclic glycopeptide libraries using other glycan-binding proteins, such as the galectin superfamily and anti-*Mycobacterium tuberculosis* antibodies. Selection of macrocyclic fucose-peptide libraries against DC-SIGN is also a direction worth trying.

5.4 Conclusion

In conclusion, I have demonstrated the use of DCO, for the one-step macrocyclization plus labeling of a model peptide, as well as disulfide-containing peptide libraries displayed on M13 phage, with biotin and carbohydrates. Our strategy bypasses the need for the isolation of intermediate and minimizes the losses of valuable starting materials, *e.g.*, peptides, proteins, or phage libraries, in multi-step reactions. The reaction is highly chemo- and regioselective. It does not interfere with the infectivity of phage libraries and maximizes the number of viable phage particles for downstream selection. The macrocyclic peptide libraries contain an additional fragment, in this work, a carbohydrate, which could be useful for selection with glycan-binding proteins.⁷⁷ Yet, the fragment is not limited to carbohydrate; any known inhibitor, enzyme substrate or small-molecule lead could be incorporated into the libraries and serve as an anchor fragment to facilitate the search for more potent compounds using genetically encoded fragment-based design.^{68,70-77} We anticipate that this strategy will find wide

application for fragment-based selection of conformationally constrained, proteolytically stable, and potent macrocyclic peptide-fragment conjugates for many protein targets.

5.5 Experimental procedures

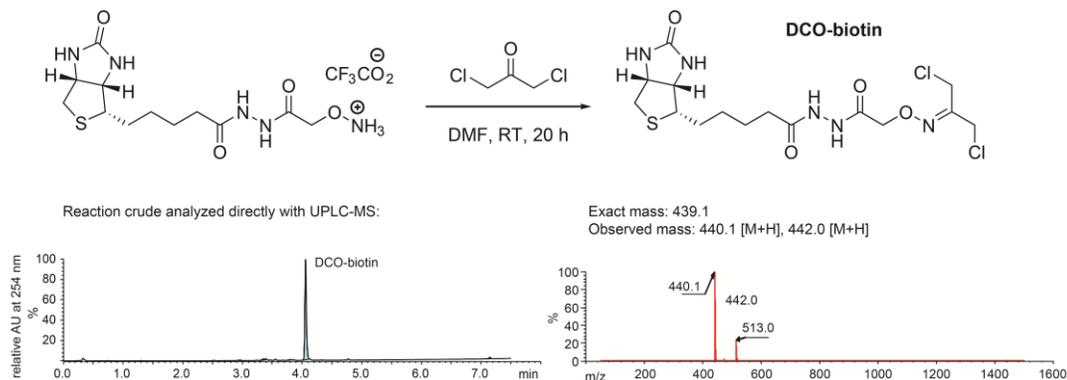
5.5.1 Materials and general information

Tris buffer was prepared as a solution of 50 mM Tris at pH of 8.5. All solutions used for phage work were sterilized by filter sterilization (0.22 μm). The peptide (SWCSC) was synthesized on solid support according to the protocol described in section section 2.4.4. Aminoxy-biotin (#10009350) was purchased from Cayman Chemical. Dichloroacetone (#168548) was purchased from Sigma-Aldrich. Biotin-PEG₂-iodoacetamide (BIA, #21334) was purchased from Thermo Fisher Scientific. Product purification was accomplished with automated chromatography machine (CombiFlash[®] Rf, Teledyne Isco, Inc.). ¹H and ¹³C NMR spectra were acquired on Agilent/Varian VNMRS 500 MHz spectrometers. Chemical shifts (δ) are reported in ppm and coupling constants (J) are given in Hz. The following abbreviations classify the multiplet peaks in the ¹H NMR spectra: s = singlet, d = doublet, m = multiplet or unresolved. HRMS (ESI) spectra were recorded on Agilent 6220 oaTOF mass spectrometer using either positive or negative ionization mode. Characterization of crude reactions was performed with UPLC-MS using a C18 column (Phenomenex Kinetex 1.7 μm EVO C18, 2.1 \times 50mm) 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. The phage libraries (SXCX₃C and SXCX₄C) was prepared according to literature protocols.¹⁷⁹

5.5.2 Synthesis of dichloro-oxime-biotin (DCO-biotin)

Slight excess of aminoxy-biotin (5 mg, 11 μ mol, 1.1 eq.) was added to 1,3-dichloroacetone (1.3 mg, 10 μ mol, 1 eq.) dissolved in DMF (100 μ L). The reaction proceeded to completion within 20 h at RT. The reaction mixture was characterized by UPLC-MS (see below). The concentration of DCO-biotin produced *in situ* here was 100 mM. This solution (100 \times concentrated stock) was used directly for the modification of peptide or phage libraries without further purification—typically by adding 1 μ L of DCO-biotin (100 mM) into 99 μ L of reaction mixture containing peptide or phage.



5.5.3 Synthesis of dichloro-oxime-carbohydrate

The carbohydrate-hydroxylamine derivatives were synthesized using methods adapted from section 2.4.3. To make a DCO-carbohydrate derivative, the hydroxylamine derivative (40 μ mol, 2 eq.) was added to 1,3-

dichloroacetone (20 μmol , 1 eq.) dissolved in MeOH (0.5 mL). After 24 h reaction at RT, the solvent was evaporated. The residue was purified with silica gel chromatography using a gradient of 0 to 10% MeOH in CH_2Cl_2 to afford the title compound as a white solid in 60–90% yield.

DCO-ManS: ^1H NMR (500 MHz, D_2O) δ = 5.49 (d, J = 1.5 Hz, 1 H), 4.51 (d, J = 12.5 Hz, 1 H), 4.39 – 4.35 (m, 3 H), 4.14 (dd, J = 1.5, 3.5 Hz, 1 H), 3.87 – 3.85 (m, 2 H), 3.78 – 3.73 (m, 2 H), 3.68 – 3.63 (m, 1H); ^{13}C NMR (125 MHz, D_2O) δ = 157.3, 102.9, 74.7, 71.6, 69.5, 67.3, 61.5, 42.7, 34.5; HRMS (ESI) calcd for $\text{C}_9\text{H}_{15}\text{Cl}_2\text{NO}_6\text{Na}$ $[\text{M}+\text{Na}]^+$ m/z = 326.0169, found 326.0162.

DCO-ManL: ^1H NMR (500 MHz, D_2O) δ = 4.88 (d, J = 1.5 Hz, 1 H), 4.42 (s, 2 H), 4.38 – 4.35 (m, 4 H), 4.01 – 3.97 (m, 1 H), 3.95 – 3.94 (m, 1 H), 3.85 (dd, J = 2.0, 12.0 Hz, 1 H), 3.80 – 3.78 (m, 2 H), 3.75 (dd, J = 5.5, 12.0 Hz, 1 H), 3.68 – 3.61 (m, 2 H); ^{13}C NMR (125 MHz, D_2O) δ = 154.7, 100.9, 74.3, 73.7, 71.5, 70.9, 67.6, 66.8, 61.8, 42.9, 34.4; HRMS (ESI) calcd for $\text{C}_{11}\text{H}_{19}\text{Cl}_2\text{NO}_7\text{Na}$ $[\text{M}+\text{Na}]^+$ m/z = 370.0431, found 370.0423.

DCO-Fuc: ^1H NMR (500 MHz, D_2O) δ = 5.53 (apparent s, 1 H), 4.59 (d, J = 12.5 Hz, 1 H), 4.44 – 4.35 (m, 3 H), 4.13 – 4.09 (m, 1 H), 3.96 (s, 2 H), 3.85 (s, 1 H), 1.19 (d, J = 6.5 Hz, 3 H); ^{13}C NMR (125 MHz, D_2O) δ = 157.4, 102.1, 72.7, 70.5, 68.8, 68.3, 42.7, 34.6, 16.2; HRMS (ESI) calcd for $\text{C}_9\text{H}_{15}\text{Cl}_2\text{NO}_5\text{Na}$ $[\text{M}+\text{Na}]^+$ m/z = 310.0219, found 310.0217.

DCO-Gal: ^1H NMR (500 MHz, D_2O) δ = 5.03 (d, J = 8.0 Hz, 1 H), 4.51 (s, 2 H), 4.41 (s, 2 H), 3.97 (d, J = 2.5 Hz, 1 H), 3.80 – 3.74 (m, 5 H); ^{13}C NMR (125 MHz,

D₂O) δ = 157.5, 105.2, 76.7, 73.6, 69.9, 69.4, 61.8, 42.4, 34.7; HRMS (ESI) calcd for C₉H₁₅Cl₂NO₆Na [M+Na]⁺ m/z = 326.0169, found 326.0163.

5.5.4 Macrocyclization of SWCSC with DCO derivatives

DCO derivatives (1 μ L, 100 mM in water or in DMF for DCO-biotin, 10 eq.) were added to a solution of SWCSC peptide (99 μ L, 0.1 mM, 1 eq.) in Tris buffer (pH 8.5) containing 0.5 mM TCEP. The reaction was incubated at RT. At each time interval, an aliquot of the mixture (30 μ L) was quenched by mixing well with acetic acid (1 μ L). The quenched mixtures were then analyzed by UPLC-MS to characterize the progress of the reaction. See Figure 5-2B and Figure 5-3 to 5-6 for the spectra. All the reactions were performed under a pseudo-first-order condition. Fitting of the kinetic curve to the equation $A_t = 1 - e^{-k[\text{DCO}]^*t}$ yielded a second-order rate constant of k , where A_t is the fraction of the product at time t and [DCO] is 0.001 M.

5.5.5 Kinetic studies of modification of phage with DCO-biotin

A mixture of phage libraries (SXCX₃C/SXCX₄C/WT in 1/1/2 ratio) was prepared at a concentration of 2×10^9 pfu mL⁻¹ in Tris buffer (pH 8.5). To the mixture of phage (99 μ L), TCEP (1 μ L, 50 mM in water) was added, followed by DCO-biotin (1 μ L, 100 mM in DMF). The reaction mixture was incubated at RT. At each time interval, the reaction was quenched by diluting 10⁵-fold an aliquot of the phage solution (10 μ L). The efficiency of the modification was quantified by biotin-capture assay as described in section 2.4.2. The studies, modification and quantification, were repeated three times on separate days

to validate the reproducibility of the experiments. The quenched mixtures were then analyzed by UPLC-MS to characterize the progress of the reaction. See Figure 5-2B and Figure 5-3 to 5-6 for the spectra. All the reactions were performed under a pseudo-first-order condition. Fitting of the kinetic curve to the equation $A_t = 1 - e^{-k[\text{DCO}]t}$ yielded a second-order rate constant of k , where A_t is the fraction of the captured phage at time t and $[\text{DCO}]$ is 0.001 M.

5.5.6 Pulse-chase reaction of SWCSC with DCO-ManL and BIA

DCO-ManL (1 μL , 100 mM in water, 10 eq.) was added to a solution of SWCSC peptide (99 μL , 0.1 mM, 1 eq.) in Tris buffer (pH 8.5) containing 0.5 mM TCEP. The reaction was incubated at RT. After pulsing for 5 or 60 min, biotin-PEG₂-iodoacetamide (BIA, 1 μL , 100 mM in DMF, 10 eq.) was added immediately to the reaction mixture. After incubating at RT for 30 min, an aliquot of the mixture (30 μL) was quenched by mixing well with acetic acid (1 μL). The quenched mixtures were then analyzed by UPLC-MS.

5.5.7 Modification of phage libraries with DCO-carbohydrate derivatives

A mixture of phage libraries (SXCX₃C/SXCX₄C/WT in 1/1/2 ratio) was prepared at a concentration of 2×10^{11} pfu mL⁻¹ in Tris buffer (pH 8.5). To the mixture of phage (99 μL), TCEP (1 μL , 50 mM solution in water) was added, followed by DCO-carbohydrate derivatives (1 μL , 100 mM in water). The reaction mixture was incubated at RT for 3 h and then an aliquot of the mixture (1 μL) was diluted immediately 100-fold with Tris buffer containing

0.5 mM TCEP. To the diluted phage solution (100 μ L), biotin-PEG₂-iodoacetamide (BIA, 1 μ L, 100 mM in DMF) was added. The reaction mixture was incubated at RT for another 30 min and then quenched by diluting 10⁵-fold. The efficiency of the modification was quantified by biotin-capture assay as described in section 2.4.2.

Chapter 6: Conclusion and outlook

6.1 Conclusion

This thesis describes an emerging technology to accelerate the discovery of inhibitors for glycan-binding proteins. Design of inhibitors for glycan-binding proteins usually involves rational design and requires multistep synthesis and complex chemical manipulations due to the structural complexity of glycans. As a result, the generation of compound analogs has generally been slow. The task is further aggravated by weak binding affinity between glycans and glycan-binding proteins, mostly due to the shallow protein surface of the binding site. Therefore, there remains an unmet need for simpler, less expensive, and more efficient routes to identify potent inhibitors for glycan-binding proteins.

Several groups have successfully designed potent inhibitors for glycan-binding proteins using fragment-based approaches (see section 1.3.3 for the examples). In this approach, the key is to identify a potential second site located in close proximity to the glycan-binding site and devise a strategy to link the second-site fragment to the glycan to afford a compound with improved binding affinity. By combining the synthetic versatility of monosaccharides and the immense diversity of phage-displayed peptide libraries ($>10^8$), we believe that the hybrid carbohydrate-peptide library would accelerate the discovery process and still adhere to the principle of fragment-based lead discovery.

In Chapter 2, I pioneered a methodology to create a hybrid carbohydrate-peptide library displayed on M13 phage. Development of reactions on phage is plagued by the lack of efficient method to characterize and monitor the progress of reactions. The maximal solubility of M13 phage (a 16MDa virus) in aqueous buffer is $\sim 2 \times 10^{13}$ PFU mL⁻¹, which gives a concentration of ~ 33 nM. Display of peptides on 5 copies of pIII coat protein slightly increases the concentration of the peptides to ~ 165 nM. To characterize the reaction with mass spectrometry, a highly sensitive instrument and high amount of phage is required ($>10^{12}$ phage particles).²²² Furthermore, the phage need to be purified with size-exclusion chromatography or diafiltration before and after the modification, and the peptides have to be cleaved off the coat proteins by trypsin digestion for 24 hr prior to analysis with mass spectrometry.²²² This lengthy process of characterization is suitable for well-established reactions, which may demand less optimization on phage. However, it would consume a large portion of the phage library and become quite tedious when new reactions are investigated to determine optimal conditions on phage. This problem was conveniently solved with a biotin-capture assay as described in Chapter 2. The quantitative assay is highly sensitive, requires only dilution, followed by 15 min capture on bead and phage titering.

In Chapter 3, I have demonstrated that a hybrid carbohydrate-peptide library could be a useful source for the rapid discovery of potent inhibitors for glycan-binding protein. By using ConA as model target, mannose attached

onto the peptides “docks” the whole library in the desired binding site and guides the selection of peptides that bind simultaneously in a synergistic fashion. Such a strategy accounts for the plasticity of protein and could potentially identify a novel secondary binding site in close proximity to the primary glycan-binding site. For example, selection of the pendant peptide in the presence of a mannose “anchor” allowed us to identify a novel second site of ConA, which is occupied by the peptide in our inhibitors. Interestingly, without the “anchor”, hits could not be identified from the selection of the non-modified peptide library. The most potent glycopeptides have comparable IC_{50} 's (4 μ M) and selectivity to a standard trimannoside ligand ($IC_{50} = 2 \mu$ M) for ConA.

Chapter 4 focuses on the selection of a glycopeptide library for DC-SIGN. Through the selection, I discovered a novel class of inhibitors for DC-SIGN, albeit with moderate enhancement in activity relative to a mannose. Modest affinity correlates well with the reports of other groups^{171,174,176} in the design of mannose-based inhibitors for DC-SIGN. These experiences and FTMap analysis suggest that mannose might not be an optimal fragment.

In Chapter 5, I describe an optimized methodology for the generation of a new class of macrocyclic glycopeptide libraries displayed on phage. In contrast to a previous report, the use of dichloro-oxime derivatives allows one-step macrocyclization and functionalization of peptide library with carbohydrate fragment. The reaction is chemo- and regioselective for the disulfide-containing phage library.

I anticipate that these libraries, both linear and cyclic glycopeptide, will have a broad range of use in fragment-based ligand discovery, “hot spot” mapping, and selection of potent ligands for many protein targets. Ligands selected from these libraries could serve as a starting point for further hit-to-lead optimizations. The technology, we term it “genetically encoded fragment-based discovery” (GE-FBD), relies on discovery accelerated with an efficient and high-throughput process. It holds promise for discovering potent ligands and novel classes of lead compounds. That being said, it is not a panacea to the challenges facing modern drug discovery, but an alternative to complement and “guide” the rational design of more potent and “drug-like” compounds, when other approaches have failed.

6.2 Future directions

GE-FBD is still in its early day; like many emerging technologies, GE-FBD requires more examples to demonstrate its general application and usefulness. Several projects have been initiated in our lab and are currently on-going. These projects include selection of galactose-containing peptide libraries for the galectin superfamily and selection of arabinose-containing peptide libraries for anti-*Mycobacterium tuberculosis* antibodies.

In the DC-SIGN project (Chapter 4), I was able to obtain inhibitors of moderately improved affinity as compared to monosaccharide and comparable affinity to a trimannoside. The reasons for the relatively unsuccessful selection could be the lack of a second site, or so-called “hot

spot”, around the glycan-binding site and/or the suboptimal design of the library. I propose that a completely different library should be made using fragment such as fucose, which is also a natural ligand for DC-SIGN. Attachment of peptide library to the α -anomeric position of fucose would allow the projection of peptide moieties into a completely different direction. Computational solvent mapping suggested that a putative fragment-binding site is available and could be accessible for fucose-peptide conjugates. This hypothesis would need to be tested with phage selection in the near future.

The next pivotal improvement of this technology is the design of library biased toward restricted conformations. This endeavor could be realized with the generation of macrocyclic glycopeptide libraries. Macrocyclizations of peptide can potentially lead to increased binding affinity by pre-organizing the ligand into its biologically active conformation and eliminating unproductive conformations. I have demonstrated a methodology to create such a library (Chapter 5), though its application for selection remains to be proven.

Macrocyclization of peptides is currently an intense topic of interest in the peptide community. Novel and better reactions for macrocyclization are developed in an ever-increasing pace. Any water-based reaction, which is selective for disulfide and/or aldehyde yet benign to phage, will provide a new avenue for further diversification of the phage library. In view of this, we see endless of possibility and potential.

References

- (1) Barthel, S. R.; Gavino, J. D.; Descheny, L.; Dimitroff, C. J. Targeting selectins and selectin ligands in inflammation and cancer. *Expert Opin. Ther. Targets* **2007**, *11*, 1473.
- (2) Dalotto-Moreno, T.; Croci, D. O.; Cerliani, J. P.; Martinez-Allo, V. C.; Dergan-Dylon, S.; Méndez-Huergo, S. P.; Stupirski, J. C.; Mazal, D.; Osinaga, E.; Toscano, M. A.; Sundblad, V.; Rabinovich, G. A.; Salatino, M. Targeting Galectin-1 Overcomes Breast Cancer-Associated Immunosuppression and Prevents Metastatic Disease. *Cancer Res.* **2013**, *73*, 1107.
- (3) Imberty, A.; Varrot, A. Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* **2008**, *18*, 567.
- (4) Ernst, B.; Magnani, J. L. From carbohydrate leads to glycomimetic drugs. *Nat. Rev. Drug Discov.* **2009**, *8*, 661.
- (5) Cummings, Richard D.; Pierce, J. M. The Challenge and Promise of Glycomics. *Chem. Biol.* **2014**, *21*, 1.
- (6) Fazio, F.; Bryan, M. C.; Blixt, O.; Paulson, J. C.; Wong, C.-H. Synthesis of Sugar Arrays in Microtiter Plate. *J. Am. Chem. Soc.* **2002**, *124*, 14397.
- (7) Bryan, M. C.; Plettenburg, O.; Sears, P.; Rabuka, D.; Wacowich-Sgarbi, S.; Wong, C.-H. Saccharide Display on Microtiter Plates. *Chem. Biol.* **2002**, *9*, 713.
- (8) Wang, D.; Liu, S.; Trummer, B. J.; Deng, C.; Wang, A. Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nat. Biotech.* **2002**, *20*, 275.
- (9) 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. Biotech.* **2002**, *20*, 1011.
- (10) Hart, G. W.; Copeland, R. J. Glycomics Hits the Big Time. *Cell* **2010**, *143*, 672.
- (11) Rillahan, C. D.; Paulson, J. C. Glycan Microarrays for Decoding the Glycome. *Annu. Rev. Biochem* **2011**, *80*, 797.
- (12) Wesener, D. A.; Wangkanont, K.; McBride, R.; Song, X.; Kraft, M. B.; Hodges, H. L.; Zarling, L. C.; Splain, R. A.; Smith, D. F.; Cummings, R. D.; Paulson, J. C.; Forest, K. T.; Kiessling, L. L. Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.* **2015**, *22*, 603.
- (13) Taylor, M. E.; Drickamer, K. Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands. *Glycobiology* **2009**, *19*, 1155.

- (14) Varki, A. Nothing in Glycobiology Makes Sense, except in the Light of Evolution. *Cell* **2006**, *126*, 841.
- (15) Apweiler, R.; Hermjakob, H.; Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database1. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1999**, *1473*, 4.
- (16) Herget, S.; Toukach, P.; Ranzinger, R.; Hull, W.; Knirel, Y.; von der Lieth, C.-W. Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): Characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC Struct. Biol.* **2008**, *8*, 35.
- (17) Bertozzi, C. R.; Rabuka, D. *Structural basis of glycan diversity; Essentials of Glycobiology*, 2nd ed.; Cold Spring Harbor Laboratory Press: New York, 2009.
- (18) Ley, K.; Kansas, G. S. Selectins in T-cell recruitment to non-lymphoid tissues and sites of inflammation. *Nat. Rev. Immunol.* **2004**, *4*, 325.
- (19) Pinho, S. S.; Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **2015**, *15*, 540.
- (20) Croci, Diego O.; Cerliani, Juan P.; Dalotto-Moreno, T.; Méndez-Huergo, Santiago P.; Mascanfroni, Ivan D.; Dergan-Dylon, S.; Toscano, Marta A.; Caramelo, Julio J.; García-Vallejo, Juan J.; Ouyang, J.; Mesri, Enrique A.; Junttila, Melissa R.; Bais, C.; Shipp, Margaret A.; Salatino, M.; Rabinovich, Gabriel A. Glycosylation-Dependent Lectin-Receptor Interactions Preserve Angiogenesis in Anti-VEGF Refractory Tumors. *Cell* **2014**, *156*, 744.
- (21) Hudak, Jason E.; Bertozzi, Carolyn R. Glycotherapy: New Advances Inspire a Reemergence of Glycans in Medicine. *Chem. Biol.* **2014**, *21*, 16.
- (22) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **2002**, *1*, 727.
- (23) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotech.* **2007**, *25*, 71.
- (24) 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*, 13528.
- (25) Lis, H.; Sharon, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* **1998**, *98*, 637.
- (26) Weis, W. I.; Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996**, *65*, 441.
- (27) Mammen, M.; Choi, S.-K.; Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem. Int. Ed.* **1998**, *37*, 2754.

- (28) Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. Synthetic Multivalent Ligands as Probes of Signal Transduction. *Angew. Chem. Int. Ed.* **2006**, *45*, 2348.
- (29) Lundquist, J. J.; Toone, E. J. The Cluster Glycoside Effect. *Chem. Rev.* **2002**, *102*, 555.
- (30) Chabre, Y. M.; Roy, R. Multivalent glycoconjugate syntheses and applications using aromatic scaffolds. *Chem. Soc. Rev.* **2013**, *42*, 4657.
- (31) Cecioni, S.; Imberty, A.; Vidal, S. Glycomimetics versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. *Chem. Rev.* **2015**, *115*, 525.
- (32) Chang, J.; Patton, J. T.; Sarkar, A.; Ernst, B.; Magnani, J. L.; Frenette, P. S. GMI-1070, a novel pan-selectin antagonist, reverses acute vascular occlusions in sickle cell mice. *Blood* **2010**, *116*, 1779.
- (33) Telen, M. J.; Wun, T.; McCavit, T. L.; De Castro, L. M.; Krishnamurti, L.; Lanzkron, S.; Hsu, L. L.; Smith, W. R.; Rhee, S.; Magnani, J. L.; Thackray, H. Randomized phase 2 study of GMI-1070 in SCD: reduction in time to resolution of vaso-occlusive events and decreased opioid use. *Blood* **2015**, *125*, 2656.
- (34) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531.
- (35) Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S. M.; Fong, D.; Zhu, Y.-L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3041.
- (36) Murray, C. W.; Carr, M. G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J. E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenamin, R.; O'Brien, M. A.; Patel, S.; Phillips, T. R.; Williams, G.; Woodhead, A. J.; Woolford, A. J. A. Fragment-Based Drug Discovery Applied to Hsp90. Discovery of Two Lead Series with High Ligand Efficiency. *J. Med. Chem.* **2010**, *53*, 5942.
- (37) Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; Fazal, L.; Feltell, R.; Figueroa, E.; Frederickson, M.; Lewis, J.; McMenamin, R.; Murray, C. W.; O'Brien, M. A.; Parra, L.; Patel, S.; Phillips, T.; Rees, D. C.; Rich, S.; Smith, D.-M.; Trewartha, G.; Vinkovic, M.; Williams, B.; Woolford, A. J. A. Discovery of (2,4-Dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydroisoindol-2-yl]methanone (AT13387), a Novel Inhibitor of the

Molecular Chaperone Hsp90 by Fragment Based Drug Design. *J. Med. Chem.* **2010**, *53*, 5956.

(38) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **2005**, *435*, 677.

(39) Souers, A. J.; Levenson, J. D.; Boghaert, E. R.; Ackler, S. L.; Catron, N. D.; Chen, J.; Dayton, B. D.; Ding, H.; Enschede, S. H.; Fairbrother, W. J.; Huang, D. C. S.; Hymowitz, S. G.; Jin, S.; Khaw, S. L.; Kovar, P. J.; Lam, L. T.; Lee, J.; Maecker, H. L.; Marsh, K. C.; Mason, K. D.; Mitten, M. J.; Nimmer, P. M.; Oleksijew, A.; Park, C. H.; Park, C.-M.; Phillips, D. C.; Roberts, A. W.; Sampath, D.; Seymour, J. F.; Smith, M. L.; Sullivan, G. M.; Tahir, S. K.; Tse, C.; Wendt, M. D.; Xiao, Y.; Xue, J. C.; Zhang, H.; Humerickhouse, R. A.; Rosenberg, S. H.; Elmore, S. W. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **2013**, *19*, 202.

(40) Verdine, G. L.; Walensky, L. D. The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. *Clin. Cancer. Res.* **2007**, *13*, 7264.

(41) <http://practicalfragments.blogspot.ca/2015/01/fragments-in-clinic-2015-edition.html>.

(42) Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **2007**, *6*, 211.

(43) Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **2009**, *1*, 187.

(44) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430.

(45) Murray, C. W.; Erlanson, D. A.; Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H.; Richmond, N. J. Validity of Ligand Efficiency Metrics. *ACS Med. Chem. Lett.* **2014**, *5*, 616.

(46) Krishnamurthy, V. M.; Estroff, L. A.; Whitesides, G. M. In *Fragment-based Approaches in Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: 2006, p 11.

(47) Hung, A. W.; Silvestre, H. L.; Wen, S.; Ciulli, A.; Blundell, T. L.; Abell, C. Application of Fragment Growing and Fragment Linking to the Discovery of Inhibitors of Mycobacterium tuberculosis Pantothenate Synthetase. *Angew. Chem. Int. Ed.* **2009**, *48*, 8452.

- (48) Filbin, M. T. Myelin-associated inhibitors of axonal regeneration in the adult mammalian CNS. *Nat. Rev. Neurosci.* **2003**, *4*, 703.
- (49) Vyas, A. A.; Blixt, O.; Paulson, J. C.; Schnaar, R. L. Potent Glycan Inhibitors of Myelin-associated Glycoprotein Enhance Axon Outgrowth in Vitro. *J. Biol. Chem.* **2005**, *280*, 16305.
- (50) Shelke, S. V.; Cutting, B.; Jiang, X.; Koliwer-Brandl, H.; Strasser, D. S.; Schwardt, O.; Kelm, S.; Ernst, B. A Fragment-Based In Situ Combinatorial Approach To Identify High-Affinity Ligands for Unknown Binding Sites. *Angew. Chem. Int. Ed.* **2010**, *49*, 5721.
- (51) Jahnke, W.; Perez, L. B.; Paris, C. G.; Strauss, A.; Fendrich, G.; Nalin, C. M. Second-Site NMR Screening with a Spin-Labeled First Ligand. *J. Am. Chem. Soc.* **2000**, *122*, 7394.
- (52) Jahnke, W.; Rüdissler, S.; Zurini, M. Spin Label Enhanced NMR Screening. *J. Am. Chem. Soc.* **2001**, *123*, 3149.
- (53) Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radić, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B. Click Chemistry In Situ: Acetylcholinesterase as a Reaction Vessel for the Selective Assembly of a Femtomolar Inhibitor from an Array of Building Blocks. *Angew. Chem. Int. Ed.* **2002**, *41*, 1053.
- (54) Hiratsuka, S.; Goel, S.; Kamoun, W. S.; Maru, Y.; Fukumura, D.; Duda, D. G.; Jain, R. K. Endothelial focal adhesion kinase mediates cancer cell homing to discrete regions of the lungs via E-selectin up-regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3725.
- (55) Egger, J.; Weckerle, C.; Cutting, B.; Schwardt, O.; Rabbani, S.; Lemme, K.; Ernst, B. Nanomolar E-Selectin Antagonists with Prolonged Half-Lives by a Fragment-Based Approach. *J. Am. Chem. Soc.* **2013**, *135*, 9820.
- (56) Chen, W. C.; Kawasaki, N.; Nycholat, C. M.; Han, S.; Pilotte, J.; Crocker, P. R.; Paulson, J. C. Antigen Delivery to Macrophages Using Liposomal Nanoparticles Targeting Sialoadhesin/CD169. *PLoS ONE* **2012**, *7*, e39039.
- (57) Nycholat, C. M.; Rademacher, C.; Kawasaki, N.; Paulson, J. C. In Silico-Aided Design of a Glycan Ligand of Sialoadhesin for in Vivo Targeting of Macrophages. *J. Am. Chem. Soc.* **2012**, *134*, 15696.
- (58) Kawasaki, N.; Vela, J. L.; Nycholat, C. M.; Rademacher, C.; Khurana, A.; van Rooijen, N.; Crocker, P. R.; Kronenberg, M.; Paulson, J. C. Targeted delivery of lipid antigen to macrophages via the CD169/sialoadhesin endocytic pathway induces robust invariant natural killer T cell activation. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 7826.
- (59) Davis, B. J.; Erlanson, D. A. Learning from our mistakes: The 'unknown knowns' in fragment screening. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2844.

- (60) Chung, S.; Parker, J. B.; Bianchet, M.; Amzel, L. M.; Stivers, J. T. Impact of linker strain and flexibility in the design of a fragment-based inhibitor. *Nat. Chem. Biol.* **2009**, *5*, 407.
- (61) Hansen, S.; Erlanson, D. 2011, US Patent Application 20110118126.
- (62) Halpin, D. R.; Harbury, P. B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol.* **2004**, *2*, 1022.
- (63) 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*, 1601.
- (64) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotech.* **2004**, *22*, 568.
- (65) Hansen, M. H.; Blakskjær, P.; Petersen, L. K.; Hansen, T. H.; Højfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. V. A Yoctoliter-Scale DNA Reactor for Small-Molecule Evolution. *J. Am. Chem. Soc.* **2009**, *131*, 1322.
- (66) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuzzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **2009**, *5*, 647.
- (67) Franzini, R. M.; Neri, D.; Scheuermann, J. DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries. *Acc. Chem. Res.* **2014**, *47*, 1247.
- (68) Wichert, M.; Krall, N.; Decurtins, W.; Franzini, R. M.; Pretto, F.; Schneider, P.; Neri, D.; Scheuermann, J. Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* **2015**, *7*, 241.
- (69) Scheuermann, J.; Neri, D. Dual-pharmacophore DNA-encoded chemical libraries. *Curr. Opin. Chem. Biol.* **2015**, *26*, 99.
- (70) Li, S.; Roberts, R. W. A Novel Strategy for In Vitro Selection of Peptide-Drug Conjugates. *Chem. Biol.* **2003**, *10*, 233.
- (71) Erlanson, D. A. Fragment-Based Ligand Discovery Meets Phage Display. *ACS Chem. Biol.* **2007**, *2*, 779.
- (72) Meyer, S. C.; Shomin, C. D.; Gaj, T.; Ghosh, I. Tethering Small Molecules to a Phage Display Library: Discovery of a Selective Bivalent Inhibitor of Protein Kinase A. *J. Am. Chem. Soc.* **2007**, *129*, 13812.

- (73) 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*, 4940.
- (74) Santoso, B.; Lam, S.; Murray, B. W.; Chen, G. A simple and efficient maleimide-based approach for peptide extension with a cysteine-containing peptide phage library. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5680.
- (75) Wang, W.; Hirano, Y.; Uzawa, T.; Liu, M.; Taiji, M.; Ito, Y. In vitro selection of a peptide aptamer that potentiates inhibition of cyclin-dependent kinase 2 by purvalanol. *MedChemComm* **2014**, *5*, 1400.
- (76) Tokunaga, Y.; Azetsu, Y.; Fukunaga, K.; Hatanaka, T.; Ito, Y.; Taki, M. Pharmacophore Generation from a Drug-like Core Molecule Surrounded by a Library Peptide via the 10BASEd-T on Bacteriophage T7. *Molecules* **2014**, *19*, 2481.
- (77) 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*, 5248.
- (78) Koniev, O.; Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **2015**, *44*, 5495.
- (79) Josephson, K.; Ricardo, A.; Szostak, J. W. mRNA display: from basic principles to macrocycle drug discovery. *Drug Discov. Today* **2014**, *19*, 388.
- (80) Ng, S.; Jafari, M. R.; Derda, R. Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry. *ACS Chem. Biol.* **2012**, *7*, 123.
- (81) Heinis, C.; Winter, G. Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.* **2015**, *26*, 89.
- (82) Shomin, C. D.; Meyer, S. C.; Ghosh, I. Staurosporine tethered peptide ligands that target cAMP-dependent protein kinase (PKA): Optimization and selectivity profiling. *Biorg. Med. Chem.* **2009**, *17*, 6196.
- (83) Tsomaia, N. Peptide therapeutics: Targeting the undruggable space. *Eur. J. Med. Chem.* **2015**, *94*, 459.
- (84) Bhat, A.; Roberts, L. R.; Dwyer, J. J. Lead discovery and optimization strategies for peptide macrocycles. *Eur. J. Med. Chem.* **2015**, *94*, 471.
- (85) Fosgerau, K.; Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discov. Today* **2015**, *20*, 122.
- (86) Bird, G. H.; Madani, N.; Perry, A. F.; Princiotta, A. M.; Supko, J. G.; He, X.; Gavathiotis, E.; Sodroski, J. G.; Walensky, L. D. Hydrocarbon double-stapling

remedies the proteolytic instability of a lengthy peptide therapeutic. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14093.

(87) White, C. J.; Yudin, A. K. Contemporary strategies for peptide macrocyclization. *Nat. Chem.* **2011**, *3*, 509.

(88) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. N-Methylation of Peptides: A New Perspective in Medicinal Chemistry. *Acc. Chem. Res.* **2008**, *41*, 1331.

(89) White, T. R.; Renzelman, C. M.; Rand, A. C.; Rezai, T.; McEwen, C. M.; Gelev, V. M.; Turner, R. A.; Lington, R. G.; Leung, S. S. F.; Kalgutkar, A. S.; Bauman, J. N.; Zhang, Y.; Liras, S.; Price, D. A.; Mathiowetz, A. M.; Jacobson, M. P.; Lokey, R. S. On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat. Chem. Biol.* **2011**, *7*, 810.

(90) Chatterjee, J.; Rechenmacher, F.; Kessler, H. N-Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angew. Chem. Int. Ed.* **2013**, *52*, 254.

(91) Smith, G. P.; Petrenko, V. A. Phage display. *Chem. Rev.* **1997**, *97*, 391.

(92) Tanaka, F.; Barbas III, C. F. Phage display selection of peptides possessing aldolase activity. *Chem. Commun.* **2001**, 769.

(93) Nelson, A. L.; Dhimolea, E.; Reichert, J. M. Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discov.* **2010**, *9*, 767.

(94) Frankel, A.; Li, S.; Starck, S. R.; Roberts, R. W. Unnatural RNA display libraries. *Curr. Opin. Struct. Biol.* **2003**, *13*, 506.

(95) Jespers, L.; Bonnert, T. P.; Winter, G. Selection of optical biosensors from chemisynthetic antibody libraries. *Protein Eng. Des. Sel.* **2004**, *17*, 709.

(96) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009**, *5*, 502.

(97) Angelini, A.; Cendron, L.; Chen, S.; Touati, J.; Winter, G.; Zanotti, G.; Heinis, C. Bicyclic peptide inhibitor reveals large contact interface with a protease target. *ACS Chem. Biol.* **2012**, *7*, 817.

(98) Angelini, A.; Heinis, C. Post-translational modification of genetically encoded polypeptide libraries. *Curr. Opin. Chem. Biol.* **2011**, *15*, 355.

(99) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **2005**, *44*, 7342.

(100) Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997**, *15*, 553.

- (101) Wilson, D. S.; Keefe, A. D.; Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3750.
- (102) Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. In vitro selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 11779.
- (103) Sandman, K. E.; Benner, J. S.; Noren, C. J. Phage display of selenopeptides. *J. Am. Chem. Soc.* **2000**, *122*, 960.
- (104) Saleh, L.; Noren, C. J. In *Phage Nanobiotechnology*; V. A. Petrenko, G. P. Smith ed.; The Royal Society of Chemistry: Cambridge, 2011, p 202.
- (105) Tian, F.; Tsao, M.-L.; Schultz, P. G. A phage display system with unnatural amino acids. *J. Am. Chem. Soc.* **2004**, *126*, 15962.
- (106) Derda, R.; Tang, S. K. Y.; Li, S. C.; Ng, S.; Matochko, W.; Jafari, M. R. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* **2011**, *16*, 1776.
- (107) Geoghegan, K. F.; Stroh, J. G. Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjugate Chem.* **1992**, *3*, 138.
- (108) Matochko, W. L.; Chu, K.; Jin, B.; Lee, S. W.; Whitesides, G. M.; Derda, R. Deep sequencing analysis of phage libraries using Illumina platform. *Methods* **2012**, *58*, 47.
- (109) Li, K.; Chen, Y.; Li, S.; Nguyen, H. G.; Niu, Z.; You, S.; Mello, C. M.; Lu, X.; Wang, Q. Chemical modification of M13 bacteriophage and its application in cancer cell imaging. *Bioconjugate Chem.* **2010**, *21*, 1369.
- (110) Jespers, L. S. A. T.; Winter, G. P.; Bonnert, T. P.; Simon, T. M. 2000, US Patent 6017732.
- (111) 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*, 834.
- (112) Love, K. R.; Swoboda, J. G.; Noren, C. J.; Walker, S. Enabling glycosyltransferase evolution: a facile substrate-attachment strategy for phage-display enzyme evolution. *ChemBioChem* **2006**, *7*, 753.
- (113) Derda, R.; Tang, S. K. Y.; Whitesides, G. M. Uniform amplification of phage with different growth characteristics in individual compartments consisting of monodisperse droplets. *Angew. Chem. Int. Ed.* **2010**, *49*, 5301.
- (114) Cordes, E. H.; Jencks, W. P. Nucleophilic catalysis of semicarbazone formation by anilines. *J. Am. Chem. Soc.* **1962**, *84*, 826.
- (115) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. Nucleophilic catalysis of oxime ligation. *Angew. Chem. Int. Ed.* **2006**, *45*, 7581.

- (116) Smith, G. P.; Petrenko, V. A. Phage Display. *Chem. Rev.* **1997**, *97*, 391.
- (117) Kadam, R. U.; Bergmann, M.; Hurley, M.; Garg, D.; Cacciarini, M.; Swiderska, M. A.; Nativi, C.; Sattler, M.; Smyth, A. R.; Williams, P.; Cámara, M.; Stocker, A.; Darbre, T.; Reymond, J.-L. A glycopeptide dendrimer inhibitor of the galactose-specific lectin LecA and of *Pseudomonas aeruginosa* biofilms. *Angew. Chem. Int. Ed.* **2011**, *50*, 10631.
- (118) Maljaars, C. E. P.; André, S.; Halkes, K. M.; Gabius, H.-J.; Kamerling, J. P. Assessing the inhibitory potency of galectin ligands identified from combinatorial (glyco)peptide libraries using surface plasmon resonance spectroscopy. *Anal. Biochem.* **2008**, *378*, 190.
- (119) Scholle, M. D.; Kehoe, J. W.; Kay, B. K. Efficient construction of a large collection of phage-displayed combinatorial peptide libraries *Comb. Chem. High Throughput Screening* **2005**, *8*, 545.
- (120) Rose, K.; Chen, J.; Dragovic, M.; Zeng, W.; Jeannerat, D.; Kamalaprija, P.; Burger, U. New cyclization reaction at the amino terminus of peptides and proteins. *Bioconjugate Chem.* **1999**, *10*, 1038.
- (121) Hudak, J. E.; Yu, H. H.; Bertozzi, C. R. Protein glycoengineering enabled by the versatile synthesis of aminoxy glycans and the genetically encoded aldehyde tag. *J. Am. Chem. Soc.* **2011**, *133*, 16127.
- (122) Sears, P.; Wong, C.-H. Carbohydrate Mimetics: A New Strategy for Tackling the Problem of Carbohydrate-Mediated Biological Recognition. *Angew. Chem. Int. Ed.* **1999**, *38*, 2300.
- (123) Bernardi, A.; Cheshev, P. Interfering with the Sugar Code: Design and Synthesis of Oligosaccharide Mimics. *Chem. Eur. J.* **2008**, *14*, 7434.
- (124) Andreini, M.; Doknic, D.; Sutkeviciute, I.; Reina, J. J.; Duan, J.; Chabrol, E.; Thepaut, M.; Moroni, E.; Doro, F.; Belvisi, L.; Weiser, J.; Rojo, J.; Fieschi, F.; Bernardi, A. Second generation of fucose-based DC-SIGN ligands : affinity improvement and specificity versus Langerin. *Org. Biomol. Chem.* **2011**, *9*, 5778.
- (125) Hauck, D.; Joachim, I.; Frommeyer, B.; Varrot, A.; Philipp, B.; Möller, H. M.; Imberty, A.; Exner, T. E.; Titz, A. Discovery of Two Classes of Potent Glycomimetic Inhibitors of *Pseudomonas aeruginosa* LecB with Distinct Binding Modes. *ACS Chem. Biol.* **2013**, *8*, 1775.
- (126) Kadam, R. U.; Bergmann, M.; Garg, D.; Gabrieli, G.; Stocker, A.; Darbre, T.; Reymond, J.-L. Structure-Based Optimization of the Terminal Tripeptide in Glycopeptide Dendrimer Inhibitors of *Pseudomonas aeruginosa* Biofilms Targeting LecA. *Chem. Eur. J.* **2013**, *19*, 17054.
- (127) Bächle, D.; Loers, G.; Guthöhrlein, E. W.; Schachner, M.; Sewald, N. Glycomimetic Cyclic Peptides Stimulate Neurite Outgrowth. *Angew. Chem. Int. Ed.* **2006**, *45*, 6582.

- (128) Borrok, M. J.; Kiessling, L. L. Non-carbohydrate Inhibitors of the Lectin DC-SIGN. *J. Am. Chem. Soc.* **2007**, *129*, 12780.
- (129) Kiessling, L. L.; Splain, R. A. Chemical Approaches to Glycobiology. *Annu. Rev. Biochem* **2010**, *79*, 619.
- (130) Johnson, M. A.; Pinto, B. M. Molecular Mimicry of Carbohydrates by Peptides. *Aust. J. Chem.* **2002**, *55*, 13.
- (131) Vyas, N. K.; Vyas, M. N.; Chervenak, M. C.; Bundle, D. R.; Pinto, B. M.; Quioco, F. A. Structural basis of peptide-carbohydrate mimicry in an antibody-combining site. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15023.
- (132) Oldenburg, K. R.; Loganathan, D.; Goldstein, I. J.; Schultz, P. G.; Gallop, M. A. Peptide ligands for a sugar-binding protein isolated from a random peptide library. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5393.
- (133) Scott, J. K.; Loganathan, D.; Easley, R. B.; Gong, X.; Goldstein, I. J. A family of concanavalin A-binding peptides from a hexapeptide epitope library. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5398.
- (134) Yu, L.; Yu, P. S.; Yee Yen Mui, E.; McKelvie, J. C.; Pham, T. P. T.; Yap, Y. W.; Wong, W. Q.; Wu, J.; Deng, W.; Orner, B. P. Phage display screening against a set of targets to establish peptide-based sugar mimetics and molecular docking to predict binding site. *Biorg. Med. Chem.* **2009**, *17*, 4825.
- (135) Matsubara, T. Potential of Peptides as Inhibitors and Mimotopes: Selection of Carbohydrate-Mimetic Peptides from Phage Display Libraries. *Journal of Nucleic Acids* **2012**, *2012*, 15.
- (136) Fukuda, M. N. Peptide-displaying phage technology in glycobiology. *Glycobiology* **2012**, *22*, 318.
- (137) Jain, D.; Kaur, K.; Sundaravadivel, B.; 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*, 16098.
- (138) 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*, 5407.
- (139) St. Hilaire, P. M.; Lowary, T. L.; Meldal, M.; Bock, K. Oligosaccharide Mimetics Obtained by Novel, Rapid Screening of Carboxylic Acid Encoded Glycopeptide Libraries. *J. Am. Chem. Soc.* **1998**, *120*, 13312.
- (140) Ying, L.; Liu, R.; Zhang, J.; Lam, K.; Lebrilla, C. B.; Gervay-Hague, J. A Topologically Segregated One-Bead-One-Compound Combinatorial Glycopeptide Library for Identification of Lectin Ligands. *J. Comb. Chem.* **2005**, *7*, 372.

- (141) 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*, 1482.
- (142) Schneider, T. D.; Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097.
- (143) Dam, T. K.; Oscarson, S.; Brewer, C. F. Thermodynamics of Binding of the Core Trimannoside of Asparagine-linked Carbohydrates and Deoxy Analogs to Dioclea grandiflora Lectin. *J. Biol. Chem.* **1998**, *273*, 32812.
- (144) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622.
- (145) Kadirvelraj, R.; Grant, O. C.; Goldstein, I. J.; Winter, H. C.; Tateno, H.; Fadda, E.; Woods, R. J. Structure and binding analysis of Polyporus squamosus lectin in complex with the Neu5Ac α 2-6Gal β 1-4GlcNAc human-type influenza receptor. *Glycobiology* **2011**, *21*, 973.
- (146) *Hydrogen bonding: The last mystery in drug design*; Kubinyi, H., Ed.; Verlag Helvetica Chimica Acta: Zürich, 2001.
- (147) Pilobello, K. T.; Agrawal, P.; Rouse, R.; Mahal, L. K. Advances in Lectin Microarray Technology: Optimized Protocols for Piezoelectric Print Conditions. *Curr. Protoc. Chem. Biol.* **2013**, *5*, 1.
- (148) van Kooyk, Y.; Geijtenbeek, T. B. H. DC-SIGN: escape mechanism for pathogens. *Nat. Rev. Immunol.* **2003**, *3*, 697.
- (149) El-Hawiet, A.; Kitova, E. N.; Klassen, J. S. Quantifying Carbohydrate-Protein Interactions by Electrospray Ionization Mass Spectrometry Analysis. *Biochemistry* **2012**, *51*, 4244.
- (150) Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I. Structural Basis for Selective Recognition of Oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **2001**, *294*, 2163.
- (151) Mandal, D. K.; Brewer, C. F. Differences in the binding affinities of dimeric concanavalin A (including acetyl and succinyl derivatives) and tetrameric concanavalin A with large oligomannose-type glycopeptides. *Biochemistry* **1993**, *32*, 5116.
- (152) Schwarz, F. P.; Puri, K. D.; Bhat, R. G.; Surolia, A. Thermodynamics of monosaccharide binding to concanavalin A, pea (*Pisum sativum*) lectin, and lentil (*Lens culinaris*) lectin. *J. Biol. Chem.* **1993**, *268*, 7668.
- (153) Noren, K. A.; Noren, C. J. Construction of High-Complexity Combinatorial Phage Display Peptide Libraries. *Methods* **2001**, *23*, 169.
- (154) 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*, 1784.

- (155) Liu, M.; Mao, X.-a.; Ye, C.; Huang, H.; Nicholson, J. K.; Lindon, J. C. Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. *J. Magn. Reson.* **1998**, *132*, 125.
- (156) Steinman, R. M.; Banchereau, J. Taking dendritic cells into medicine. *Nature* **2007**, *449*, 419.
- (157) Palucka, K.; Banchereau, J. Cancer immunotherapy via dendritic cells. *Nat. Rev. Cancer* **2012**, *12*, 265.
- (158) Trumpfheller, C.; Finke, J. S.; López, C. B.; Moran, T. M.; Moltedo, B.; Soares, H.; Huang, Y.; Schlesinger, S. J.; Park, C. G.; Nussenzweig, M. C.; Granelli-Piperno, A.; Steinman, R. M. Intensified and protective CD4⁺ T cell immunity in mice with anti-dendritic cell HIV gag fusion antibody vaccine. *J. Exp. Med.* **2006**, *203*, 607.
- (159) Soares, H.; Waechter, H.; Glaichenhaus, N.; Mougneau, E.; Yagita, H.; Mizenina, O.; Dudziak, D.; Nussenzweig, M. C.; Steinman, R. M. A subset of dendritic cells induces CD4⁺ T cells to produce IFN- γ by an IL-12 - independent but CD70-dependent mechanism in vivo. *J. Exp. Med.* **2007**, *204*, 1095.
- (160) Unger, W. W. J.; van Beelen, A. J.; Bruijns, S. C.; Joshi, M.; Fehres, C. M.; van Bloois, L.; Verstege, M. I.; Ambrosini, M.; Kalay, H.; Nazmi, K.; Bolscher, J. G.; Hooijberg, E.; de Gruijl, T. D.; Storm, G.; van Kooyk, Y. Glycan-modified liposomes boost CD4⁺ and CD8⁺ T-cell responses by targeting DC-SIGN on dendritic cells. *J. Controlled Release* **2012**, *160*, 88.
- (161) Geijtenbeek, T. B. H.; Torensma, R.; van Vliet, S. J.; van Duijnhoven, G. C. F.; Adema, G. J.; van Kooyk, Y.; Figdor, C. G. Identification of DC-SIGN, a Novel Dendritic Cell-Specific ICAM-3 Receptor that Supports Primary Immune Responses. *Cell* **2000**, *100*, 575.
- (162) Engering, A.; Geijtenbeek, T. B. H.; van Vliet, S. J.; Wijers, M.; van Liempt, E.; Demarex, N.; Lanzavecchia, A.; Fransen, J.; Figdor, C. G.; Piguet, V.; van Kooyk, Y. The Dendritic Cell-Specific Adhesion Receptor DC-SIGN Internalizes Antigen for Presentation to T Cells. *The Journal of Immunology* **2002**, *168*, 2118.
- (163) Drickamer, K.; Taylor, M. E. Recent insights into structures and functions of C-type lectins in the immune system. *Curr. Opin. Struct. Biol.* **2015**, *34*, 26.
- (164) Aarnoudse, C. A.; Bax, M.; Sánchez-Hernández, M.; García-Vallejo, J. J.; van Kooyk, Y. Glycan modification of the tumor antigen gp100 targets DC-SIGN to enhance dendritic cell induced antigen presentation to T cells. *Int. J. Cancer* **2008**, *122*, 839.
- (165) Singh, S. K.; Stephani, J.; Schaefer, M.; Kalay, H.; García-Vallejo, J. J.; den Haan, J.; Saeland, E.; Sparwasser, T.; van Kooyk, Y. Targeting glycan modified

OVA to murine DC-SIGN transgenic dendritic cells enhances MHC class I and II presentation. *Mol. Immunol.* **2009**, *47*, 164.

(166) Fehres, C. M.; van Beelen, A. J.; Bruijns, S. C. M.; Ambrosini, M.; Kalay, H.; Bloois, L. v.; Unger, W. W. J.; Garcia-Vallejo, J. J.; Storm, G.; de Gruijl, T. D.; Kooyk, Y. v. In situ Delivery of Antigen to DC-SIGN+CD14+ Dermal Dendritic Cells Results in Enhanced CD8+ T-Cell Responses. *J. Invest. Dermatol.* **2015**, *135*, 2228.

(167) Geijtenbeek, T. B. H.; Kwon, D. S.; Torensma, R.; van Vliet, S. J.; van Duijnhoven, G. C. F.; Middel, J.; Cornelissen, I. L. M. H. A.; Nottet, H. S. L. M.; KewalRamani, V. N.; Littman, D. R.; Figdor, C. G.; van Kooyk, Y. DC-SIGN, a Dendritic Cell-Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells. *Cell* **2000**, *100*, 587.

(168) Martínez-Ávila, O.; Bedoya, L. M.; Marradi, M.; Clavel, C.; Alcamí, J.; Penadés, S. Multivalent Manno-Glyconanoparticles Inhibit DC-SIGN-Mediated HIV-1 Trans-Infection of Human T Cells. *ChemBioChem* **2009**, *10*, 1806.

(169) Sattin, S.; Daggetti, A.; Thépaut, M.; Berzi, A.; Sánchez-Navarro, M.; Tabarani, G.; Rojo, J.; Fieschi, F.; Clerici, M.; Bernardi, A. Inhibition of DC-SIGN-Mediated HIV Infection by a Linear Trimannoside Mimic in a Tetravalent Presentation. *ACS Chem. Biol.* **2010**, *5*, 301.

(170) Bernardi, A.; Jimenez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T.; Fieschi, F.; Finne, J.; Funken, H.; Jaeger, K.-E.; Lahmann, M.; Lindhorst, T. K.; Marradi, M.; Messner, P.; Molinaro, A.; Murphy, P. V.; Nativi, C.; Oscarson, S.; Penades, S.; Peri, F.; Pieters, R. J.; Renaudet, O.; Reymond, J.-L.; Richichi, B.; Rojo, J.; Sansone, F.; Schaffer, C.; Turnbull, W. B.; Velasco-Torrijos, T.; Vidal, S.; Vincent, S.; Wennekes, T.; Zuilhof, H.; Imberty, A. Multivalent glycoconjugates as anti-pathogenic agents. *Chem. Soc. Rev.* **2013**, *42*, 4709.

(171) Reina, J. J.; Sattin, S.; Invernizzi, D.; Mari, S.; Martínez-Prats, L.; Tabarani, G.; Fieschi, F.; Delgado, R.; Nieto, P. M.; Rojo, J.; Bernardi, A. 1,2-Mannobioside Mimic: Synthesis, DC-SIGN Interaction by NMR and Docking, and Antiviral Activity. *ChemMedChem* **2007**, *2*, 1030.

(172) Timpano, G.; Tabarani, G.; Anderluh, M.; Invernizzi, D.; Vasile, F.; Potenza, D.; Nieto, P. M.; Rojo, J.; Fieschi, F.; Bernardi, A. Synthesis of Novel DC-SIGN Ligands with an α -Fucosylamide Anchor. *ChemBioChem* **2008**, *9*, 1921.

(173) Garber, K. C. A.; Wangkanont, K.; Carlson, E. E.; Kiessling, L. L. A general glycomimetic strategy yields non-carbohydrate inhibitors of DC-SIGN. *Chem. Commun.* **2010**, *46*, 6747.

(174) Obermajer, N.; Sattin, S.; Colombo, C.; Bruno, M.; Švajger, U.; Anderluh, M.; Bernardi, A. Design, synthesis and activity evaluation of mannose-based DC-SIGN antagonists. *Mol. Divers.* **2011**, *15*, 347.

- (175) Prost, L. R.; Grim, J. C.; Tonelli, M.; Kiessling, L. L. Noncarbohydrate Glycomimetics and Glycoprotein Surrogates as DC-SIGN Antagonists and Agonists. *ACS Chem. Biol.* **2012**, *7*, 1603.
- (176) Varga, N.; Sutkevičiute, I.; Guzzi, C.; McGeagh, J.; Petit-Haertlein, I.; Gugliotta, S.; Weiser, J.; Angulo, J.; Fieschi, F.; Bernardi, A. Selective Targeting of Dendritic Cell-Specific Intercellular Adhesion Molecule-3-Grabbing Nonintegrin (DC-SIGN) with Mannose-Based Glycomimetics: Synthesis and Interaction Studies of Bis(benzylamide) Derivatives of a Pseudomannobioside. *Chem. Eur. J.* **2013**, *19*, 4786.
- (177) Mangold, S. L.; Prost, L. R.; Kiessling, L. L. Quinoxalinone inhibitors of the lectin DC-SIGN. *Chem. Sci.* **2012**, *3*, 772.
- (178) Ordanini, S.; Varga, N.; Porkolab, V.; Thepaut, M.; Belvisi, L.; Bertaglia, A.; Palmioli, A.; Berzi, A.; Trabattoni, D.; Clerici, M.; Fieschi, F.; Bernardi, A. Designing nanomolar antagonists of DC-SIGN-mediated HIV infection: ligand presentation using molecular rods. *Chem. Commun.* **2015**, *51*, 3816.
- (179) Ng, S.; Tjhung, K.; Paschal, B.; Noren, C.; Derda, R. In *Peptide Libraries*; Derda, R., Ed.; Springer New York: 2015; Vol. 1248, p 155.
- (180) Tomašić, T.; Hajšek, D.; Švajger, U.; Luzar, J.; Obermajer, N.; Petit-Haertlein, I.; Fieschi, F.; Anderluh, M. Monovalent mannose-based DC-SIGN antagonists: Targeting the hydrophobic groove of the receptor. *Eur. J. Med. Chem.* **2014**, *75*, 308.
- (181) Deng, L.; Sun, N.; Kitova, E. N.; Klassen, J. S. Direct Quantification of Protein–Metal Ion Affinities by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 2170.
- (182) Kozakov, D.; Hall, D. R.; Napoleon, R. L.; Yueh, C.; Whitty, A.; Vajda, S. New Frontiers in Druggability. *J. Med. Chem.* **2015**.
- (183) Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621.
- (184) 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. Protocols* **2015**, *10*, 733.
- (185) Sörme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J. Structural and Thermodynamic Studies on Cation– Π Interactions in Lectin–Ligand Complexes: High-Affinity Galectin-3 Inhibitors through Fine-Tuning of an Arginine–Arene Interaction. *J. Am. Chem. Soc.* **2005**, *127*, 1737.
- (186) Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D. A.; Alvarez, R.; Blixt, O.; Taylor, M. E.; Weis, W. I.; Drickamer, K. Structural basis for distinct ligand-

binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* **2004**, *11*, 591.

(187) Thépaut, M.; Guzzi, C.; Sutkeviciute, I.; Sattin, S.; Ribeiro-Viana, R.; Varga, N.; Chabrol, E.; Rojo, J.; Bernardi, A.; Angulo, J.; Nieto, P. M.; Fieschi, F. Structure of a Glycomimetic Ligand in the Carbohydrate Recognition Domain of C-type Lectin DC-SIGN. Structural Requirements for Selectivity and Ligand Design. *J. Am. Chem. Soc.* **2013**, *135*, 2518.

(188) Feinberg, H.; Castelli, R.; Drickamer, K.; Seeberger, P. H.; Weis, W. I. Multiple Modes of Binding Enhance the Affinity of DC-SIGN for High Mannose N-Linked Glycans Found on Viral Glycoproteins. *J. Biol. Chem.* **2007**, *282*, 4202.

(189) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849.

(190) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery – an underexploited structural class. *Nat. Rev. Drug Discov.* **2008**, *7*, 608.

(191) Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco Jr, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. How proteins bind macrocycles. *Nat. Chem. Biol.* **2014**, *10*, 723.

(192) Marsault, E.; Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* **2011**, *54*, 1961.

(193) Giordanetto, F.; Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *J. Med. Chem.* **2014**, *57*, 278.

(194) Szewczuk, Z.; Gibbs, B. F.; Yue, S. Y.; Purisima, E. O.; Konishi, Y. Conformationally restricted thrombin inhibitors resistant to proteolytic digestion. *Biochemistry* **1992**, *31*, 9132.

(195) Clark, R. J.; Fischer, H.; Dempster, L.; Daly, N. L.; Rosengren, K. J.; Nevin, S. T.; Meunier, F. A.; Adams, D. J.; Craik, D. J. Engineering stable peptide toxins by means of backbone cyclization: Stabilization of the α -conotoxin MII. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13767.

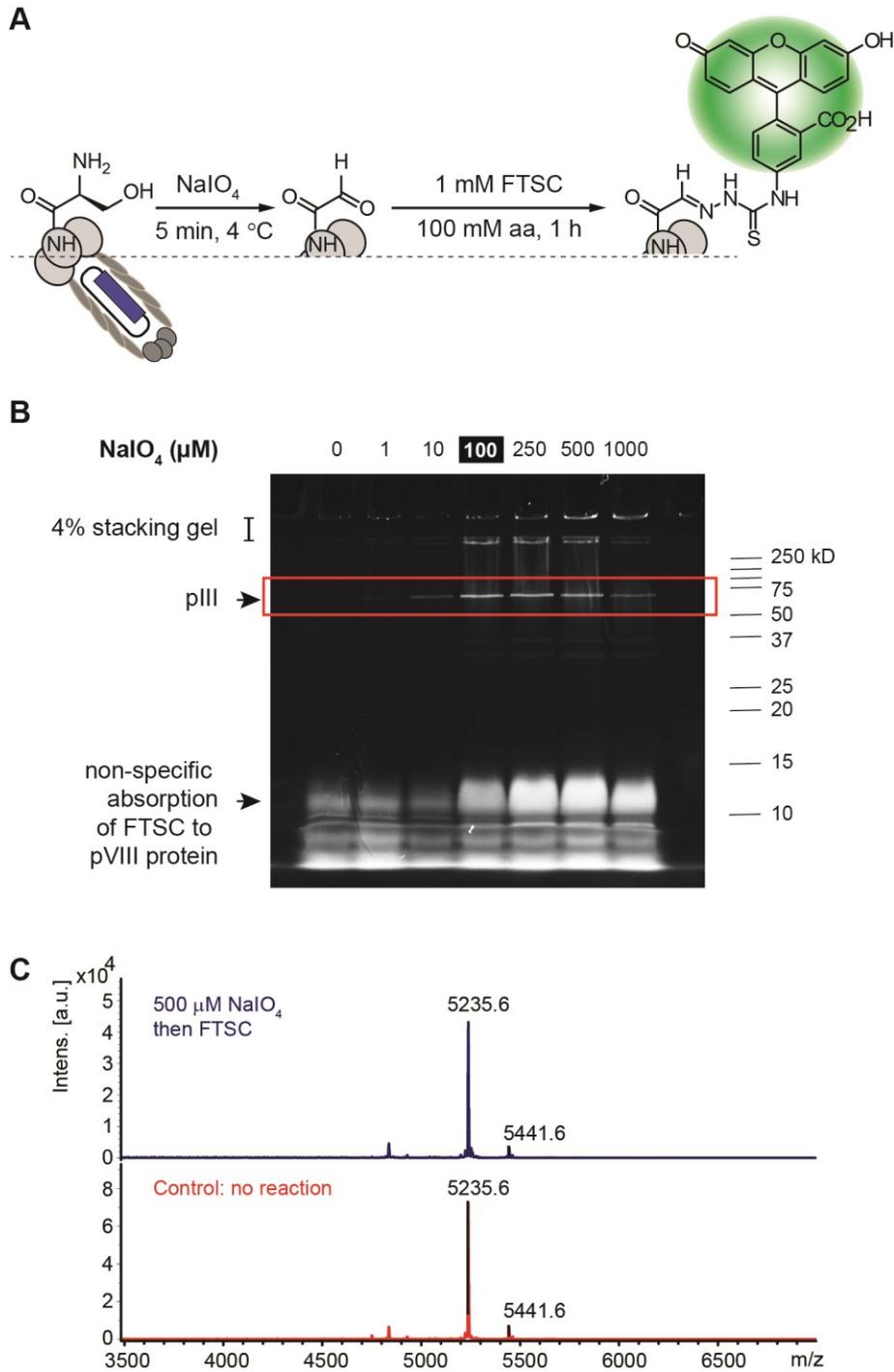
(196) Lau, Y. H.; de Andrade, P.; Quah, S.-T.; Rossmann, M.; Laraia, L.; Skold, N.; Sum, T. J.; Rowling, P. J. E.; Joseph, T. L.; Verma, C.; Hyvonen, M.; Itzhaki, L. S.; Venkitaraman, A. R.; Brown, C. J.; Lane, D. P.; Spring, D. R. Functionalised staple linkages for modulating the cellular activity of stapled peptides. *Chem. Sci.* **2014**, *5*, 1804.

(197) Kessler, H. Conformation and Biological Activity of Cyclic Peptides. *Angew. Chem. Int. Ed.* **1982**, *21*, 512.

- (198) Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; Wagner, G.; Verdine, G. L.; Korsmeyer, S. J. Activation of Apoptosis in Vivo by a Hydrocarbon-Stapled BH3 Helix. *Science* **2004**, *305*, 1466.
- (199) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nat. Chem. Biol.* **2014**, *10*, 716.
- (200) Hill, T. A.; Shepherd, N. E.; Diness, F.; Fairlie, D. P. Constraining Cyclic Peptides To Mimic Protein Structure Motifs. *Angew. Chem. Int. Ed.* **2014**, *53*, 13020.
- (201) Alex, A.; Millan, D. S.; Perez, M.; Wakenhut, F.; Whitlock, G. A. Intramolecular hydrogen bonding to improve membrane permeability and absorption in beyond rule of five chemical space. *MedChemComm* **2011**, *2*, 669.
- (202) Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S. Testing the Conformational Hypothesis of Passive Membrane Permeability Using Synthetic Cyclic Peptide Diastereomers. *J. Am. Chem. Soc.* **2006**, *128*, 2510.
- (203) Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. Conformational Flexibility, Internal Hydrogen Bonding, and Passive Membrane Permeability: Successful in Silico Prediction of the Relative Permeabilities of Cyclic Peptides. *J. Am. Chem. Soc.* **2006**, *128*, 14073.
- (204) Georghiou, G.; Kleiner, R. E.; Pulkoski-Gross, M.; Liu, D. R.; Seeliger, M. A. Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles. *Nat. Chem. Biol.* **2012**, *8*, 366.
- (205) 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*, 10469.
- (206) Passioura, T.; Katoh, T.; Goto, Y.; Suga, H. Selection-Based Discovery of Druglike Macrocylic Peptides. *Annu. Rev. Biochem* **2014**, *83*, 727.
- (207) Lam, K. S.; Liu, R.; Miyamoto, S.; Lehman, A. L.; Tuscano, J. M. Applications of One-Bead One-Compound Combinatorial Libraries and Chemical Microarrays in Signal Transduction Research. *Acc. Chem. Res.* **2003**, *36*, 370.
- (208) Gartner, Z. J.; Liu, D. R. The Generality of DNA-Templated Synthesis as a Basis for Evolving Non-Natural Small Molecules. *J. Am. Chem. Soc.* **2001**, *123*, 6961.
- (209) Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985**, *228*, 1315.
- (210) Lin, H.; Thayer, D. A.; Wong, C.-H.; Walsh, C. T. Macrolactamization of Glycosylated Peptide Thioesters by the Thioesterase Domain of Tyrocidine Synthetase. *Chem. Biol.* **2004**, *11*, 1635.

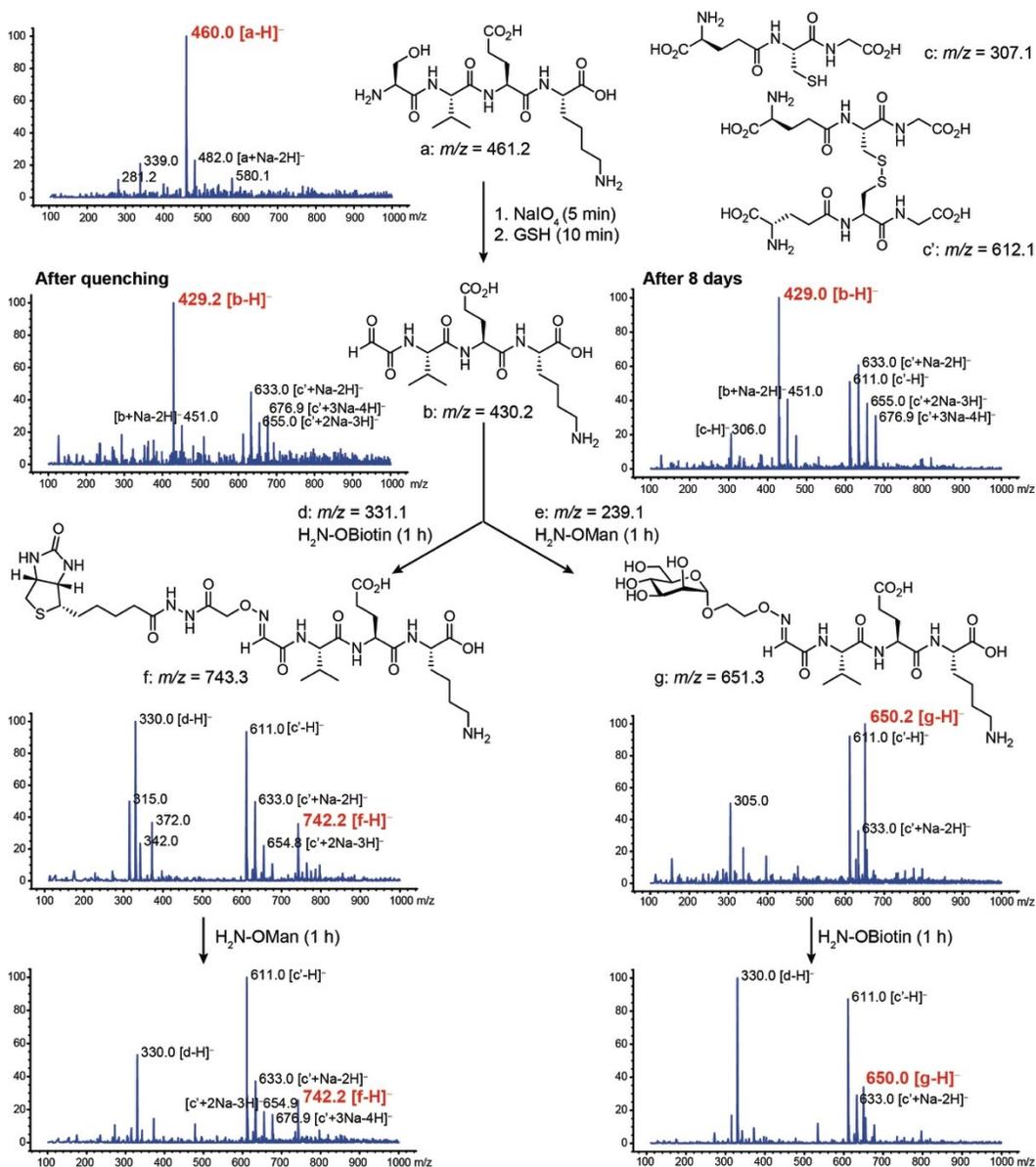
- (211) Hu, H.; Xue, J.; Swarts, B. M.; Wang, Q.; Wu, Q.; Guo, Z. Synthesis and Antibacterial Activities of N-Glycosylated Derivatives of Tyrocidine A, a Macrocyclic Peptide Antibiotic. *J. Med. Chem.* **2009**, *52*, 2052.
- (212) Kay, B. K.; Adey, N. B.; Yun-Sheng, H.; Manfredi, J. P.; Mataragnon, A. H.; Fowlkes, D. M. An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **1993**, *128*, 59.
- (213) McConnell, S.; Uveges, A.; Fowlkes, D.; Spinella, D. Construction and screening of M13 phage libraries displaying long random peptides. *Mol. Divers.* **1996**, *1*, 165.
- (214) Wind, T.; Kjær, S.; Clark, B. F. C. Display of Ras on filamentous phage through cysteine replacement. *Biochimie* **1999**, *81*, 1079.
- (215) Chen, S.; Rentero Rebollo, I.; Buth, S. A.; Morales-Sanfrutos, J.; Touati, J.; Leiman, P. G.; Heinis, C. Bicyclic Peptide Ligands Pulled out of Cysteine-Rich Peptide Libraries. *J. Am. Chem. Soc.* **2013**, *135*, 6562.
- (216) Chen, S.; Heinis, C. In *Peptide Libraries*; Derda, R., Ed.; Springer New York: 2015; Vol. 1248, p 119.
- (217) 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*, 443.
- (218) Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. Peptide Ligands Stabilized by Small Molecules. *Angew. Chem. Int. Ed.* **2014**, *53*, 1602.
- (219) Assem, N.; Ferreira, D. J.; Wolan, D. W.; Dawson, P. E. Acetone-Linked Peptides: A Convergent Approach for Peptide Macrocyclization and Labeling. *Angew. Chem. Int. Ed.* **2015**, *54*, 8665.
- (220) 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*, 8149.
- (221) Saito, F.; Noda, H.; Bode, J. W. Critical Evaluation and Rate Constants of Chemoselective Ligation Reactions for Stoichiometric Conjugations in Water. *ACS Chem. Biol.* **2015**, *10*, 1026.
- (222) Chen, S.; Touati, J.; Heinis, C. Tracking chemical reactions on the surface of filamentous phage using mass spectrometry. *Chem. Commun.* **2014**, *50*, 5267.

Appendix A: Supporting information for chapter 2

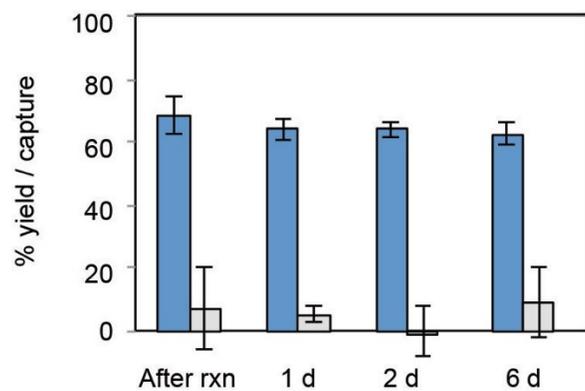


Appendix A-1. Monitoring the progress of reactions on phage with fluorescent densitometry in SDS-PAGE yields poorly reproducible results and it could lead to misinterpretation of the coupling results. (A) The oxidation of

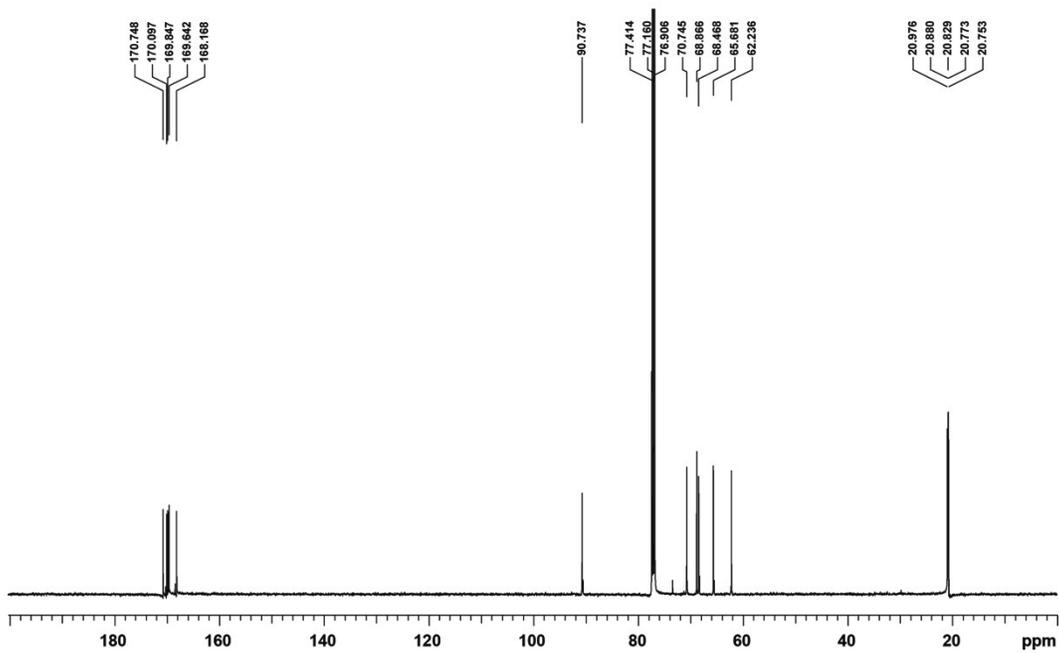
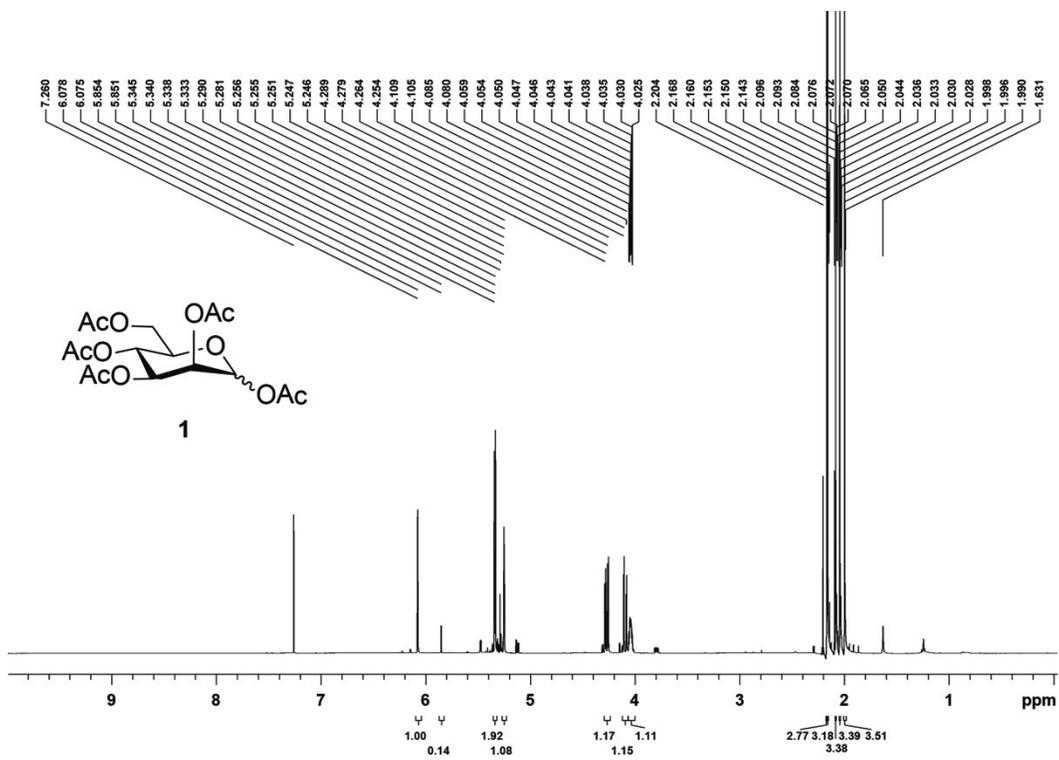
SVEK phage with different concentrations of NaIO_4 . The intermediate glyoxyl was ligated with fluorescein-5-thiosemicarbazide (FTSC). Abbreviation: aa = anilinium acetate (pH 4.7). (B) SDS-PAGE analysis of pIII protein of phage labeled with FTSC. The fluorophores were removed by gel filtration on Zeba spin desalting column (7k MWCO, Thermo Scientific). Phage was then denatured by boiling in Laemmli buffer followed by separation on 13% SDS-PAGE gel. Labeling of pIII protein was dependent on the concentration of NaIO_4 , and the trend I observed here was similar to optimization performed in Figure 2-4 (chapter two). Although SDS-PAGE was useful to map out reaction conditions, overall, I found this technique inconvenient and poorly reproducible. The main problem was low intensity of the band that corresponds to minor coat proteins pIII. It was marginally visible when $>10^{11}$ phage particles were loaded on the gel. Increasing this intensity was impossible because gel was "overloaded" by major coat protein pVIII. It is known that, in SDS-PAGE, significant fraction of the proteins does not enter the gel and remains as aggregates both at the well and stacking gel. Interpretation was plagued by multiple bands, which could correspond to aggregates of pVIII. I observed that FTSC non-specifically binds to pVIII, and thus leading to false-positive interpretation about the regioselectivity of the reaction. (C) MALDI-TOF spectra showing the monoisotopic mass of pVIII from phage treated with $500 \mu\text{M}$ NaIO_4 followed by FTSC (top spectrum); the spectra is identical to a spectra of non-treated phage (bottom spectrum) showing the mass of pVIII. Calculated monoisotopic mass for pVIII $[\text{M}+\text{H}]^+$ $m/z = 5235.7$, found 5235.6. No peak was found to contain FTSC bound covalently to the pVIII ($\text{M}+421$ Da).



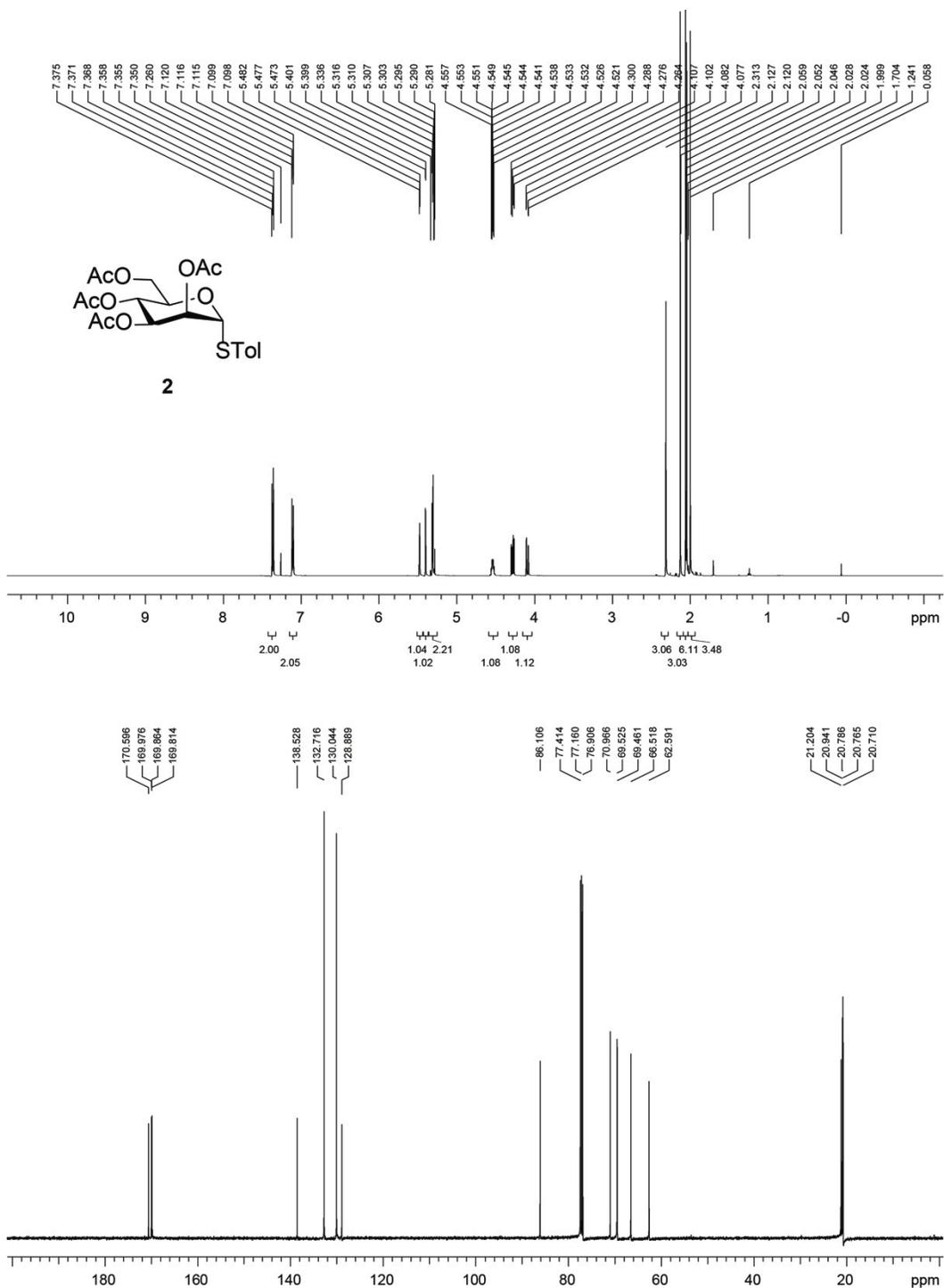
Appendix A-2. Oxidation and oxime ligation of tetrapeptide (H₂N-SVEK-CO₂H) and verification of their corresponding products with ESI-MS. In these experiments, 1 mM SVEK in PBS was oxidized with 2 mM NaIO₄ for 5 min on ice and subsequently quenched with 5 mM GSH for 10 min at room temperature. The resulting mixture was injected directly into Agilent 1100 mass spectrometer. Conversion to aldehyde was quantitative and this intermediate did not show any sign of degradation after 8 days. Treatment of the aldehydes with 1 mM aminoxy-reagents in 100 mM anilinium acetate gave the corresponding products. Once the oxime linkage was formed, adding a different aminoxy-reagent did not compromise the initial ligation product and only the mass of one peptide-conjugate could be found on ESI-MS.



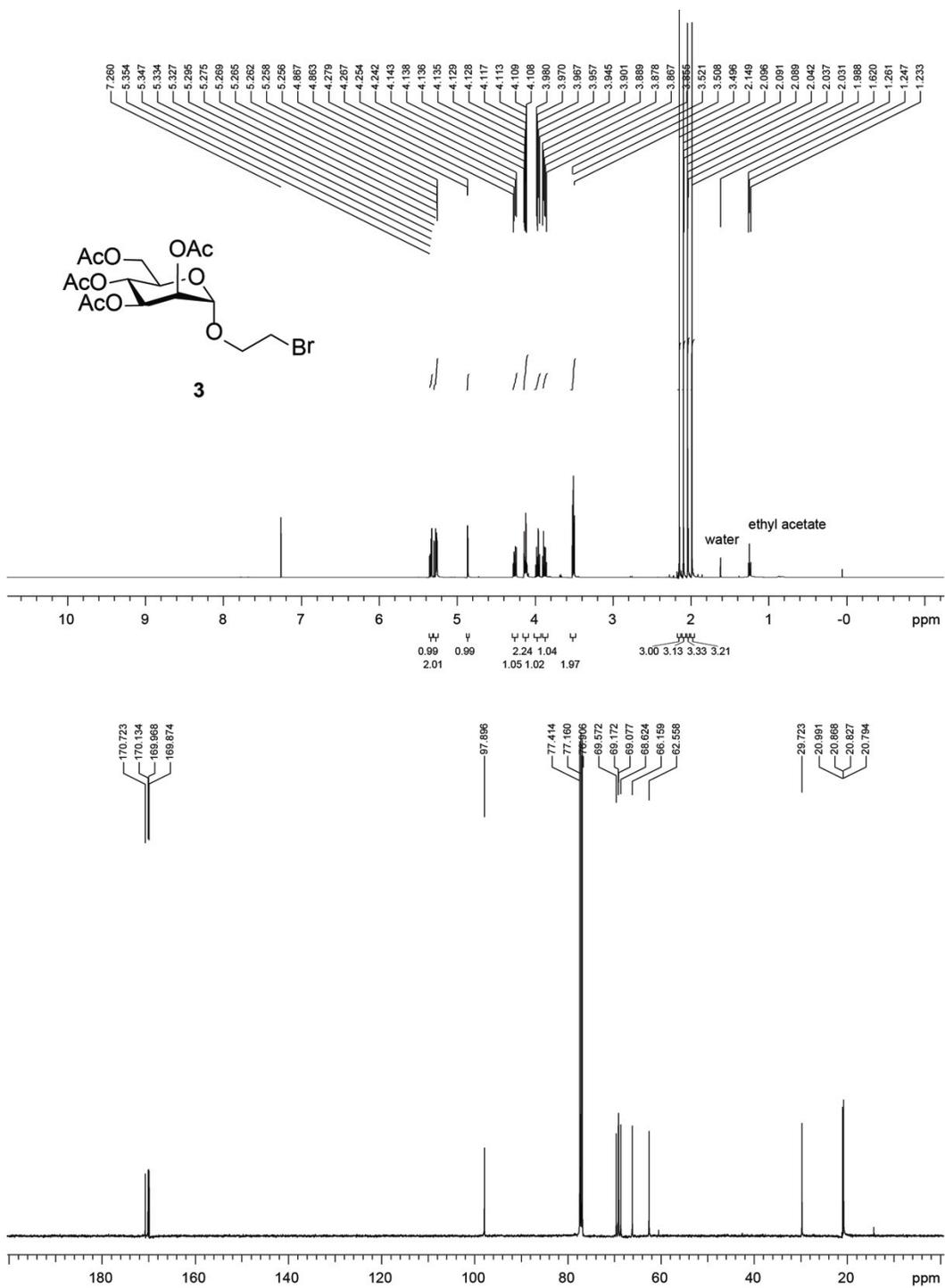
Appendix A-3. The oxime linkage showed no significant sign of degradation for at least 6 days when phage was stored in neutral solution at 4 °C. Blue bars are biotinylated phage. Grey bars are unreacted wild-type phage. The value is an average of three biotin-capture experiments. Error bar is equal to one standard deviation.



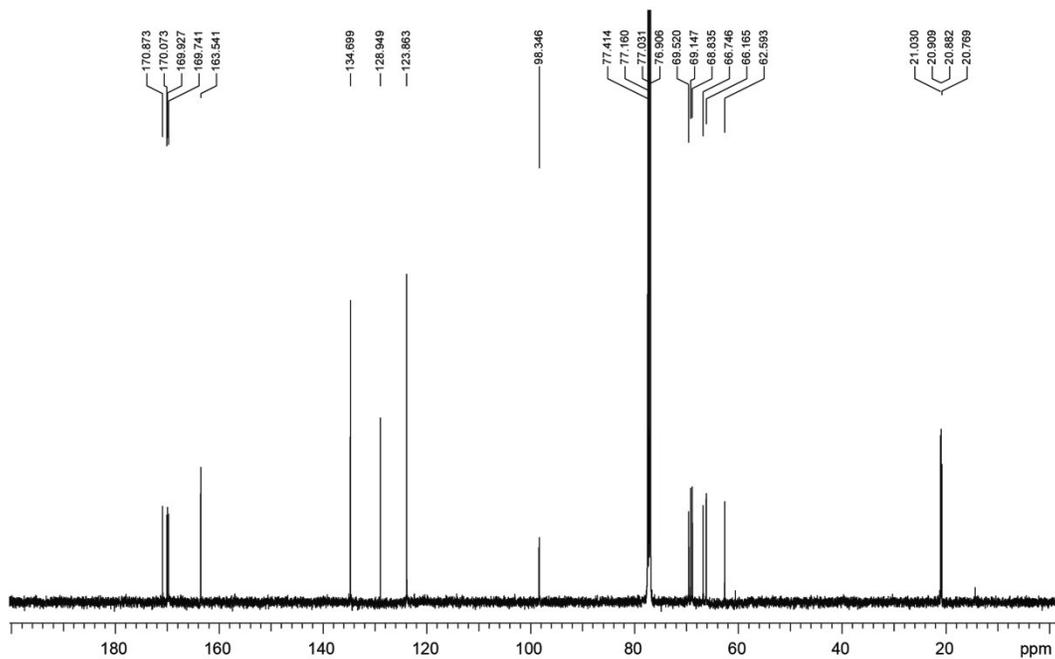
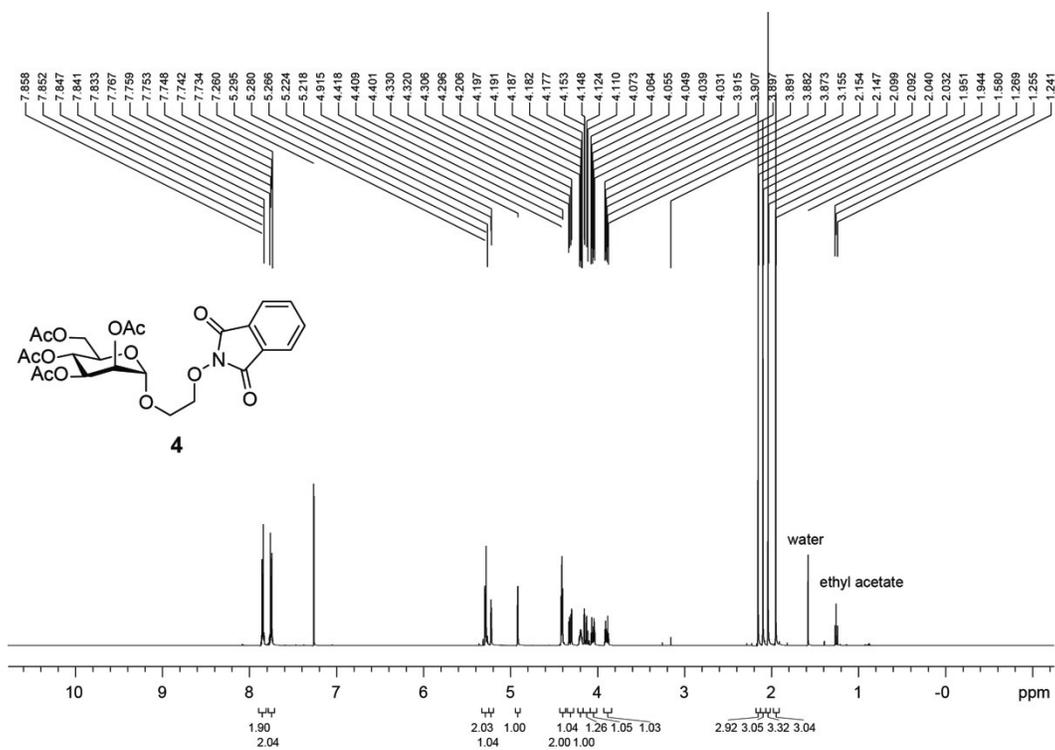
Appendix A-4. ¹H NMR (500 MHz, CDCl₃) and ¹³C NMR (126 MHz, CDCl₃) spectra of 1,2,3,4,6-penta-*O*-acetyl- α/β -D-mannopyranose (**1**).



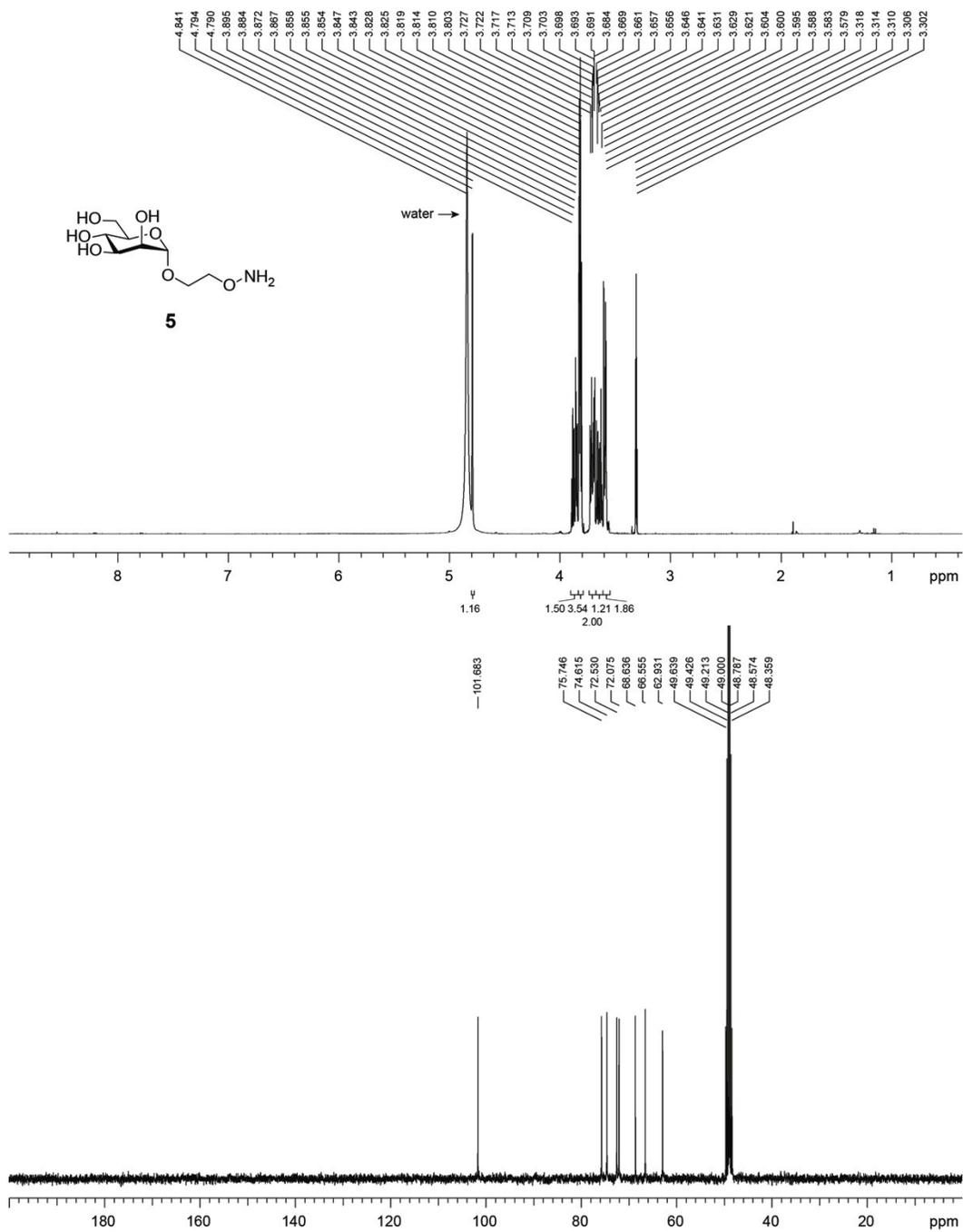
Appendix A-5. ¹H NMR (500 MHz, CDCl₃) and ¹³C NMR (126 MHz, CDCl₃) spectra of 4-methylphenyl 2,3,4,6-tetra-*O*-acetyl-1-thiol- α -D-mannopyranoside (**2**).



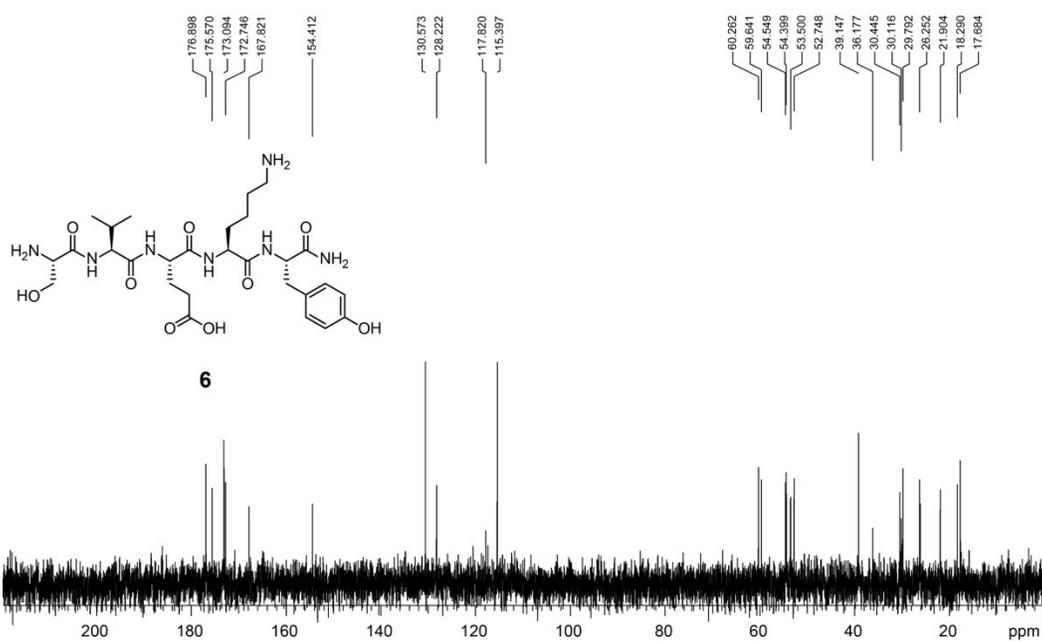
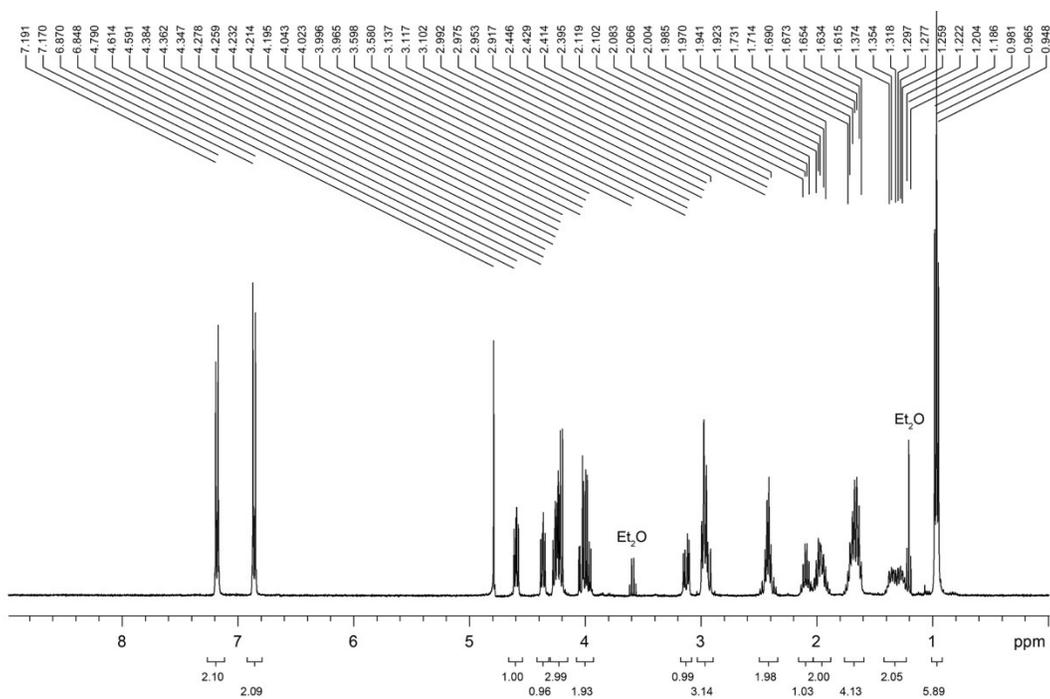
Appendix A-6. ^1H NMR (500 MHz, CDCl_3) and ^{13}C NMR (126 MHz, CDCl_3) spectra of 2-Bromoethyl 2,3,4,6-tetra-*O*-acetyl- α -D-mannopyranoside (**3**).



Appendix A-7. ¹H NMR (500 MHz, CDCl₃) and ¹³C NMR (126 MHz, CDCl₃) spectra *N*-2-[2-[(2,3,4,6-tetra-*O*-acetyl- α -D-mannopyranosyl)oxy]ethoxy]phthalimide (**4**).



Appendix A-8. ¹H NMR (400 MHz, CD₃OD) and ¹³C NMR (101MHz, CD₃OD) spectra of 2-(aminooxy)ethyl α-D-mannopyranoside (**5**).

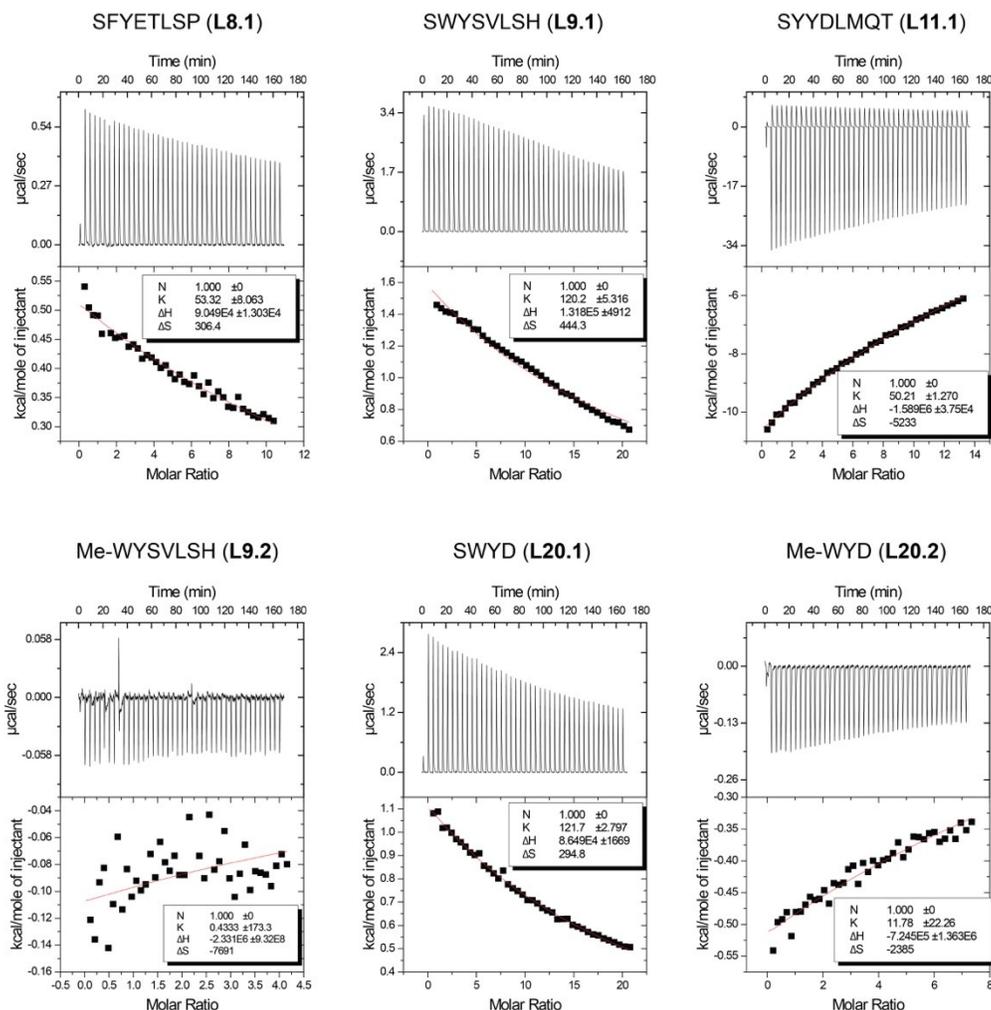


Appendix A-9. ¹H NMR (400 MHz, D₂O) and ¹³C NMR (101 MHz, D₂O) spectra of H₂N-SVEKY-CONH₂.

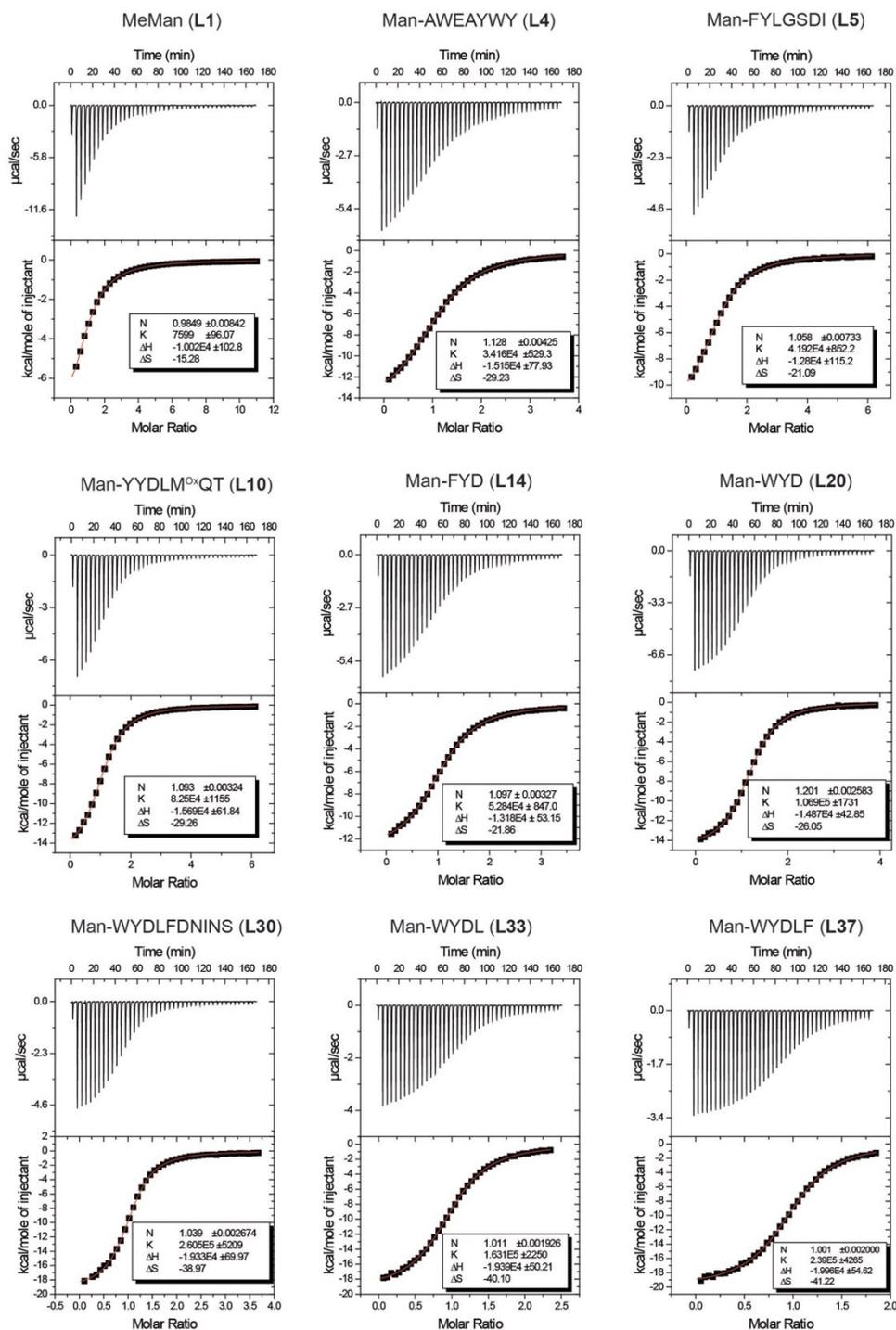
Appendix B: Supporting information for chapter 3

$K_D \pm S.D.$ (μM)	$IC_{50} \pm S.E.$ (μM)	86 Hit peptide sequences	Rank in A screen	Log(mean fraction)	Ratio of A/B	Ratio of A/C	Ratio of A/D
29.2		SAWEAYWY	1	-1.31	197	11	12
42.6	34.4	SFYSTTSR	2	-1.38	2852	12	1984
14.1	15.2	SWYSVLSH	5	-1.72	1285	963	894
15.9	14.2	SFYETLSP	7	-1.88	890	666	619
	15.2	SWYNSFGT	9	-1.97	729	546	507
		STYAWLDV	12	-2.25	547	57	6
		SLYDMNPL	14	-2.28	359	269	33
23.9		SFYLGSDI	16	-2.32	326	40	12
55.2		SYWEFTSL	19	-2.34	308	97	9
		SFYDQTYL	21	-2.36	293	220	204
20.9	16.1 \pm 0.4	SYYHNPNP	39	-2.62	163	122	20
		SWYSHTLK	68	-2.94	78	58	54
11.1		SYDDLQMT	87	-3.06	59	44	41
		SYYSTSLA	89	-3.07	57	43	40
		QSHHWQA	98	-3.1	54	7	6
		SSLALQMP	107	-3.12	51	38	36
		SWLTHDHT	108	-3.12	51	38	17
		SFYARSHS	119	-3.15	48	36	33
		SYYETNRI	120	-3.17	46	34	32
		SLPMHIKI	126	-3.19	44	17	31
		SFYDRFNS	137	-3.24	39	29	27
		STQGVGSD	142	-3.25	38	28	26
		SMMSKATT	150	-3.3	22	33	24
17.2	10.9 \pm 0.2	SFYDTPID	171	-3.35	30	23	21
		SWFQPLNL	189	-3.41	27	20	19
		SWYQITAS	196	-3.43	25	19	17
		SWYSVHQP	214	-3.47	23	17	16
		SWYSSPLY	233	-3.5	22	16	15
		SWYRDMDV	236	-3.51	21	16	15
		SWYSTIRT	273	-3.57	18	14	13
		SRGPAHQY	288	-3.59	17	13	12
		SIELFFYK	289	-3.59	17	13	12
		SWYASIFI	294	-3.6	17	13	12
		SWYSLMDK	296	-3.6	17	13	12
		SWFESALT	298	-3.61	17	13	12
		SWYHLNSA	299	-3.61	17	13	12
		STPFYVRY	300	-3.61	17	13	12
		SWLQMARD	314	-3.63	16	12	11
		SWYTLPSQ	325	-3.65	15	11	11
		SWYRTPVI	345	-3.67	15	11	10
		SWYMTQFG	346	-3.67	15	11	10
		SWYDHATT	354	-3.68	14	11	10
		SWYDSLDP	356	-3.68	14	11	10
		SWYSFRTP	363	-3.69	14	10	10
		SYYDRPYS	370	-3.69	14	10	10
		SWYEQTSR	373	-3.69	14	10	10
		SFYAVQPA	379	-3.7	14	10	10
		SWYLDHSD	381	-3.7	14	10	9
		SKVTFPYS	382	-3.7	13	10	9
		SWYHMLP	394	-3.72	13	10	9
		SYYDRASN	396	-3.72	13	10	9
		SPYSIFSA	402	-3.72	13	10	8
		SWYHQVHP	410	-3.72	13	10	9
		SYYESSLF	442	-3.76	12	9	8
		SYYESLSH	446	-3.76	12	9	8
		SWYALPRT	455	-3.77	12	9	8
		SWYHRTPI	456	-3.77	11	9	8
		SWYSTLAA	470	-3.79	11	8	8
		SWYGARPQ	473	-3.79	11	8	8
		SSWPALYG	484	-3.8	11	8	7
		SWYASATA	487	-3.81	11	8	7
		SRTIPHTD	494	-3.82	10	8	7
		SWYSQLLV	500	-3.83	10	8	7
		SWYEVARV	510	-3.84	10	7	7
		SWYADTRT	516	-3.85	10	7	7
		SWYQVVHP	521	-3.85	10	7	7
		SWYMNPLT	523	-3.85	9	7	7
		SWYADALV	528	-3.86	9	7	7
		SWYSEPMT	532	-3.86	9	7	6
		SWYAYPPD	533	-3.86	9	7	6
		SWYETPSA	566	-3.89	9	6	6
		SWYNLTAS	568	-3.89	9	6	6
		SWYKNPMS	594	-3.92	8	6	6
		SWYDTSHP	599	-3.92	8	6	6
		SYYEMYPS	611	-3.93	8	6	5
		SWYEVVSN	613	-3.94	8	6	5
		SYYEQVTL	614	-3.94	8	6	5
		SWYDDPFH	626	-3.94	8	6	5
		SWYDTLAG	633	-3.94	8	6	5
		SSSSDFPY	644	-3.96	8	6	5
		STLTGELV	654	-3.96	7	6	5
		SWYSHNTN	659	-3.96	7	6	5
		SNPDPYTP	661	-3.96	7	6	5
		SPSIPPTPI	669	-3.97	7	6	5
		SWYAVPHN	671	-3.97	7	5	5

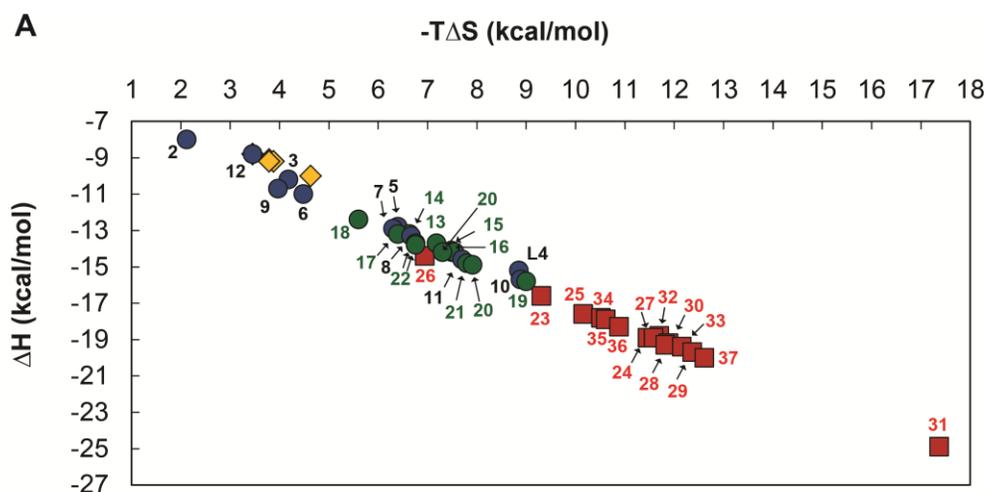
Appendix B-1. The list of 86 peptide sequences identified from three sets of volcano plot.



Appendix B-2. Raw ITC data of peptide fragments. Raw data obtained for 42 injections of ligands (2 mM for **L9.2**, **L20.2**; 5 mM for **L8.1**; and 10 mM for **L9.1**, **L11.1** and **L20.1**) into a solution of ConA (0.1 mM–0.20 mM) at 4-min intervals and 30 °C. The integrated curve showed experimental points (■) and the best fit (–) to the points by a nonlinear least-squares regression algorithm. The fitting were performed by fixing the stoichiometry of binding to one. It is important to note that, the determination of K_a , ΔH and ΔS in these experiments are inaccurate since they were performed with c values much lower than one, because the preparation of ConA with concentration $\gg 1$ mM is difficult due to solubility issue. In these cases, it is safe to assume that the K_D values of these peptide fragments are much greater than 1 mM.



Appendix B-3. Examples of raw ITC data of ConA binders. Selected examples of raw data obtained for 42 injections of ligands (2 mM for **L4**, **L5**, **L10**, **L14**, and **L20**; 1 mM for **L30**, **L33** and **L37**; and 10 mM for **L1**) into a solution of ConA (0.07 mM–0.30 mM) at 4-min intervals and 30 °C. The integrated curve showed experimental points (■) and the best fit (–) to the points by a nonlinear least-squares regression algorithm.



B

Ligand	K_D	Ligand	K_D	Ligand	K_D
L1 MeMan	137.4 ± 6.1	L13 Man-YYD	20.1	L23 Man-WYDANHSKPL	6.0
L2 Man-YWEFTSL	55.2	L14 Man-FYD	18.9	L24 Man-WYDRQETRFR	4.6
L3 Man-FYSTTSR	42.6	L15 Man-WY-OH	17.1 ± 0.4	L25 Man-WYDLHHSRTR	4.5
L4 Man-AWEAYWY	29.2	L16 Man-WYS	14.3	L26 Man-WYDLYHPVQH	4.3
L5 Man-FYLGSDI	23.9	L17 Man-WYA	13.5	L27 Man-WYELDDDDIT	5.3
L6 Man-YYHNPNA	20.9	L18 Man-WYH	12.7	L28 Man-WYDQFPPLHQ	5.1
L7 Man-FYDTIPD	17.2	L19 Man-WY	12.0	L29 Man-WYDNFDTIFA	5.0
L8 Man-FYETLSP	15.9	L20 Man-WYD	10.0 ± 0.9	L30 Man-WYDLFDNINS	3.8
L9 Man-WYSVLSH	14.1	L21 Man-WYE	9.6	L31 Man-WYDRFPPHES	3.7
L10 Man-YYDLM ^Q QT	12.0	L22 Man-WYG-OH	9.8 ± 1.1	L32 Man-WYDR	7.5
L11 Man-YYDLMQT	11.1			L33 Man-WYDL	6.1
L12 Man-HTHDSVE	151.5			L34 Man-WYDFF	5.8
				L35 Man-WYEIF	5.3
				L36 Man-WYDRF	4.9
				L37 Man-WYDLF	4.2

Appendix B-4. Summary of thermodynamic data and affinity. (A) Enthalpy and entropy of the binding of glycopeptides and MeMan to ConA measured with ITC. Orange diamonds (◆) are MeMan ($n = 5$). Blue circles (●) are ligands selected from first round. Green circles (●) are truncations of ligands from the first round. Red squares (■) are ligands selected from second round (affinity maturation) and their truncations. (B) Summarized table of the K_D of ligand interaction with ConA derived from the ITC data.

Appendix B-5. Molecular dynamics simulations and the analysis.

A crystal structure of the ConA protein bound to trimannoside (PDB ID: 1CVN) has been reported at 2.3 Å resolution. The mannopyranosyl (Man) residue of the synthetic ligand used in the present study and Man-240 (the 1,6-linked Man) in the endogenous ligand, bind to the protein at the same site (Figure 3-6B) and in the same 4C_1 conformation. These two Man residues displayed equivalent hydrogen-bonding patterns (Arg228 NH to O3, Asn14 side chain NH to O4, Leu99 NH to O5, and Tyr100 NH to O6). In contrast to the endogenous trimannoside, the synthetic ligand forms few additional hydrogen bonds, and prefers to van der Waals contacts with the protein surface (Figure 3-6C). This is as shown by a higher contact area of synthetic ligand (662 Å) compared to the trisaccharide Man3 (204 Å).

A number of structural studies show that a conserved water molecule plays an important role in facilitating the ConA-carbohydrate interaction, by forming hydrogen bonds with ARG 228, ASP 16, ASP 14 and MAN 241. A crystallized water molecule is present at the same location in the present protein structure, but because of the absence of MAN 241 in the synthetic ligand, this water does not mediate hydrogen bond formation with the ligand.

To examine the stabilities and strengths of these interactions, and to study the dynamics of the ligand in the binding pocket of the protein, a 50 ns molecular dynamics (MD) simulation was performed with the AMBER/GLYCAM force field and explicit water. The ligand-protein complex remained stable over the course of the simulation (average backbone RMSD = 0.69 (0.03) Å, average ligand displacement RMSD = 1.62 (0.32) Å).

Key interactions between protein and ligand

1. Hydrogen bonds

The MAN residue forms two stable hydrogen bonds with ASN 14 and ASP 208. TYR in the ligand also participates in the hydrogen bonding with PRO 206. The bonds with ASP 208 and PRO 206 are not present in the crystal structure, whereas the hydrogen bonds of MAN with LEU 100, ARG 228 and TYR 100, which exist in the crystal structure, are not formed during the MD simulation.

2. Hydrophobic interactions

HIS 205 in the protein interacts with TYR in the ligand via parallel displaced aromatic stacking. The histidine ring flips through the simulation maintaining this interaction, but it might not be strong enough to prevent it from flipping. TYR 100 in the protein also interacts with TYR in the ligand via CH/ π interaction. This interaction can provide binding energy of 3.54 kcal/mol, as estimated by QM calculation at MP2/aug-cc-pVQZ, between two benzene rings in this conformation. These interactions were characterized by the angle (θ) between the normals to the ring planes, and the distance (R)

between their centroids. For a stacking conformation, θ should be around 180° or 0° , and for CH/ π , it should be around 90° . In case of HIS 205 the average θ over the flipped and non-flipped arrangement is 138° and 21° at an average distance of 4.8 Å, while with TYR 100, the average θ is 93° at 5.1 Å.

Table. Intermolecular hydrogen bonds^[a] between protein and ligand

Donor		Acceptor		MD Occupancy ^[b]	MD Distance ^[c]	X-ray
Residue	Atom	Residue	Atom			
MAN	O4	ASN 14	Nδ2	100	2.9 (0.1)	2.98
ASP 208	Oδ2	MAN	O4	100	2.6 (0.1)	2.52
PRO 206	O	TYR (ligand)	OH	100	2.7 (0.1)	2.45
MAN	O3	ARG 228	N	100	3.0 (0.1)	2.88
MAN	O6	TYR 100	N	99	3.1 (0.1)	3.22
MAN	O5	LEU 99	N	97	3.2 (0.2)	2.93

^[a] Based on a distance cut-off of 3.5Å.

^[b] In percent of total simulation.

^[c] Average distance in Å.

Binding free energy analysis

The results from the binding free energy analysis were sorted and ranked on the basis of the per-residue total energy contributions (see Table at next page). Using this information, the residues significant to the ligand binding were identified. Residues involved in hydrogen bonds and hydrophobic interactions are amongst the top contributors, and stabilize the binding mainly via electrostatic and van der Waals interactions respectively. TYR 100 was involved in a hydrogen bond with the ligand in the crystal structure, but that bond does not exist during the simulation and largely its contribution to binding is through van der Waals. Most of the high ranked residues are hydrophobic which suggests that the binding interaction could mainly be non-polar. With regard to the ligand, the Man residue appears to contribute over 50% of the interaction energy, with the remainder coming principally from the TYR and TRP residues.

Table. Binding free energy analysis estimated by MD simulation.

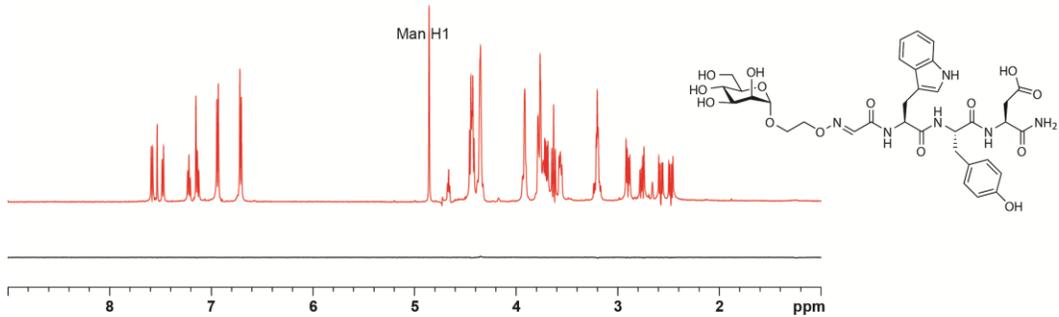
Average per-residue binding energy^[a] contributions for key^[b] residues in the protein and the ligand

Residue	van der Waals	Electrostatic	Polar Desolvation	Nonpolar Desolvation	Total
Protein					
Tyr100	-3.9	-2.9	3.6	-0.4	-3.6
Leu99	-2.5	-2.1	1.4	-0.3	-3.5
Asp208	1.1	-9.4	5.5	0.0	-2.9
Tyr12	-4.1	-1.4	3.3	-0.5	-2.7
Pro206	-0.3	-4.3	2.4	-0.1	-2.3
Gly98	-1.1	-2.8	1.8	-0.1	-2.2
Arg228	-1.1	-10.0	9.3	-0.1	-2.0
Pro13	-2.3	-0.2	0.9	-0.3	-1.8
Asn14	-0.3	-3.1	1.9	-0.1	-1.5
His205	-2.7	-2.4	4.0	-0.3	-1.5
Gly227	-0.9	-1.1	0.7	-0.1	-1.5
Ala207	-0.8	0.5	-0.5	0.0	-0.8
Ser21	-0.8	-0.3	0.8	-0.2	-0.5
Thr97	-0.4	0.1	-0.2	0.0	-0.5
Ligand					
Man	-8.2	-43.0	37.3	-1.6	-15.5
TYR	-6.4	-7.7	9.0	-1.2	-6.2
TRP	-5.8	-3.6	5.4	-1.0	-4.9
Linker	-2.9	11.2	-8.7	-0.3	-1.0
ASP	-0.5	42.2	-40.3	-0.1	1.3

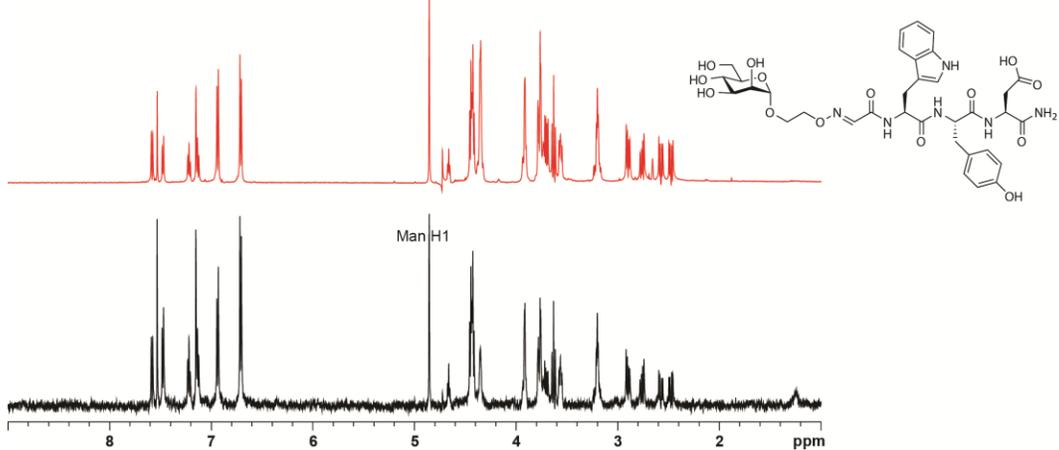
^[a] In kcal/mol

^[b] Residues that contribute greater than 0.5 kcal/mol to the total binding energy.

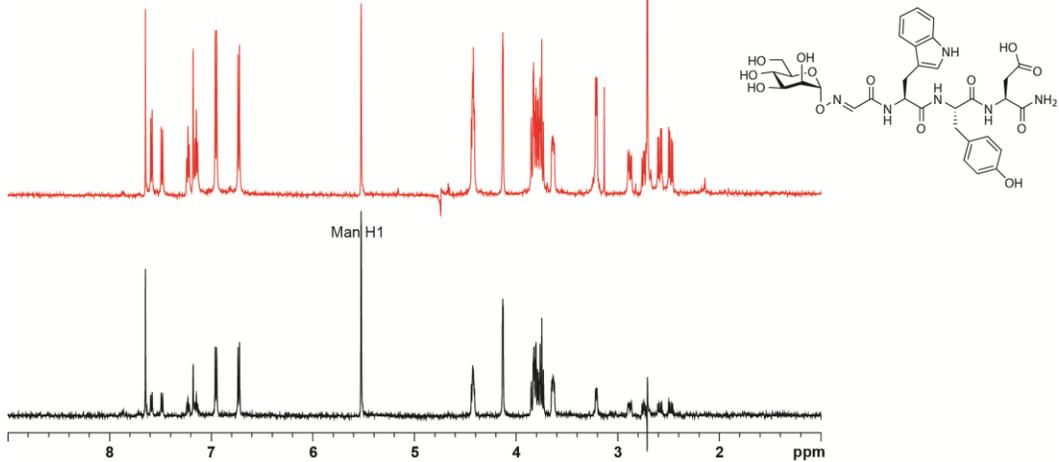
Control: Man-WYD (**L20**) in the absence of ConA



Man-WYD (**L20**) in the presence of ConA



Man-NL-WYD (**L20.3**) in the presence of ConA



Red traces: reference spectra
Black traces: STD-NMR spectra

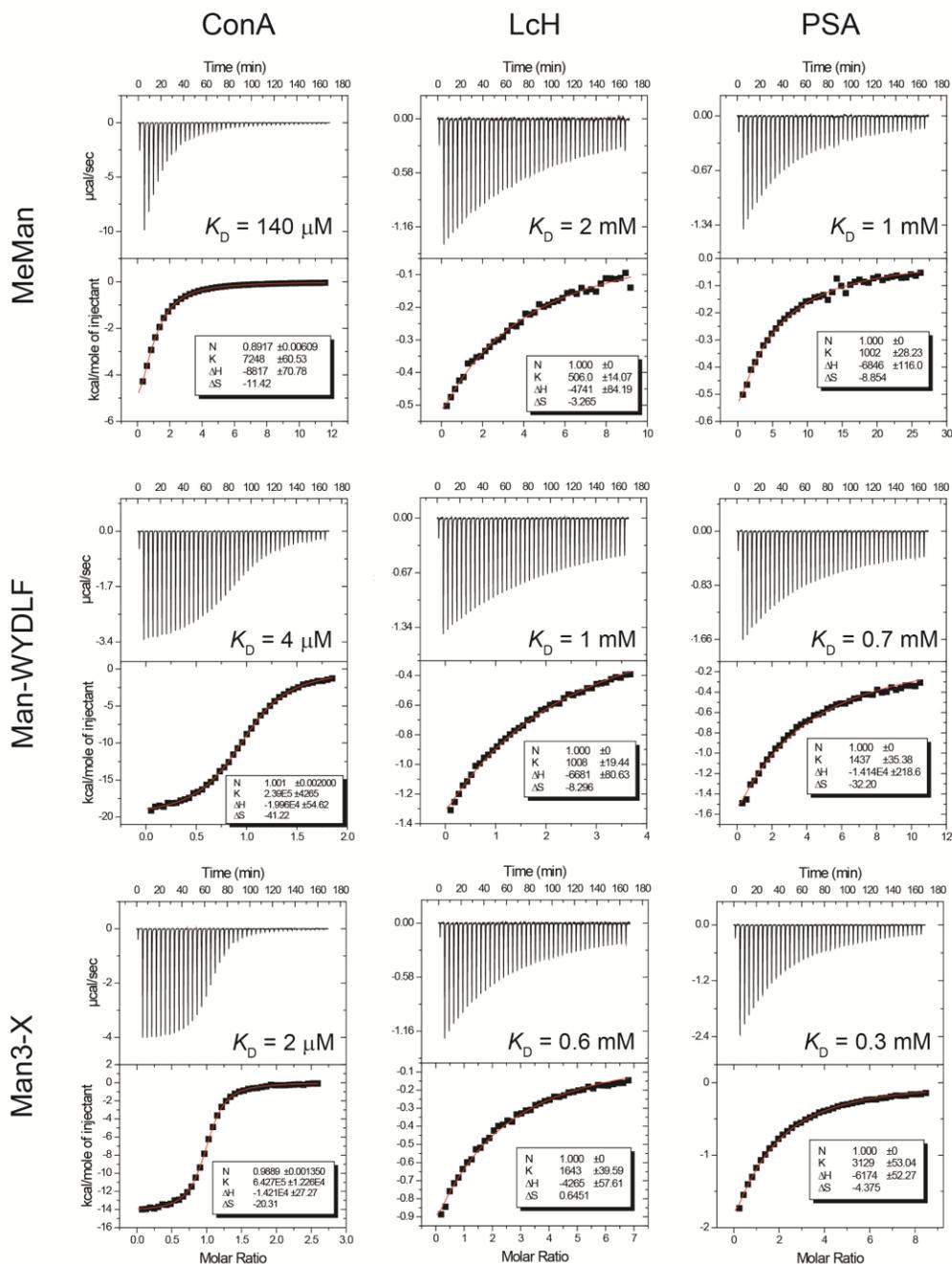
Appendix B-6. Raw spectra of STD-NMR experiments.

Appendix B-7. List of 85 lectins printed on microarray.

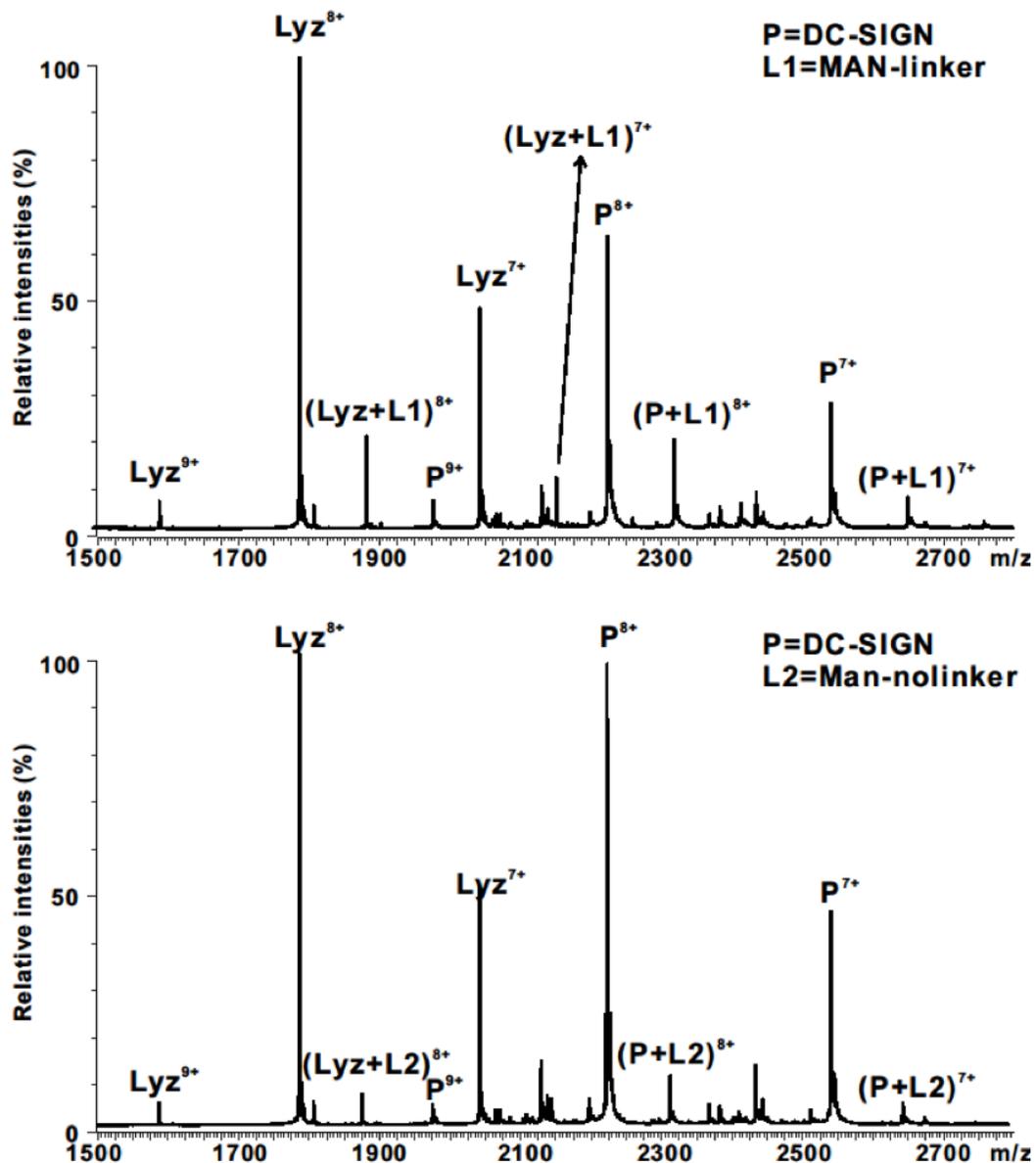
Abbreviation	Lectin source	Print conc. $\mu\text{g/ml}$	Inhibitory mono-saccharide	Rough specificity
AAA	<i>Anguilla anguilla</i>	1000	Fuc	α -Fuc
AAL	<i>Aleuria aurantia</i>	500	Fuc	Fuc
AOL	<i>Aspergillus oryzae</i>	1000	Fuc	Fuca1-6 (core fucosylation), Fuca1,2Gal
UEA-I	<i>Ulex europaeus I</i>	500	Fuc	Fuca1-2Gal β 1-4GlcNAc
PTLII	<i>Psophocarpus tetragonolobus</i>	1000	Gal	β -GalNAc, Type II Blood H
TJA-II	<i>Trichosanthes japonica</i>	500	Lac	Fuca1-2Gal β 1-3/4GlcNAc, GalNAc β 1-4Gal β 1
PSA	<i>Pisum sativum</i>	500	Man	Man
CCA	<i>Cancer antennarius</i>	1000	Lac	9-O-Acetyl NeuAc and 4-O-Acetyl NeuAc
LFA	<i>Limax flavus</i>	500	Lac	α -NeuAc (O-glycans)
LPA	<i>Limulus polphemus</i>	500	Lac	α -NeuAc
MAA	<i>Maackia amurensis</i>	500	Lac	NeuAc α 2-3LacNAc
MAL-I	<i>Maackia amurensis - I</i>	1000	Lac	NeuAc α 2-3LacNAc
MAL-II	<i>Maackia amurensis - II</i>	1000	Lac	NeuAc α 2-3LacNAc
PSL	<i>Polyporus Squamosus</i>	500	Lac	NeuAc α -2-6LacNAc
SNA	<i>Sambucus nigra</i>	500	Lac	NeuAc α -2-6, (Lac core)
TJA-I	<i>Trichosanthes japonica</i>	1000	Lac	NeuAc α 2-6LacNAc or 6-Sulfo LacNAc.
PHA-L	<i>Phaseolus vulgaris-L</i>	500	Gal	Complex triantennary N-linked glycans
ECA	<i>Erythrina cristagalli</i>	500	Gal	GalNAc β 1-4GlcNAc, Gal β 1-4GlcNAc
RCA	<i>Ricinus communis agglutinin B</i>	1000	Lac	Terminal β -Gal, terminal LacNAc
PHA-E	<i>Phaseolus vulgaris-E</i>	500	Lac	Complex N-linked (bisecting GlcNAc)
CA	<i>Colchicum autumnale</i>	1000	Gal	Terminal β -Gal, α - and β -GalNAc
BPA	<i>Bauhinia purpurea</i>	500	Gal	Gal β 1-3 or GalNAc β 1-4 more weakly
APA	<i>Abrus precatorius</i>	500	Gal	Gal β -1,3GalNAc (TF antigen) > Gal
GS-I	<i>Griffonia simplicifolia I</i>	1000	Gal	α -Gal, some GalNAc
APP	<i>Aegopodium podagraria</i>	500	GalNAc	GalNAc>Lacose>Galactose
BDA	<i>Bryonia dioica</i>	500	Gal	GalNAc
Blackbean	<i>Black bean</i>	1000	Lac	GalNAc

CAA	<i>Caragana arborescens</i>	500	Gal	GalNAc/Gal (monosaccharides best)
CSA	<i>Cystisus scoparius</i>	500	Gal	β -GalNAc, terminal
IRA	<i>Iris Hybrid</i>	1000	Gal	GalNAc(GalNAc α -1,3)Gal>GalNAc>Gal
VVA	<i>Vicia villosa</i>	500	Gal	α -Linked terminal GalNAc, GalNAc α -1,3 Gal
WFA	<i>Wisteria floribunda</i>	500	Gal	GalNAc
ASA	<i>Allium sativum</i>	1000	Man	Mannose
AMA	<i>Allium moly</i>	1000	Man	D-Mannose
Calsepa	<i>Calystegia sepium</i>	500	Man	Man/maltose
Con A	<i>Canavalia ensiformis</i>	1000	Man	Branched and terminal mannose [High-Man, Man α -1,6 (Man α -1,3) Man]
CVN	<i>Cyanovirin</i>	1000	Man	α -1,2 Mannose
GNA	<i>Galanthus nivalis</i>	1000	Man	Terminal α -1,3mannose
GRFT	<i>Griffithsin</i>	1000	Man	Mannose,GlcNAc
HHL	<i>Hippeastrum Hybrid</i>	1000	Man	α -1,3 Mannose and α -1,6 mannose
NPA	<i>Narcissus pseudonarcissus</i>	1000	Man	Terminal and internal Man
MNA-M	<i>Moringa M</i>	500	Man	Branched and terminal mannose
SVN	<i>Scytovirin</i>	500	Man	α -1,2 Mannose
TL	<i>Tulipa sp.</i>	1000	GlcNAc	Man3 core, bi- and tri-antennary complex-type N-glycan, GalNAc
PMA	<i>Polugonatum multiflorum</i>	500	Man	Mannan
UDA	<i>Urtica dioica</i>	1000	GlcNAc	GlcNAc β -1,4 GlcNAc oligomers and high mannose epitopes
VVA (man)	<i>Vicia villosa</i>	500	Man	Man
VFA	<i>Vicia faba</i>	500	Man	Man>Glc>GlcNAc
LcH	<i>Lens culinaris</i>	500	Man	Complex
Jacalin, AIA	<i>Artocarpus integrifolia</i>	500	Gal	O-glycosidically linked oligosaccharides, preferring the structure galactosyl (β -1,3) N-acetylgalactosamine
EEL	<i>Eunonymus europaeus</i>	1000	Lac	Blood group B antigen, Gal α -1,3 Gal
MPL	<i>Maclura pomifera</i>	1000	Gal	Gal β -1,3 GalNAc, GalNAc
GHA	<i>Glechoma hederacea</i>	500	Gal	multivalent GalNAc α -Ser/Thr
PNA	<i>Arachis hyogaea</i>	500	Gal	Terminal Gal β -OR
ACA	<i>Amaranthus Caudatus</i>	1000	Lac	Gal β -1,3 GalNAc (the T antigen)
DBA	<i>Dolichos biflorus</i>	500	Gal	GalNAc α -OR (GalNAc α -1,3 GalNAc) and Blood group A antigen,

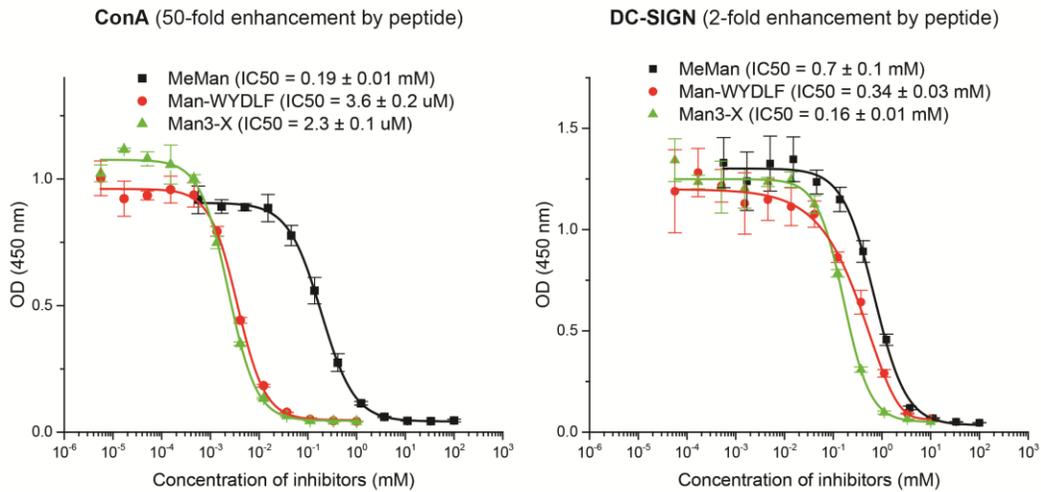
SBA	<i>Glycine max</i>	500	Gal	α - or β -Linked terminal GalNAc, GalNAc α -1,3 Gal
SJA	<i>Sophora japonica</i>	500	Gal	GalNAc
HPA	<i>Helix pomatia</i>	500	Gal	α -Linked terminal GalNAc
SNA-II	<i>Sambucus nigra</i>	500	Gal	GalNAc linked alpha to C-2, C-3 or C-6 hydroxyl group of galactose
PTLI	<i>Psophocarpus tetragonolobus</i>	1000	Gal	α -GalNAc, A-antigen
PTA	<i>Psophocarpus tetragonolobus</i>	500	Gal	Gal
LBA	<i>Phaseolus lunatus</i>	1000	Gal	GalNAc α -1,3 [Fuc α -1,2]Gal
DSA	<i>Datura stramonium</i>	500	Lac	GlcNAc β -1,4 GlcNAc oligomers and LacNAc (Gal β 1-4 GlcNAc)
GS-II	<i>Griffonia simplicifolia</i>	500	GlcNAc	terminal GlcNAc
UEA-II	<i>Ulex europaeus II</i>	1000	GlcNAc	Oligomers of β -1,4 GlcNAc
RPA	<i>Robinia pseudoacacia</i>	500	GlcNAc	Complex
LEA	<i>Lycopersicon esculentum</i>	500	GlcNAc	β -1,4GlcNAc oligomers
STA	<i>Solanus tuberosum</i>	500	GlcNAc	GlcNAc oligomers, LacNAc
WGA	<i>Triticum vulgare</i>	1000	GlcNAc	β -GlcNAc, sialic acid, GalNAc
Cholera Toxin	<i>Cholera Toxin from Vibrio cholerae</i>	2000	Lac	glycolipid
MOA	<i>Marasmius oreades</i>	500	Gal	Gal α -1,3 Gal and Gal α -1,3 Gal β -1,4GlcNAc
MNA-G	<i>Morniga sp.</i>	1000	Gal	GalNAc α , Tn antigen
Lotus	<i>Lotus tetragonolobus</i>	500	Fuc	Terminal α -Fuc, Lewis x
IAA	<i>Iberis amara</i>	500	GalNAc	GalNAc
Ricin B Chain		500	Gal	β -Gal
GS-I	<i>Griffonia simplicifolia I</i>	1000	Gal	α -Gal, some GalNAc
DC-SIGN		100	Man	High Mannose
PA-IL [†]	<i>Pseudomonas Aeruginosa</i>	500	Gal	Terminal α -Gal
7LE	antibody	100		Lewis a
SPM110	antibody	100		Sialyl Lewis a
LeX P12	antibody	100		Lewis x
LeX R&D	antibody	100		Lewis x
Lewis Y	antibody	100		Lewis Y
Lewis Y IgM	antibody	100		Lewis Y



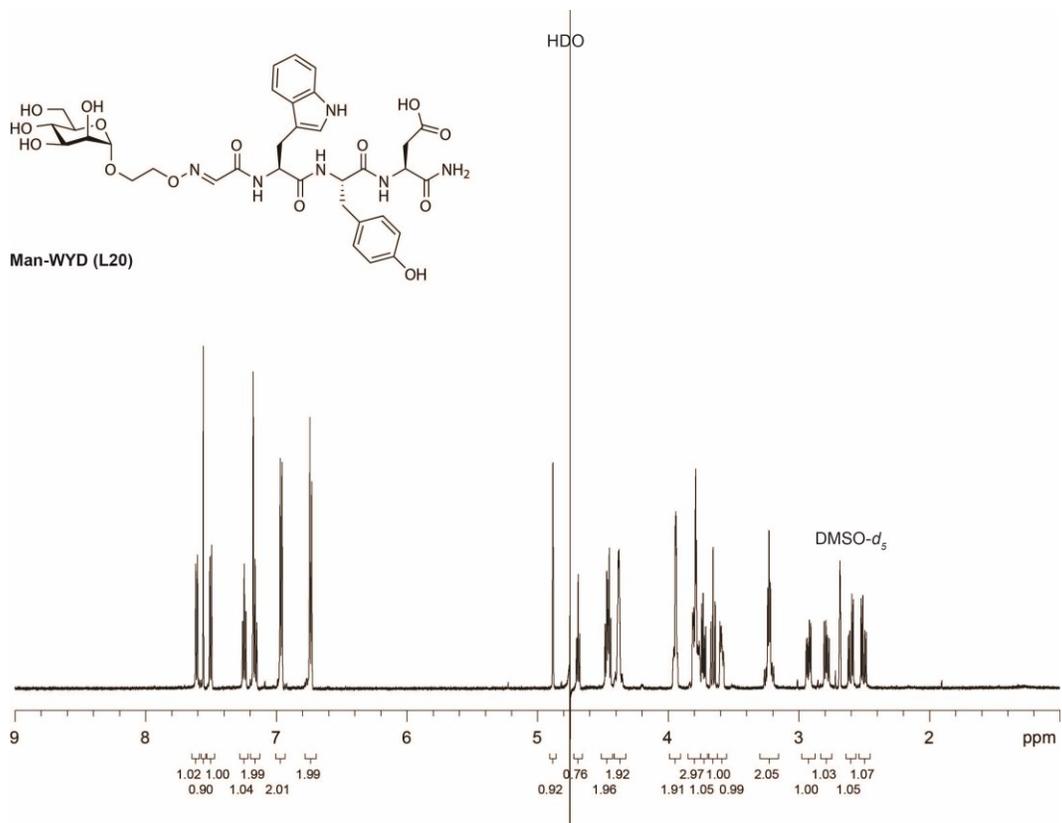
Appendix B-8. Raw ITC data of MeMan, Man-WYDLF, and Man3-X binding to ConA, LcH and PSA. Raw data obtained for 42 injections of ligands (10 mM for MeMan; 1–4 mM for Man-WYDLF; 1–5 mM for Man3-X) into a solution of ConA (0.19 mM), LcH (0.24 mM) or PSA (0.09 mM) at 4-min intervals and 30 °C. The integrated curve showed experimental points (■) and the best fit (–) to the points by a nonlinear least-squares regression algorithm. The stoichiometry parameter, N, was fixed at 1.0 for $c \leq 1$ but allowed to float freely for $c > 1$.



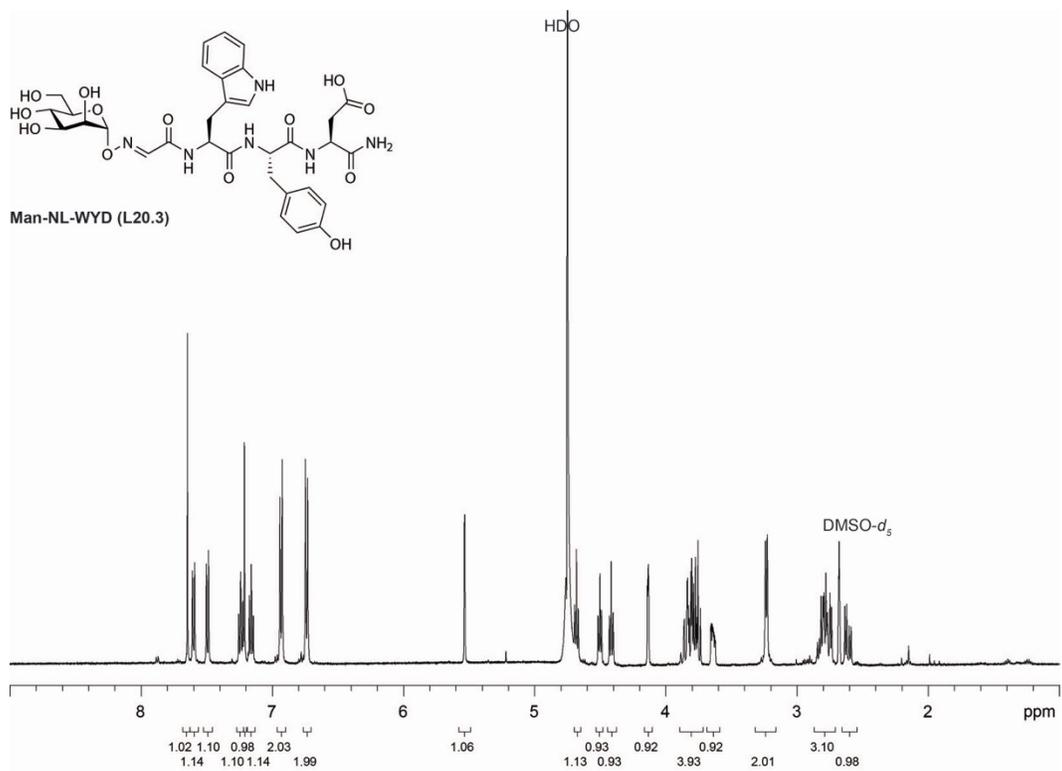
Appendix B-9. ESI-MS binding measurement for DC-SIGN. The mass spectra of ESI-MS binding measurement for DC-SIGN. Top spectrum: Man-WYD (L20); bottom spectrum: Man-NL-WYD (L20.3). The measured K_D of L20 and L20.3 are $600 \pm 7 \mu\text{M}$ and $1890 \pm 40 \mu\text{M}$ respectively.



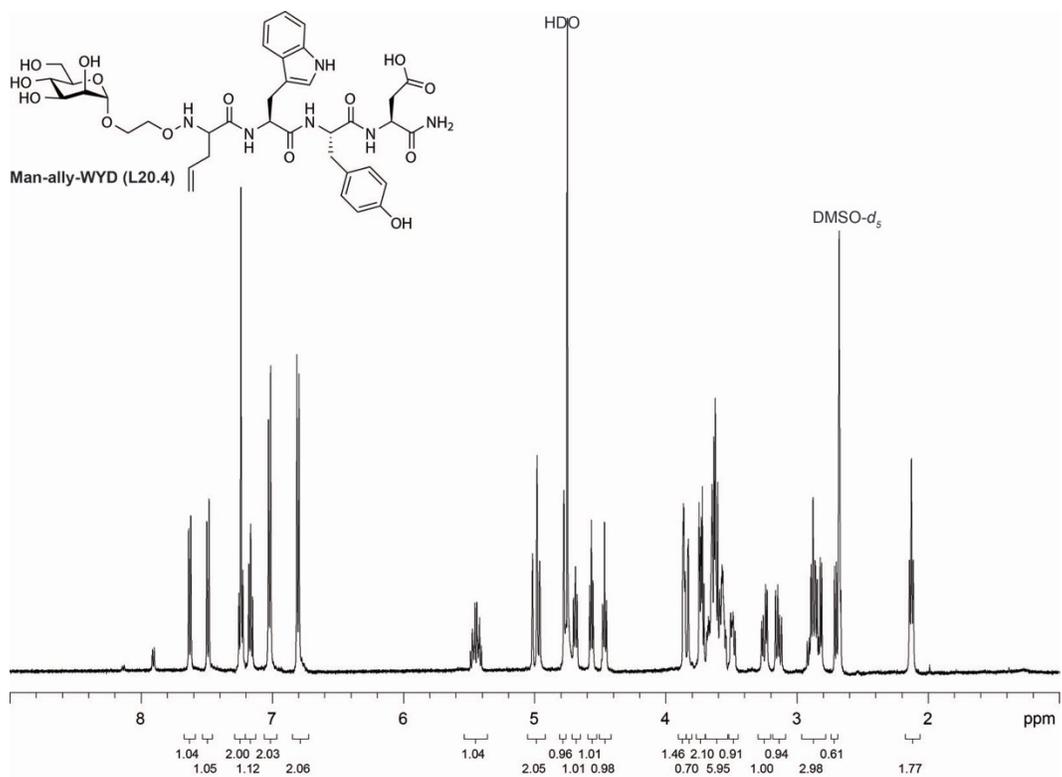
Appendix B-10. MeMan, Man-WYDLF, and Man3-X competitively inhibit the binding of glycoprotein probe, horseradish peroxidase (HRP), which contains trimannoside core, to immobilized ConA (left panel) or immobilized tetrameric extracellular domain of DC-SIGN (right panel). The monosaccharide couples synergistically with the peptide fragment and decreases the half maximal inhibitory concentration (IC_{50}) by more than 50-fold for ConA but only 2-fold for DC-SIGN. The results indicate a selective activity enhancement of the peptide fragment for ConA. The absorbance at 450 nm (OD = optical density) is a mean value measured in three independent wells with the error represents one standard deviation of the mean.



Appendix B-11. ¹H NMR (600MHz , deuterated PBS + 0.1% (v/v) DMSO-*d*₆) spectrum of Man-WYD (L20).



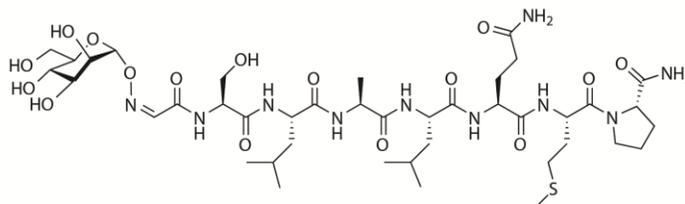
Appendix B-12. ¹H NMR (500MHz, D₂O + 0.1% (v/v) DMSO-d₆) spectrum of Man-NL-WYD (L20.3)



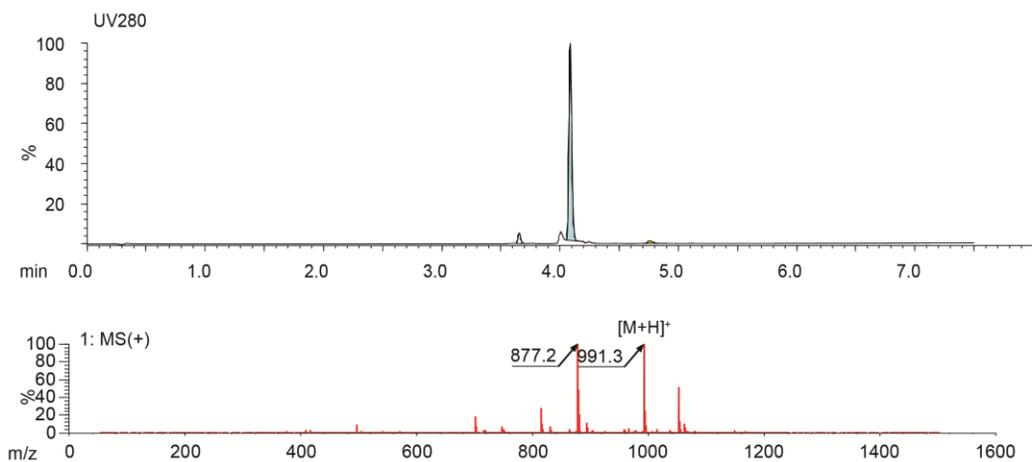
Appendix B-13. $^1\text{H NMR}$ (500MHz, $\text{D}_2\text{O} + 0.1\%$ (v/v) $\text{DMSO-}d_6$) spectrum of Man-allyl-WYD (L20.4).

Appendix C: Supporting information for chapter 4

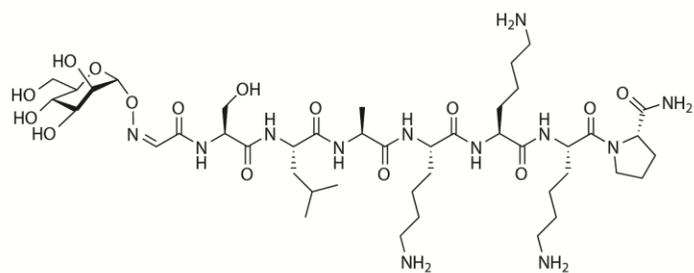
Man-SLALQMP



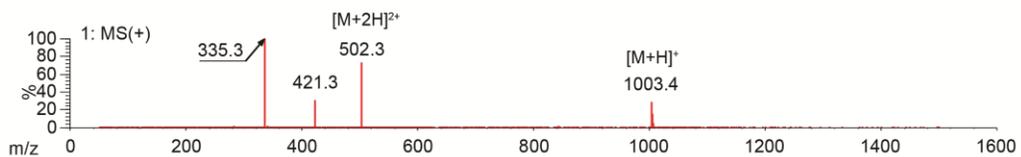
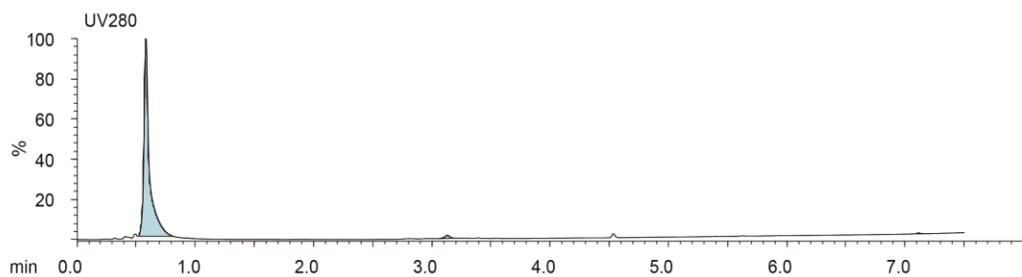
HRMS: m/z calc'd for $C_{41}H_{71}N_{10}O_{16}S$ ($[M+H]^+$): 991.4765; found: 991.4779



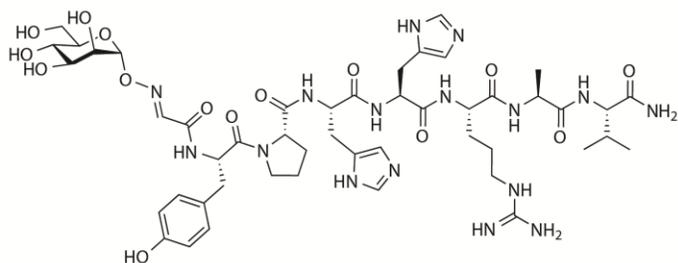
Man-SLAKKKP



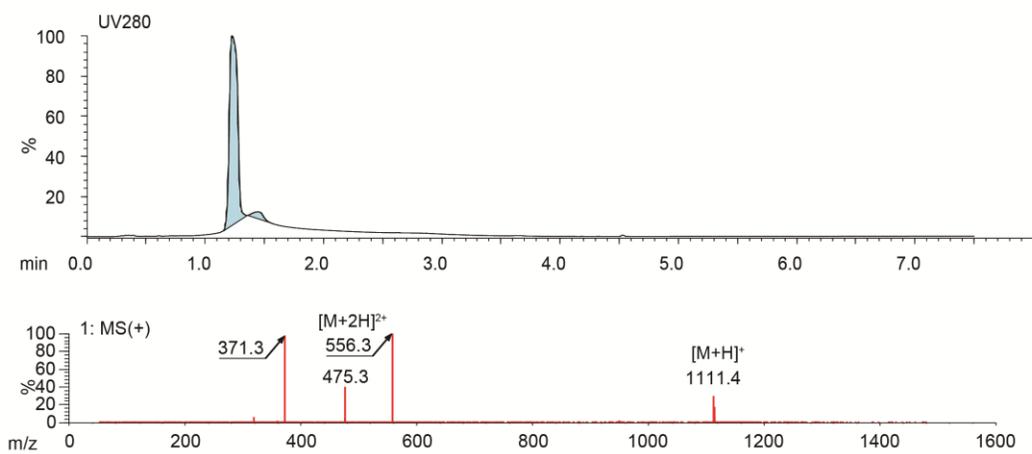
HRMS: m/z calc'd for $C_{43}H_{79}N_{12}O_{15}$ ($[M+H]^+$): 1003.5782; found: 1003.5787



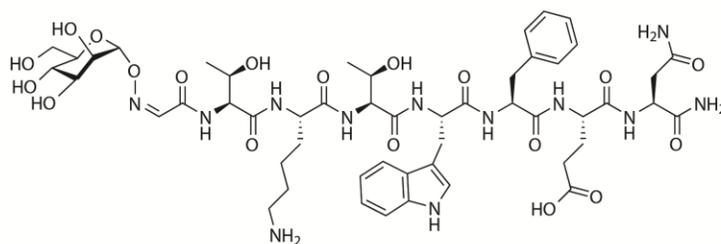
Man-YPHHRAV



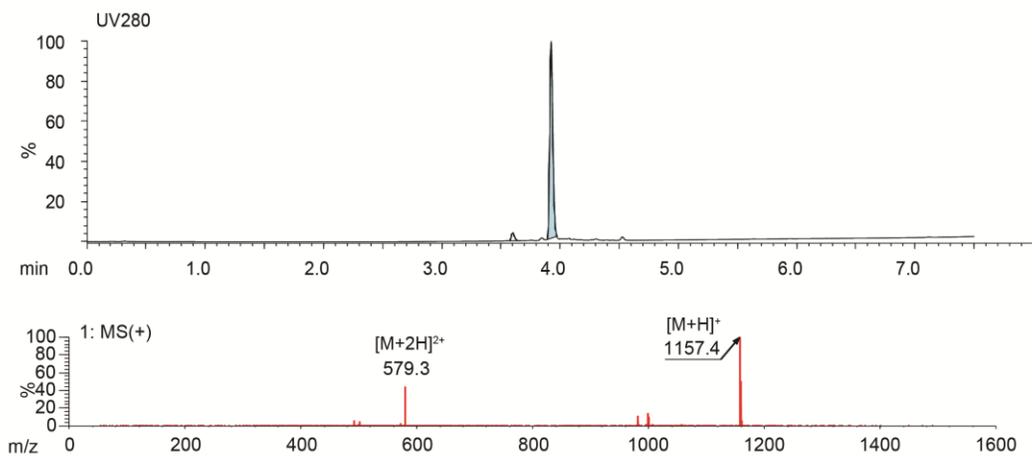
HRMS: m/z calc'd for $C_{48}H_{71}N_{16}O_{15}$ ($[M+H]^+$): 1111.5279; found: 1111.53

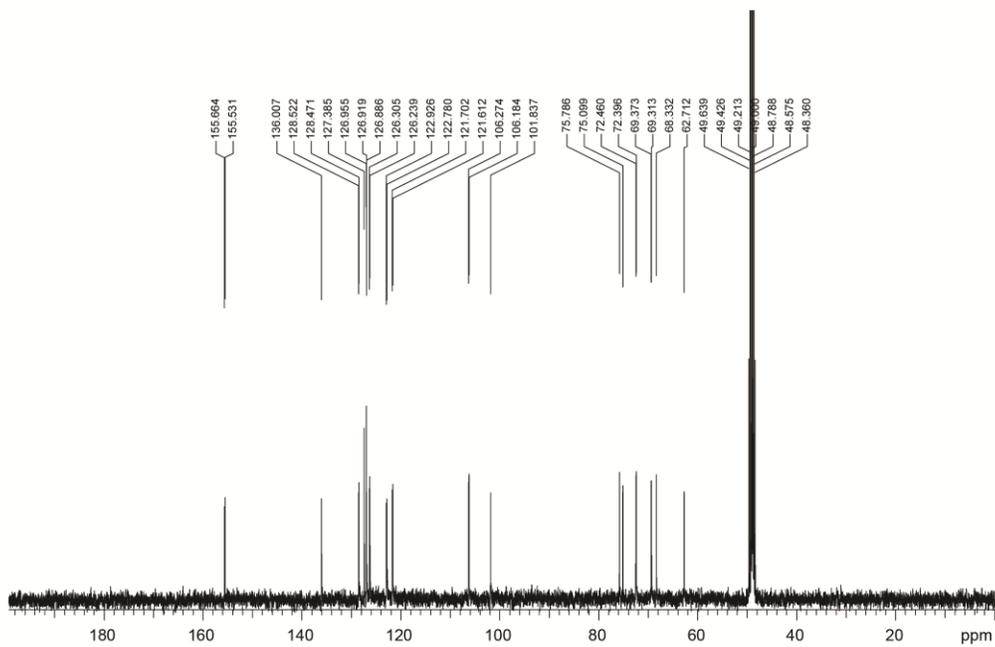
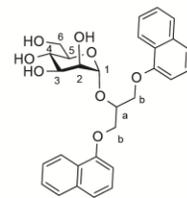
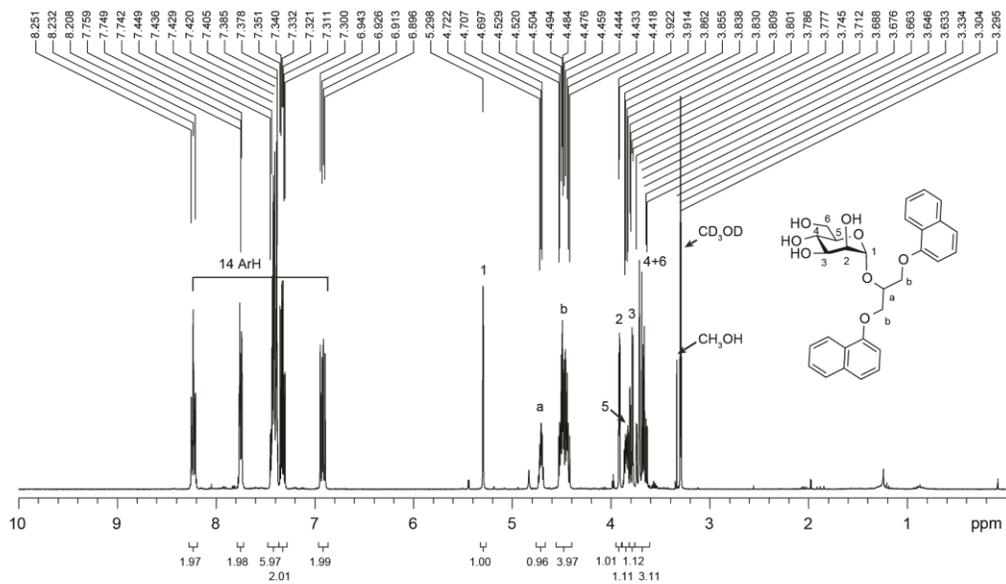


Man-TKTWFEN

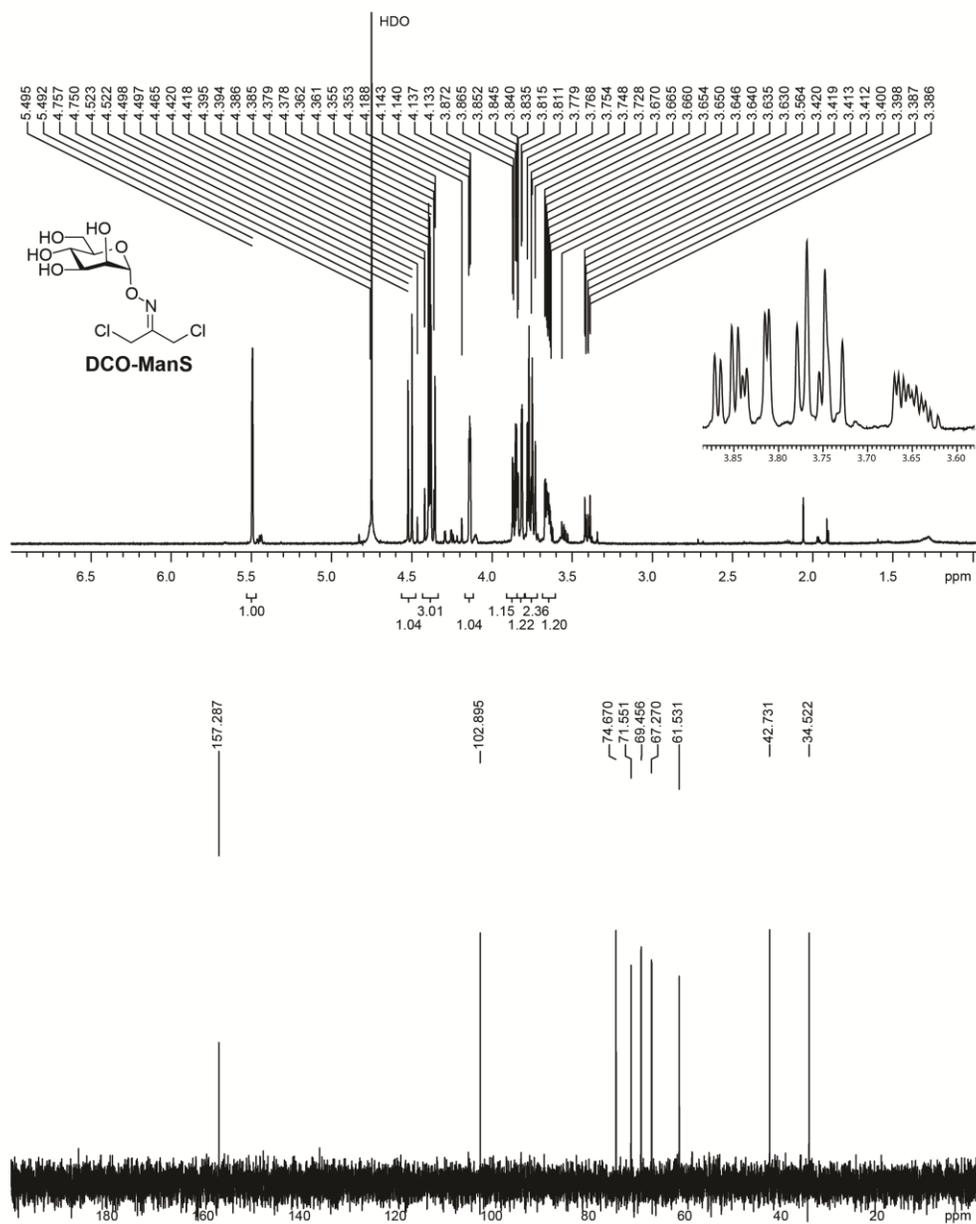


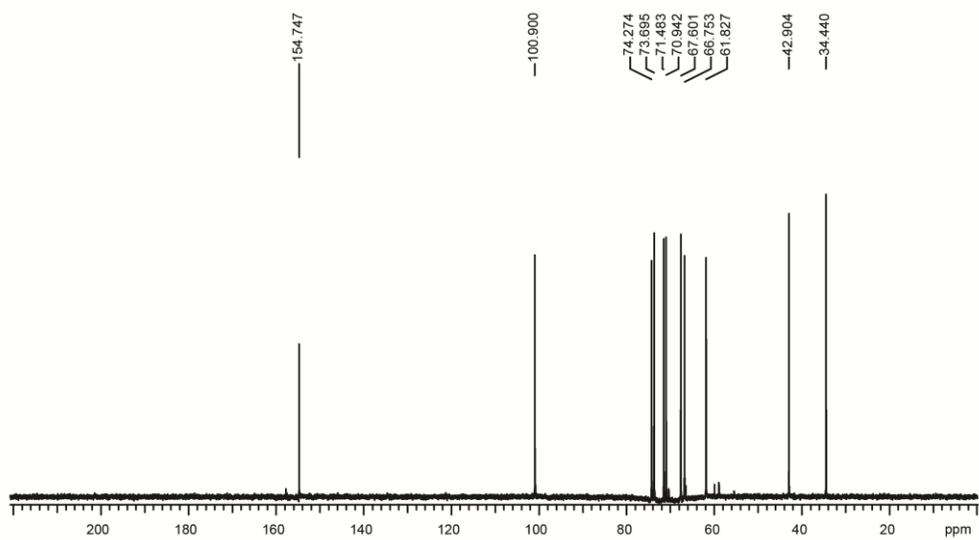
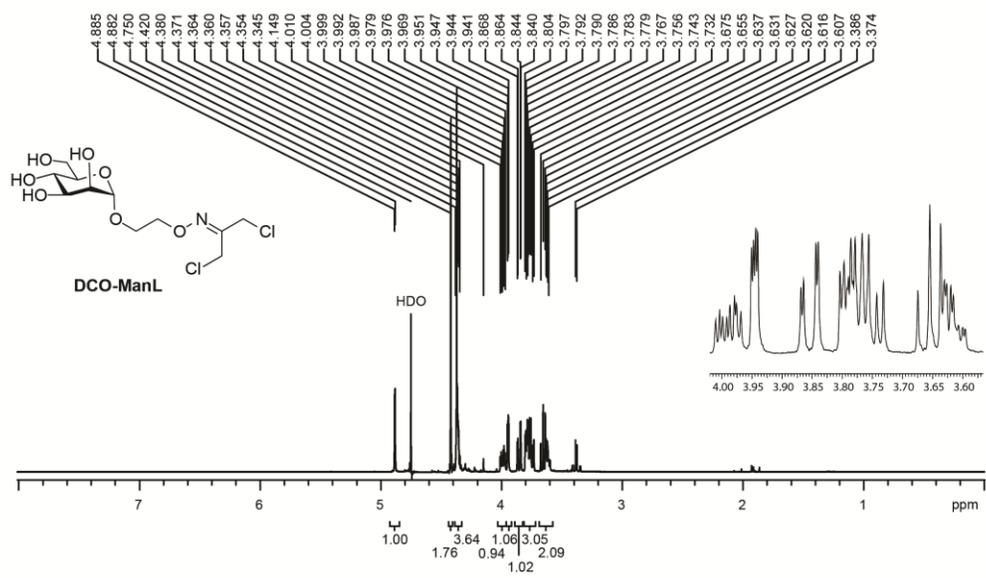
HRMS: m/z calc'd for $C_{51}H_{73}N_{12}O_{19}$ ($[M+H]^+$): 1157.5109; found: 1157.513

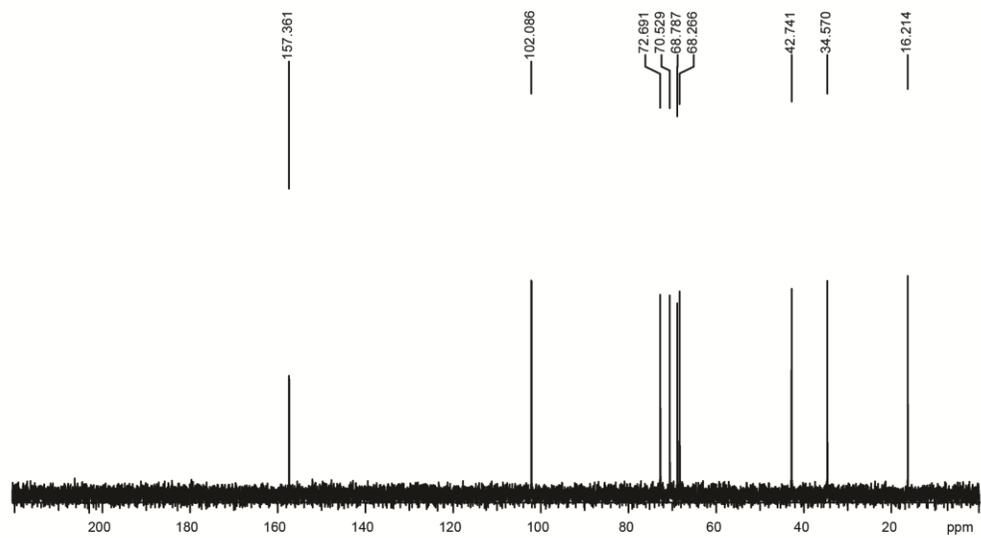
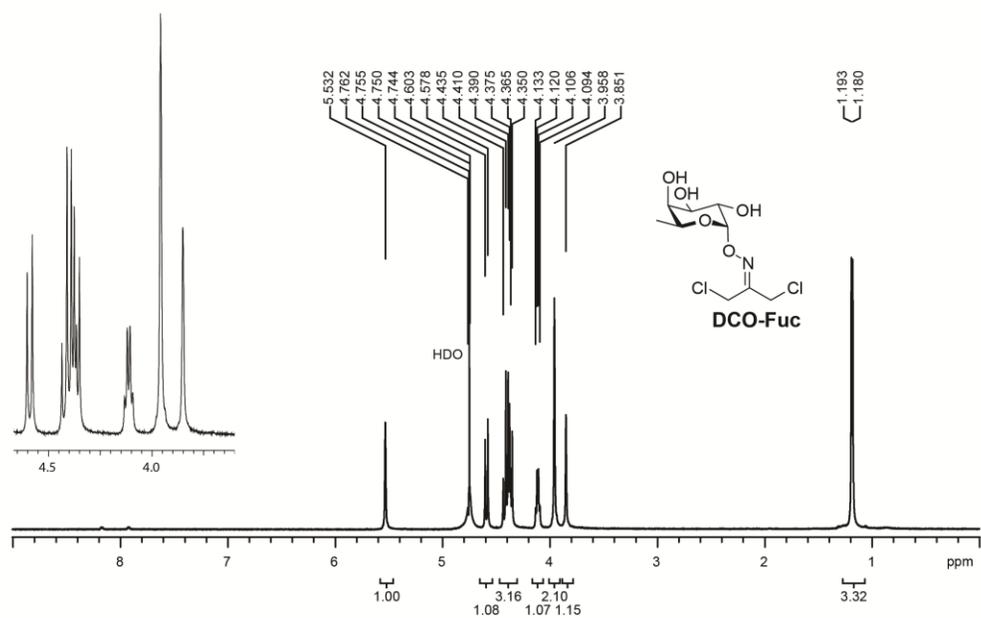


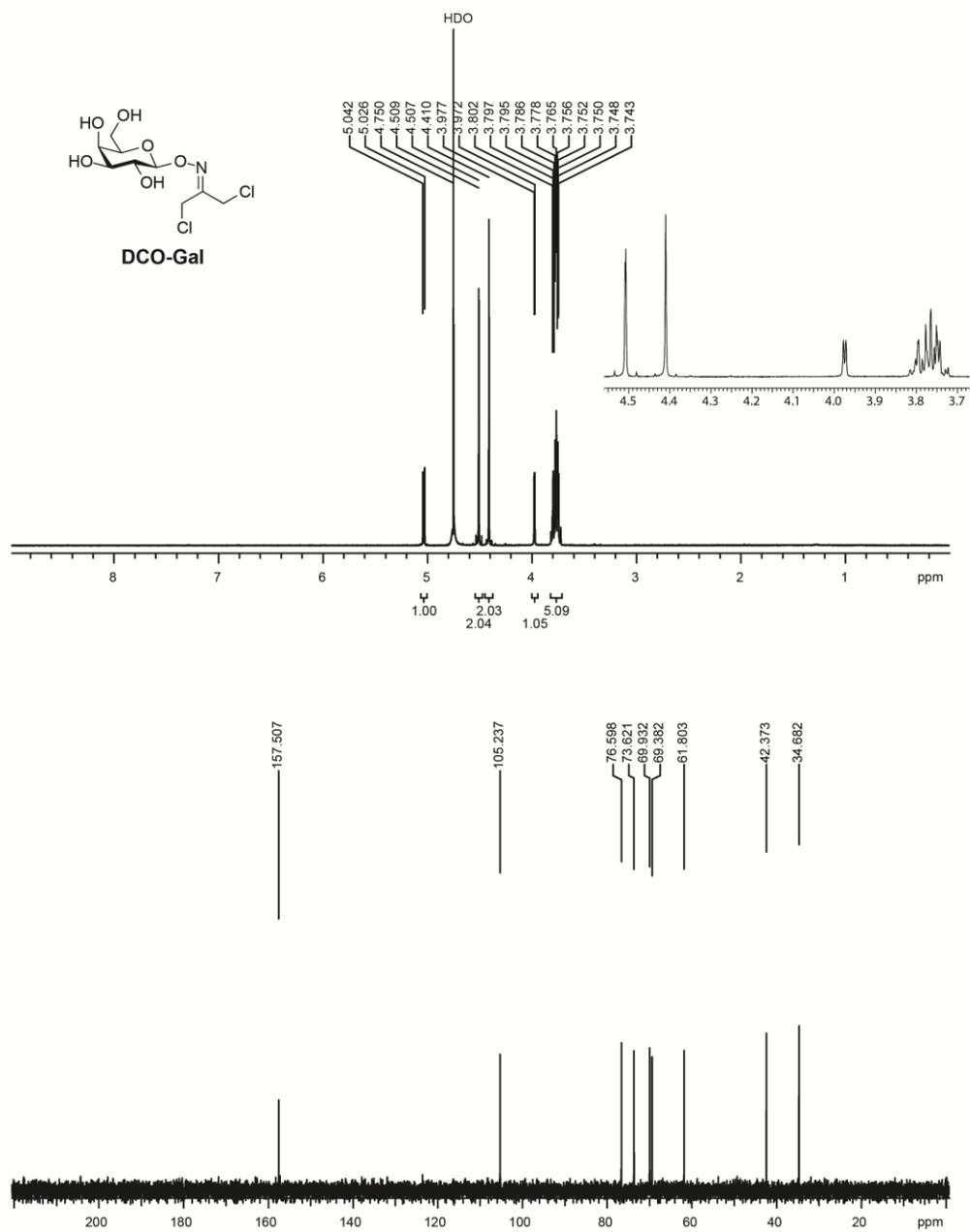


Appendix D: Supporting information for chapter 5



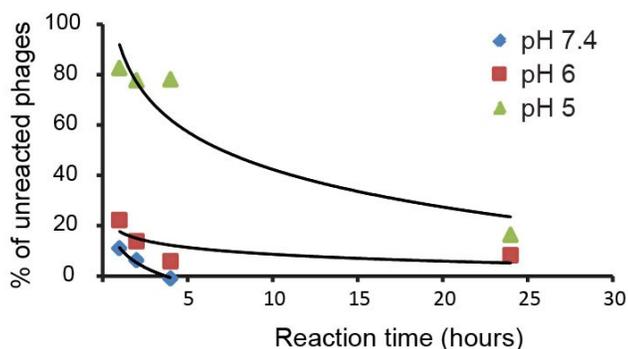
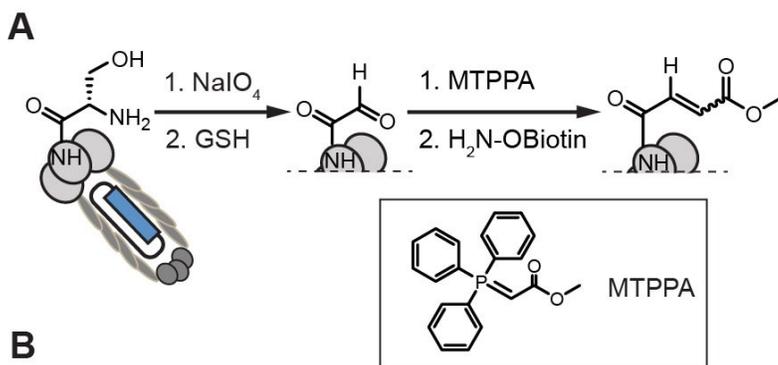




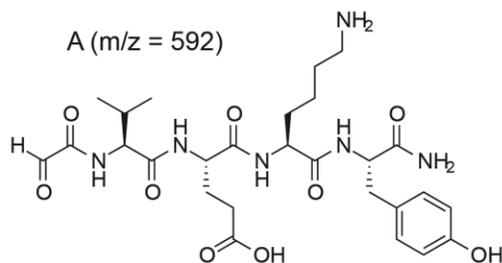
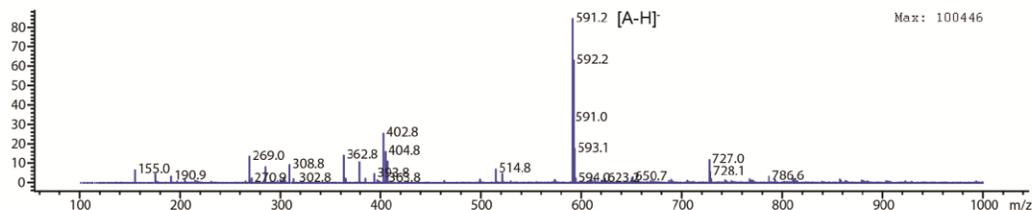


Appendix E: Wittig reaction on phage

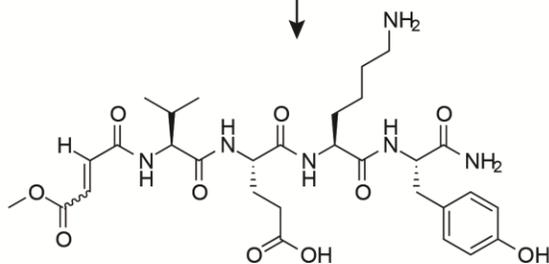
We postulate that the aldehyde generated on SVEK phage by periodate oxidation would react with a ylide, such as methyl (triphenylphosphoranylidene)acetate (MTPPA) to give the Wittig product (Appendix E-1A). This reaction is special in a way that it forms a carbon-carbon bond, which is quite different from the reactions we usually performed on phage. At pH 7.4, all SVEK phage reacted with MTPPA (1 mM) within 4 hours (Appendix E-1B), and no phage was biotinylated when chasing with aminoxy-biotin. The reaction slowed down as pH decreased. These pulse-chase reactions suggested that the free aldehydes were converted to the Wittig product. These experiments, however, do not prove the selectivity of the reaction. To do so, a biotin-ylide probe is required. Direct Wittig reaction between the biotin-ylide probe and a mixture of library and wild-type phage would unambiguously confirm the selectivity of the reaction.



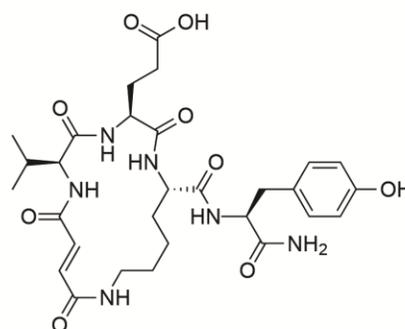
Appendix E-1. Wittig reaction on phage. (A) Scheme illustrating Wittig reaction between MTPPA and SVEK phage. (B) After pulsing the aldehyde-phage with MTPPA for 1, 2, 4 or 24 hours, the unreacted phage was chased with aminoxy-biotin and the biotinylated phage was quantified by biotin-capture assay. The percentage of unreacted phage was measured by the following equation: $[(\text{number of phage before capture} - \text{the number of phage after capture}) / \text{number of phage before capture}] \times 100\%$.



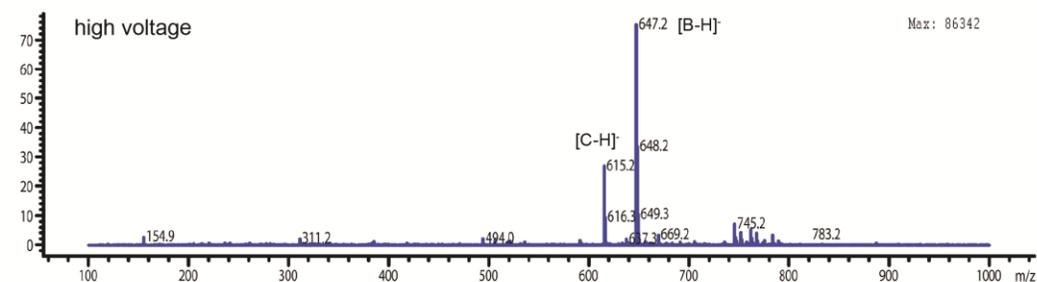
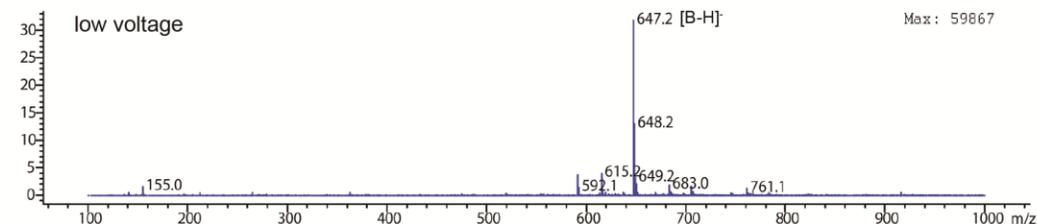
MTPPA (1.1 eq.)



B (m/z = 648)



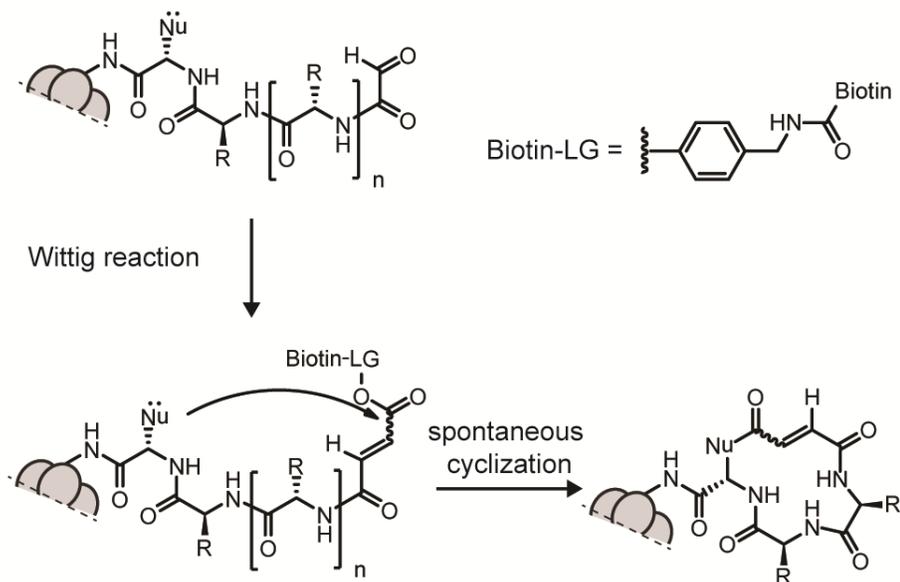
C (m/z = 616)



Appendix E-2. Wittig reaction on model peptide. Glyoxyl-VEKY-CONH₂ (50 μL, 0.17 mM in PBS, pH 7.4) was mixed with MTPPA (50 μL, 0.19 mM in

DMSO). The reaction mixture was analyzed by ESI-MS. At low voltage, we observed the desired Wittig product ($[M-H]^-$ $m/z = 647$). However, at high voltage, we observed another significant peak ($[M-H]^-$ $m/z = 615$) along with the desired Wittig product. The extra peak could correspond to a cyclized product which is formed by eliminating a methanol ($m/z = 32$). This cyclized product could be an artifact though, due to the fragmentation induced by the high voltage.

One important thing to note is that the Wittig product is formed cleanly between the purified glyoxyl-peptide and MTPPA. More works are required to assess the possibility to perform oxidation and Wittig reaction in one-pot, without isolating the aldehyde intermediate. Two important questions have to be answered: (i) is the GSH quencher compatible with the Wittig reagent, such as MTPPA? (ii) could the thiol of excess GSH undergo Michael addition to the newly formed Michael acceptor of the Wittig product? Analyzing the reaction crude of the one-pot reaction using UPLC-MS instrument should address these questions easily.

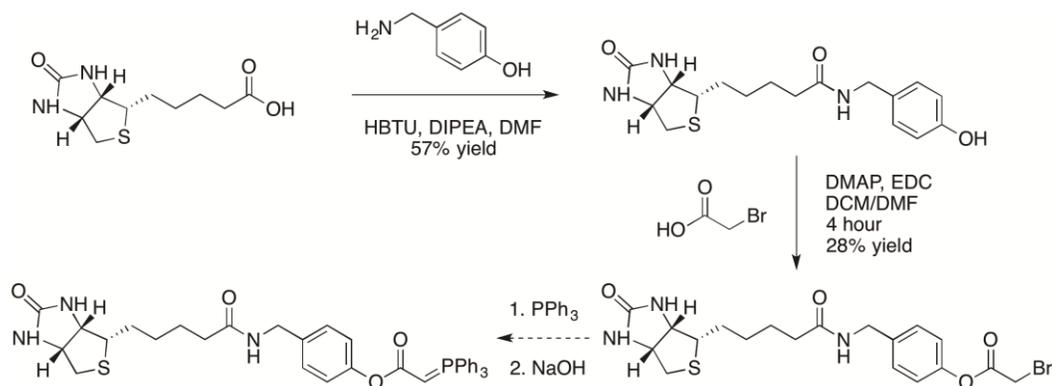


Appendix E-3. We hypothesized that the Wittig product formed on phage might be able to undergo cyclization with nucleophilic side chain of peptide. Although we did not observe or isolate any cyclized Wittig product using the model peptide (SVEKY) and MTPPA, by tuning the pH and the leaving group of the ester, it should be possible to induce a spontaneous intramolecular cyclization after the Wittig reaction. Furthermore, we could apply this idea on phage-displayed peptide libraries and select for peptide sequences that were able to undergo the cyclization. This idea could be realized by capturing the biotinylated phage (the Wittig product), followed by eluting phage clones (specific peptide sequences) that could undergo the intramolecular

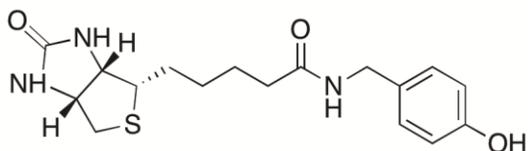
cyclization at elevated pH. The idea sounds sexy, but remains difficult to execute. Phage could be eluted, *i.e.*, released from the bead, by eliminating the biotin-leaving group through hydrolysis of the ester bond instead of intramolecular cyclization. Design of control experiments that rule out this possibility remains elusive. Furthermore, the *cis/trans* isomer of the Wittig product and the potential Michael acceptor of the Wittig product further complicate the design.

Protocol for monitoring the Wittig reaction on phage

To the solution of SVEK phage (99 μL , $\sim 10^{11}$ PFU mL^{-1}) in PBS, NaIO_4 (1 μL , 60 μM in water) was added. The reaction was incubated on ice for 5 min, followed by quenching with GSH (1 μL , 50 mM in water) for 10 min. An aliquot of the reaction mixture (20 μL) was transferred to a fresh microcentrifuge tube and was diluted with $2\times$ NaOAc buffer (25 μL , 200 mM, pH 5/6/7.4). To the resulting mixture (45 μL), MTPPA (5 μL , 10 mM in DMSO) was added and the mixture was incubated at RT for 1, 2, 4 or 24 hours. At each time interval, an aliquot of the reaction mixture (10 μL) was mixed with aminoxy-biotin (10 μL , 2 mM in 200 mM anilinium acetate, pH 4.7). The resulting mixture was incubated at RT for 1 hour. The reaction was quenched by diluting the mixture and the fraction of biotinylated phage was quantified by biotin-capture assay.

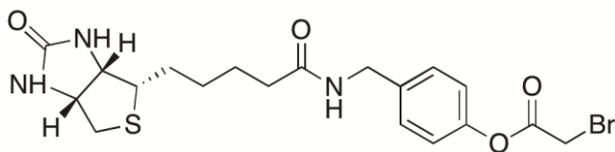


Appendix E-4. Reaction scheme for the synthesis of biotin-tagged phosphorilide that might allow intramolecular cyclization of the Wittig product with reactive side chains of peptide.



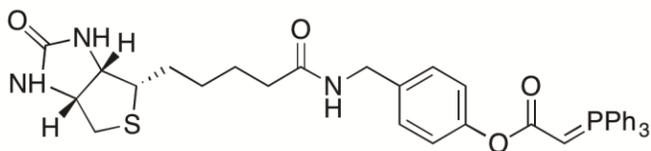
Biotin-phenol conjugate

D-(+)-biotin (0.50 g, 2.05 mmol, 1 eq.) was dissolved in DMF (41mL). HBTU (0.78 g, 2.05 mmol, 1 eq.) was added to the solution. Then, 4-hydroxybenzylamine (0.30 g, 2.46 mmol, 1.2 eq.) was added, followed by DIPEA (0.81 mL, 4.92 mmol, 2.4 eq.). After stirring overnight at room temperature, DMF was evaporated in high-performance vacuum. Then, the solution was washed with water (2 × 50mL) and HCl (1M, 1 × 50mL), and dried over Na₂SO₄. The drying agent and solvent was removed. Then, the crude product was further purified on silica gel (40 g) with a gradient of 0–10% methanol in DCM using CombiFlash® Rf. Removal of solvent *in vacuo* yielded the product as white solid (478 mg, 57.0% yield).



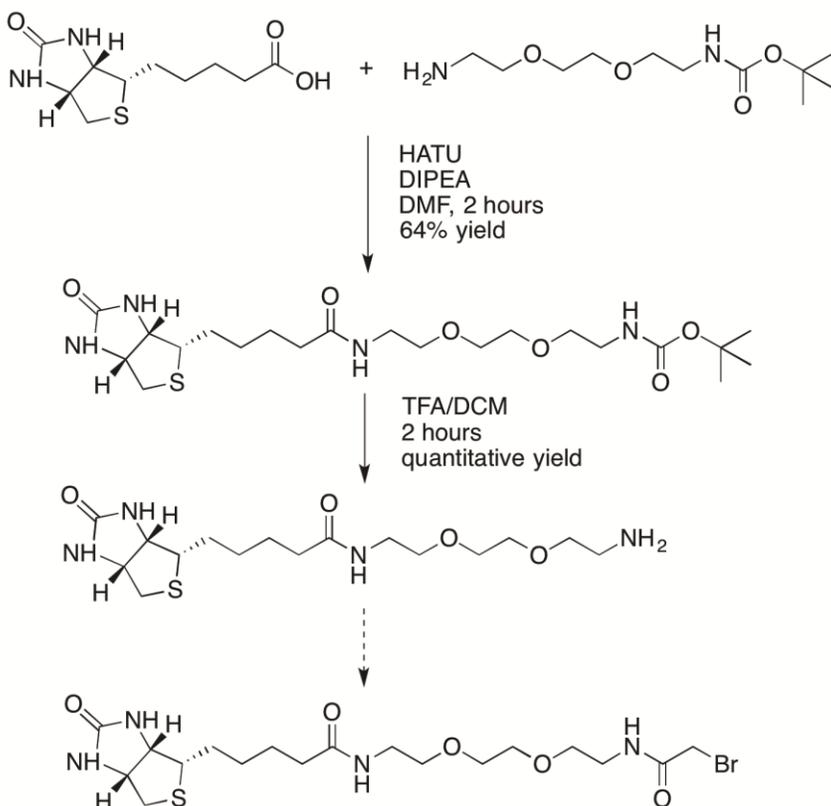
Biotin-bromoacetate conjugate

Biotin-phenol conjugate (100 mg, 0.29 mmol, 1 eq.), bromoacetic acid (48 mg, 0.34 mmol, 1.2 eq.), DMAP (42 mg, 0.34 mmol, 1.2 eq.) were dissolved in a mixture of DCM and DMF (7.5 mL and 2.5 mL). EDC (110 mg, 0.57 mmol, 2 eq.) was added. The reaction mixture was stirred for 4 hours at room temperature. Then, the solution was washed with water (2 × 10 mL) and HCl (1M, 1 × 10 mL), and dried over with Na₂SO₄. The solvent was evaporated *in vacuo*. White crystalline product was isolated (28 mg, 28.0% yield). Earlier workup (3-4 hours) gave product of higher yield. The ester bond is labile and susceptible to attack by moisture or methanol, as evidence by complicated NMR spectra upon storage of the product in CD₃OD overnight.

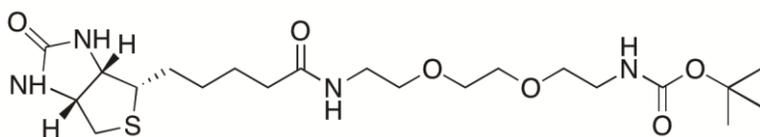


Ester-phosphor-ylide (unsuccessful)

Biotin-bromoacetate conjugate (22 mg, 0.047 mmol) was dissolved in dry THF (5 mL). Triphenyl phosphine (17 mg, 0.052 mmol, 1.1 eq.) dissolved in dry THF (5 mL) was added to the mixture. The reaction was stirred overnight at room temperature under nitrogen. The solvent was removed. Then, the reaction crude was purified by separation using water and DCM. The product should have dissolved into aqueous phase since it is in salt form. We observed a new spot in the aqueous phase and suspected that as the precursor of the ylide. Therefore, aqueous layer was isolated and treated with NaOH (~10 drops, 2M) until reaching pH 14. DCM was added to extract the product. The product dot disappeared. This is probably due to low concentration or instability of both starting material and product.

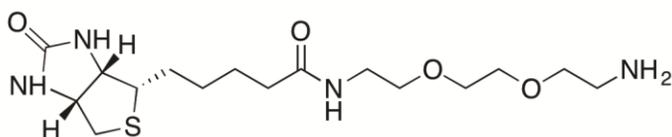


Appendix E-5. Reaction scheme for the synthesis of biotin-tagged phosphor-ylide that doesn't allow intramolecular cyclization of the Wittig product with reactive side chains of peptide.



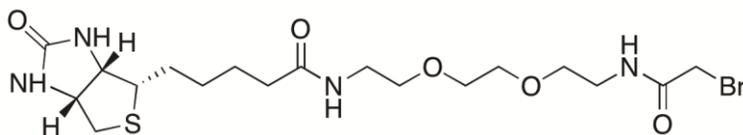
Biotin-PEG-NHBoc

D-(+)-biotin (3.07g, 12.6 mmol, 1.2 eq.) and HATU (4.78 g, 12.6 mmol, 1.2 eq.) was dissolved in DMF (200 mL). The H₂N-PEG-NHBoc (2.6 g, 10.5 mmol, 1 eq.) was dissolved separately in DCM (100 mL). After mixing the two solutions, DIPEA (4.15 mL, 25.2 mmol, 2.4 eq.) was added. The reaction mixture was stirred at room temperature for 2 hours. The solvent was then evaporated *in vacuo*. The crude was dissolved in DCM (200 mL), washed with HCl (1M, 1 × 200 mL), followed by brine (1 × 200 mL), and then dried with Na₂SO₄. The drying agent and solvent were removed. The crude product was further purified using silica gel (120 g) with a gradient of 0–20% methanol in DCM using CombiFlash[®] Rf. Removal of solvent *in vacuo* yielded the product as white solid (3.8 g, 64% yield).



Biotin-PEG-NH₂

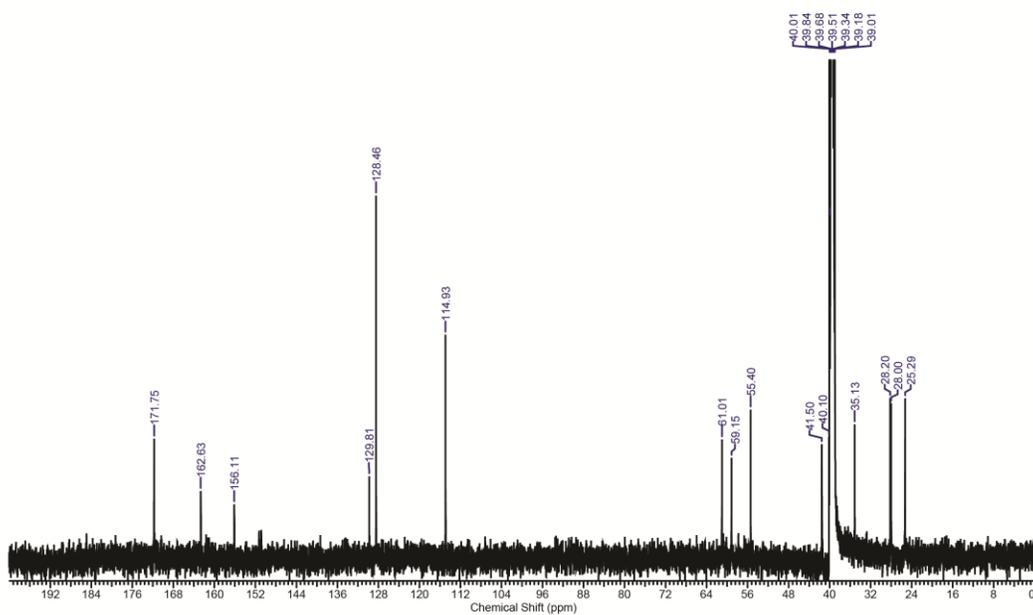
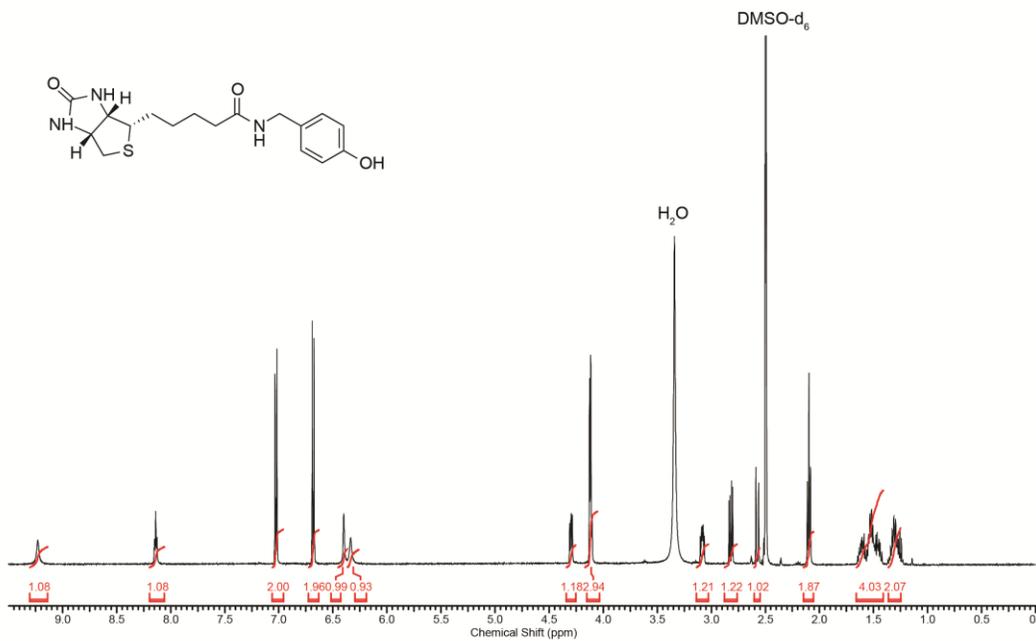
Biotin-PEG-NHBoc (3.8 g, 8 mmol) was dissolved in TFA/DCM mixture (1:1) and was stirred at room temperature for 2 hours. The solvent was evaporated. The reaction crude was dissolved in water and washed with toluene. The aqueous solution was lyophilized to give the product as TFA salt (quantitative yield).

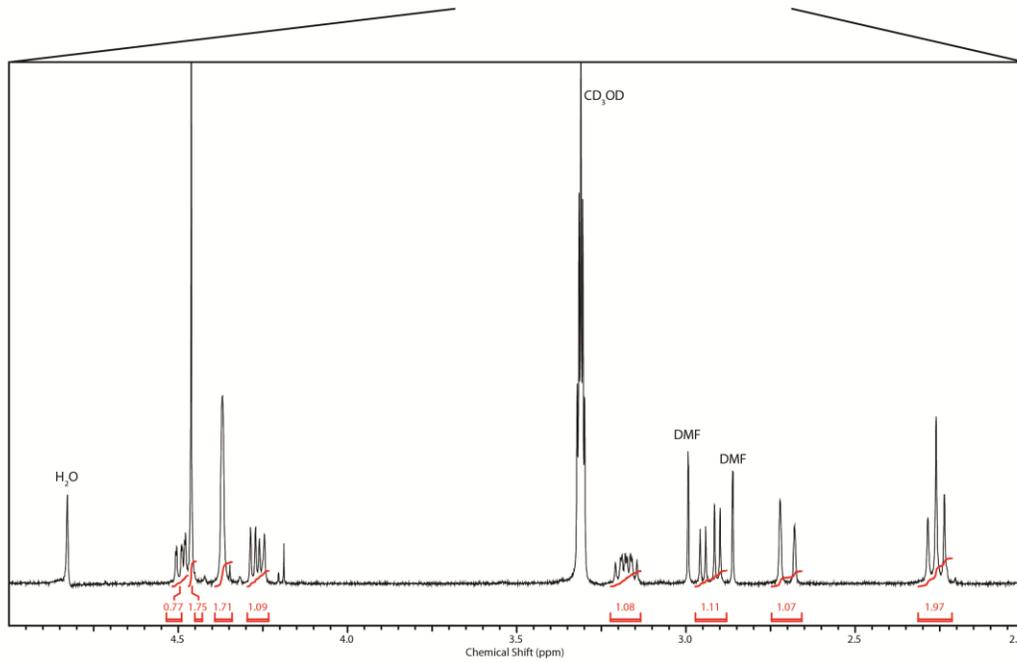
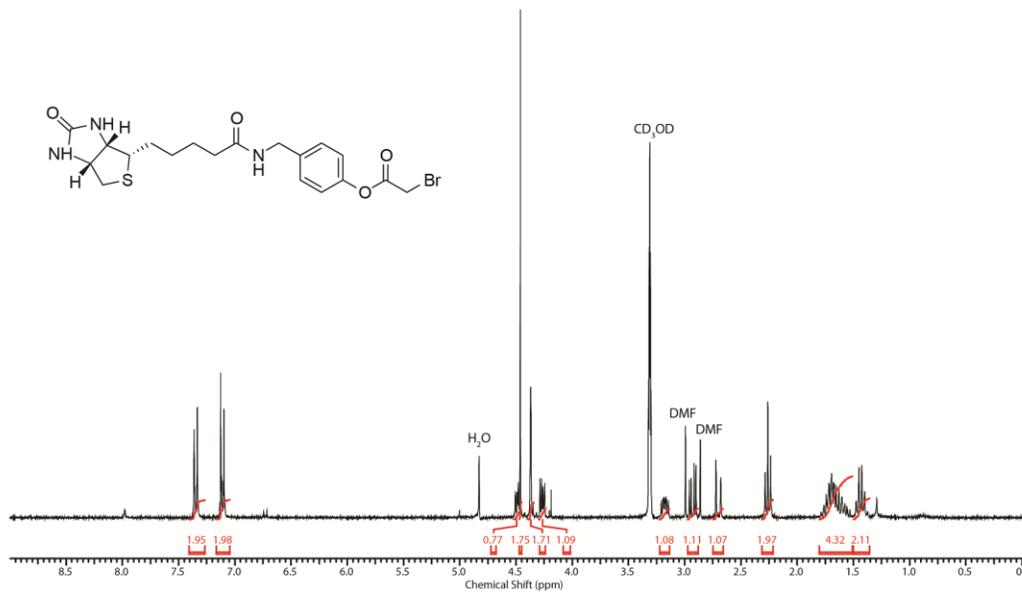


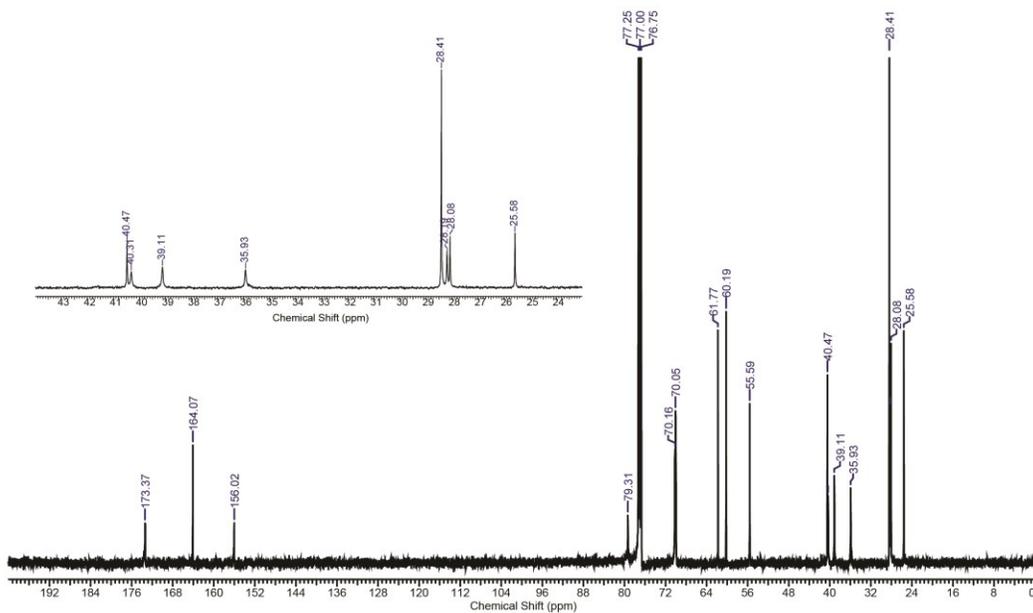
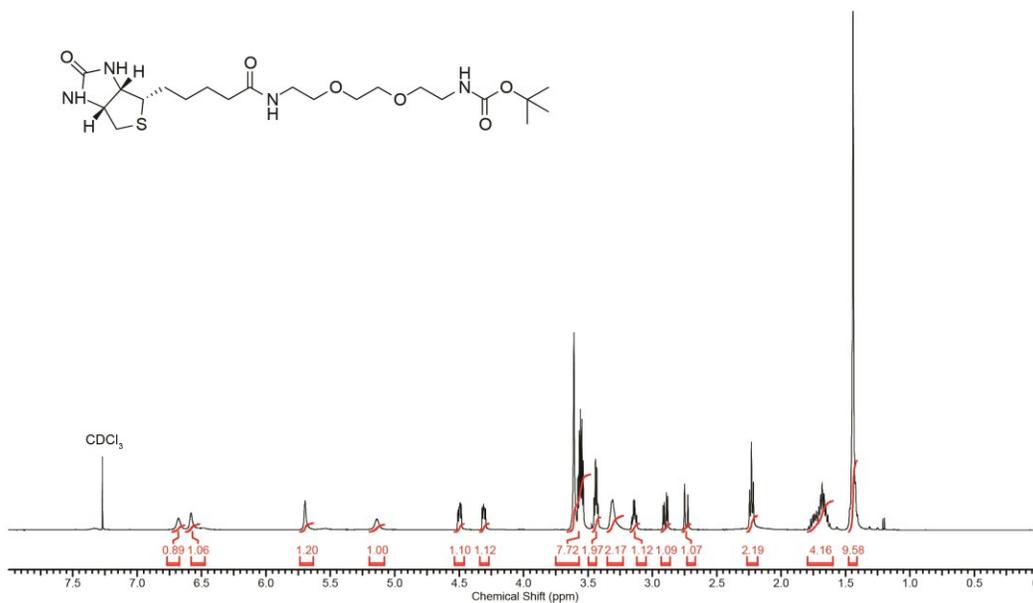
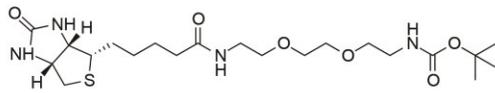
Biotin-PEG-bromoacetate (unsuccessful)

Biotin-PEG-NH₂ (500 mg, 1.34 mmol, 1 eq.) was dissolved in DMF (40 mL). Bromoacetic acid was dissolved separately in minimal amount of DMF (~10 mL). HATU (1.02 g, 2.68 mmol, 2 eq.) was added to the bromoacetic acid solution. The two solutions were combined. DIPEA (2.2 mL, 13.4 mmol, 10 eq.) was added (excess base was added to neutralize the carried-over TFA).

The reaction mixture was stirred at room temperature for 3 hours. Removal of DMF was critical for the purification of the product; otherwise, DMF will elute the product and no separation will be achieved on the column. The high-performance evaporation was not available at that time due to instrument breakdown. The direct coupling with bromoacetyl bromide instead of bromoacetic acid and activator was faster; however, the desired product could not be isolated. I suspect that there is too much side product and the desired product is very unstable in the reaction mixture.

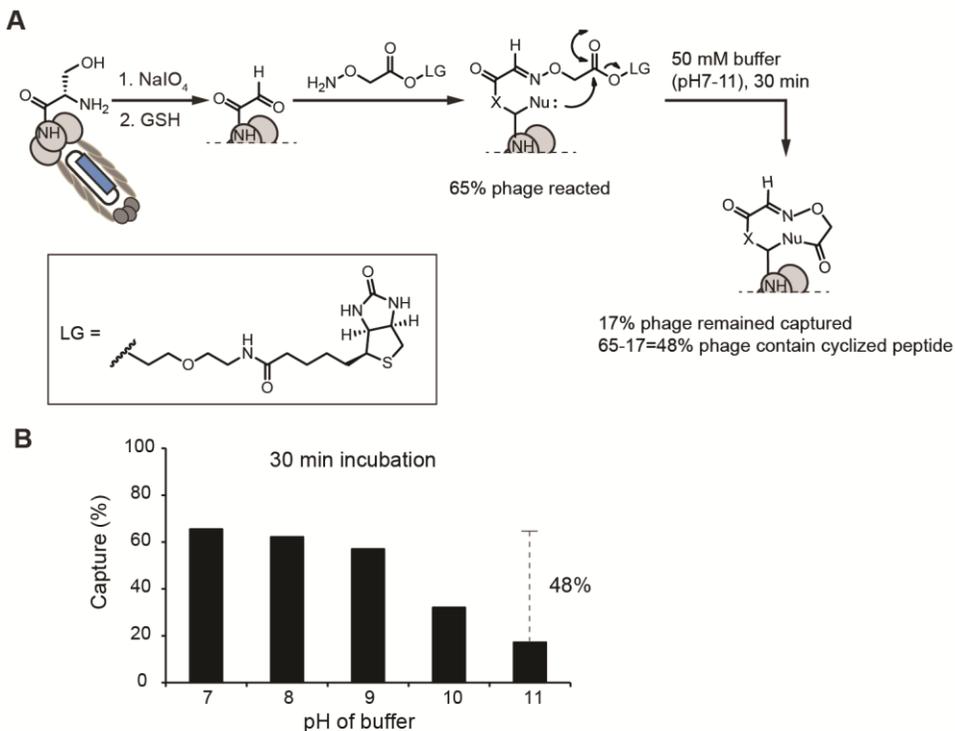




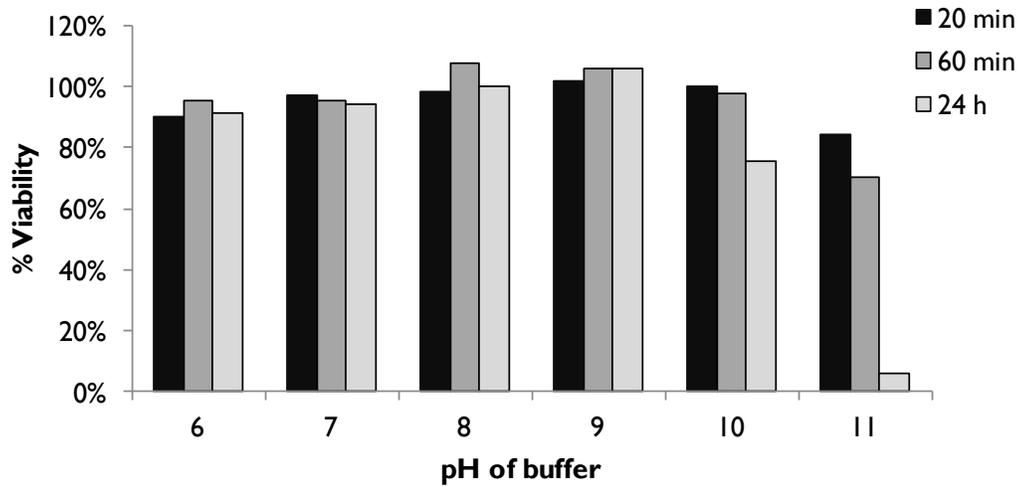


Appendix F: Spontaneous cyclization on phage

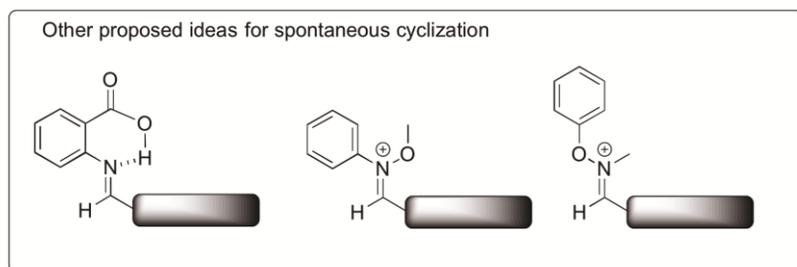
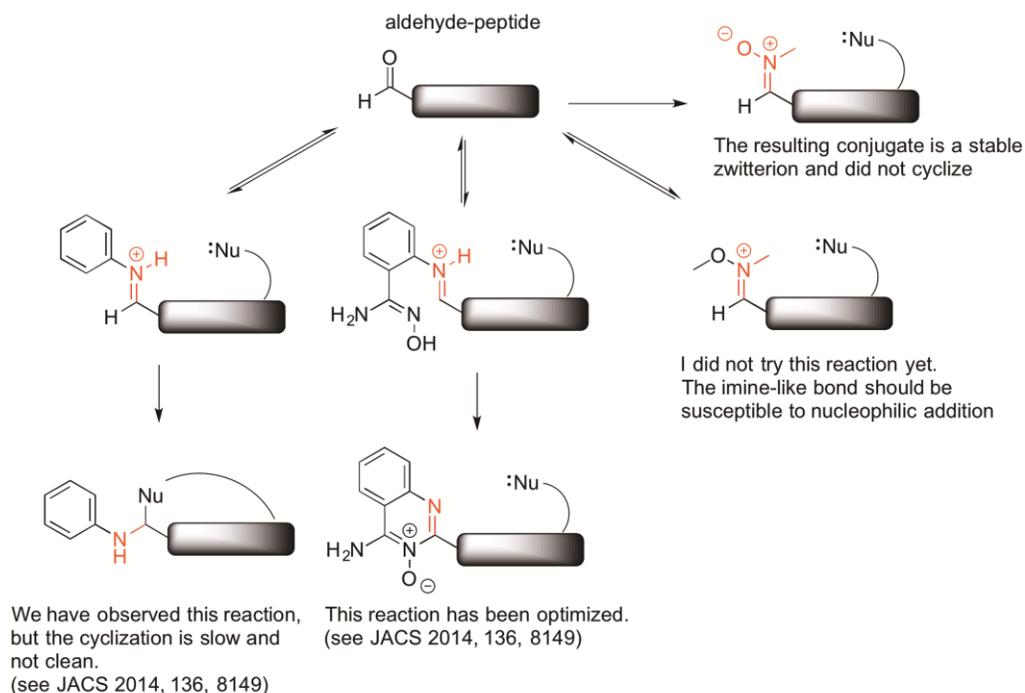
We propose to select for efficient cyclization that use native nucleophiles in phage-displayed peptides. We propose to use a linker that contains two reactive groups: (i) group for site-specific conjugation of the linker to phage. It could be, for example, Wittig reaction described in previous section or simply the oxime ligation of an aminoxy linker, and (ii) a leaving group containing the biotin tag.



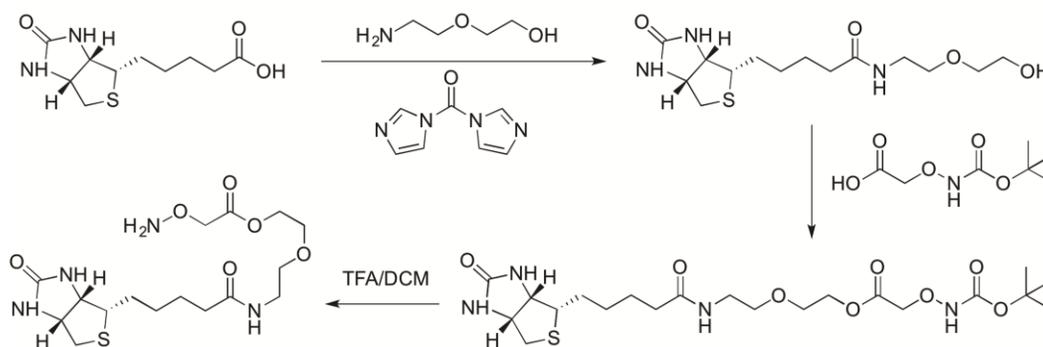
Appendix F-1. The phage-displayed peptide (SVEKNDQKTYHA here) reacted with a biotin-tagged linker might undergo intramolecular cyclization upon increase of pH. (A) The ester is activated by a leaving group and the α -electronegative atom ($-\text{ON}=\text{C}$). (B) Increase of pH enhances the nucleophilicity of the potential nucleophile and might induce the spontaneous cyclization through addition-elimination of the ester. This hypothesis is supported by the decrease of phage capture due to the loss of biotin tag, after incubation in buffer of higher pH for 30 min. The possibility that elimination of the biotin tag through hydrolysis, however, could not be ruled out.



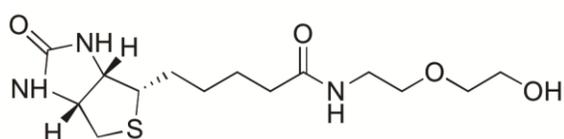
Appendix F-2. The stability of SVEK phage in buffer of different pH. Phage remained infective up to pH 9 even after incubation for 24 hours. At pH 10, infectivity of phage began to decrease upon prolonged incubation and the situation is exacerbated at pH 11.



Appendix F-3. Other than cyclization through addition-elimination of the ester, we postulate that an iminium-like intermediate might undergo spontaneous cyclization with reactive nucleophile of native peptide. This idea is supported by evidences that, glyoxyl-VEKY displayed on phage or the peptide itself, when exposed to aniline in the absence of aminoxy-biotin, led to the loss of aldehyde and irreversible loss of reactivity toward aminoxy-biotin (see Kitov et. al *JACS* 2014). Although the stable intermediate was not isolated, we believe it is a cyclized product. Therefore, the iminium-like intermediate could be intercepted by a nucleophile through intramolecular cyclization. For example, 6-endo-trig cyclization of the iminium intermediate formed between 2-amino benzamidoxime and N-terminal glyoxal-peptide has been optimized (see Kitov et. al *JACS* 2014). I suggest several other reagents, as shown in the figure, that are worth trying. Spontaneous cyclization by nucleophiles present in native peptide is valuable because the reaction could potentially allow the generation of macrocyclic peptide library of different ring size and conformational structure. One-step reaction then greatly expands the diversity of the library.

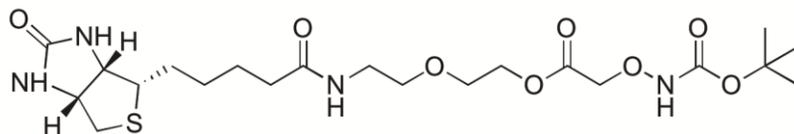


Appendix F-4. Reaction scheme for the synthesis of biotin-PEG-ONH₂ used in Appendix F-1.



Biotin-PEG

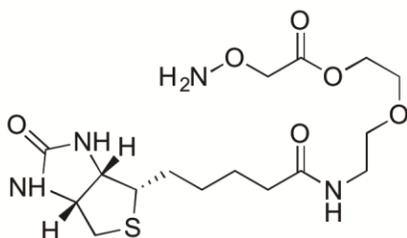
D-(+)-biotin (334 mg, 1.37 mmol, 1 eq.) was dissolved in DMF (7.5 mL) while heating at 55 °C. The solution was then allowed to cool to room temperature. Carbonyldiimidazole (375 mg, 2.31 mmol, 1.7 eq.) was added as a solution in DMF (1.5 mL). The reaction mixture was stirred for 4 hours at room temperature. To the reaction mixture was then added 2-(2'-aminoethoxy)ethanol (565 mg, 5.38 mmol, 3.9 eq.) as a solution in DMF (3.0 mL). The reaction mixture was stirred for 16 hours. The solvent was removed *in vacuo* and the residue was purified on silica gel (80 g) with a gradient of 0–7.5% methanol in DCM using CombiFlash® Rf. Removal of solvent *in vacuo* yielded the product (300 mg, 66% yield).



Biotin-PEG-ONHBoc

Biotin-PEG (100 mg, 0.3 mmol, 1 eq.), (Boc-aminooxy)acetic acid (173 mg, 0.9 mmol, 3 eq.), DCC (187 mg, 0.9 mmol, 3 eq.) and DMAP (11 mg, 0.09 mmol, 0.3 eq.) were added to an oven-dried flask containing molecular sieve (3 Å). The flask was evacuated and then filled with nitrogen (repeat three times). To the mixture, anhydrous DCM (25 mL) was added. The reaction mixture was stirred at room temperature for 24 h and then filtered through Celite to remove the molecular sieve. The organic phase was washed with 1 N HCl (50 mL), brine (50 mL), and then dried with Na₂SO₄. After removal of Na₂SO₄ and the solvent, the residue was purified on silica gel (40 g) with a

gradient of 0–5% methanol in DCM using CombiFlash® Rf. Removal of solvent *in vacuo* yielded the product (68 mg, 45% yield).



Biotin-PEG-ONH₂

Biotin-PEG-ONHBoc (25 mg, 0.05 mmol) was dissolved in TFA/DCM mixture (1:1) and was stirred at room temperature for 1 hour. The solvent was evaporated. The reaction crude was dissolved in water and washed with toluene. The aqueous solution was lyophilized to give the product as TFA salt (quantitative yield).

