

**Toward the use of Ankyrins and Affibodies as Scaffolds for Glycan
Binding Proteins: A Directed Evolution and Computational Approach**

Ruben Warkentin

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Biology) at

Concordia University, Montreal Quebec, Canada

August 2022

©Ruben Warkentin, 2022

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Ruben Warkentin

Entitled: Toward the use of Ankyrins and Affibodies as Scaffolds for Glycan Binding Proteins: A Directed Evolution and Computational Approach

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Chair
Dr. Malcolm Whiteway

Examiner
Dr. David Kwan

Examiner
Dr. Peter Pawelek

Examiner
Dr. Steve Shih

Thesis Supervisor
Dr. David Kwan

Approved by

Graduate Program Director

Dr. Robert Weladji

Dean of Arts & Science

Dr. Pascale Sicotte

Date: August 15th, 2022

ABSTRACT

Toward the use of Ankyrins and Affibodies as Scaffolds for Glycan Binding Proteins: A Directed Evolution and Computational Approach

Ruben Warkentin

Glycans are present in all domains of life and have broad physiological functions that are implicated not only in normal, healthy biological functions, but also in various diseases such as cancer, making them diagnostic and therapeutic targets. The study of glycans is currently limited in part by a lack of specific tools to target glycans of interest, with available glycan binding proteins often having low affinity and specificity. Thus, there is a need to develop new tools that can accelerate the development of novel glycan binding proteins. Here we discuss the application of directed evolution by mRNA display using new binding protein scaffold libraries, in combination with molecular dynamics simulations, as tools for developing novel glycan binding proteins.

We designed mRNA display compatible libraries of affibodies and designed ankyrin repeat protein (DARPin) to be screened for binding against sialyl Lewis X (SLe^X), a tumour associated carbohydrate antigen that is overexpressed on cell surfaces of various cancers. We developed an improved enzymatic synthesis protocol for SLe^X and describe a click chemistry immobilization method coupled with a fluorescent-based lectin assay to test glycan immobilization. The affibody library was successfully created for future selection, whereas the DARPin library assembly needs to be optimized further. Future mRNA display selection should also be complemented with the molecular dynamics probing method developed here. We demonstrate that short, computationally inexpensive probing simulations were able to identify the binding site of a nucleoside sugar dTDP-Qui3N in an ankyrin domain of an *N*-formyltransferase. Future approaches could simulate protein variants selected by display methods to identify glycan-protein interactions that may be used to improve the protein design. In combination, the tools developed here provide a framework to accelerate the discovery and production of future glycan binding proteins with applications in cancer diagnostics and therapeutics.

ACKNOWLEDGEMENTS

This work could not have been done without the people that supported me along the way. I would like to thank the SynBioApps program and the Department of Biology at Concordia University. It has been a pleasure to work under Dr. David Kwan's supervision over the last two years – he provided an environment that enabled me to explore my scientific curiosities and gave the guidance I needed along the way.

I thank Dr. Marcos Di Falco for training me on mass spectrometry and Dr. Chris Hipolito and Dr. Li Yongqi for providing their mRNA display expertise. I would also like to thank my committee members, Dr. Malcolm Whiteway and Dr. Peter Pawelek, for their insightful and encouraging feedback during the committee meetings. Special thanks to everyone in the Kwan Lab, Dr. Ifthiha Mohideen and Dr. Lan Huong for helping me get started in the Kwan lab and (Dr.) Mohamed Nasr for his nods of encouragement throughout the project. Trisha Ghosh and Sara Ouadhi for providing insightful knowledge about lab techniques and helping me with my coding troubles, respectively. Special thanks to Jacob Sichei, Dulce Valdez, and Keegan Scott, I will miss our chats and thank you for helping me remain caffeinated throughout the years. This project would not have been possible without Brandon Albert and Pariya Yousefi, two incredibly talented students that I had the privilege of supervising throughout their undergraduate theses.

This work was supported through NSERC-CGS-M, FRQNT Master's (B1X), and PROTEO scholarships. Computational work was enabled in part by Calcul Québec and the Digital Research Alliance of Canada. Molecular dynamics simulations used NAMD, which was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. Structural data of proteins was obtained from the Protein Data Bank.

I would like to also thank my parents Larissa and Andreas Warkentin and my sisters Emily, Lea, Jenny, and Debora for their emotional support throughout the years (das heißt jetzt Master Klugscheißer).

Lastly, I thanks my partner Gabrielle Godbille-Cardona, I would not have been able to do this without you. Thank you for all the love and support you have given me over the years.

CONTRIBUTIONS

This project was conceived by Ruben Warkentin^{1,2} and Dr. David Kwan^{1,2,3}. Dr. Mostafa Nategh conceived project details related to steered molecular dynamics. Brandon Albert helped with the synthesis of SLe^X-PEG₃-Azide and Pariya Yousefi helped with the DARPIn library segment synthesis as well as the recombinant protein expression and purification of WlaRB and WlaRG. Dr. Marcos Di Falco⁴ performed some of the MS experiments analyzing SLe^X-PEG₃-Azide synthesis reactions. Trisha Ghosh¹ and Dr. Christopher Hipolito helped in the mRNA display method development. All other experiments were designed, performed, and analyzed by Ruben Warkentin.

Chapter 2 is a first author review article published in *Molecules*, by Ruben Warkentin and David H. Kwan, which contains detailed background information for this thesis:

Warkentin, R.; Kwan, D. H. Resources and Methods for Engineering “Designer” Glycan-Binding Proteins. *Molecules* **2021**, *26* (2), 380. <https://doi.org/10.3390/molecules26020380>.

1. Department of Biology, Centre for Applied Synthetic Biology, and Centre for Structural and Functional Genomics, Concordia University, 7141 Sherbrooke Street West, Montreal, QC H4B 1R6, Canada
2. PROTEO, Quebec Network for Research on Protein Function, Structure, and Engineering, Quebec City, QC G1V 0A6, Canada
3. Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke Street West, Montreal, QC H4B 1R6, Canada
4. Centre for Structural and Functional Genomics, Concordia University, 7141 Sherbrooke Street West, Montreal, QC, H4B 1R6 Canada

TABLE OF CONTENTS

List of Figures	x
List of Tables	xii
List of Abbreviations	xiii
CHAPTER 1: INTRODUCTION	1
1.1 Glycans and Their Role Human Disease	1
1.2 Glycan Binding Protein Scaffolds and Directed Evolution	3
1.3 mRNA Display.....	4
1.4 Molecular Dynamics Simulations of Glycans.....	5
1.5 Research Outline.....	6
CHAPTER 2: RESOURCES AND METHODS FOR ENGINEERING “DESIGNER” GLYCAN-BINDING PROTEINS.....	7
2.1 Abstract.....	7
2.2 Introduction.....	8
2.3 Glycan-Binding Protein Scaffolds	10
2.3.1 Lectins.....	11
2.3.2 Carbohydrate Binding Modules	12
2.3.3 Pseudoenzymes	12
2.3.4 Carbohydrate-Active Enzymes (CAZymes)	13
2.3.5 Antibody-based Scaffolds	14
2.3.6 Summary of Available GBP Scaffolds	15
2.4 Mutagenesis Methods for Library Generation	18
2.4.1 Sequence Agnostic Random Mutagenesis	18
2.4.2 Rational and Semi-Rational Mutagenesis	20
2.4.3 Computational Tools for Rational and Semi-rational Mutagenesis	22
2.5 Library Selection and Screening Methods for GBPs	24
2.5.1 Phage display	25
2.5.2 Yeast display	28
2.5.3 Ribosome display	29
2.5.4 mRNA display	32
2.5.5 Glycan Immobilization Strategies for GBP Selection.....	34
2.6 Binding Characterization of GBPs.....	35
2.6.1 Frontal Affinity Chromatography	35
2.7 Surface Plasmon Resonance	36
2.7.1 Titration Calorimetry	37

2.8 Current Limitations in Lectin Engineering	37
2.8.1 Glycan Availability	38
2.8.2 Glycan Binding Protein Scaffolds.....	38
2.9 Conclusion	39
CHAPTER 3: MATERIALS AND METHODS	40
3.1 Recombinant Protein Expression and Purification	40
3.1.1 General Recombinant Protein Expression and Harvesting Protocol.....	40
3.1.2 Affinity Chromatography Protein Purification using Ni-NTA	40
3.1.3 Affinity Chromatography Protein Purification using Amylose Resin	41
3.1.4 Expression and Purification of Enzymes used in the SLe ^X -PEG ₃ -Azide Synthesis.....	41
3.1.5 Expression and Purification of Ankyrin, DARPin, and MtDARPin.....	41
3.1.6 Expression and Purification of Enzymes used in the dTDP-Qui3N Synthesis Reaction....	41
3.2 Enzymatic Synthesis and Purification of SLe ^X -PEG ₃ -Azide	42
3.3 Immobilization and Detection of SLe ^X -PEG ₃ -Azide	43
3.3.1 Click Chemistry immobilization of SLe ^X -PEG ₃ -Azide on Magnetic Beads.....	43
3.3.2 Fluorescent assay for immobilized SLe ^X -PEG ₃ -Azide Detection.....	43
3.4 DARPin and Affibody Library Design	43
3.4.1 DAPRin library design.....	43
3.4.2 DAPRin Library Assembly by Type IIS Digestion and T7 ligation	45
3.4.3 Affibody Library design.....	46
3.5 mRNA Display Protocols	47
3.5.1 <i>In vitro</i> Transcription of the Affibody Library	47
3.5.2 Puromycin Linkage of the Affibody Library	47
3.6 Enzymatic Synthesis and Purification of dTDP-Qui3N.....	49
3.7 Rational Design of Ankyrin Based Carbohydrate Binding Protein	49
3.8 Molecular Dynamics	50
3.8.1 Equilibration of Proteins and dTDP-Qui3N.....	50
3.8.2 Steered Molecular Dynamics Probing Simulation	50
3.8.3 Data Analysis for Steered Molecular Dynamics Probing Simulations	51
CHAPTER 4: RESULTS	52
4.1 Enzymatic Synthesis of SLe ^X -PEG ₃ -azide.....	52
4.2 SLe ^X -PEG ₃ -Azide immobilization using Click Chemistry	56
4.3 DARPin and Affibody Library Design for mRNA Display.....	58
4.3.1 DARPin Library Design and Assembly.....	58
4.3.2 Affibody Library Design and Assembly	61

4.3.3	Affibody Library Transcription, and Puromycin Linkage	62
4.4	Rational Design of Ankyrin Based Carbohydrate Binding Proteins.....	64
4.5	Enzymatic Synthesis of dTDP-Qui3N	65
4.6	Steered Molecular Dynamics of Ankyrin Proteins with dTDP-Qui3N.....	66
CHAPTER 5: DISCUSSION.....		70
5.1	Enzymatic Synthesis of SLe ^X -PEG ₃ -Azide.....	70
5.2	SLe ^X -PEG ₃ -Aide Immobilization on Magnetic Beads.....	71
5.3	Site Saturated DARPin and Affibody Library Design.....	72
5.4	mRNA Display.....	Error! Bookmark not defined.
5.5	<i>In Silico</i> Study of Glycan-Protein Interactions	75
5.6	Enzymatic Synthesis of dTDP-Qui3N	78
CHAPTER 6: CONCLUSION AND FUTURE DIRECTIONS.....		78
REFERENCES		80
APPENDIX.....		89

LIST OF FIGURES

Figure 1.1: Physiological roles of glycosylation in cancer.	3
Figure 2.1: Valency and selectivity of protein scaffolds with glycan binding sites.	11
Figure 2.2: A) Example phagemid design for phage display.....	27
Figure 2.3: Construct designs for yeast display systems.....	29
Figure 2.4: Overview of the ribosome display method.....	31
Figure 2.5: Overview of the mRNA display method.	33
Figure 2.6: Overview of SPR with glycan labelled surfaces.	37
Figure 3.1 Sequence alignment of the ankyrin domain and DARPin.	49
Figure 4.1: Enzymatic synthesis pathway of click-chemistry enabled SLe ^X -PEG ₃ -Azide.	54
Figure 4.2: Mass spectrometry of intermediate SLe ^X -PEG ₃ -Azide reaction steps to remove unreacted precursors.	55
Figure 4.3: Mass spectrometry of the purified product from the enzymatic synthesis SLe ^X -PEG ₃ -Azide.	55
Figure 4.4: Click-chemistry bead labelling and detection protocol.	57
Figure 4.5: Fluorescent bead labelling assay of immobilized SLe ^X -PEG ₃ -Azide on magnetic DBCO beads.....	58
Figure 4.6: DARPin library design and synthesis approach.	60
Figure 4.7: DARPin library PCR amplification and assembly	61
Figure 4.8: Affibody library design and synthesis.....	62
Figure 4.9: Urea-PAGE of the puromycin linked, and unlinked, mRNA affibody library. The band of size of the puromycin-mRNA (P-mRNA) was higher than the unlinked mRNA library. Note that 1 µL of 10 mM P-mRNA and unlinked mRNA were loaded into the wells. Samples were run on an 8 % urea-PAGE and the gel was incubated with SYBR safe for visualization.	63
Figure 4.10: Ankyrin (PDB: 4XCZ) and DARPin (PDB: 2XEE) structural alignment and mutagenesis. .	64
Figure 4.11: Enzymatic synthesis reaction of dTDP-Qui3N and mass spectrometry.....	66
Figure 4.12: Steered Molecular Dynamics (SMD) approach to detect protein-glycan interactions.	67

Figure 4.13: Hydrogen bond heatmaps generated from the SMD simulations of dTDP-Qui3N with ankyrin, DARPin, and MtDARPin.....	68
Figure 4.14: Overlay of the simulated and crystal structure interactions between dTDP-Qui3N and the ankyrin domain.....	69
Figure 5.1: Directed evolution by mRNA display.	75
Figure 5.2: Hydrogen bond requirements.	78
SFigure 6.1: Mass spectrometry of wrongly fucosylated intermediate in SLe ^x -PEG ₃ -Azide reaction.	91
SFigure 6.2: DARPin library 1 st PCR amplification of DARPin library segments.	92
SFigure 6.3: DARPin library PCR amplification following ligation of modules.	92
SFigure 6.4: SDS-PAGE of the recombinant protein purification of WlaRB and WlaRG.	93
SFigure 6.5: Aromatic and phosphate interaction heatmaps between dTDP-Qui3N and ankyrin, DARPin, or MtDARPin.	94
SFigure 6.6: Rotational Autocorrelation Approach and Results.....	95
SFigure 6.7: Mass spectra of dTDP-Qui3N prior to freeze drying.....	96
SFigure 6.8: SDS-PAGE of the recombinant protein purification of DARPin and MtDARPin.....	96

LIST OF TABLES

Table 2.1 Carbohydrate binding proteins (CBPs) with available structural and ligand binding information.	16
Table 2.2 Overview of random mutagenesis methods.....	20
Table 2.3 Overview of site-directed mutagenesis techniques.....	22
Table 2.4 Example of computational approaches used in protein engineering.....	24
Table 3.1: PCR mix for amplifying site saturated DARPin oligos.....	44
Table 3.2: PCR mix for adding Type IIS cut sites to the site saturated DARPin oligos.....	44
Table 3.3: Thermocycler protocols for the DARPin library amplification.....	44
Table 3.4: Reaction mix used in the digestion and ligation of the DARPin library.....	45
Table 3.5: Thermocycler protocols for the DARPin library digestion and ligation.....	45
Table 3.6: PCR mix for of the first affibody library amplification.....	46
Table 3.7: Thermocycler protocols for the affibody library amplification.....	46
Table 3.8: PCR mix for of the second Affibody library amplification.....	47
Table 3.9: Reaction mix for the puromycin linkage to the mRNA affibody library.....	48
STable 6.1: Oligos used in the 1 st PCR for the DARPin library assembly.....	89
STable 6.2: Primers used in the 2 nd PCR for the DARPin library assembly.....	89
STable 6.3: All oligos and primers used in the project.....	89

LIST OF ABBREVIATIONS

AAL: *Aleuria aurantia* Lectin

ACN: Acetonitrile

BCA: Bicinchoninic acid

BSA: Bovine serum albumin

BgaA: Exo- β -galactosidase from *Streptococcus pneumoniae*

BME: β -Mercaptoethanol

Cu(I): Copper (I)

DBCO: Dibenzocyclooctyne

DARPin: Designed ankyrin repeat protein

DMSO: Dimethyl sulfoxide

FPLC: Fast protein liquid chromatography

Fuc: Fucose

Gal: Galactose

GalNAc: N-acetylgalactosamine

GBP: Glycan-binding proteins

GDP: Guanosine-5'-diphosphate

GDP-Fucose: guanosine 5'-diphospho- β -L-fucose

GlcNAc: N-acetylglucosamine

GTP: Guanosine-5'-triphosphate

HPLC-MS: High performance liquid chromatography – mass spectrometry

IPTG: Isopropyl β -D-1-thiogalactopyranoside

IMAC: Immobilized affinity chromatography

LB: Luria broth

MBP: Maltose binding protein

MD: Molecular dynamics

MS: Mass Spectrometry

NAMD: Nanoscale Molecular Dynamics

PCR: Polymerase chain reaction

RPM: Rotation per minute

RNA: ribonucleic acid

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

TACA: Tumour associated carbohydrate antigen

TB: Terrific broth

SLe^X: Sialyl-Lewis X

SMD: Steered molecular dynamics

SPAAC: Strain-promoted alkyne-azide cycloadditions

SpHex: exo- β -N-acetylhexosaminidase from *Streptomyces plicatus*

VMD: Visual Molecular Dynamics

Chapter 1: Introduction

1.1: Glycans and Their Role Human Disease

CHAPTER 1: INTRODUCTION

Here we provide a brief overview of the current literature on glycans, their implications in diseases, and current work on engineering new glycan binding proteins (GBPs). Glycans are implicated not only in normal, healthy biological functions, but also in a variety of diseases, making them diagnostic and therapeutic targets. Current advances in the field of glycobiology are leading to the rapid discovery of novel disease associated glycan epitopes; however, there is a lack of specific and high affinity GBPs to target these antigens. New proteins, such as affibodies and designed ankyrin repeat proteins (DARPs), could be explored as scaffolds for glycan binding proteins. Probing new scaffolds for glycan binding activity using directed evolution approaches coupled with molecular dynamics simulations may accelerate the production of designer GBPs. More detailed background information is found in **Chapter 2:**, a review article we published last year covering advances GBP engineering [1].

1.1 Glycans and Their Role Human Disease

Glycans (oligosaccharides and polysaccharides) are the third class of biopolymers besides proteins and nucleic acids. Glycans exist in a free form, or conjugated to macromolecules, such as proteins (glycoproteins), lipids (glycolipids), and RNAs [2,3]. Glycosyltransferases are responsible for glycan synthesis, by transferring activated monosaccharides to acceptor macromolecules, which occurs in the endoplasmic reticulum (ER), or the Golgi apparatus. Glycan synthesis is influenced by enzymes and substrate availability, but is not template-based like the synthesis of DNA, RNA, and proteins. Due to the wide variety of physiological functions that involve glycans, including cell signaling, energy storage, and cellular structure, they are implicated in a wide variety of diseases – making them diagnostic and therapeutic targets. For example, aberrant glycosylation is a hallmark of malignant tumours and is associated with various types of cancers [4]. Glycans associated with various cancers have been linked to immune system evasion [5], increasing cell proliferation [6,7], causing tumour promoting inflammation [8], and activating metastasis [9] (**Figure 1.1A**).

Chapter 1: Introduction

1.1: Glycans and Their Role Human Disease

Here we focus on the tumour associated carbohydrate antigen (TACA), sialyl Lewis X (SLe^X). SLe^X, a tetrasaccharide composed of N-acetylglucosamine (GlcNAc), Galactose (Gal), N-acetylneuraminic acid (Neu5Ac), and Fucose (Fuc) (**Figure 1.1B**). SLe^X is essential for immune cell function and mediating cell adhesion, is often over-expressed on cancer cell surfaces [10–12]. Increased sialylation, including SLe^X, is generally associated with malignant cells. Cancer cell adhesion is mediated by SLe^X recognition by E-selectin, a receptor on endothelial cell surfaces, which enables the migration of cancer cells along vascular endothelium. SLe^X is available for purchase (Sigma Aldrich); however, it is prohibitively expensive, which is why enzymatic synthesis protocols for SLe^X have been developed [13]. The role of glycans in various diseases, is further discussed in section **2.2**.

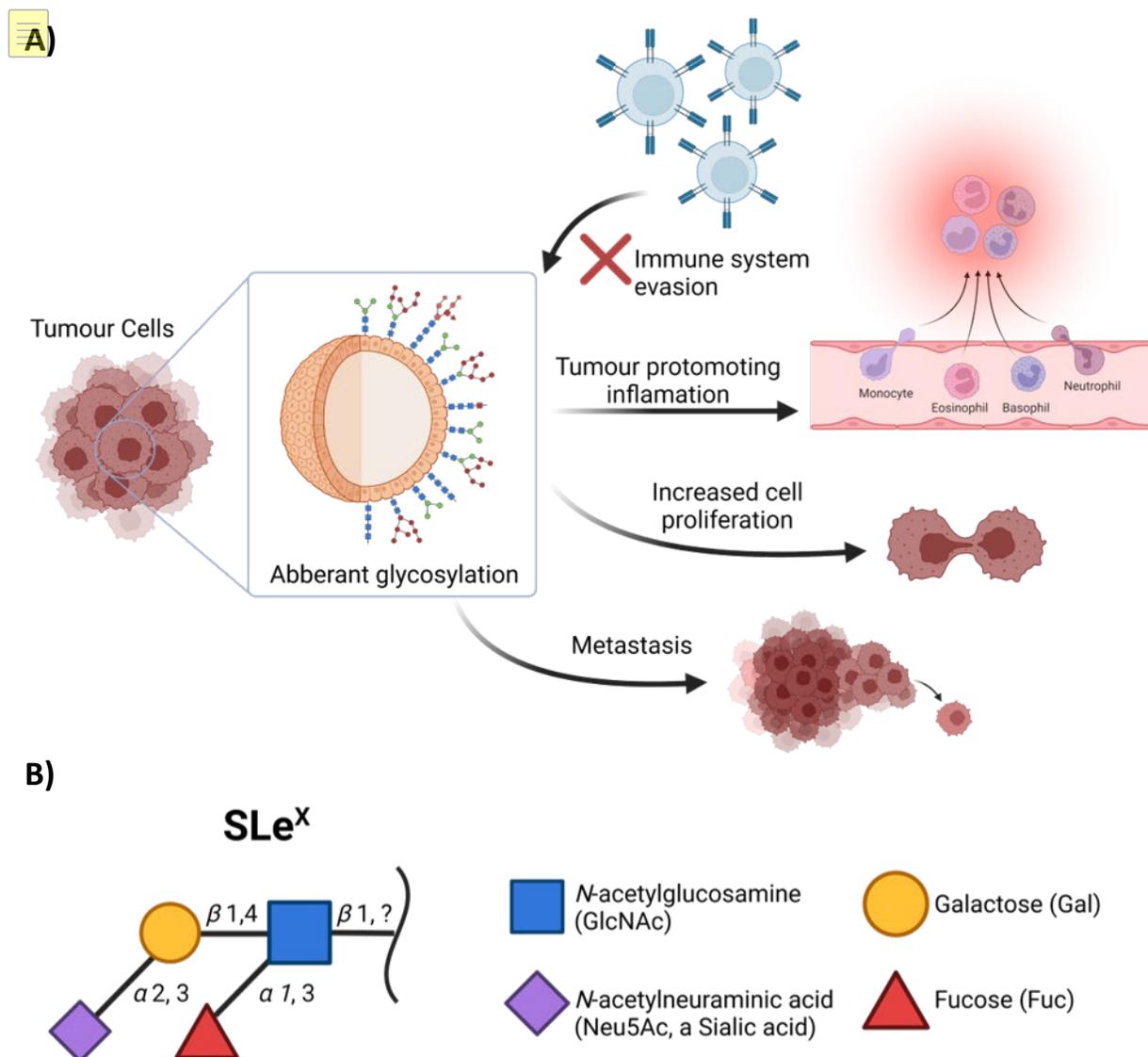


Figure 1.1: Physiological roles of glycosylation in cancer. **A)** Glycans impact cancer progression in various forms, including immune system evasion, tumour promoting inflammation, increased cell proliferation, and activating metastasis. **B)** Sialyl Lewis X (SLe^x) is a tumour associated tetrasaccharide composed of *N*-acetylglucosamine (GlcNAc), Galactose (Gal), *N*-acetylneuraminic acid (Neu5Ac), and Fucose (Fuc).

1.2 Glycan Binding Protein Scaffolds and Directed Evolution

Glycans are found in all domains of life and their complex physiological functions have made them diagnostic and therapeutic targets, which in turn spurred interest in the development of glycan binding proteins (GBPs). Advances in the field of glycobiology are leading to the rapid discovery of new glycan epitopes that could be targeted for potential diagnostics and therapeutics, but there is a lack of sensitive and specific GBPs. Directed evolution approaches are one method that could be used to increase the number of

Chapter 1: Introduction

1.3: mRNA Display

available GBPs. Detailed information on GBPs and directed evolution approaches are covered in sections **2.3** and **2.5**, respectively.

Directed evolution approaches can accelerate the discovery of new binding proteins; however, the starting scaffold used in directed evolution can limit its success. Naturally occurring GBPs have been engineered to bind different ligands, but they tend to have low affinity and specificity to their targets. Carbohydrate-protein interactions are often weak [14], with dissociation constants in the mM range [15]. In nature, the avidity of carbohydrate-protein interactions is increased due to multivalent interactions between larger complexes [16].

Designed ankyrin repeat proteins (DARPin) are a new class of small, heat stable proteins that contain modular internal repeats, allowing for the design of multivalent binding proteins [17,18]. DARPins have never been engineered to bind carbohydrates, but the DARPin design is based on ankyrin proteins found in nature. In 2015, a binding site for a nucleoside sugar, dTDP-Qui3N, was found in the ankyrin domain of *Providencia alcalifaciens* N-formyltransferase [19]. Another promising scaffold are affibodies. Typically affibodies are used for protein-protein interactions, but recent work demonstrated that affibodies can be engineered to bind polysialic acid (an anionic polysaccharide) with low nanomolar affinity [20].

1.3 mRNA Display

Display techniques are directed evolution methods used to screen large peptide or protein libraries. The phenotype of a protein (*i.e.* binding activity) is physically linked to the genotype (*i.e.* coding sequence), allowing for the isolation and identification of proteins from vast variant libraries. Display methods include phage, yeast, ribosome, and mRNA display, which are discussed in detail in section **2.5**. Here we focus on mRNA display. mRNA display allows for cell free screening of protein libraries with up to 10^{14} variants [21,22]. In brief, mRNA display involves the *in vitro* transcription of a DNA library, followed by linkage of the resultant mRNA to puromycin. In the subsequent translation step, puromycin mimics a charged amino-acyl tRNA and enters the ribosome, which creates a covalent bond between the mRNA-linked puromycin and the nascent peptide. The mRNA-peptide complex can then be reverse transcribed to produce

Chapter 1: Introduction

1.4: Molecular Dynamics Simulations of Glycans

a more stable RNA-DNA hybrid, which in turn is incubated with an immobilized target. During each round of selection unbound proteins are washed away while the bound proteins can be recovered by their interaction with the immobilized target and their covalently attached coding mRNAs can then be reverse transcribed and PCR amplified for enrichment and then sequenced enabling the identification of successfully binding partners. More detailed background and figures of the mRNA display cycles are in section **2.5.4**.

1.4 Molecular Dynamics Simulations of Glycans

Molecular Dynamics (MD) simulations are computational techniques that generate atomic trajectories of a system based on Newton's equation of motion. MD programs rely on force fields, which are sets of mathematical formula and parameters that determine atomic coordinates and energies of a molecule in a simulation. The ubiquity and diversity of glycan physiology have spurred development of carbohydrate force fields that enable accurate glycan MD simulations. The MD program NANoscale Molecular Dynamics (NAMD) uses the Chemistry at Harvard Macromolecular Mechanics (CHARMM) force field, which contains carbohydrate specific parameters [23,24]. For information on other available MD programs see section **2.4.3**.

MD simulations provide insights into glycan-protein interactions that can be used to rationally alter the binding affinity or specificity of the protein. MD simulations are well suited for the characterization of glycans, as their inherent flexibility makes their study by X-ray crystallography challenging [25]. Advances in technologies and software allows sampling of larger conformational spaces for hundreds of nanoseconds, which was previously too computationally expensive. Due to the physiological roles of glycan-protein interactions there is interest in applying MD simulations for the characterization interactions that could aid in the rational design of diagnostics and therapeutics [26,27]. MD probing simulations, which simulate a protein of interest with a ligand to identify interactions, have been used extensively to identify protein-small molecule interaction for drug design [28]. Heatmaps of the protein surface and glycan interactions can be generated from the MD probing simulations to identify "hot-spots" on the protein where interactions with

Chapter 1: Introduction

1.5: Research Outline

the ligand of interest are stronger. Yet, probing simulations have not been used to identify glycan-protein interactions. Glycan-protein interactions are often weak with their interactions arising from weak electrostatic and hydrophobic interactions, and hydrogen bonding [15]. Analyzing these interactions in MD simulations can aid the rational design of glycan binding proteins.

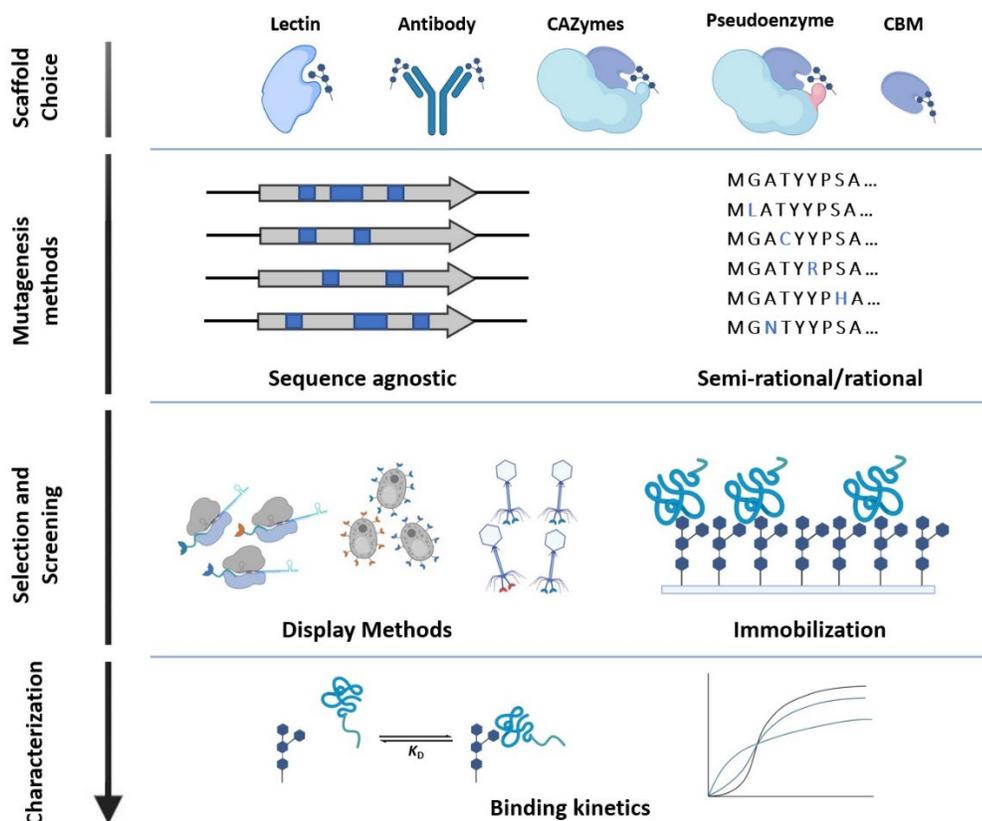
1.5 Research Outline

The study of glycans has vast applications in the medical field and is currently limited by a lack of specific glycan-binding proteins (GBPs). One factor limiting the study of glycans is a scarcity of high affinity and specific tools that can be used to detect glycan structures [29]. Here we aim to provide a framework for the design of glycan binding proteins using a combination of directed evolution and computational methods. Specifically, we aim to create site saturated affibody and designed ankyrin repeat protein (DARPin) libraries to be screened for GBP activity using mRNA display. We intend to test our approach using the glycan antigen SLe^x, a cancer biomarker. We optimized the enzymatic synthesis and purification of an azide labelled SLe^x (SLe^x-PEG₃-Azide) that can be immobilized and used in mRNA display. Additionally, we developed a fluorescent lectin assay for the detection of immobilized SLe^x. However, mRNA display can only provide sequences of proteins that are associated with binding affinity and does not elucidate the binding site location and interactions. Therefore, we aim to supplement the mRNA display approach using molecular dynamics (MD) simulations. To achieve this, we designed MD simulations that probe for interactions between the nucleoside sugar dTDP-Qui3N and the ankyrin domain of the *Providencia alcalifaciens* N-formyltransferase. The variant libraries and molecular dynamics methods developed here can be used to generate novel binding proteins for SLe^x. These methods can also be applied to other glycan antigens to test the versatility of the affibodies and DARPins as glycan scaffolds.

CHAPTER 2: RESOURCES AND METHODS FOR ENGINEERING “DESIGNER” GLYCAN-BINDING PROTEINS

2.1 Abstract

This review provides information on available methods for engineering glycan-binding proteins (GBP). Glycans are involved in a variety of physiological functions and are found in all domains of life and viruses. Due to their wide range of functions, GBPs have been developed with diagnostic, therapeutic, and biotechnological applications. The development of GBPs has traditionally been hindered by a lack of available glycan targets and sensitive and selective protein scaffolds; however recent advances in glycobiology have largely overcome these challenges. Here we provide information on how to approach the design of novel “designer” GBPs, starting from the protein scaffold to the mutagenesis methods, selection, and characterization of the GBPs.



2.2 Introduction

Glycans – a broad term describing carbohydrates including oligosaccharides and polysaccharides – are a third class of important biological macromolecules, following nucleic acids and proteins. All domains of life and viruses contain glycans – they can exist as free sugars, but are more commonly found as glycoconjugates including proteoglycans, glycoproteins, and glycolipids. Glycans are involved in a wide variety of physiological functions and have implications in numerous infectious and non-infectious disease, making them diagnostic and therapeutic targets. Additionally, glycans are involved in various biotechnological and industrial applications. The broad applications of glycans have spurred interest in the generation of glycan binding proteins (GBPs).

GBPs can be categorized into lectins, antibodies, pseudoenzymes, and carbohydrate binding modules (CBMs). Lectins are non-immunoglobulin proteins that contain at least one non-catalytic domain that exhibits reversible carbohydrate binding. CBMs are similar to lectins, but are small binding domains typically found on lectins or carbohydrate active enzymes (CAZymes). CAZymes can be further classified into glycoside hydrolases, glycosyltransferases, polysaccharide lyases, and carbohydrate esterases – detailed information on these enzymes is available through the Carbohydrate Active Enzymes (CAZy) database [30]. Over time, some CAZymes have evolved into pseudoenzymes that have lost their catalytic activity but retain their glycan binding properties; these have been categorized separately from lectins as their overall structure is distinct. Antibodies against glycans are also found in nature, however glycans are generally poorly immunogenic, leading to low binding affinities and specificity of anti-glycan antibodies. The aforementioned protein categories have been used to create a variety of GBPs and their use in GBP engineering is discussed in **section 2.3**.

One application for GBP engineering is towards the use in diagnostics and therapeutics. Glycan recognition is involved in a variety of bacterial and viral infections, which has led to the production of several diagnostic and therapeutic GBPs. For example, the glycan epitopes displayed on the envelope spike protein of the human immunodeficiency virus type-1 (HIV-1) are involved in immune system evasion [31]. More than a

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.2: Introduction

dozen lectins have been identified to exhibit anti-HIV properties and some are currently being further explored for therapeutic potential [32]. Similar to HIV-1, glycans are involved in immune system evasion in various other viral and microbial diseases – an excellent review has been published on the therapeutic value of GBPs in microbial infections [33].

GBPs also have the potential to target glycans that are involved in a variety of non-infectious diseases, such as diabetes [34], arthritis [35], and cancer – with glycans of the latter being the most well studied as biomarkers. Aberrant glycan profiles are a hallmark of malignant tumour transformations and are commonly targeted for cancer diagnostics and therapeutics [36]. The exact glycosylation patterns of tumours varies greatly, but include glycan epitopes that can be categorized as T-antigens, poly-*N*-acetylglucosamines (PLAs), Lewis antigens, and glycosaminoglycans, among others [4]. Cancer glycan epitopes have been recently reviewed elsewhere, and therefore will not be discussed in greater detail here [4].

GBP applications are not limited to the medical field and have seen broad applications in industrial settings and in biotechnologies. Lectins have been used as biological insecticides in genetically engineered crops and it has been suggested that this approach may reduce the use of synthetic insecticides – reducing risk to human health and the environment [37]. On another note, the CBM category of GBPs have applications in bioprocessing methods for affinity purification of biomolecules using them in an immobilized selection step, or as affinity tags [38]. CBMs have also been used in textile industries to increase the efficiency of polysaccharide degrading enzymes – allowing for more efficient dyeing and printing on fabrics [39,40].

We have provided a summary of applications for engineered GBPs, but this list is by no means extensive – recent advances in glycomics techniques are expanding the known glycans that can be targeted for biotechnological, industrial, or medicinal purposes. In particular, development in laboratory techniques like lectin-arrays are leading to faster discovery of glycan biomarkers in human disease [41]. The rapidly growing literature on glycan targets has created a need for the development of novel diagnostic and therapeutic GBPs.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

GBPs can be modified chemically or by using protein engineering techniques – here we will focus on the latter. The applications of GBPs range from diagnostics, therapeutics, and industrial settings to biotechnologies; however, the lack of known GBPs for specific glycans has created a need for GBPs with novel or altered binding specificities. Here we provide information that serves as a starting point for GBP engineering, with a focus on high-throughput directed evolution approaches. We address the topics of scaffold choice (**Section 2.3**), mutagenesis methods (**Section 2.4**), library screening and selection (**Section 2.5**), and characterization (**Section 2.6**), with a focus on how these methods have been used, and have yet to be used, for glycan binding protein engineering.

2.3 Glycan-Binding Protein Scaffolds

Choosing the right protein scaffold is a crucial aspect of engineering a GBP with improved or novel binding properties. In this section we discuss examples of several protein scaffolds available for GBP engineering including: lectins, carbohydrate binding modules (CBMs), pseudoenzymes, carbohydrate-associated enzymes (CAzymes), and antibody-based scaffolds (**Figure 2.1**). The definition of lectins has changed over the years [42,43], but can generally be defined as proteins that bind carbohydrates. Hence, most GBPs can be categorized as a lectin, however for the purpose of this review we have categorized certain GBPs separately from lectins due to their distinct characteristic folds and properties. A summary of the scaffolds discussed in this section, along with example scaffolds that have structural and binding data available, is available in **Table 2.1**.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

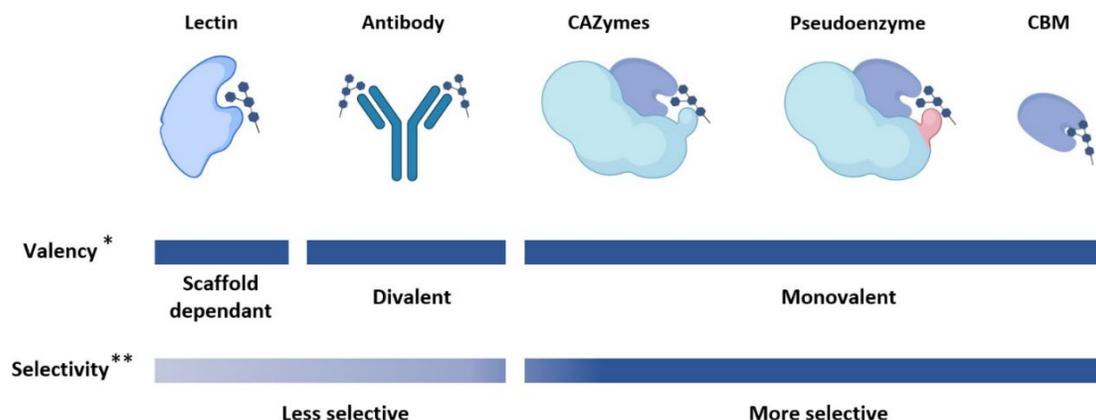


Figure 2.1: Valency and selectivity of protein scaffolds with glycan binding sites. The valency of lectins and antibody-based scaffold varies, as some lectins contain tandem repeat units and antibody-based scaffolds can be designed to contain only a single, or multiple variable fragments. Similarly, CBMs can be designed in tandem to increase valency. The selectivity of antibody-based scaffold is affected by the poor immunogenicity of carbohydrates. *It should be noted that the valency of lectins, antibodies, and CBMs can be altered with protein engineering. **There are exceptions to this trend and the selectivity can be affected by valency.

2.3.1 Lectins

Lectins are carbohydrate binding proteins that are placed into sub-categories based on their folds and function: P-type, I-type, L-type, R-type, C-type, and galectins. Lectins display a wide variety of physiological functions and have biotechnological and biomedical applications – lectins have already been used in the detection and targeted treatments of human diseases such as cancer [44,45]. Here we provide a brief overview of lectins and some examples in GBP engineering. An excellent resource for detailed information on the various sub-categories of lectins can be found in the comprehensive text, *Essentials of Glycobiology* (specifically chapters 28 to 38) [46].

Generally, lectins have relatively low affinities for their glycan targets, with dissociation constants in the micromolar range [47,48]. This may be explained by the shallow binding interface that is observed in most lectins, causing more competitive solvent interactions. The shallow binding interface may also explain the promiscuous binding observed in lectins – glycans with similar structures often bind similar lectins. In nature, the low affinity problem is overcome by oligomerization and multivalency, in biological settings lectins tend to assemble into oligomeric structures containing multiple binding sites, allowing for higher

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

affinities to be reached. The relatively low affinities and promiscuity of lectins in the monomeric state must be considered when selecting scaffolds for GBP engineering; however lectins with improved binding specificity and affinity have been developed [49]. One advantage of using lectins over other protein scaffold is that databases like UniLectin3D are available that can search for lectin scaffolds based on the glycan target [50].

2.3.2 Carbohydrate Binding Modules

Carbohydrate binding modules (CBMs) [51], also known as carbohydrate binding domains (CBDs), are non-catalytic protein domains generally found on carbohydrate-active enzymes (CAZymes). There is low sequence identity between CBMs [51], but there are conserved tertiary folds that are categorized based on their binding site topology as types A, B, or C [52]. The topologies of CBMs are characterized in type A by a planar hydrophobic surface, in type B by an extended binding cavity, and in type C by a short binding pocket – for more information on the structures of CBM types please see the extensive review by Armenta *et. al* [51]. For the purposes of GBP engineering, type A CBMs are suitable for binding insoluble, crystalline carbohydrates, due to the exposed planar binding interface [53]. In contrast, type B CBMs bind oligosaccharides [54], and type C CBMs bind mono and di-saccharides [55]. One attractive aspect of CBMs as GBP scaffolds is their modularity, due to their small size CBMs can be designed in tandem to increase specificity or allow for multiple binding targets. Additionally, there is a variety of well characterized CBMs that can be used as scaffolds – not surprisingly, CBMs have been used to engineer a variety of GBPs with altered binding characteristics [54,56–58].

2.3.3 Pseudoenzymes

In nature, a number of GBPs have evolved from enzymes through the loss of catalytic activity while retaining binding function. These can be defined as pseudoenzymes, which are catalytically inactive proteins related to ancestral enzymes [59]. Pseudoglycosidases are a type of pseudoenzyme that evolved from glycosidases (glycoside hydrolases). These proteins, which bind glycans but cannot hydrolyze glycosidic linkages, can also be characterized as lectins since glycan-binding is their primary function. A

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

few notable examples of pseudoglycosidases that act as GBPs have been observed in nature. In animals, chitinase-like proteins such as the human YKL-39 are pseudoglycosidases (GH18 homologues) with enigmatic biological functions that have been shown to bind to chitooligosaccharides as part of their apparent role in modulating the innate immune response [60,61]. Another example in animals is found in α - and β -klotho proteins, which each make up part of a receptor complex responsive to fibroblast growth factors (FGFs), wherein catalytically inactive GH1-like tandem repeats of the klotho proteins bind to “sugar-mimicking motifs” of FGF19 and FGF21 [62]. In protozoans, the CyRPA protein of *Plasmodium falciparum* – part of the invasion complex that allows the malaria-causing parasite to bind and enter red blood cells – appears to be a catalytically inactive pseudoglycosidase related to GH33 sialidases [63–65]. Pseudoenzymes evolved from other types of enzymes can also bind to glycans. For example, PgaB in *E. coli* is a deacetylase that is involved in the formation of the partially deacetylated poly-1,6-*N*-acetylglucosamine component of the bacterium’s biofilm coat, and the protein consists of two tandem domains related to carbohydrate esterase family 4 (CE4), with the C-terminal domain being a catalytically inactive pseudoesterase involved in binding poly-1,6-*N*-acetylglucosamine [66].

Although pseudoenzymes can, in theory, be used as glycan binding scaffolds, there are no published works on engineering pseudoenzyme scaffolds into novel GBPs as of 2022. This may be due to a lack of known pseudoenzymes scaffolds but may also be due to the prevalence of mutagenesis techniques that allow for inactivation of enzymes. The use of enzymes as GBP scaffolds is discussed in greater detail in the following section.

2.3.4 Carbohydrate-Active Enzymes (CAZymes)

Carbohydrate-active enzymes (CAZymes) catalyze reactions that break down, assemble, or modify saccharides, and they are categorized based on their activities and further subdivided into families based on sequences. Categories include glycosyltransferase, glycoside hydrolase, polysaccharide lyase, carbohydrate esterase, and auxiliary activity families. The examples of pseudoenzymes from nature demonstrate that inactivation of CAZymes can result in proteins that bind to glycans but do not catalytically turn them over.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

Naturally, this suggests that inactivating CAZymes through artificial mutations may be an effective method to engineer novel GBPs. Generating a GBP from a CAZyme requires the inactivation of catalytic residues, which sometimes only requires the mutation of a single amino acid. This may make CAZymes an attractive scaffold for GBP engineering. An example of a nanomolar affinity GBP engineered by inactivation of a CAZyme can be seen in the site-specific mutation of a glycoside hydrolase from *E. coli* K1 bacteriophages. The GH58 endosialidase, Endo-NF, was mutated to generate a catalytically inactive GBP that still binds to polysialic acid with a dissociation constant (K_D) of 191 nM [67,68]. This engineered GBP has been applied as a very sensitive tool for detecting polysialic acid [69,70]. In another example, mutation of a CE2 carbohydrate esterase from *Clostridium thermocellum* has also been shown to produce a catalytically inactive GBP with micromolar affinity [71]. A single amino acid replacement of the *Ct*CE2 enzyme not only abolished esterase activity, but increased the affinity to cellooligosaccharides nearly 8-fold, with the mutant binding to cellohexaose with a K_D of 4.1 μ M.

The strategy of engineering GBPs by inactivating CAZymes has been developed and commercialized most notably by the biotech company Lectenz Bio, who have produced a variety of catalytically inactivated CAZymes, which they have dubbed “Lectenz®” (lectins engineered from enzymes) [72]. The company has produced several Lectenz® through site-directed mutagenesis and computationally guided directed evolution. One advantage of using CAZyme scaffolds is that carbohydrate-processing enzymes tend to be more specific for their ligands than lectins, although this will vary among proteins.

2.3.5 Antibody-based Scaffolds

Antibody-based scaffolds consist of immunoglobulin or immunoglobulin-like protein folds. A variety of antibody-based scaffolds are found in animals, but the most commonly used for developing antigen binding proteins are immunoglobulin G (IgG), and more recently, camelid antibodies [73]. The production of naturally occurring antibodies is time consuming and costly as it requires the immunization of an animal; however, antibody-based scaffolds have been engineered that circumvent the use of animals. These include, but are not limited to, antigen binding fragments (F_{ab}) [74], single chain variable fragments (ScFvs) [75],

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

diabodies [76], monobodies, and nanobodies [77]. There has been a concerted effort to produce antibodies against tumour associated carbohydrate antigens (TACAs) – in total, antibodies have been designed for about 250 distinct glycan targets [78]. Antibody scaffolds offer certain advantages over lectins, including a larger binding interface for longer glycan epitopes, and generally more selective binding due to the complementary determining regions. However, producing an anti-glycan antibody can be costly, labour intensive, and time consuming. Production of an antibody generally requires the immunization of an animal with the glycan, yet glycans often illicit poor immunogenicity [78]. Additionally, anti-glycan antibodies generally have lower affinities (K_D in the micromolar range) than protein-targeting antibodies (K_D in the nanomolar range). Within the last decade, phage display has provided methods for overcoming some of these limitations, resulting in antibodies with higher affinity for their glycan targets [79]. However, this approach still requires an initial scaffold obtained through immunization to be used as the base scaffold for improving the affinity and selectivity.

2.3.6 Summary of Available GBP Scaffolds

Here we discussed the available protein scaffolds and some of their respective challenges and considerations when applied to GBP engineering. The scaffold that is chosen for GBP engineering will influence which mutagenesis techniques and selection methods are most appropriate. This brief overview provides a resource for glycobiochemists who aim to design novel GBPs for specific glycan targets. **Table 2.1** is by no means a complete scaffold list; it serves as a list of example scaffolds that are available and characteristics that need to be taken into consideration. Finding scaffolds ideal for a glycan of choice can be challenging and we recommend using UniLectin3D or equivalent GBP databases as starting point for finding potential scaffolds [50].

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

Table 2.1 Carbohydrate binding proteins (CBPs) with available structural and ligand binding information.

Scaffold Category	Scaffold Sub-Category	Description	Origin	Example Protein (EP)	EP Length	EP Ligand	EP Oligomeric State	EP Multivalency
Lectins	P-type	Lectin that binds to mannose 6-phosphate	Animal	Bovine CD-MPR binding domain [55]	154 aa	Mannose 6-Phosphate	Dimer	Monovalent
	I-type	Protein that is homologous to the immunoglobulin superfamily (IgSF)	Vertebrata	hCD22 domains 1-3 [56]	324 aa	Sialoglycans	Monomer	Monovalent
	L-type	Proteins that are structurally similar to lectins found in the seeds of leguminous plants	All domains of life and viruses	Concanavalin A [57,58]	237 aa	Trimannoside containing-oligosaccharides [58]	Oligomer	Divalent
	R-type	Proteins that are structurally similar to the carbohydrate recognition domain (CRD) in ricin	All domains of life and viruses	Ricin [59]	267 aa	β 1,4 galactose, <i>N</i> -acetylgalactosamine	Dimer	Divalent
	C-type	Ca ²⁺ dependant proteins that share a primary and secondary homology in their CRDs	Animal	C-type domain of murine DCIR2 [60]	129 aa	<i>N</i> -glycans	Monomer	Monovalent
	Galectin	Globular proteins that share primary structural homology in their CRDs	Animal	hGalectin-3 [61]	146 aa	<i>N</i> -acetyllactosamine	Monomer	Monovalent
Carbohydrate Binding Modules (CBMs)	Type A	Protein domain that binds to crystalline surfaces of cellulose and chitin	All domains of life and viruses	CBM from Cel7A [62]	36 aa	Cellulose	Monomer	Monovalent
	Type B	Protein domain that binds endo-glycan chains	All domains of life and viruses	CBM4-2 from xylanase [37]	150 aa	Xylans, β -glucans	Monomer	Monovalent
	Type C	Protein domain that binds exo-type glycan chains	All domains of life and viruses	Cp-CBM 32 of hexosaminidase [63]	150 aa	<i>N</i> -acetyllactosamine	Monomer	Monovalent
Pseudo-enzymes	Pseudo-glycosidase	Carbohydrate binding proteins that evolved from glycosidases but are no longer catalytically active	Possibly all domains of life*	hYKL-39 [40]	365 aa	Chitooligosaccharides	Monomer	Monovalent

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.3: Glycan-Binding Protein Scaffolds

	Pseudo-esterase	Carbohydrate binding proteins that evolved from carbohydrate esterases but are no longer catalytically active	Possibly all domains of life*	C-terminal domain of PgaB [46]	367 aa	Poly-1,6-N-acetylgluco-samine	Monomer	Monovalent
Carbohydrate-Active Enzymes (CAZymes)	Glycoside hydrolase	Enzymes that cleave glycosidic linkages	All domains of life and viruses	Endo-NF (GH58) [48]	811 aa	Polysialic acid	Trimer	Multivalent
	Carbohydrate esterase	Enzymes that hydrolyze ester linkages of acyl groups attached to carbohydrates	All domains of life and viruses	CtCE2 [51]	333 aa	Cellooligosaccharides	Monomer	Monovalent
	Other CAZymes (glycosyl-transferase, polysaccharidase, auxiliary activities)	Enzymes involved in the assembly, break-down, and modification of carbohydrates	All domains of life and viruses	-	-	-	-	-
Antibodies	N/A	Naturally or synthetically produced proteins with an immunoglobulin, or derived from an immunoglobulin-like structure	Vertebrata	hu3S193 [64]	LC: 219 aa HC: 222 aa	Lewis ^Y	Dimeric	Divalent

* Ancestral enzymes found in all domains of life

2.4 Mutagenesis Methods for Library Generation

When a suitable protein scaffold is chosen it can be engineered into a novel glycan binding protein (GBP) using random, rational, or semi-rational mutagenesis. Random mutagenesis can generate large libraries for directed evolution approaches without structural information, whereas semi-rational and rational mutagenesis require structural data. There are various mutagenesis methods that fall into random, semi-rational, and rational mutagenesis, some of which can be used in combination. Here we provide a brief overview of these methods and examples of their uses for engineering GBPs.

2.4.1 Sequence Agnostic Random Mutagenesis

Random mutagenesis techniques that do not require any structural information and allow for mutation at any position within the protein-coding region of a gene can be considered “sequence agnostic”. Sequence (and structure) agnostic mutagenesis approaches are the methods of choice for library generation when no structural data is available. The mutant libraries generated in this way can be used for directed evolution, in combination with high-throughput screening or selection techniques (**Section 2.5**). Methods for random mutagenesis include error prone PCR (epPCR) [80], DNA shuffling [81], and *in vivo* mutagenesis using mutator strains [82], and external mutagens [83] (**Table 2.2**).

A powerful and versatile yet straightforward technique, epPCR is the most common method for creating mutant libraries of a single gene. In epPCR, conditions are chosen to allow for a relatively high mutation rate by the DNA polymerase (*i.e.* low fidelity of replication). This can typically be achieved by adjusting the concentration of DNA polymerase and MgCl₂, adding MnCl₂, and adjusting the ratio of dNTPs, or by using an engineered DNA polymerase mutant with reduced fidelity [80]. As it is the most common mutagenesis technique, it comes as no surprise that epPCR has been applied to engineer novel GBPs. In one example from 2007, Yabe *et al.* cloned an earthworm galactose-binding lectin, EW29Ch, as the starting point for directed evolution wherein variants from successive generations were selected from mutant libraries generated by epPCR. This approach produced an engineered GBP specific for α 2,6-sialic acid, a ligand not recognized by the parent protein [84].

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.4: Mutagenesis Methods for Library Generation

In other examples, epPCR can also be combined with other mutagenesis techniques, such as DNA shuffling. DNA shuffling involves recombination of a population of homologous genes that have diverged either naturally or through laboratory mutagenesis of a parent (*e.g.* by epPCR). In this technique, random fragmentation of genes in a library (*e.g.* by DNase I digestion) is followed by PCR-based reassembly of overlapping fragments with sufficient homology, which effectively recombines mutations within the gene library [85]. Examples of DNA shuffling in combination with epPCR are seen in protein engineering efforts that have introduced mutations to the CBM of cyclodextrin glucanotransferase from *Bacillus* sp., and the glycan-binding regions of *N*-oligosaccharyltransferase from *Campylobacter jejuni* resulting in increased specificity and efficiency of those enzymes [86,87].

Alternative to the *in vitro* mutagenesis methods described above, one can perform *in vivo* mutagenesis on a target gene. These *in vivo* methods involve manipulating the DNA replication and repair machinery of the organism in which the target gene is cloned. For example, in mutator strains like *E. coli* XL1-red, which is deficient in three of the primary DNA repair pathways (carrying mutations *mutS*, *mutD*, and *mutT*), imperfect replication of DNA results in the accumulation of mutations in the cloned gene (along with the vector) [82]. In an example from 2011, Mendonça and Marana used *in vivo* mutagenesis in *E. coli* XL1-Red to alter the specificity of a β -glycosidase, *S* β Gly, from *Spodoptera frugiperda* [88]. Mutants from a library generated in the mutator strain were screened for their specificity towards fucosides vs. glucosides, and several variants were identified that differed from the parent enzyme in their substrate preference. Given that glycosidases can serve as scaffolds for GBPs through their catalytic inactivation (**Section 2.3.4**), this can be a useful strategy towards engineering novel GBPs. The advantage of mutator strains is that their use involves simple protocols, generally involving transformation of the mutator strain by a plasmid followed by propagation and plasmid recovery. However, mutator strains get progressively sick as they divide due to the deficiencies in their DNA repair mechanisms, and consequently this mutagenesis method often requires frequent re-transformations. Other *in vivo* mutagenesis methods use external mutagens such as UV

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
2.4: Mutagenesis Methods for Library Generation

radiation or mutagenic chemicals (*e.g.* ethyl methanesulfonate) which can avoid some of the challenges of maintaining mutator strains.

Regardless of the sequence agnostic random mutagenesis techniques used, one disadvantage is that the produced libraries only cover a small fraction of the possible mutations and require considerable effort to screen. Rational and semi-rational mutagenesis can be more efficient at producing effective mutations based on the structural and functional information when it is available.

Table 2.2 Overview of random mutagenesis methods.

Method	Definition	Pros	Cons
<i>Error prone PCR (epPCR)</i>	epPCR relies on increased error rate of the polymerase	Efficient amplification of mutants	Library size is limited by cloning efficiency
<i>DNA shuffling</i>	DNA shuffling randomly recombines point mutations during PCR	Can be followed up with epPCR	Mutation efficiency is highly dependant on the shuffled library
<i>In vivo mutagenesis</i>	Mutations are introduced in bacteria using chemical or physical means, chemical mutagens or mutators strains	Wider variety of mutations without bias	More labor intensive, mutator strains get progressively sick from mutations

2.4.2 Rational and Semi-Rational Mutagenesis

Rational and semi-rational mutagenesis can produce smaller libraries than sequence agnostic random mutagenesis techniques, while simultaneously focusing on mutations that are more likely to impact protein function in a desirable way. Rational mutagenesis, as defined herein, involves making precise amino acid substitutions to a protein scaffold based on its structural data, whereas semi-rational mutagenesis uses the structural data to target specific sites that are then randomized. Some degree of rationality is always used to limit the sequence space that can be covered in mutagenesis, making it difficult to clearly distinguish between semi-rational and rational design – therefore we have grouped these techniques together. Both techniques make use of various site-directed mutagenesis (SDM) methods (**Table 2.3**).

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.4: Mutagenesis Methods for Library Generation

SDM methods are applied to engineering binding proteins when detailed structural information is known about the binding site, or binding determining regions of the scaffold protein. The most common and efficient way to target specific sites is by PCR – desired single point mutations are included in the primers that amplify the gene of interest. PCR based SDM is one of the most common SDM methods applied in protein engineering and variations of PCR site-directed mutagenesis have been successfully applied in engineering GBPs in several examples including altering the binding specificity of a fucose-binding lectin PA-IIL [89] and increasing the affinity of a tail spike protein Sf6 to the glycans comprising bacterial O-antigens [90]. SDM has also been used to inactivate a streptococcal endo-*N*-acetylglucosaminidase (EndoS) [91] and an *E. coli* K1 bacteriophage endosialidase (EndoNF) to dissociate binding from hydrolytic activity [67,68] and produce novel GBPs for specific glycan detection [69,70].

One drawback of standard SDM approaches is that they only sample a few defined mutations that may or may not result in the desired protein function; hence a sub-category of SDM, site-saturation mutagenesis (SSM) is often used to increase the sample space. Instead of a straightforward replacement of one amino acid for another, SSM randomizes a specific codon, or short sequence of codons, to produce libraries of mutants with all possible amino acid substitutions (or a subset of possible substitutions) at the targeted positions [92]. Screening of such libraries allows for the identification of ideal amino acid replacements for those positions [93]. Although there are various SSM techniques, they all rely on site-directed mutagenesis PCR using degenerate codons. SSM has been successfully used to alter the glycan-binding specificity of a galectin towards $\alpha(2,3)$ -linked sialic acid in a single mutagenesis step [94]. This demonstrates the efficiency of SSM when compared to sequence agnostic random mutagenesis techniques that often require multiple cycles of mutagenesis to produce a desired mutant.

While directed mutagenesis can result in site-specific replacement of codons for targeted amino acid positions, it can also be employed to replace longer sequences of DNA for motif- or domain-swapping mutagenesis (substituting long strings of amino acids). Replacement DNA cassettes can be introduced by restriction digestion and ligation, or by overlap extension PCR to substitute a parental DNA fragment. In

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
2.4: Mutagenesis Methods for Library Generation

one example, motif-swapping has been used to alter the binding specificity of galactose-specific *Bauhinia purpurea* lectin by switching nine amino acids in the binding region with the corresponding nine residues present in the mannose-specific *Lens culinaris* lectin, generating a chimeric lectin that has a unique carbohydrate-binding specificity not observed in the parental proteins [95].

Table 2.3 Overview of site-directed mutagenesis techniques.

Method	Definition	Pros	Cons
PCR site-directed mutagenesis [78]	Primers containing the desired mutation(s) are used to alter the original gene	Not limited by the availability of nearby restriction enzyme cut sites	Primer design can be complicated when introducing multiple mutations
Site-saturation mutagenesis	A set of codons is substituted with every amino acid using degenerate codons	Allows for the screening of ideal amino acids at different positions	Mutation bias from the degenerate codon
Motif- and domain-swapping	Uses a "cassette" DNA fragment containing the mutations, which replaces the unmutated segment in the original gene	High mutation efficiency	Is limited by the domains/motifs that are used.

2.4.3 Computational Tools for Rational and Semi-rational Mutagenesis

In rational and semi-rational approaches, application of the aforementioned mutagenesis methods is guided by various computational approaches to identify positions for directed mutagenesis and predict beneficial mutations. These include homology modeling, molecular dynamics, deep learning, and tools for designing focused libraries (**Table 2.4**). Here we provide a brief overview of these computational methods and their applications in GBP engineering.

The 3D structure of a protein provides valuable information to guide its engineering, allowing one to identify residues that interact (directly or indirectly) with a ligand and to focus mutations at positions where they may be most effective. For proteins without solved structures, a homology model may be used where possible as an imperfect substitute. Homology modelling is one of the most common computational methods used for protein engineering, as it allows for the construction of an atomic resolution model for a target protein of unknown structure, provided that the structure of a sequence-homologous protein has been

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.4: Mutagenesis Methods for Library Generation

solved. The model is based on the structural data of homologous proteins and is generally considered reliable if there is more than 50 % sequence identity between the target and homologue. In certain cases, homology models allow for precise editing of a protein. Using a homology model, Lienenmann *et al.* were able to identify and inactivate the catalytic residue in the *Trichoderma harzianum* chitinase, Chit42 – turning the enzyme into a GBP [96]. However, X-ray crystal structures and homology models only provide a snapshot of the protein – unlike molecular dynamics (MD) simulations.

Molecular dynamics simulations generate a set of possible conformations based on the protein structure. One advantage of MD simulations is that they can identify flexible regions that may alter the protein activity. There are plenty of MD program packages available to use, with some of the most common being YASSA, MOE, Enlighten2, and GROMACS. In recent work on GBP engineering, Kunstmann *et al.* used MD simulations generated with the GROMACS 4.5.5 package to accurately link mutations in the tail-spike protein of bacteriophage Sf6 to varying affinities towards glycans of the O-antigens on *Shigella flexneri* (which is host to the phage) [90]. Yet being able to make such predictions accurately is rare and it is more common to identify beneficial mutations through screening focused libraries.

The insights gathered from analysis of protein structures, homology models, and MD simulations can be used in designing focused libraries of mutants. Focused libraries – in which specific positions are targeted for mutation (*e.g.* by SSM) – can be created with the assistance of computer tools for identifying amino acid sequences that are more likely to be involved in stability, catalytic activity or specificity. These enriched libraries improve the efficiency of directed evolution by reducing the library size and have been used to design proteins with increased selectivity [97,98] and specificity [99]. Multiple tools are available for focused library design, such as CASTER and HotSpot Wizard 2.0, with the latter being a more accessible web-based tool that requires less bioinformatical knowledge.

One method of creating focused libraries that has been advancing rapidly is deep learning. Deep learning algorithms are a form of machine learning that trains artificial neural networks to recognize patterns in complex datasets such as sequencing data to predict properties of novel or uncharacterized proteins.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
2.5: Library Selection and Screening Methods for GBPs

Although deep learning is a fairly new technique, it has already been applied to lectin engineering. In 2012 Stephen *et al.* used a deep learning algorithm called CPred to produce a starch-binding domain with altered selectivity by predicting the effects of circular permutation [100,101].

Here we have only briefly covered some computational methods that are available for protein engineering of GBP – for more details on computational approaches to protein design, Kuhlman and Bradley provide an excellent review [102].

Table 2.4 Example of computational approaches used in protein engineering.

Computational Approach	Definition	Examples	Utility in Protein Engineering
Homology model	Constructs an atomic resolution model of a protein based on available structural data of related homologous proteins.	Phyre2, ROBETTA, SWISS-MODEL	Predicted structures can be used in other computational methods, such as docking and molecular dynamics simulations.
Molecular Dynamics [86]	Predicts the conformational energy landscape available to a protein based on the structure.	YASARA, Enlighten2[87], MOE, GROMACS	Dynamics can indicate how certain mutations can affect the behavior of protein such as folding, stability, ligand interaction and enzymatic activity.
Deep learning	Uses known protein sequences and properties to predict the properties of uncharacterized or novel proteins.	CPred [83], UniRep [88]	Allows for the generation of more efficiency mutant libraries, by focusing on the most promising candidates.

2.5 Library Selection and Screening Methods for GBPs

Library display techniques such as phage, yeast, ribosome, and mRNA display, allow for the high-throughput selection of specific binding proteins from libraries of mutant variants. Regardless of the method used, there are three general steps involved: 1) library generation, 2) biopanning, and 3) characterization of variants. This section will focus on biopanning, since the library generation was previously described (Section 2.4), and characterization is dependant on the target protein (Section 2.6).

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.5: Library Selection and Screening Methods for GBPs

2.5.1 Phage display

Phage display is a high throughput selection technique for large libraries of mutant proteins that links the protein library to the coat proteins of bacteriophages (most commonly M13 filamentous phage, T4, T7, and λ phage). This method displays protein libraries on the surface of bacteriophages by encoding the library into a phage-derived circular DNA vector – a phagemid. The phagemid encodes a coat protein fused to a mutant protein of interest, which links the phenotype of the protein (binding ability, enzymatic activity etc.) to its genotype. Typical elements of phagemids include bacterial and phage origins, a selection marker, the gene of the displayed protein appended to a phage protein-coding gene, which may also encode a periplasmic localization signal depending on the phage protein (**Figure 2.2A**). Displayed proteins can be selected for desired function then genotyped using high-throughput selection methods followed by sequencing. This section provides a brief overview of phage display, an extensive review is published in Chapter 3, volume 580 of *Methods of Enzymology* [103].

A phagemid library encoding a pool of mutants produced by one or more mutagenesis methods (**Section 2.4**) is used to transfect *E. coli* – the transformation efficiency limits the protein library diversity between 10^7 to 10^9 unique protein sequences [104]. Typically, the phagemid does not encode viral proteins that are needed for viral propagation. Hence, a helper phage that expresses viral proteins essential for efficient viral propagation is used to ensure high copy numbers of the phage library. Methods have been developed that eliminate the need for helper phages, but they are not as commonly used [105]. Regardless of whether helper phages are used, the amplification of the phage library is the beginning of the biopanning cycle that involves binding, washing, elution, reinfection and phage amplification steps (**Figure 2.2B**). The phage library can then be used to test for binding to the target of interest and successful binders may be used to repeat the cycle using phage reinfection and amplification.

With regards to glycan binding proteins, phage display has been used to produce a single chain variable fragment (ScFv) against T-antigens – a glycan marker of many adenocarcinomas [106]. The anti T-antigen ScFv was selected from a library of human ScFvs and successfully bound T-antigens with micromolar

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.5: Library Selection and Screening Methods for GBPs

affinities. Likewise CBM4-2 of *Rhodothermus marinus* xylanase Xyn10A (18kDa) has been used as a scaffold in phage display to engineer new binding properties towards xylan, and the glycoprotein IgG4 [57].

An advantage of phage display over the other display methods is that it has been well optimized as it is the oldest and most common display technique. However, phage display does not allow for simple incorporation of many post-translational modifications (*e.g.* protein glycosylation) without chemical manipulation of the phage after amplification [107] – in contrast to yeast display.

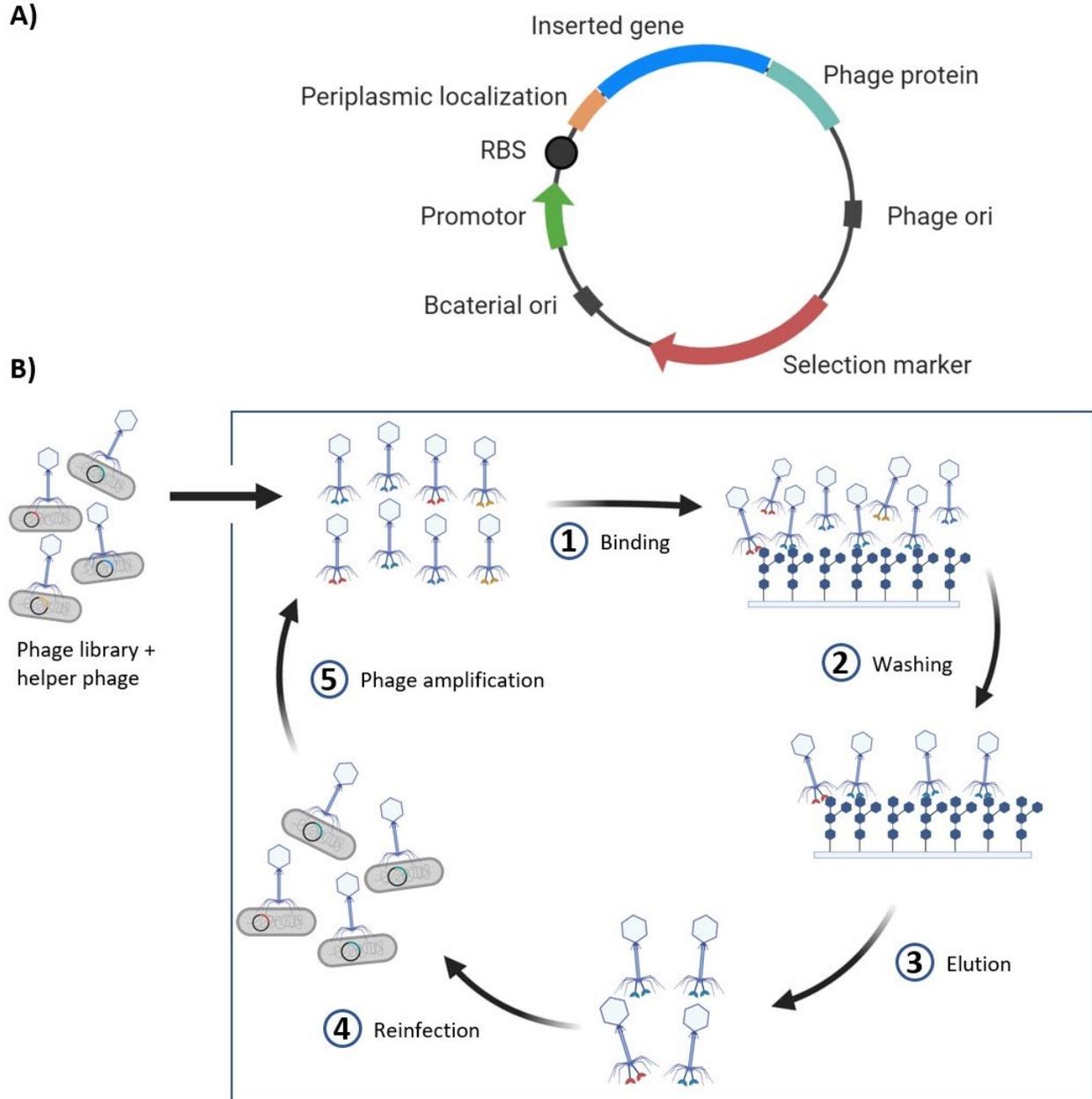


Figure 2.2: A) Example phagemid design for phage display. The phage plasmid requires two origins of replications, one for *E. coli* (the host) and one for the phage. Additionally, attaching a periplasmic localization signal to the target gene may improve the display, although this may depend on the appended phage protein. The phage protein is drawn attached to the C-terminus; however, it may be required to append it by the N-terminus – linkage is dependent on the phage protein. **B) Simplified schematic of phage display.** The protein library is transformed into *E. coli* and amplified using helper phages, resulting in phages displaying the protein library. The phages can then be tested for antigen binding, using surfaces or beads that contain the antigen. A washing step removes unbound phages, and positive binders can be eluted and used for reinfection and amplification of the selected phages. This bio-panning cycle is repeated for multiple rounds and remaining phages are characterized and recombinantly expressed.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.5: Library Selection and Screening Methods for GBPs

2.5.2 Yeast display

In yeast display the proteins of interest are fused to cell surface proteins, most commonly α -agglutinin (Aga1), a-agglutinin (Aga1p-Aga2p), or flocculin (**Figure 2.3**). Depending on the anchor protein, the target protein can be fused to the N- or C-terminus, resulting in display of up to 100,000 copies [108]. The target protein is generally tethered to the anchor protein at the terminus farthest from the functional region, or binding region of the protein, to avoid interference. Additionally, the target protein can be flanked by protein tags for simplified purification and detection methods.

Yeast display is commonly used for antibody-like proteins, such as ScFvs, Fabs, and monobodies. In 2013, Hong *et al.* used yeast display to produce monoclonal lamprey antibodies (lambodies) against several tumor associated carbohydrate antigens, Lewis antigens, and several glycoproteins [109]. More recently, a fully human ScFv was produced that could bind an epithelial tumor associated *N*-glycoform of periostin, further demonstrating the value of yeast display in engineering GBPs [110]. One advantage of yeast display over phage and *in vitro* display methods is the ability to include native (or similar to native) post-translational modifications, which can impact the protein’s fold and function. Additionally, yeast display can take advantage of fluorescence-activated cell sorting (FACS), for efficient library screening (after the cells bind to a fluorescently tagged target ligand). However, yeast display is more time intensive than *in vitro* display methods, and the mutant library size is restricted to 10^7 unique mutants – the smallest out of all display methods discussed in this review.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
2.5: Library Selection and Screening Methods for GBPs

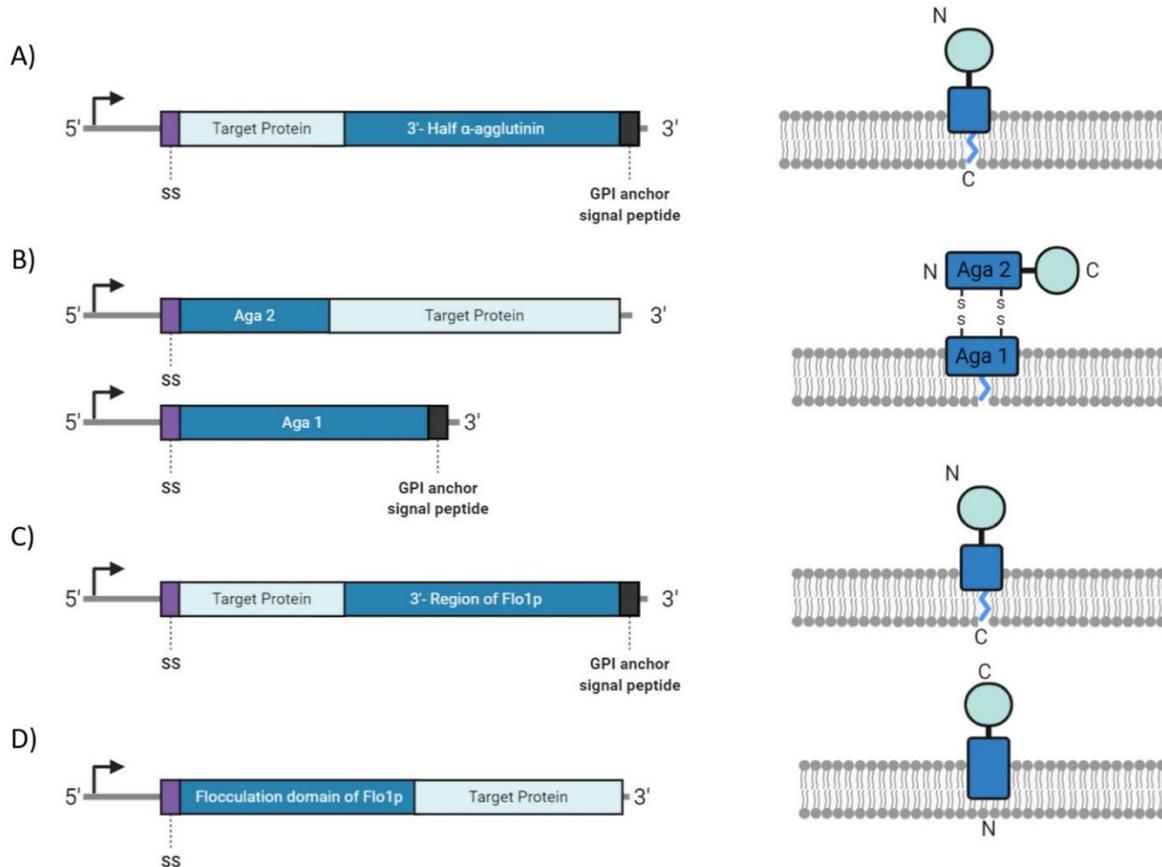


Figure 2.3: Construct designs for yeast display systems. A) α -agglutinin is fused to the C-terminus of the target protein and anchored to the cell membrane by a glycosylphosphatidylinositol (GPI) anchor. **B)** α -agglutinin (Aga) is a dimer, with Aga2 linked to the N-terminus of the target protein. Aga1 is membrane bound by a GPI anchor and forms disulfide bonds with Aga2. **C)** The C-terminal domain of Flo1p is linked to the C-terminal region of the target protein. **D)** The N-terminal domain of the Flo1p is linked to the N-terminal region of the target protein.

2.5.3 Ribosome display

Unlike yeast display and phage display, ribosome display is performed in a cell-free process that involves *in vitro* transcription and translation, and it is not limited by transformation efficiencies. In ribosome display the target protein is linked to its mRNA in a ribosome-mRNA-target protein complex. This complex is established during *in vitro* translation, as the mRNA of the target protein lacks a stop codon – preventing the dissociation of the mRNA and protein from the ribosome. The design of the DNA construct, including the gene coding for the target protein, requires an upstream ribosomal binding site (RBS) and a downstream spacer (**Figure 2.4A**). The DNA construct is transcribed *in vitro* to mRNA, and the spacer sequence on the mRNA allows the protein of interest to sit outside of the ribosome tunnel, and fold properly, while

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
2.5: Library Selection and Screening Methods for GBPs

remaining bound to the ribosome after it is translated *in vitro*. It is recommended that the construct include regions that will be transcribed into stemloops upstream of the RBS and downstream of the spacer – the secondary structure of the mRNA stemloops can prevent mRNA degradation and increase ribosome efficiency up to 15-fold [111]. **Figure 2.4B** provides an overview of the entire mRNA display cycle.

In one example of ribosome display applied to GBP engineering, a sialic acid-binding lectin for use in analytical microarrays has been engineered from a galactose-binding lectin using ribosome display in combination with epPCR [84]. One advantage of cell-free display methods like ribosome display is the ease with which PCR-based mutagenesis methods can be implemented after each selection round – allowing for multiple generations of mutants to be evolved through iterative cycles of randomization and selection, mimicking the natural evolutionary process. However, during ribosome display the protein of interest is bound to an enormous 2.7 MDa ribosome complex, which may interfere with the target proteins characteristics. A cell-free display technique that does not link the target protein to a macromolecule and may have less interference is mRNA display.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
 2.5: Library Selection and Screening Methods for GBPs

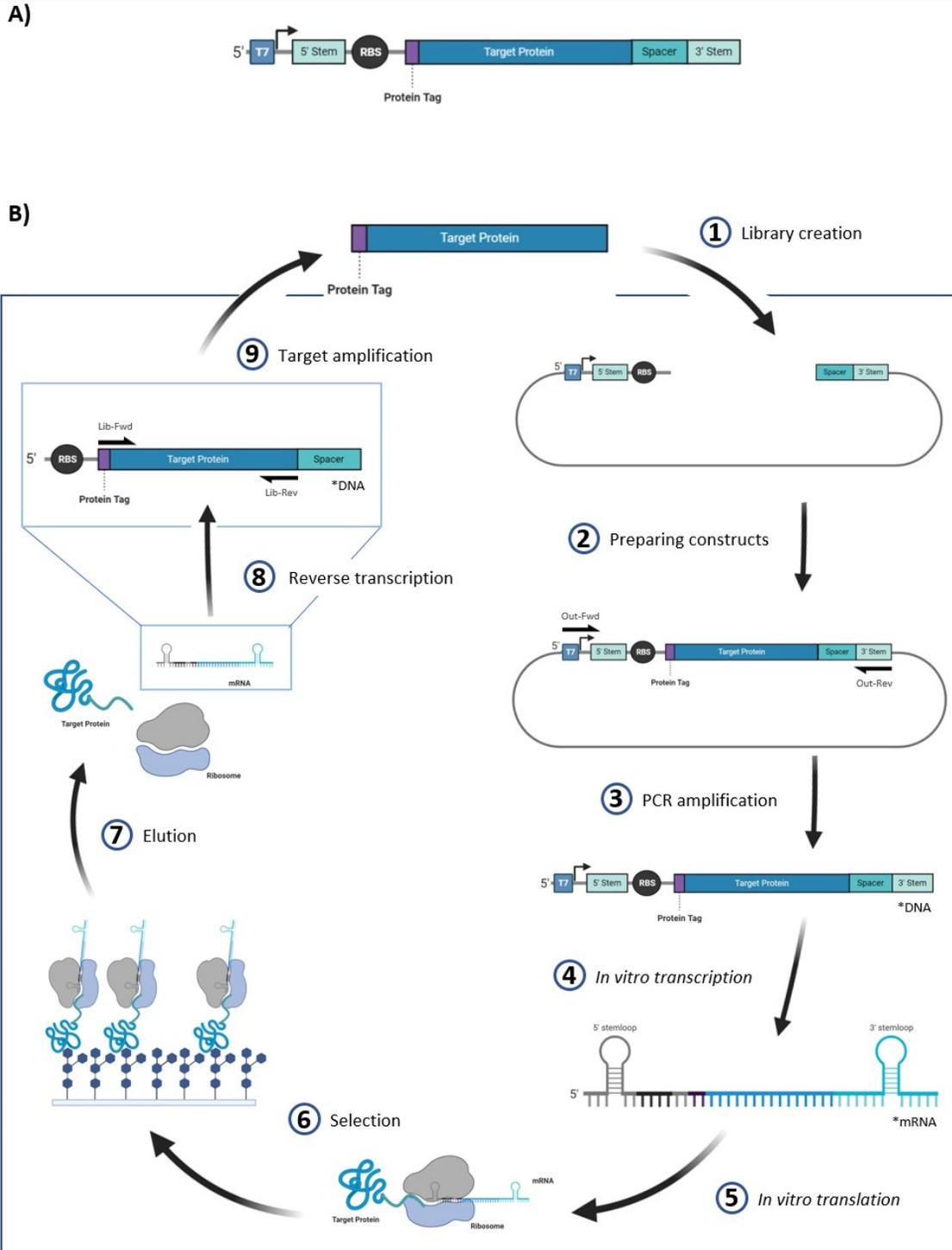


Figure 2.4: Overview of the ribosome display method. A) Construct design for ribosome display. The sequence that is transcribed (4) during the ribosome display contains 5' and 3' stem loop regions that help prevent degradation of the subsequent mRNA. **B)** Cyclic representation of ribosome display. The plasmid used for ribosome display typically includes the promoter, stem loop regions, RBS, and spacer region.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.5: Library Selection and Screening Methods for GBPs

2.5.4 mRNA display

mRNA display is a cell-free display method in which the protein of interest is covalently linked to the 3' end of its own mRNA. DNA constructs for mRNA display require a promoter (commonly the T7 promoter for an *E. coli*-derived system but may vary depending on the cell-free expression system) to recruit RNA polymerase for *in vitro* transcription, and a ribosomal binding site in the 5' UTR, allowing for *in vitro* translation of the protein. The mRNA is transcribed *in vitro*, then enzymatically or photochemically ligated to a DNA-puromycin linker [112,113]. During *in vitro* translation, the puromycin mimics an aminoacyl-tRNA and forms a bond with the nascent peptide. The resulting protein-mRNA complex can then be used in selection methods to identify binding interactions – an overview of the mRNA display process is provided in **Figure 2.5**.

The mRNA-protein linkage allows for simplified sequence detection following selection – the mRNA is reverse transcribed allowing for double-stranded cDNA to then be amplified by PCR. The resulting DNA can then be sent for next generation sequencing (NGS) – the sequencing data can be used to determine which proteins are enriched. As previously mentioned, one advantage of mRNA display over ribosome display is that it does not require the ribosome complex to be bound to the protein; hence interference is less likely. A review on mRNA display has recently been published, and discusses the topic in much greater detail than we will cover here [21]. However, we do note that as of 2020, no reports of mRNA display applied to the engineering novel GBPs have been published. It should also be noted that cell-free expression systems, like mRNA display and ribosome display, do not typically include protein folding chaperones, which may decrease the yield of properly folded proteins. Chaperones such as the *E. coli* proteins DnaK and GroEL can be added to cell free expression systems to increase the yield of functional proteins, however they may not act as chaperones for every protein product and optimization would be required [114].

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins
2.5: Library Selection and Screening Methods for GBPs

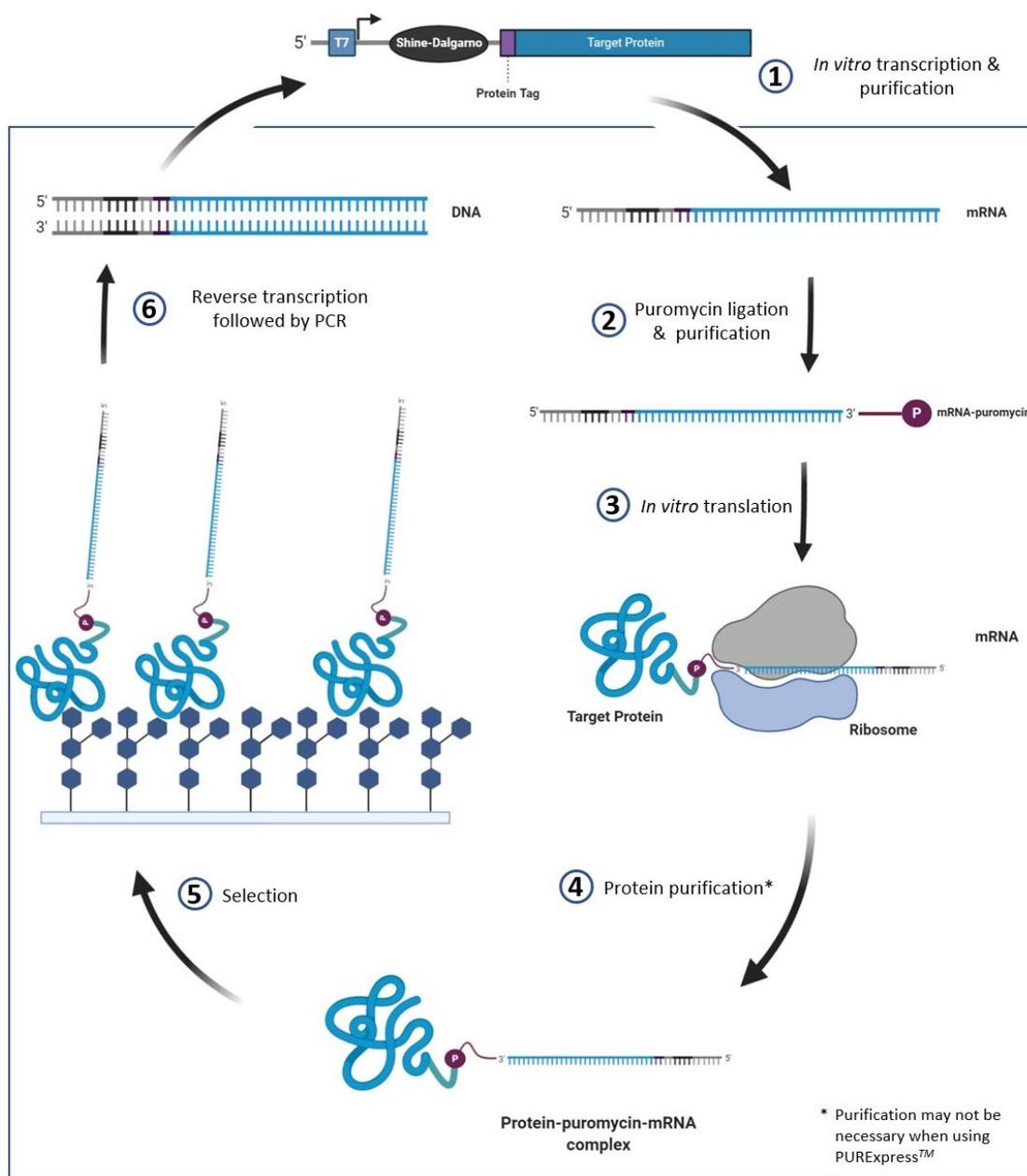


Figure 2.5: Overview of the mRNA display method. The initial construct design has to contain a promoter (e.g. T7) and RBS (e.g. Shine Dalgarno) specific to the cell free expression system used. The puromycin mimics an aminoacyl tRNA which causes the mRNA to be linked to the nascent protein during *in vitro* translation. Once translated, the mRNA-protein complex is screened for glycan binding. The mRNA attached to the bound proteins is reverse transcribed and PCR amplified – allowing for simplified sequence detection.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.5: Library Selection and Screening Methods for GBPs

2.5.5 Glycan Immobilization Strategies for GBP Selection

Common to the range of selection methods to identify binding proteins, which involve different display techniques, is the need for an immobilized target. Various glycan immobilization strategies may be used for the selection and purification of GBPs. Glycan targets can be bound to a variety of solid supports including agarose resin [84], polymer-coated superparamagnetic particles (*e.g.* Dynabeads) [115], or the wells of standard microtiter plates [69,116]. This can be done using a number of chemical approaches such as amine coupling and click-chemistry.

Glycoproteins may serve as a convenient source of target glycans which can be fixed to a solid support. The entire protein along with its glycan modifications may be immobilized by covalent attachment to a chromatographic resin such as agarose. Glycoproteins such as bovine fetuin are readily available materials for cell biology and can be linked by amine coupling to resins that have been functionalized with, for example, aldehyde groups, cyanate esters, or *N*-hydroxysuccinamide esters. Fetuin-agarose produced in this fashion has been applied in the selection cycle for the directed evolution of a novel sialic acid-binding protein in a process involving ribosome display (**Section 2.5.3**) [84].

Coated magnetic particles, such as Dynabeads, can serve as another sort of solid support for target glycans. These versatile beads are used (often with the aid of automated magnetic particle handling robots) in the pull-down of proteins and nucleic acids. The use of glycan-coated Dynabeads has been demonstrated for specific pull-down of GBPs. Hence, Dynabead-linked carbohydrates can be used in a selection process for engineered GBPs. In a study by Liebroth *et al.* [115], Lewis^X and *N*-acetyllactosamine were immobilized to Dynabeads through a bovine serum albumin carrier, and these coated beads were then used to test the binding of TAG-1, Contactin, and NCAM120 (three lectin-like neuronal receptors). In pull-down assays carried out on a mixture of cellular proteins, it was determined that TAG-1 makes specific interactions with Lewis^X (but not *N*-acetyllactosamine), while the other two proteins do not. In a similar fashion, glycan-functionalized Dynabeads could be used to pull down binders to a target glycan from a library of GBPs, removing non-binders as part of a selection process.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.6: Binding Characterization of GBPs

A useful approach to link target glycans to solid support resins or surfaces involves “click” chemistry. Click chemistry involves a set of water-compatible, biocompatible reactions that can link two appropriately tagged reagents together in high yield (*e.g.* using azide-alkyne cycloaddition). Here we will focus on azide-alkyne based click chemistry, since it is a widely adapted tool in glycobiology research, seeing broad uses in applications including *in vivo* glycoengineering [117,118], and glycan labelling [116]. Notably click chemistry has been used to immobilize functionalized glycans onto the surface of wells in microtitre plates. This has enabled high-throughput plate-based assays for glycosyltransferases exhibiting, for example, fucosyltransferase [119] or polysialyltransferase activity [69] on immobilized oligosaccharide substrates, detected through specific binding of the enzyme-product by tagged GBPs. Solid supports coated with immobilized glycans such as these could also be used for selection of GBPs.

2.6 Binding Characterization of GBPs

The most common binding characterization tools used for engineered GBPs are surface plasmon resonance (SPR) and titration calorimetry. SPR and titration calorimetry are both non-destructive methods for determining binding characteristics; however, SPR requires the binding partner to be immobilized on a surface whereas titration calorimetry is done in solution. Specific applications of affinity chromatography using immobilized ligands have also proven useful in measuring GBP binding affinity, although this method is less accurate.

2.6.1 Frontal Affinity Chromatography

Using chromatographic resins functionalized with immobilized ligands, frontal affinity chromatography is an analytical technique that can be used to measure the binding interactions between molecular species. Application of this technique can be used to determine the binding constants of binding proteins of a wide range of equilibrium constants [120]. In particular, it has proven a useful tool in measuring protein-glycan interactions [121]. As a solution containing a GBP flows through a column packed with a glycan-functionalized resin, the degree to which the GBP is slowed by its interaction with the immobilized glycan can be used to measure binding interactions. Custom affinity resins with specific glycans can be generated,

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.7: Surface Plasmon Resonance

for example, as described in this review (Section 4.5). This approach has been demonstrated by Yabe et al. to characterize the binding of an engineered lectin to the sialic acid ligand to which it was tailored [84]. One advantage of affinity chromatography is that it can also be used to select and purify a protein target, yet the binding kinetics determined are not as accurate as SPR or titration calorimetry.

2.7 Surface Plasmon Resonance

Surface plasmon resonance (SPR) techniques measure the frequency of the electromagnetic oscillations on a metal surface, by exciting the surface electrons using a light source. The angle of the reflected light is influenced by the mass at the surface, hence mass changes on the surface can be measured based on the change in the reflected angle [122]. Glycan-labelled metal surfaces can therefore be measured for protein binding based on the change of the reflected angle (**Figure 2.6**). Moreover, SPR can measure binding interactions in real time – allowing for the determination of kinetic parameters, such as binding constants.

SPR has already been applied in glycobiology research to screen GBPs for glycan binding [123], produce a mannose biosensor capable of detecting nM concentrations of lectins [124], and for comparing glycan binding of lectin mutants [96]. The main advantage of SPR over other analytical techniques is the ability to provide real-time kinetic data for glycan-protein interactions, without the need for labelling methods. However, the equipment and specialized knowledge needed to apply SPR methods can prohibit the use of these techniques. One of the main challenges is attaching the glycans or GBPs to the surface. For glycans, protocols have been developed to attach phenoxy-derived sugars [125]. Immobilization of GBPs to the surface can be done in various approaches, including amine coupling, nickel affinity for His-tagged proteins, and streptavidin-binding for biotinylated proteins. Here we do not cover the various SPR techniques in detail – for a more detailed review on SPR and available surface labelling methods please refer to the Handbook of Surface Plasmon Resonance [126,127].

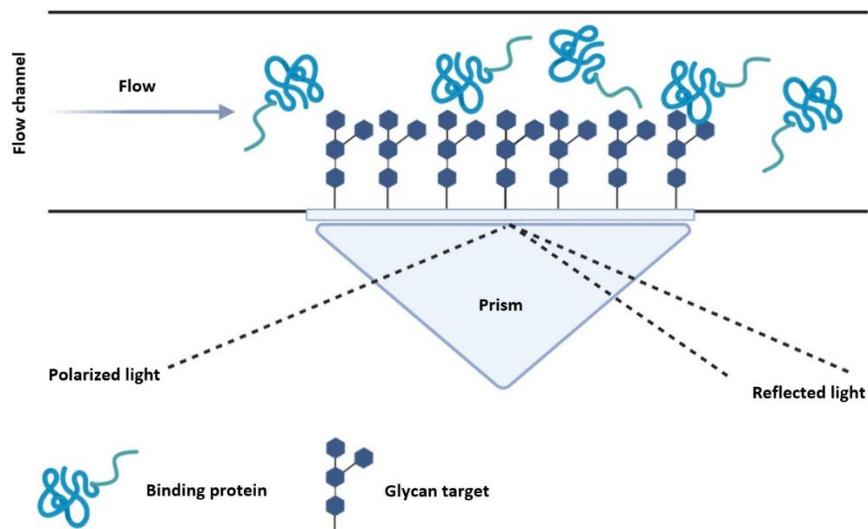


Figure 2.6: Overview of SPR with glycan labelled surfaces. Glycans are bound to a metallic surface inside a flow chamber, a GBP solution is introduced allowing the protein to bind. A light source excites the metallic surface, and the angle of the reflected light is relative to the mass at the surface, which enables binding kinetics to be determined.

2.7.1 Titration Calorimetry

Isothermal titration calorimetry (ITC) can be used to screen GBPs for glycan binding by measuring changes in heat that occur during binding. Based on the heat changes, the enthalpy, entropy, stoichiometry, and binding constants can be determined. ITC has been used to screen the altered binding affinities of a mutated fucose lectin PA-IIL – providing evidence for the location of the specificity binding loop of the protein [89]. Additionally, ITC was used to determine the binding affinities of a mutated starch binding domain (CP90) for longer carbon chain starches [101]. Overall, this technique provides a non-destructive way to determine binding kinetics; however, ITC is not suitable with high-throughput approaches since each mutant requires a separate ITC chamber.

2.8 Current Limitations in Lectin Engineering

Despite the available techniques for GBPs engineering, there are several factors that bottleneck the process, including the availability of glycans, the glycan binding interactions with the scaffolds, and the lack of available glycosylation profiles for cells and proteins.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.8: Current Limitations in Lectin Engineering

2.8.1 Glycan Availability

One limiting factor is the availability of glycans – screening and selection methods to engineer a GBP for a specific glycan require the glycan to be available in high purity and quantity. Due to the large variety and complexity of glycans, only some glycan epitopes are available for purchase through vendors. Unlike oligopeptides and oligonucleotides synthesis, there are no general protocols for the synthesis of complex glycans – the development of chemical synthesis protocol for a single glycan can be time consuming and expensive. However, recent advances in enzymatic and chemoenzymatic synthesis of glycans using recombinant enzymes is providing a path for affordable and efficient production of glycans. Chemoenzymatic synthesis of glycans requires synthetic precursors, a series of glycosyltransferases (GTs) and protection, de-protection steps [128] – this approach has been shown to be more affordable than purely chemical approaches [129]. Additionally, the expanding database of characterized glycosyltransferases opens the possibilities for more glycan structures to be produced enzymatically. Enzymatic synthesis may expand on the available glycan targets available for GBP engineering, yet another bottleneck is in the interactions that a GBP scaffold needs to make with a glycan target.

2.8.2 Glycan Binding Protein Scaffolds

Glycans contains subtle stereochemical and regiochemical differences between isomers that would need to be distinguished by the binding protein. Additionally, glycans form extensive hydrogen bonding networks in water that would have to be broken by the GBP, making the interactions less favorable. In nature, lectins are used to bind complex carbohydrate structures with high specificity, however lectins with higher avidity for their targets are generally multivalent. The multivalency allows for stronger binding with its target but results in larger protein complexes that are more difficult to produce and hence less ideal for biotechnology applications. Unlike multivalent lectins, peptide aptamers and cyclic peptides are smaller and easier to produce, but do not provide large enough interfaces for complex glycan interactions, making them less ideal for glycan targets. Hence, there is a need in GBP engineering for protein scaffolds with large binding interfaces that are adaptable for biotechnological applications.

Chapter 2: Resources and Methods for ENGINEERING “Designer” Glycan-Binding Proteins

2.9: Conclusion

As high throughput protein engineering methods become more common, future efforts may provide protein scaffolds for GBP engineering that bind glycans with higher affinity and avidity than what is currently available. Recently, designed ankyrin repeat proteins (DARPs) have been produced as highly stable, small molecular weight, modular scaffolds [130]. DARPs are based on ankyrin proteins, some of which have glycan binding pockets [131]; however, as of 2020, DARPs have yet to be used for GBP engineering. Due to their modular design that allows for multiple binding interfaces, these proteins may be ideal for glycan targets.

2.9 Conclusion

Engineered GBPs have broad applications in a wide variety of fields, including diagnostic, therapeutics, and biotechnology. GBPs that have been engineered so far have already been used in all the previously mentioned fields and as the field of glycobiology advances we expect more applications to become apparent. Although current methods are available to engineering GBPs, we are still limited by the number of available glycan structures that can be produced in pure quantities, and in the specificity and selectivity of the scaffolds. Additional research is needed to focus on glycan synthesis and on engineering novel scaffolds for glycan binding that provide larger binding interfaces to increase the specificity and selectivity of GBPs.

Chapter 3: Materials and methods

3.1: Recombinant Protein Expression and Purification

CHAPTER 3: MATERIALS AND METHODS

3.1 Recombinant Protein Expression and Purification

3.1.1 General Recombinant Protein Expression and Harvesting Protocol

E. coli strain BL21 (DE3) were used for recombinant protein expression, unless otherwise specified. Chemically competent BL21 (DE3) cells were transformed with the vector of interest, plated on antibiotic LB agar plates, and incubated overnight at 37 °C. A single colony was used to inoculate 50 mL of antibiotic containing terrific broth (TB) to prepare a seed culture, which was incubated at 37 °C overnight at 220 rpm. 10 mL of the seed culture were used to inoculate 800 mL of TB supplemented with antibiotic and grown at 37 °C at 220 rpm. The culture was induced in 0.5 mM IPTG at log phase ($OD_{600nm} = 0.4 - 0.8$) and incubated overnight at 18 °C at 220 rpm.

The cells were harvested by centrifugation at (4,200 x g, 30 min, 4 °C) and the pellet was suspended in 15 mL lysis/wash buffer (50 mM sodium phosphate pH 7.4 – 8.0, 500 mM sodium chloride, 0 – 20 mM imidazole, DNase I, RNase A, and 5 µg/mL of lysozyme. The cells were lysed by sonication for 3 minutes (5 sec pulse on, 15 sec pulse, 25 % amplitude; Fisherbrand™ Model 505 Sonic Dismembrator) and pelleted by centrifugation (15,000 x g for 30 min at 4 °C). The soluble fraction was then filter sterilized (0.22 µm filter).

3.1.2 Affinity Chromatography Protein Purification using Ni-NTA

Ni-NTA resin columns (1 mL) was equilibrated with 10 CVs of wash buffer and filtered (0.22 µm filter) was loaded onto the column. The column was washed with 10 CVs of wash buffer and eluted by gravity with 8 CVs of elution buffer with increasing concentration of imidazole (20 – 500 mM). Fractions containing protein were pooled and concentrated using a spin column (Vivaspin) and stored in storage buffer. Protein concentrations were determined by Bradford's assay or BCA assay. Aliquots were flash frozen and stored at -80 °C.

Chapter 3: Materials and methods

3.1: Recombinant Protein Expression and Purification

3.1.3 Affinity Chromatography Protein Purification using Amylose Resin

Filter sterilized (0.22 μm filter) cell lysate was loaded onto pre-equilibrated amylose resin column (1 mL). The column is washed with 10 CVs of column buffer (20 mM Tris-HCl, pH 7.4, 0.2 mM NaCl, 1 mM EDTA) and eluted with 4 CVs of column buffer plus 10 mM maltose. Fractions containing the protein of interest were pooled and concentrated using a spin column (Vivaspin) and stored in storage buffer. Protein concentrations were determined by Bradford's assay or BCA assay. Aliquots were flash frozen and stored at -80 °C.

3.1.4 Expression and Purification of Enzymes used in the SLe^X-PEG₃-Azide Synthesis

The enzymatic synthesis of SLe^X-PEG₃-Azide used *H. pylori* galactosyltransferase (*HpGalT*; pCW-HP0826), *C. jejuni* sialyltransferase (*Cst-I*; pET28-*CstI*), and *H. pylori* fucosyltransferase (*HpFucT*; pET21-*FucT*). An intermediate step using *Streptococcus pneumoniae* Exo- β -galactosidase from (*BgaA*; pET28-*BgaA*) and *Streptomyces plicatus* exo- β -N-acetylhexosaminidase (*SpHex*; p3AHEX-1.8).

HpGalT, *Cst-I*, *HpFucT*, *BgaA*, and *SpHex* were expressed as in the general protocol in **3.1.1** with minor modifications. *HpGalT* was expressed in *E. coli* AD202 cells at 25 °C and the protein was not purified, crude cell lysate was used for *HpGalT*. No modifications to the expression protocol (**3.1.1**) were made for *Cst-I* and *HpFucT* and they were purified as in section **3.1.2**. *Cst-I* was prepared by H. Wu. Plasmid for *BgaA* and *SpHex* were expressed and purified by M. Soroko as per sections **3.1.1** and **3.1.2**, respectively.

3.1.5 Expression and Purification of Ankyrin, DARPin, and MtDARPin

The ankyrin (pET28a-ankyrin), DARPin (pET28a-DARPin), and MtDARPin (pET28a-MtDARPin) were expressed as per section **3.1.1** and purified using a Ni-NTA column as per section **3.1.2**.

3.1.6 Expression and Purification of Enzymes used in the dTDP-Qui3N Synthesis Reaction

WlaRB (CJV-28) and WlaRG (CJV-24) plasmids were obtained from the Michel Gilbert lab (National Research Council Canada, Ottawa). The proteins were expressed in *E. coli* AD202 and induced with 1 mM

Chapter 3: Materials and methods

3.2: Enzymatic Synthesis and Purification of SLe^X-PEG₃-Azide

IPTG grown overnight at 25 °C, otherwise the expression protocol in section 3.1.1 was followed. WlaRB and WlaRG were purified as in section 3.1.3.

3.2 Enzymatic Synthesis and Purification of SLe^X-PEG₃-Azide

The enzymatic synthesis protocol for SLe^X-PEG₃-azide was adapted from Haoyu [13]. A one-pot, 3-step reaction was set-up using *HpGalT* (crude lysate), Cst-I, and *HpFucT* to transfer galactose, sialic acid, and fucose, respectively, to the GlcNAc-PEG₃-Azide substrate. Freshly prepared reaction buffer contains 5 mM of GlcNAc-PEG₃-Azide (Sigma-Aldrich, SMB00394-25MG), UDP-Gal, CMP-Sia, MgCl₂, MnCl₂, BME, and 150 U/mL alkaline phosphatase dissolved in 50 mM Tris-HCl (pH 7.5). *HpGalT* and Cst-I were added to final concentrations of 0.3 mg/mL and 0.2 mg/mL, respectively, and the reaction was incubated at 37 °C overnight. The following day, *SpHex* and BgaA were added to a final concentration 1.0 µg/mL and 2.7 µg/mL, respectively, and incubated at 37 °C for 30 minutes. Next, GDP-Fuc and *HpFucT* were added to final concentrations of 5 mM and 0.25 mg/mL and incubated at 37 °C for 4 hours. The reaction product was analyzed by LC/MS using an Agilent Technologies (Santa Clara, CA) 6546 LC/Q-TOF.

A hydrophobic Bond Elut C18 SPE column (bed wt. 100 mg; Agilent Technologies) was primed with 1 mL of methanol and equilibrated with 2 mL of water. The SLe^X-PEG₃-Azide reaction was then added to the column and eluted three times with 400 µL of water. The collected flowthrough was combined and further purified on an ENVI-Carb™ column (bed wt. 100 mg; Sigma Aldrich). 15 mL of acetonitrile (ACN) followed by 5 mL of 50 % acetonitrile were used to prime the column. Then, the column was equilibrated using 15 mL of water, after which the C18 semi-purified SLe^X-PEG₃-Azide solution was added. For the elution, 1.5 mL solutions of 40 %, 60 %, and 80 % ACN to ammonium formate (50 mM; pH 4.4) were added sequentially and the flow through was collected. The resulting fractions were analyzed by MS/LC-MS (negative ion mode) and fractions containing SLe^X-PEG₃-Azide, corresponding to the 60 % ANC elutions, were freeze dried.

Chapter 3: Materials and methods

3.3: Immobilization and Detection of SLe^X-PEG₃-Azide

3.3 Immobilization and Detection of SLe^X-PEG₃-Azide

3.3.1 Click Chemistry immobilization of SLe^X-PEG₃-Azide on Magnetic Beads

20 μL of 10 $\mu\text{g}/\mu\text{L}$ DBCO-magnetic nanoparticles (DBCO-MNPs) were pulled down on a magnetic rack and the supernatant was removed. 20 μL of 2 mM, 1 mM, 0.5 mM, or 0 mM SLe^X-PEG₃-Azide dissolved in ddH₂O were added to the beads and the solution was incubated for 1 hour at room temperature while rotating. Next, the labelled beads were pulled down on a magnetic rack, washed three times with 200 μL of ddH₂O, and resuspended in 200 μL ddH₂O, resulting in a 1 $\mu\text{g}/\mu\text{L}$ solution of labelled beads.

3.3.2 Fluorescent assay for immobilized SLe^X-PEG₃-Azide Detection

200 μL of 20 $\mu\text{g}/\text{mL}$ of *Aleuria aurantia* Lectin (AAL) lectin, dissolved in 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, was added to 10 μL of labelled DBCO-MNPs (1 $\mu\text{g}/\mu\text{L}$) and incubated for 30 min at room temperature. The DBCO-MNPs are pulled down on a magnetic rack, the supernatant was removed, and were washed three times in 200 μL of PBS, 0.05 % Tween 20TM, and 0.1 % BSA. The DBCO-MNPs were then incubated with 200 μL of 1x streptavidin-HRP dissolved in 10 mM sodium phosphate, 150 mM NaCl, pH 7.4 and incubated for 30 minutes on a rotor at room temperature. The beads were then pulled down on a magnetic rack and washed three times in 200 μL of PBS, 0.05 % Tween 20TM, 0.1 % BSA. A peroxidase substrate mix (20 μM AmplexTM Red, 98 μM H₂O₂, in PBS + 0.05% TweenTM 20) was then incubated with the at a mix 1:1 ratio with the DBCO-MNPs and the fluorescence signal was read (λ_{ex} : 572 nm, λ_{em} : 586 nm) on a CLARIOstar (BMG Labtech).

3.4 DARPin and Affibody Library Design

3.4.1 DAPRin library design

The DARPin scaffold design is based on previous research by Grütter et al [132], and the location of site saturation is based on previous work from the Plückthun lab [133]. The library was ordered as 5 site saturated oligos that translate for the N-term, I1, I2, I3, and C-term segments of the DARPin library. The DARPin library was amplified in two steps, firstly, the site saturated segments were amplified with a PCR mix as indicated in **Table 3.1**. and following the thermocycler settings in **Table 3.3**. See **STable 6.1** for the

Chapter 3: Materials and methods
3.4: DARPin and Affibody Library Design

primers used to amplify the N-term, I1, I2, I3, and N-term segments. After amplifying the site saturated oligos, the PCR is used as a template in a subsequent PCR with primers that contain type IIS restriction cut sites that ensure complementarity once digested. See **Table 6.2** for the primers used to add the restriction enzyme cut sites.

Table 3.1: PCR mix for amplifying site saturated DARPin oligos.

Reagent	Concentration	Volume (μL)
Phusion Buffer	5x	4
dNTPs	25 mM	0.4
Fwd Primer	100 μM	2
Rev Primer	100 μM	2
Phusion DNA polymerase	2 Units/ μL	0.2
MQ Water	-	11.4

Table 3.2: PCR mix for adding Type IIS cut sites to the site saturated DARPin oligos.

Reagent	Concentration	Volume (μL)
Phusion Buffer	5x	10
dNTPs	25 mM	1
Template (1 st PCR)	-	0.5
Fwd Primer	100 μM	0.25
Rev Primer	100 μM	0.25
Phusion DNA polymerase	2 Units/ μL	0.5
MQ Water	-	37.5

Table 3.3: Thermocycler protocols for the DARPin library amplification.

Cycles	Step	Temperature ($^{\circ}\text{C}$)	Time (sec)
1	Initial Denaturation	98	30
5	Denaturation	98	40
	Annealing	50-58*	40

Chapter 3: Materials and methods
3.4: DARPIn and Affibody Library Design

	Extension	72	60
1	Hold	4	Infinite

*The annealing temperature is dependent on the oligos

3.4.2 DARPIn Library Assembly by Type IIS Digestion and T7 ligation

The DARPIn oligos created from the PCR amplifications in section **3.4.1** were assembled using type IIS restriction enzymes and T7 ligase. Annealing of segments was done in pieces, following the reaction set in **Table 3.4** and the thermocycler protocol in **Table 3.5**.

Table 3.4: Reaction mix used in the digestion and ligation of the DARPIn library.

Reagent	Concentration	Volume (μL)
MQ Water	-	2
Fast Digest buffer	10x	1
ATP	10 μM	1
DARPIn_N	450 fmol/ μL	1
DARPIn_I1	450 fmol/ μL	1
BpiI	10 U/ μL	2
T7 Ligase	3,000 U/ μL	2

Table 3.5: Thermocycler protocols for the DARPIn library digestion and ligation.

Cycles	Step	Temperature ($^{\circ}\text{C}$)	Time (min)
5	Digestion	37	5
	Ligation	16	5
1	Final Digestion	37	15
1	Enzyme inactivation	85	15
1	Hold	4	infinite

Chapter 3: Materials and methods
3.4: DARPin and Affibody Library Design

3.4.3 Affibody Library design

The affibody library design was provided by Christopher Hipolito (University of Tsukuba). Briefly, the library is assembled in a 2-step overlap extension PCR. First, the primers Affi_NNK13_F129 and Affi_R105 (S**Table 6.3**) amplify the affibody library containing 13 NNK sites (**Table 3.6**). Secondly, the site saturated affibody oligo is amplified with the primers Affi_T7_F73 and CGS3_R36 are used to add the T7 promoter region and a complimentary site for puromycin linkage, respectively (**Table 3.8**). The thermocycler conditions for the first and second PCR were set as in **Table 3.7**.

Table 3.6: PCR mix for of the first affibody library amplification.

Reagent	Concentration	Volume (μ L)	Final Concentration
Phusion Buffer	5x	5	1x
dNTPs	25 mM	0.2	200 μ M
Fwd Primer: Affi-NNK13-F129	100 μ M	2.5	250 pmol
Rev Primer: Affi-R105	100 μ M	2.5	250 pmol
Phusion DNA polymerase	2,000 Units/mL	0.25	0.5 units
MQ Water	-	14.55	-

Table 3.7: Thermocycler protocols for the affibody library amplification.

Cycles	Step	Temperature ($^{\circ}$ C)	Time (sec)
1	Initial Denaturation	98	30
5/4*	Denaturation	98	40
	Annealing	55	40
	Extension	72	60
1	Hold	4	Infinite

*4 cycles were used for the second PCR

Table 3.8: PCR mix for of the second Affibody library amplification.

Reagent	Concentration	Volume (μL)	Final Concentration
Phusion Buffer	5x	200	1x
dNTPs	25 mM	8	200 μM
Fwd Primer: Affi-T7-F73	100 μM	5	1250 pmol
Rev Primer: CGS3-R36	100 μM	5	1250 pmol
Phusion DNA polymerase	2,000 Units/mL	5	0.5 units
First-step extension mix	-	10	100 pmol
MQ Water	-	767	-

3.5 mRNA Display Protocols

3.5.1 *In vitro* Transcription of the Affibody Library

The affibody library synthesized as in section 3.4.3 was transcribed and linked to puromycin in preparation for mRNA display. We used the NEB HiScribe kit (E2040S) for the transcription, using 1 μg of library DNA, 10 μL of NTP Buffer Mix, 2 μL of T7 RNA polymerase Mix, and up to a final concentration of 30 μL with water. The reaction was incubated overnight at 37 $^{\circ}\text{C}$ in a thermocycler to prevent evaporation. After the incubation, 20 μL of MQ water and 2 μL of DNase I (2000 U/mL) were added, and the reaction was incubated for 15 minutes at 37 $^{\circ}\text{C}$ to degrade the template DNA. Then, 52 mL 0.6 M NaCl, 10 mM EDTA were added, followed by 0.8 volumes (83.4 μL) of -20 $^{\circ}\text{C}$ isopropanol. The mixture was incubated at -20 $^{\circ}\text{C}$ for 5 min to allow for mRNA precipitation, followed by centrifugation at 4 $^{\circ}\text{C}$, 12,000 xg, for 15 min. We washed the pellet with 100 μL of 70 % EtOH (kept at -20 $^{\circ}\text{C}$) and centrifuged the mixture at 4 $^{\circ}\text{C}$, 15, 200 xg, for 3 min. The supernatant was discarded, and the pellet was dried before resuspending in 10 μL of MQ water.

3.5.2 Puromycin Linkage of the Affibody Library

The purified mRNA library was then linked to a puromycin oligo using T4 RNA ligase as in **Table 3.9**. The reaction was stopped using 1 volume (20 μL) of 0.6 NaCl, 10 mM EDTA. We extracted the puromycin-mRNA using 1 volume (40 μL) of phenol:chloroform:isoamyl (25:24:1) alcohol followed by a second

Chapter 3: Materials and methods
3.5: mRNA Display Protocols

extraction using 1 volume (40 μ L) of chloroform:isoamyl alcohol (24:1). To aid with mRNA precipitation, we added 0.4 μ L of UltraPure™ glycogen (20 μ g/ μ L) and 80.8 μ L of EtOH (stored in -20 °C) to the extracted aqueous layer and incubated the solution at -20 °C for 5 minutes. Then the solution was centrifuged at 4 °C, 15200 xg for 15 minutes. The supernatant was discarded, and the pellet was washed with 20 μ L of 70 % EtOH, followed by centrifugation at 4 °C, 15200 xg for 15 minutes. The supernatant was removed and the pellet was air dried and redissolved in 3 μ L of water. To check the sample for linkage, 1 μ L of the mRNA-puromycin library and 10 μ M of the starting mRNA library were run on an 8 % Urea-PAGE, at 25 mA per plate. For visualization, the gel was stained with SYBR safe (add 5 μ L of SYBR safe to 50 mL of 1x TBE) for one hour.

Table 3.9: Reaction mix for the puromycin linkage to the mRNA affibody library.

Reagents	Concentration	1x	Final Concentration
Milli Q water	-	2.45	-
T4 RNA ligase buffer	10x	2	1x
10 mM ATP	10 mM	2	1 mM
DMSO	-	2	10%
Puromycin linker	7.5 μ M	6	45 pmol
T4 RNA ligase1	200 U/ μ L	0.3	60 units
PNK	10 U/ μ L	2.25	22.5 units
Affibody library mRNA	10 μ M	3	30 pmol
Total		20	

Chapter 3: Materials and methods

3.8: Molecular Dynamics

3.8 Molecular Dynamics

3.8.1 Equilibration of Proteins and dTDP-Qui3N

Computational jobs were run on computer clusters through Calcul Québec and the Digital Research Alliance of Canada [136,137]. All molecular dynamics (MD) simulations were prepared using VMD and NAMD [23,138]. The ankyrin domain structure used in the simulations is composed of residues 269-397 of *P. alcalifaciens* N-formyltransferase (PDB: 4XCZ).[19] The DARPin structure (PDB: 2XEE) was unaltered and the MtDARPin structure was designed from the DARPin structure as in in section 3.7. The ligand structure of dTDP-Qui3N was taken from the protein data bank (PDB: T3Q).

All proteins were solvated and ionized using VMD, with a box padding of 25 and neutralizing NaCl concentration of 0.15 mol/L. The structure of dTDP-Qui3N required parametrization prior to solvation and ionization, which was done using the CHARMM-GUI ligand reader and modeler [139,140]. The solvated and ionized structures of the proteins and dTDP-Qui3N were then simulated for 100 ns at 300 K using NAMD to equilibrate the systems. The protein simulations used the CHARMM36 force-field [141,142] and the dTDP-Qui3N used the CHARMM36 carbohydrate force-field [24]. Additional details can be found in the NAMD run scripts on GitHub [143].

3.8.2 Steered Molecular Dynamics Probing Simulation

Steered molecular dynamics (SMD) probing simulations were set up with the equilibrated structures from section 3.8.1. Using VMD, the ankyrin, DARPin, or MtDARPin were aligned with the z-axis and dTDP-Qui3N was placed around the proteins in a cylindrical pattern (**Figure 4.12**). The distance at which the ligand was placed from the protein was determined by the radial maximum of residues in the protein and the endpoint was determined by the radial minimum. The number of simulations to be carried out depends on the radius at which the ligand has “full coverage” of the protein.

For example, in the centered ankyrin structure the radial minimum was 5.7 Å and the height was 54.4 Å. In order to oversample the protein, we set the radius at which the ligand has full coverage to twice the minimum radius – 11.4 Å. Then we placed the ligand around the protein in a pattern of a cylinder mosaic

Chapter 3: Materials and methods

3.8: Molecular Dynamics

with a radius of 11.4 Å and a height of 54.4 Å. Given that the radius of gyration of dTDP-Qui3N 4.98 Å, each molecule can be considered to take up a surface area of 49 Å². If we divide the surface area of the cylinder, 3896.6 Å² by the surface area of the dTDP-Qui3N mosaic piece, we find the number of ligands that saturate the protein at a radius of 11.4 Å is 79, which is equal to the number of probing simulations that are needed. Since the circumference of the cylinder is (72 Å) is 1.32 times the height (54.3 Å), the number of boxes along the circumference ($1.32 \times n = 79$) is 10 and the number of boxes around the height is 8. Therefore, the mosaic cylinder was divided into 10 segments along the circumference (18°, 54°, 90°, 126°, 162°, 198°, 234°, 270°, 306°, 342°) and 8 segments along the height (-23.80, -17.00, -10.20, -3.40, 3.40, 10.20, 17.00, 23.80). For more details on the setup of SMD probing simulations see the GitHub repository [143].

3.8.3 Data Analysis for Steered Molecular Dynamics Probing Simulations

The SMD probing simulations for ankyrin, DARPin, and MtDARPin with dTDP-Qui3N were analyzed using VMD and MDAnalysis [138,144,145]. Hydrogen bonds between the proteins and dTDP-Qui3N were analyzed using the VMD HBonds plugin, whereas ionic and aromatic interactions were analyzed using custom scripts made for MDAnalysis. The distance cutoff at which bonds are recorded for hydrogen bonds, ionic interactions, and aromatic interactions, are 3.0 Å, 3.2 Å, and 4.5 Å. For more details on the setup of SMD probing simulations see the GitHub repository [143].

Chapter 4: RESULTS

4.1: Enzymatic Synthesis of SLe^X-PEG₃-azide

CHAPTER 4: RESULTS

In this project we aimed to create new methods and tools for the development of glycan binding proteins (GBPs) using a combination of directed evolution and computational approaches. For the directed evolution approach, we aimed to create GBPs for the tumour associated carbohydrate antigen (TACA) SLe^X. To this end, we refined an enzymatic synthesis protocol for SLe^X-PEG₃-azide that can be immobilized using click chemistry (**Section 4.1**). Additionally, we created an immobilization protocol and a lectin assay to detect SLe^X immobilized on magnetic nanoparticles (MNPs) (**Section 4.2**). Two site-saturated protein libraries, based on designed ankyrin repeat proteins (DARPin) and affibodies (**Sections 4.3.1, 4.3.2**), were designed to be screened for SLe^X binding using mRNA display. However, only the affibody library could be assembled and is ready for screening by mRNA display (**Section 4.3.3**).

To complement the directed evolution approaches, we aimed to create a computational approach to rationally improve the design of selected GBPs. For this aim, we developed a steered molecular dynamics (SMD) simulation approach that identifies glycan-protein interactions. To test this approach, we screened for glycan interactions between the nucleoside sugar dTDP-Qui3N and three proteins: ankyrin, DARPin, and MtDARPin. The ankyrin protein has a known dTDP-Qui3N binding site, whereas DARPin is a negative control. MtDARPin was designed to recreate the dTDP-Qui3N binding site in ankyrin on the more stable DARPin scaffold (**Section 4.4**). Additionally, we aimed to enzymatically synthesize dTDP-Qui3N to enable future verification of the simulation binding data; however, pure dTDP-Qui3N could not be prepared (**Section 4.5**). In contrast, the SMD approach was successful, as it correctly identified the dTDP-Qui3N binding site in the ankyrin protein based on hydrogen bond interactions (**Section 4.6**).

4.1 Enzymatic Synthesis of SLe^X-PEG₃-azide

Increased production of SLe^X on cell surfaces is associated with various types of cancers [10–12], which makes it a target for developing novel diagnostic tools and therapies, as we hope to do by evolving specific binding proteins by mRNA display. Despite SLe^X being available for purchase (Sigma Aldrich), it is prohibitively expensive, and does not contain any functional groups that would allow for immobilization

Chapter 4: RESULTS

4.1: Enzymatic Synthesis of SLe^X-PEG₃-azide

of the ligand for mRNA display. Hence, *in vitro* methods have been developed to produce SLe^X-PEG₃-azide enzymatically [13] – allowing for the immobilization of the antigen using an azide-alkyne click chemistry reaction. Here, we improve on the enzymatic synthesis protocol of SLe^X-PEG₃-azide by selectively degrading intermediate products.

The enzymatic synthesis pathway of SLe^X-PEG₃-Azide used the bacterial glycosyltransferases *HpGalT*, *Cst-I*, and *HpFucT* to transfer galactose, sialic acid, and fucose to GlcNAc-PEG₃-Azide from their nucleotide-activated donors in a three-step reaction (**Figure 4.1**). Prior to the fucosylation step the reaction mixture was incubated with *SpHex* (*N*-acetylhexosaminidase) and *BgaA* (galactosidase) to remove unsialylated products that can be fucosylated by *HpFucT*. The digestion of unreacted intermediates also simplifies the purification steps, as glycosides can be difficult to separate from one another. For further details on the recombinant protein expression, and enzymatic synthesis of SLe^X-PEG₃-Azide, see sections **3.1.4** and **3.2**, respectively.

After each enzyme incubation reaction samples were taken for MS/LC-MS (negative mode) analysis to monitor the reaction progress. Following the incubation of the reaction mixture with *HpGalT* and *Cst-I*, the expected *m/z* ratio of 3'-SLN-PEG₃-Azide [-H] is present at *m/z* 830.32 (**Figure 4.2A**). However, unreacted precursors are corresponding to LacNAc-PEG₃-Azide ([M - H]⁻; *m/z* = 539.22) and LacNAc-PEG₃-Azide ([M + HCOO]⁻; *m/z* = 585.23) are present, which would also be fucosylated once *HpFucT* is added. Following the addition of *SpHex* and *BgaA* there was a reduction in the pre-cursors (**Figure 4.2B**). The reaction was then incubated with *HpFucT* and purified as in section **3.2** using a Bond EluC18 SPE and an ENVI-CarbTM column. The purified product was analyzed by MS (negative ion mode) and the major peak of *m/z* 976.37 corresponds to the expected *m/z* of SLe^X-PEG₃-Azide ([M - H]⁻) (**Figure 4.3**). From a 1 mL reaction we obtained 3.0 mg of 3'-SLe^X-PEG₃-Azide, a 61 % yield, based on the theoretical yield of a complete reaction. In order to use the purified SLe^X-PEG₃-azide in display methods, an immobilization approach is needed, such as click chemistry immobilization on MNPs.

Chapter 4: RESULTS

4.1: Enzymatic Synthesis of SLe^x-PEG₃-azide

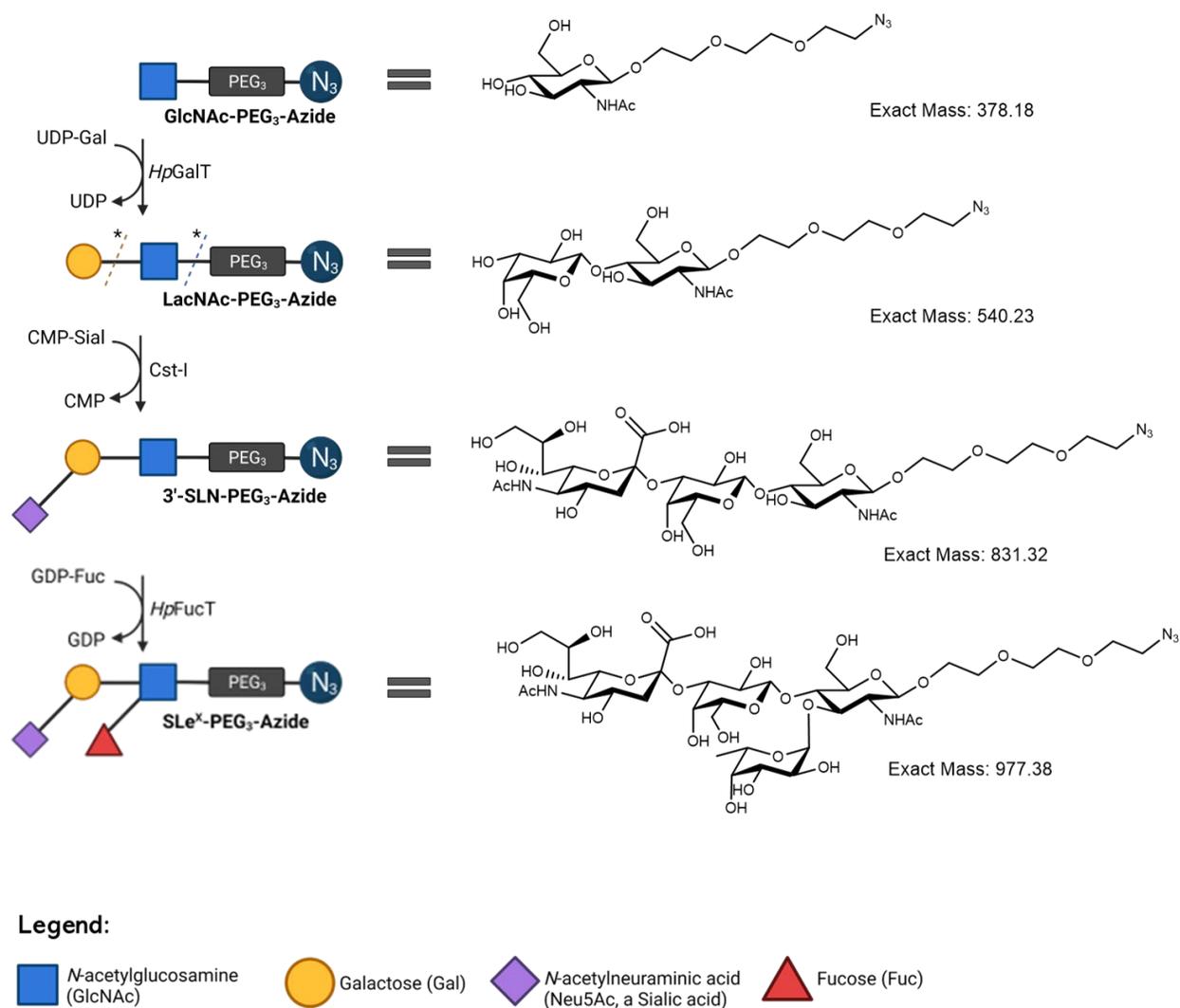


Figure 4.1: Enzymatic synthesis pathway of click-chemistry enabled SLe^x-PEG₃-Azide. The bacterial glycosyltransferases *HpGalT*, *Cst-I*, and *HpFucT*, are used to transfer galactose, sialic acid, and fucose, respectively. UDP-Gal, CMP-Sia, and GDP-Fuc, were used as sugar donors. Following the addition of sialic acid, the glycosidases *SpHex* (*N*-acetylhexosaminidase) and *BgaA* (galactosidase) were used to remove unsialylated products before adding fucose, since *HpFucT* can also act on the previous substrates.

Chapter 4: RESULTS

4.1: Enzymatic Synthesis of SLe^X-PEG₃-azide

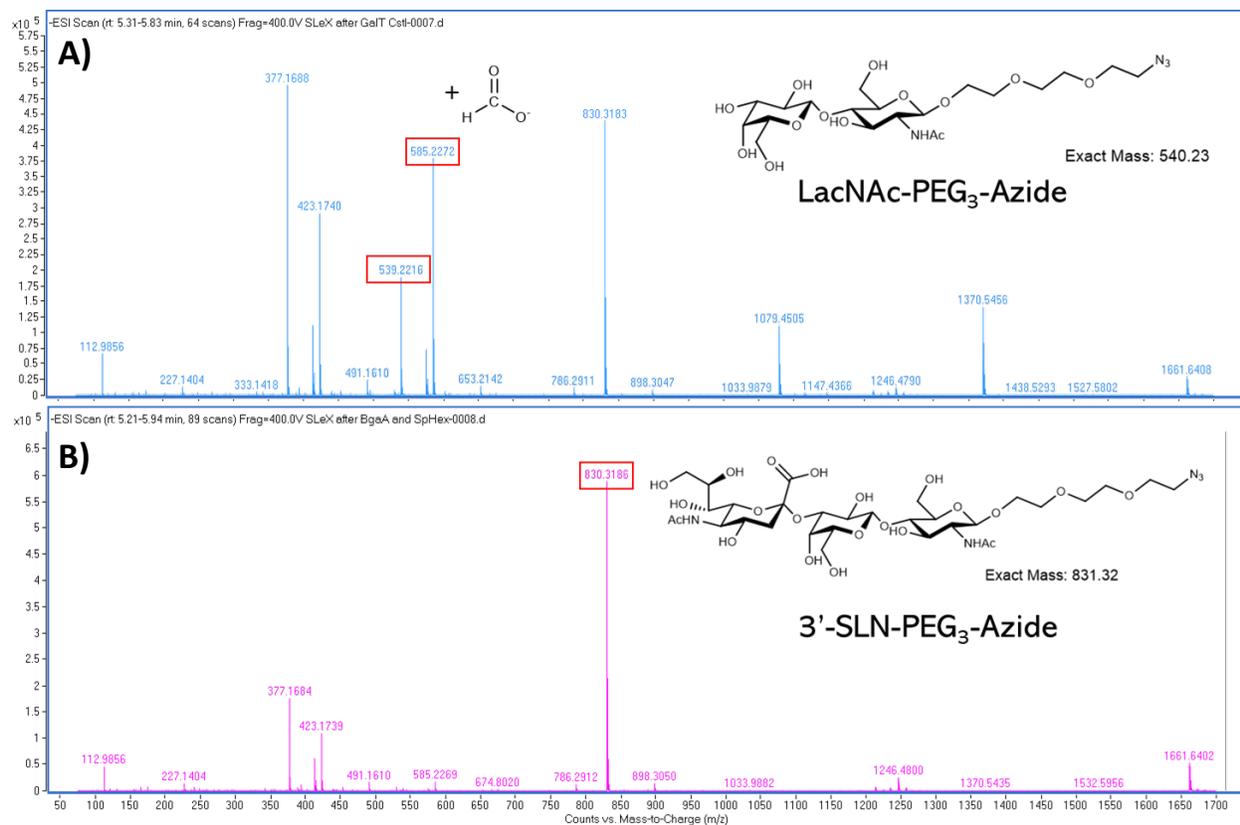


Figure 4.2: Mass spectrometry of intermediate SLe^X-PEG₃-Azide reaction steps to remove unreacted precursors. **A)** Mass spectra following incubation with the glycosyltransferases *HpGalT* and *Cst-I*. Unsialylated product was present, corresponding to LacNAc-PEG₃-Azide ($[M - H]^-$, m/z 539.22) and LacNAc-PEG₃-Azide $[M + HCOO]^-$, (m/z 585.23). **B)** Mass spectra after the addition of the glycosidases *SpHex* and *BgaA*, following the incubation with *HpGalT* and *Cst-I*. The sialylated product, 3'-SLN-PEG₃-Azide ($[M - H]^-$), has a m/z of 830.32. Samples were analyzed in negative ion mode.

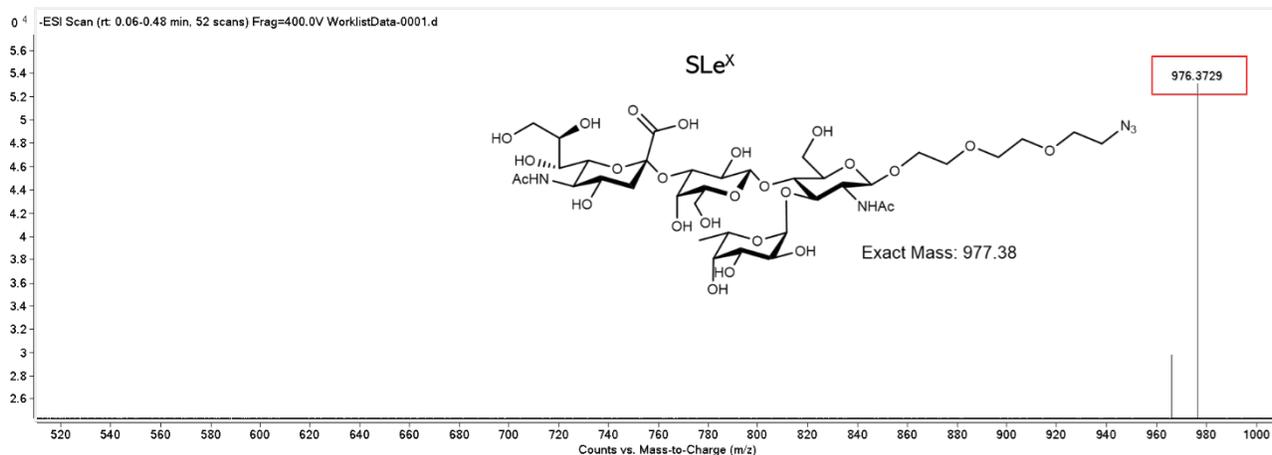


Figure 4.3: Mass spectrometry of the purified product from the enzymatic synthesis SLe^X-PEG₃-Azide. The peak at m/z 976.37 matches the expected ratio for SLe^X-PEG₃-Azide ($[M - H]^-$). The sample was analyzed in negative ion mode.

Chapter 4: RESULTS

4.2: SLe^X-PEG₃-Azide immobilization using Click Chemistry

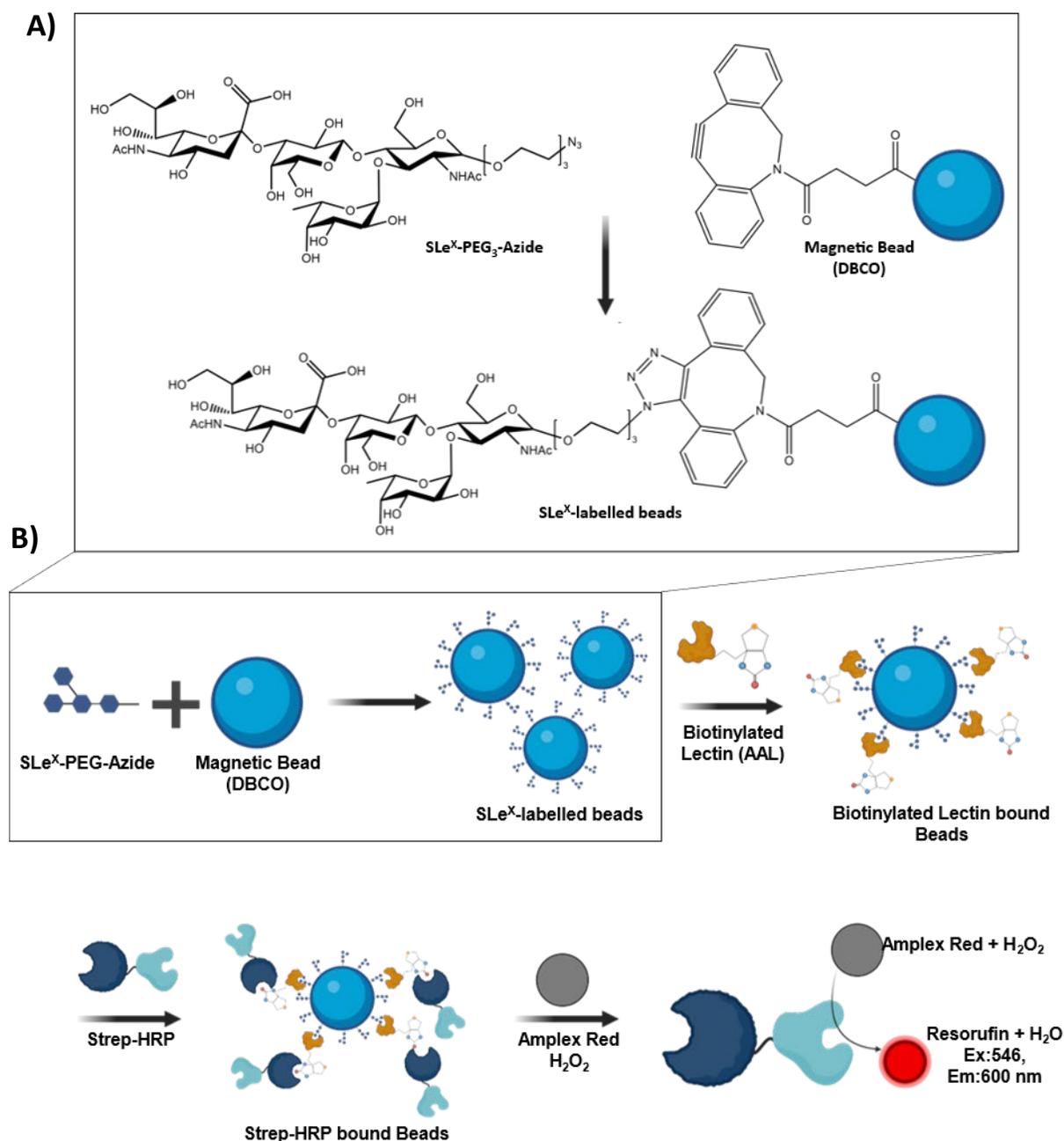
4.2 SLe^X-PEG₃-Azide immobilization using Click Chemistry

Click chemistry reactions encompasses a class of simple, high-yielding reactions that are commonly used to attach biomolecules [146]. One common type of click chemistry is azide-alkyne cycloaddition, which is typically catalyzed by Cu(I). However, strain-promoted alkyne-azide cycloadditions (SPAACs) allow for a copper free click chemistry reaction. Here, we developed a protocol for immobilizing, and detecting, SLe^X-PEG₃-Azide on dibenzocyclooctyne (DBCO) labelled magnetic beads. To test the immobilization of SLe^X-PEG₃-Azide onto the beads we designed a fluorescent lectin assay.

SLe^X-PEG₃-Azide synthesized as in section 3.2 was attached to magnetic DBCO beads as in section 3.3.1. In brief, the magnetic DBCO labelled nanoparticles were incubated with SLe^X-PEG₃-Azide in an aqueous suspension with agitation to allow the spontaneous SPAAC reaction to occur. After washing the beads, a lectin-based assay was used to confirm labeling. The nanoparticles were incubated with the biotinylated AAL lectin, which binds to fucose in the SLe^X tetrasaccharide. Carrying out washing after each step, streptavidin-HRP and AmplexTM Red were then added and the fluorescent intensity was measured (**Figure 4.4**). The fluorescent intensity of nanoparticles measured with various concentrations of SLe^X-PEG₃-Azide were measured at λ 600 nm (**Figure 4.5**). SLe^X-PEG₃-Azide concentration significantly affected ($p < 0.001$) fluorescent intensity (Two-way ANOVA, $\alpha = 0.05$, $n = 3$). The immobilized SLe^X-PEG₃-azide can now be used in the selection step of display methods such as mRNA display. We aim to design site saturated DARPIn and affibody libraries that can be screened for SLe^X binding using mRNA display.

Chapter 4: RESULTS

4.2: SLe^x-PEG₃-Azide immobilization using Click Chemistry



Chapter 4: RESULTS

4.3: DARPin and Affibody Library Design for mRNA Display

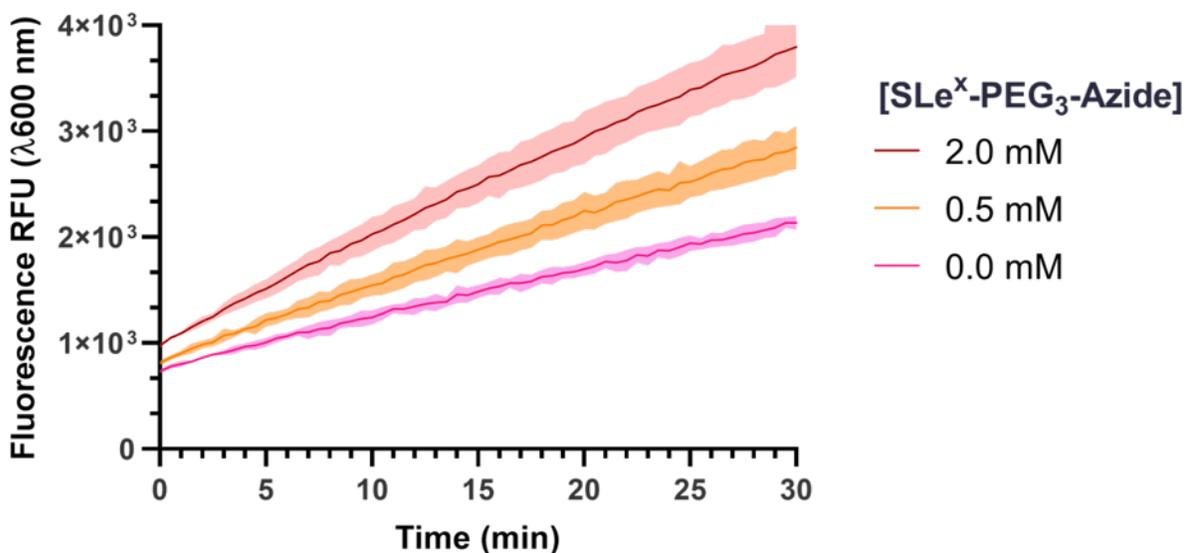


Figure 4.5: Fluorescent bead labelling assay of immobilized SLe^X-PEG₃-Azide on magnetic DBCO beads. The fluorescent intensity ($\lambda 600$ nm) of beads incubated with various concentrations of SLe^X-PEG₃-Azide (2.0 mM; red, 0.5 mM; orange, 0 mM; pink) was measured over 30 minutes. The standard deviations are indicated by the transparent area. SLe^X-PEG₃-Azide concentration significantly affected ($p < 0.001$) fluorescent intensity (Two-way ANOVA, $\alpha = 0.05$, $n = 3$).

4.3 DARPin and Affibody Library Design for mRNA Display

4.3.1 DARPin Library Design and Assembly

Designed ankyrin repeat proteins (DARPins) are highly stable proteins that can be engineered by directed evolution to generate binding proteins. Plückthun et al. previously identified residues in the DARPin scaffold for site saturation that have minimal impact on the stability of the protein [132]. In that same work, a modular library design for the DARPin library was created enabling internal repeats to be exchanged. Additionally, a DARPin library has also been previously successfully used in mRNA display to create a binding protein for Her2 [147]. Here we designed a modular DARPin library to be used in mRNA display.

The assembly method used here was inspired by the library design used by Plückthun et al. but was altered for mRNA display. Similar to the protocol created by Plückthun et al., we designed the library to be assembled using type IIS restriction enzymes; however, we keep the library linear and do not ligate it into a plasmid. The assembly method of the DARPin library is explained in section 3.4.1. In brief, the library

Chapter 4: RESULTS

4.3: DARPin and Affibody Library Design for mRNA Display

design is completed in 3 steps: site saturation amplification, addition of type IIS cut sites, and ligation (**Figure 4.6**). Site saturated oligos with 21 bp of complementarity were PCR amplified to create the site saturated modules. Then, the modules are amplified using primers with BpiI cut sites that lead to unique overhangs. The library can then be assembled by BpiI digestion and ligation by T7 ligase.

The initial site saturation amplification of the N, C, I1, I2, and I3 oligos resulted in single bands for all fragments (**Figure 6.2**). Unique type IIS cut sites were added to the N, C, I1, I2, and I3 fragments by PCR, resulting in the expected band sizes of 171, 145, 137, 124, and 141 bp (**Figure 4.7A**). The digestion and ligation of the segments is done with femtomoles of DNA, which cannot be visualized by agarose gel electrophoresis. Therefore, the digestion and ligation reaction is PCR amplified using the forward and reverse primers compatible with the N and C terminal fragments, respectively. However, the digestion and ligation resulted in multiple bands, with none matching the expected 595 bp size of the fully assembled DARPin library (**Figure 4.7B**).

Chapter 4: RESULTS

4.3: DARPin and Affibody Library Design for mRNA Display

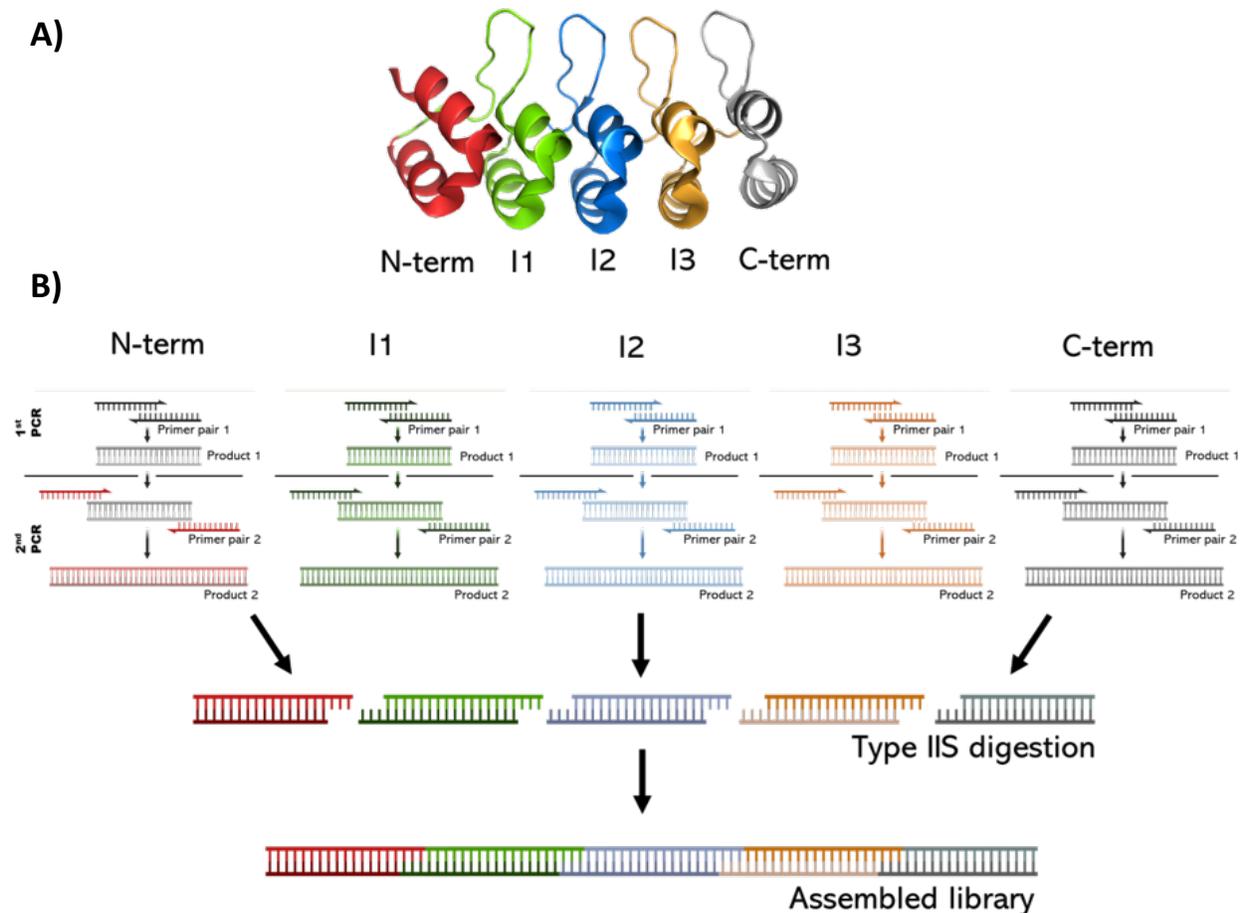


Figure 4.6: DARPin library design and synthesis approach. A) DARPin structure with modular segments color coded: N-term (red), I1 (green), I2 (blue), I3 (orange), C-term (grey). The loop regions within segments I1 to I3 are site saturated. B) DARPin library synthesis by PCR and ligation. The DARPin library is created in a three-step process. The first PCR amplifies the site saturated segments, whereas the second PCR uses primers with unique type IIS restriction enzyme cut sites. The library is then assembled using type IIS restriction enzymes.

Chapter 4: RESULTS

4.3: DARPin and Affibody Library Design for mRNA Display

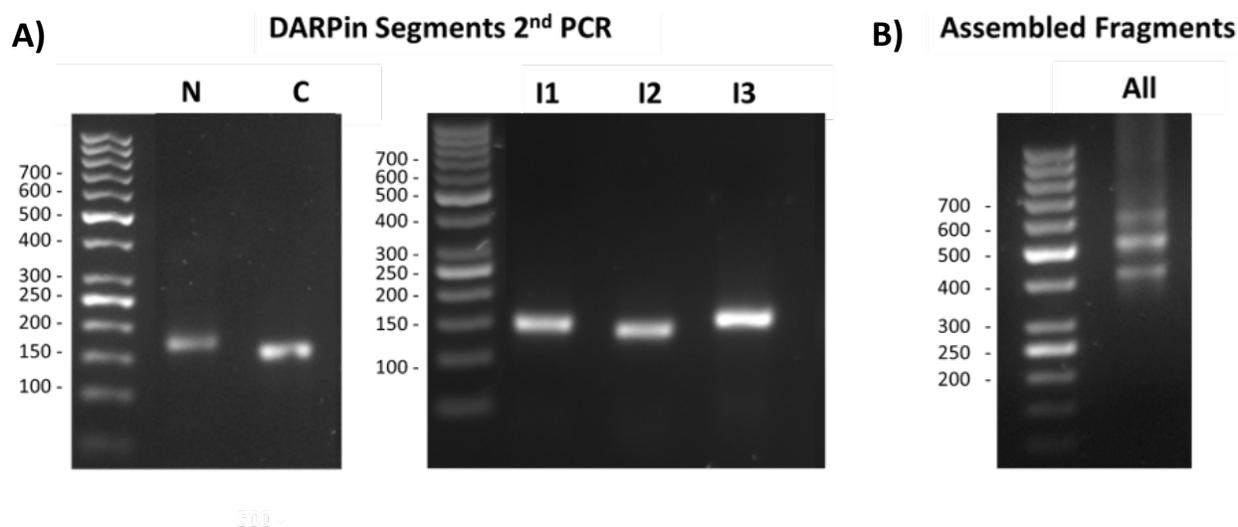


Figure 4.7: DARPin library PCR amplification and assembly. A) PCR amplification of DARPin library segments. The expected size of the N, C, I1, I2, and I3, segments are 171, 145, 137, 124, and 141 bp. Each fragment was amplified with primers that add unique type IIS restriction enzyme cut sites. B) Assembly of the DARPin library. The DARPin segments were incubated with the type IIS restriction enzyme BbsI and ligated using T7 ligase. The expected size of the assembled library is 595 bp.

4.3.2 Affibody Library Design and Assembly

Affibody are binding proteins derived from the staphylococcal Protein A and have been engineered to bind a wide variety of targets. Residues selected for site saturation were selected based on previous work by Trisha Ghosh with an affibody library [148], which are visualized in **Figure 4.8B**. Here we designed an affibody library to be used in mRNA display. Details on the library synthesis protocols are found in section **3.4.3**. In brief, the library was synthesized in two consecutive overlap extension PCRs, which amplify the site saturation and then the necessary promoter region for mRNA display (**Figure 4.8A**). The assembled affibody library size of 259 bp matches the band present in **Figure 4.8C**.

Chapter 4: RESULTS

4.3: DARPin and Affibody Library Design for mRNA Display

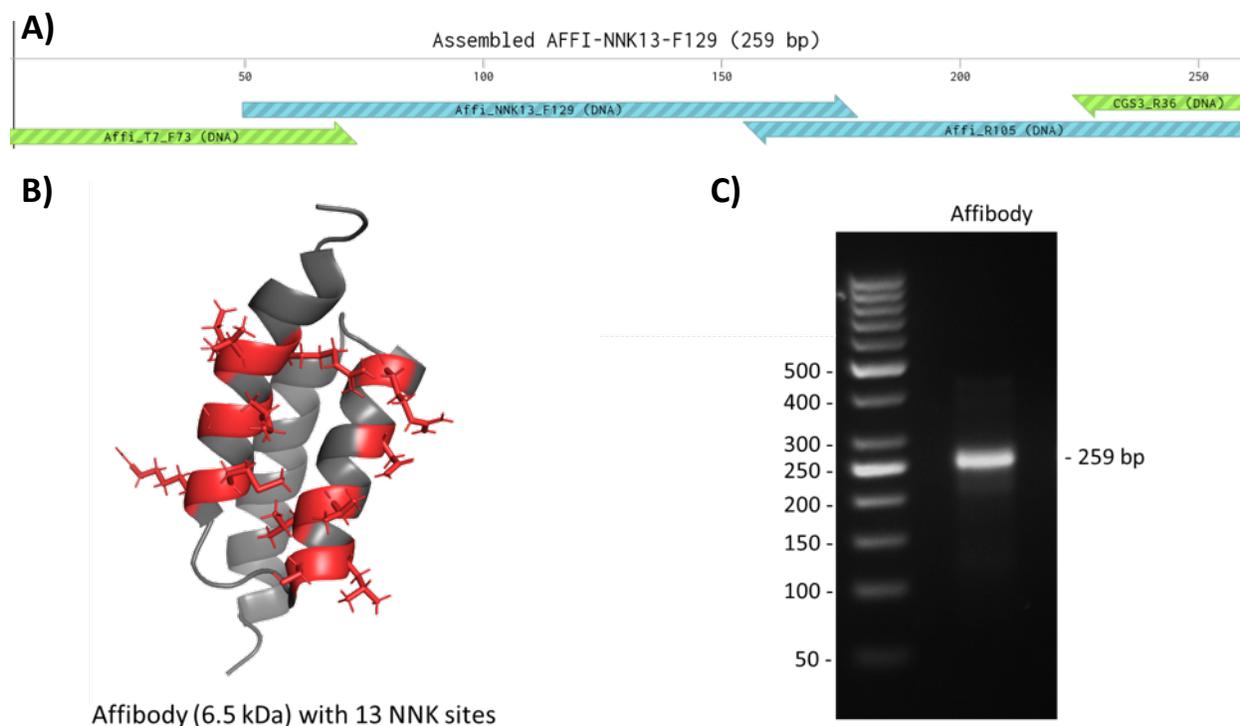


Figure 4.8: Affibody library design and synthesis. **A) Primer design for the site saturated affibody library.** The affibody library is made through two consecutive PCRs. The first reaction uses primers Affi_NNK13_F129 and Affi_R105, while the second uses Affi_T7_F73 and CGS3_R36. **B) Site saturation within the affibody structure.** The affibody (PDB: 2b89) residues in red represent the site saturated amino acid, placed in on the helix front. **C) PCR reaction of the site saturated affibody library.** The library size is expected to be 259 bp. A DNA standard is indicated in bp on the left-hand side.

4.3.3 Affibody Library Transcription, and Puromycin Linkage

The assembly of the DARPin library was unsuccessful and the mRNA display protocols for *in vitro* transcription and puromycin linkage have yet to be carried out. In contrast, the affibody library was assembled successfully. The assembled affibody library was transcribed *in vitro* as in section 3.5.1. From a 30 μL transcription reaction we purified 10 μL of mRNA at a concentration of 9100.6 $\text{ng}/\mu\text{L}$. The library was diluted to 10 μM for the puromycin linkage reaction, as in section 3.5.2. The expected size of the puromycin linked mRNA (P-mRNA) is larger than the unlinked mRNA library, which matches the band pattern observed in Figure 4.9. Once completed the mRNA display protocol would provide protein sequences that are associated with SLe^x binding affinity. However, mRNA display will not provide insight

Chapter 4: RESULTS

4.3: DARPin and Affibody Library Design for mRNA Display

on the binding site or interactions. We aim to complement mRNA display data using a computational approach that can identify glycan-protein interactions.

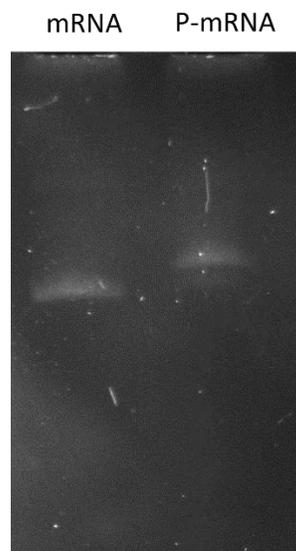


Figure 4.9: Urea-PAGE of the puromycin linked, and unlinked, mRNA affibody library. The band of size of the puromycin-mRNA (P-mRNA) was higher than the unlinked mRNA library. Note that 1 μ L of 10 mM P-mRNA and unlinked mRNA were loaded into the wells. Samples were run on an 8 % urea-PAGE and the gel was incubated with SYBR safe for visualization.

Chapter 4: RESULTS

4.4: Rational Design of Ankyrin Based Carbohydrate Binding Proteins

4.4 Rational Design of Ankyrin Based Carbohydrate Binding Proteins

We aim to create a computational approach that can complement data obtained from display methods such as mRNA display. To that end, we created a steered molecular dynamics (SMD) approach that labels glycan-protein interactions (**Section 4.6**). The SMD approach was tested for binding between dTDP-Qui3N (discussed further in **Section 4.5**) and three proteins: ankyrin, DARPin, and MtDARPin. The ankyrin protein has a known binding site for dTDP-Qui3N, whereas the DARPin does not contain a binding site. The MtDARPin was designed to recreate the dTDP-Qui3N binding pocket of ankyrin on the more stable DARPin scaffold. Ankyrin binding proteins typically have affinity for other proteins; however, Woodford et al. discovered an ankyrin domain (PDB: 4XCZ, resid: 269-397) with a binding pocket for a nucleotide sugar [19]. Here we grafted the nucleotide sugar binding site of the ankyrin domain onto the more stable DARPin structure. The mutated DARPin (MtDARPin) was designed using 9 mutations (Y81W, R89Y, E90H, D110N, D112K, Y114T, L119Y, E123R, L140Y), which were selected based on the alignment of the ankyrin and DARPin structures. The MtDARPin protein was mutated in PyMOL and the structure was determined by RoseTTAFold (**Figure 4.10**). For more details on the mutagenesis approach see section 3.7. The binding affinity of the MtDARPin to dTD-Qui3N was intended to be experimentally tested, which required us to synthesize dTDP-Qui3N.

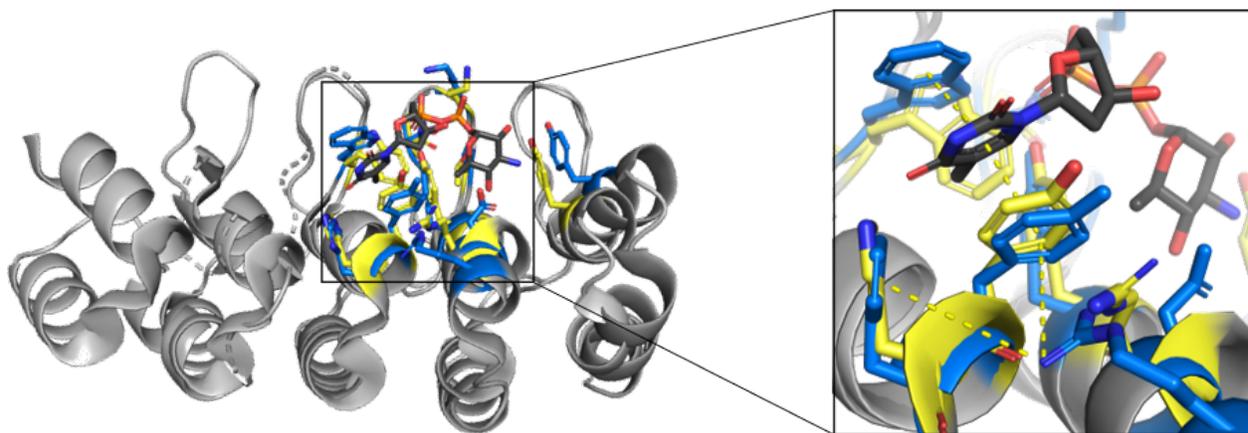


Figure 4.10: Ankyrin (PDB: 4XCZ) and DARPin (PDB: 2XEE) structural alignment and mutagenesis. The ankyrin domain with dTDP-Qui3N (PDB: T3Q) is highlighted. Ankyrin residues involved in binding are colored in blue, and mutated DARPin residues (Y81W, R89Y, E90H, D110N, D112K, Y114T, L119Y, E123R, L140Y) are colored yellow.

Chapter 4: RESULTS

4.5: Enzymatic Synthesis of dTDP-Qui3N

4.5 Enzymatic Synthesis of dTDP-Qui3N

The current only ankyrin protein known to bind a carbohydrate structure is the ankyrin domain of *Providencia alcalifaciens* N-formyltransferase (ankyrin), which binds the nucleotide sugar dTDP-Qui3N [19]. Here we enzymatically synthesized and purified dTDP-Qui3N following a previously described protocol [133], explained in section 3.5. The enzymatic synthesis of dTDP-Qui3N from dTDP-4-keto-6-deoxyglucose uses the 3,4-ketoisomerase WlaRB, followed by the aminotransferase WlaRG (**Figure 4.11A**). The reaction was purified and lyophilized as in section 3.5 and a sample was analyzed by MS (**Figure 4.11B**). dTDP-Qui3N has an expected m/z of 546.09, which was present, but the compound was not successfully produced in high quantities, as indicated by low peak intensity. In the future we also aim to experimentally test the binding interactions of dTDP-Qui3N with the ankyrin domain, DARPin, and MtDARPin. To this end, we recombinantly expressed and purified the DARPin and MtDARPin (**SFigure 6.7**), but the ankyrin domain could not be successfully expressed. Although the synthesis of both ankyrin and dTDP-Qui3N was unsuccessful, we can study the binding interactions between dTDP-Qui3N and ankyrin using molecular dynamics simulations.

Chapter 4: RESULTS

4.6: Steered Molecular Dynamics of Ankyrin Proteins with dTDP-Qui3N

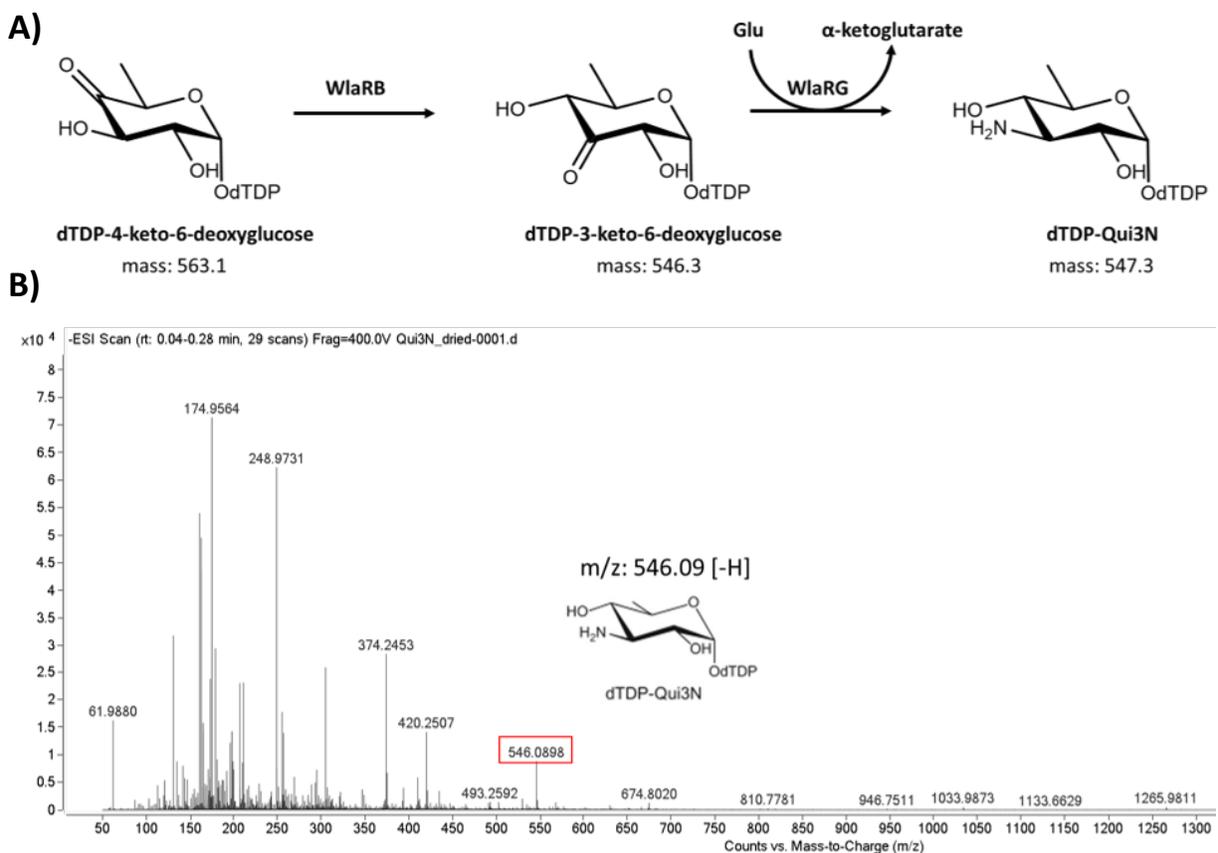


Figure 4.11: Enzymatic synthesis reaction of dTDP-Qui3N and mass spectrometry. A) Enzymatic synthesis of dTDP-Qui3N. A one-pot enzymatic reaction using WlaRB (ketoisomerase) and WlaRG (aminotransferase) produces dTDP-Qui3N from dTDP-4-keto-6-deoxyglucose. **B) Mass spectra of purified dTDP-Qui3N.** The reaction was purified using size exclusion chromatography and lyophilized. The m/z 546.09 corresponds to dTDP-Qui3N [-H]. Sample was analyzed in negative ion mode.

4.6 Steered Molecular Dynamics of Ankyrin Proteins with dTDP-Qui3N

Steered Molecular Dynamics (SMD) simulations apply a force to an atom or group of atoms to “steer” its trajectory. SMD simulations can provide insight on the interaction of small molecules and proteins [149]. Here we set up an array of SMD simulations that probe for nucleotide sugar-protein interactions between dTDP-Qui3N and ankyrin, DARPin, or MtDARPin.

Details on the ankyrin, DARPin, and MtDARPin structures and the simulation set-up can be found in section 3.8.1. The simulations were setup by aligning the protein along the z-axis and approximating the shape of the protein using a cylinder. Using DARPin as an example, the blue and green cylinders in **Figure 4.12A** indicate the starting and end points, respectively, of dTDP-Qui3N. The blue cylinder is split into a

Chapter 4: RESULTS

4.6: Steered Molecular Dynamics of Ankyrin Proteins with dTDP-Qui3N

mosaic, with tile sizes of 49 \AA^2 , as calculated from the radius of gyration of dTDP-Qui3N in section 3.8.2.

Each mosaic tile represents an independent SMD simulation with dTDP-Qui3N starting at the center of the

mosaic tile **Figure 4.12B**.

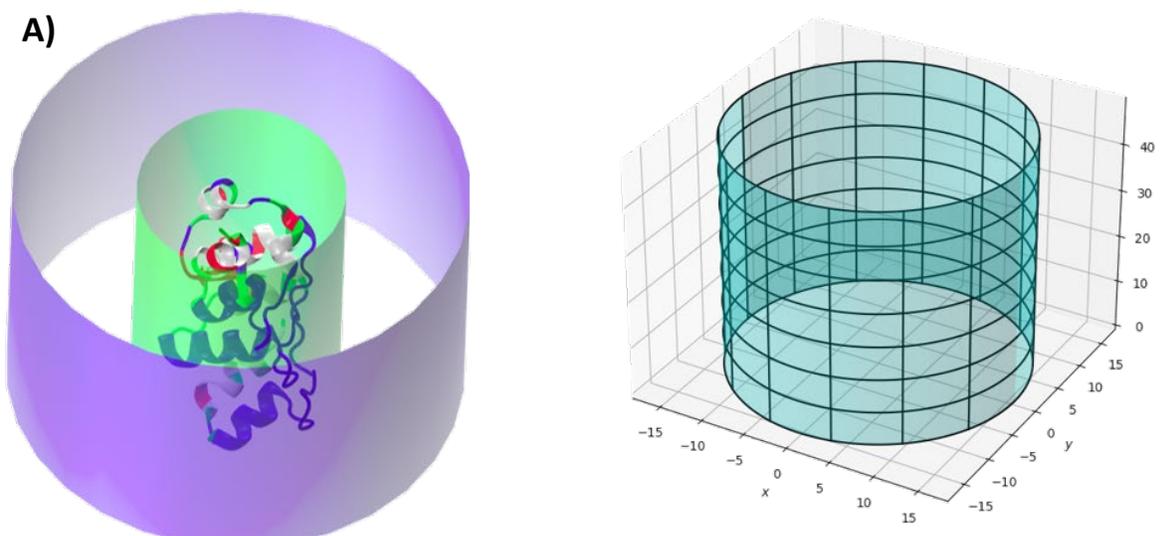


Figure 4.12: Steered Molecular Dynamics (SMD) approach to detect protein-glycan interactions. A) Simulation box setup for DARPin (PDB: 2XEE). The ligand, dTDP-Qui3N (PDB: T3Q), starts on the outer cylinder (blue) and is pushed towards the inner cylinder (green). **B) Partitioning of the cylinder into individual simulations.** The outer cylinder is divided into a mosaic, with the dTDP-Qui3N placed at the center of each mosaic tile.

The total number of bonds in the steered molecular dynamics simulations were analyzed using VMDs Hbonds plugin and heatmaps of the total number of hydrogen bonds were created using python (**Figure 4.13**). The dTDP-Qui3N binding site in the ankyrin domain is located at z and θ value of -3.4 and 162 , respectively, which corresponds to the maximum in hydrogen bond heatmap (**Figure 4.13A**). The dTDP-Qui3N binding site discovered by the SMD simulation is near the crystal structure binding site, as can be seen by the alignment of the simulated structure at the maximum on the heatmap ($z = -3.4$, $\theta = 162$) with the known crystal structure of the ankyrin domain (**Figure 4.14**). In contrast, the hydrogen bond heatmaps for DARPin and MtDARPin have interactions at a lower intensity (**Figure 4.13B, C**). The MtDARPin, which has been designed to replicate the dTDP-Qui3N binding site of the ankyrin, contains a no hot spot at the designed binding site ($z: -3.62$, $\theta: 165$).

Chapter 4: RESULTS

4.6: Steered Molecular Dynamics of Ankyrin Proteins with dTDP-Qui3N

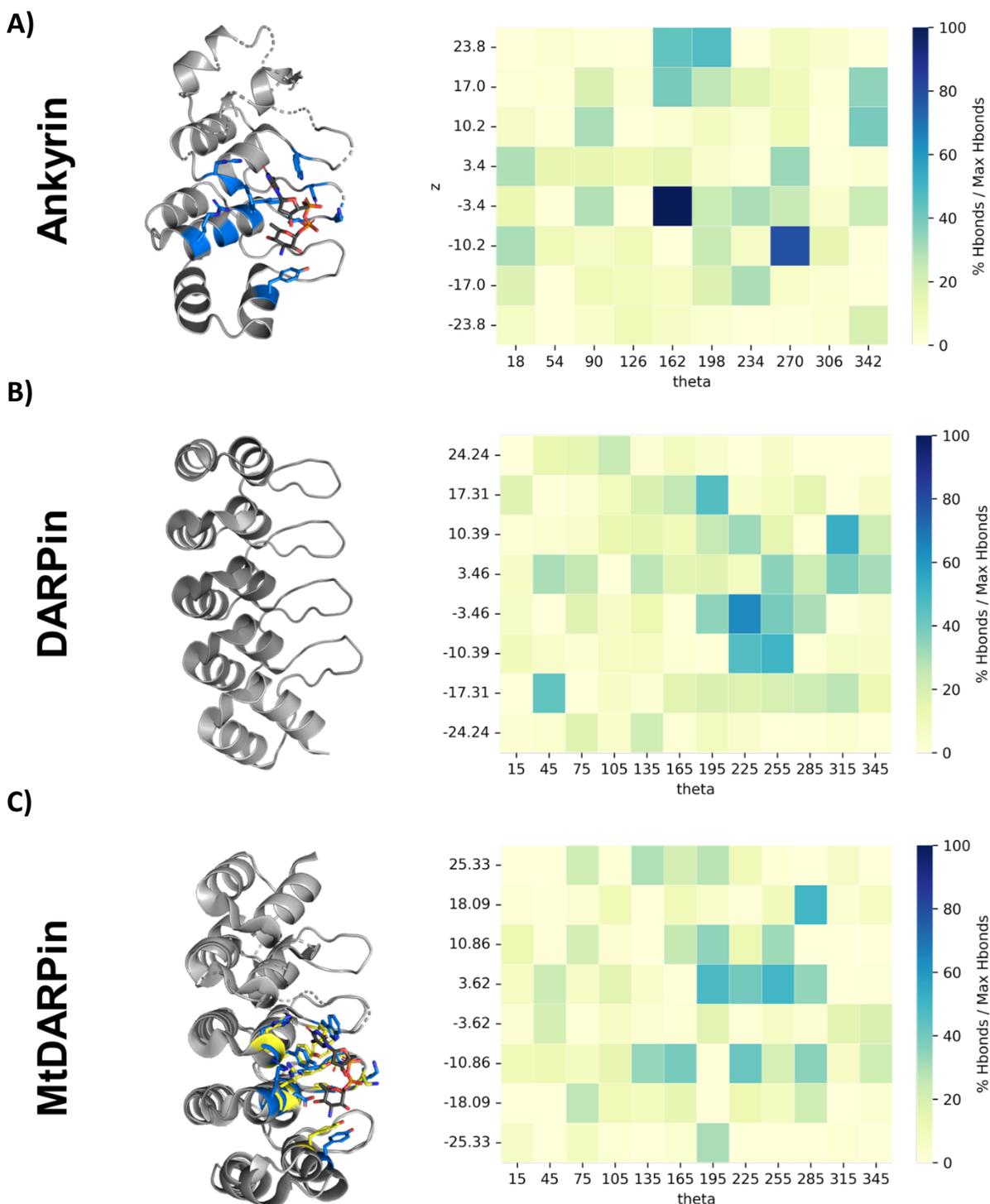


Figure 4.13: Hydrogen bond heatmaps generated from the SMD simulations of dTDP-Qui3N with ankyrin, DARPin, and MtDARPin. A) Hydrogen bond heatmap of ankyrin simulations. The heatmap maximum ($z = -3.4$, $\theta = 162$) corresponds to the dTDP-Qui3N binding site. B) Hydrogen bond heatmap of DARPin simulations set up as a negative control. C) Hydrogen bond heatmap of MtDARPin simulations. These simulations were set up to test grafting the ankyrin binding site onto the DARPin scaffold. Each tile in the heatmap represents a single SMD simulation. Theta and z refer to the angle around the protein, and the cartesian z coordinate, respectively, at which dTDP-Qui3N is placed relative to the protein.

Chapter 4: RESULTS

4.6: Steered Molecular Dynamics of Ankyrin Proteins with dTDP-Qui3N

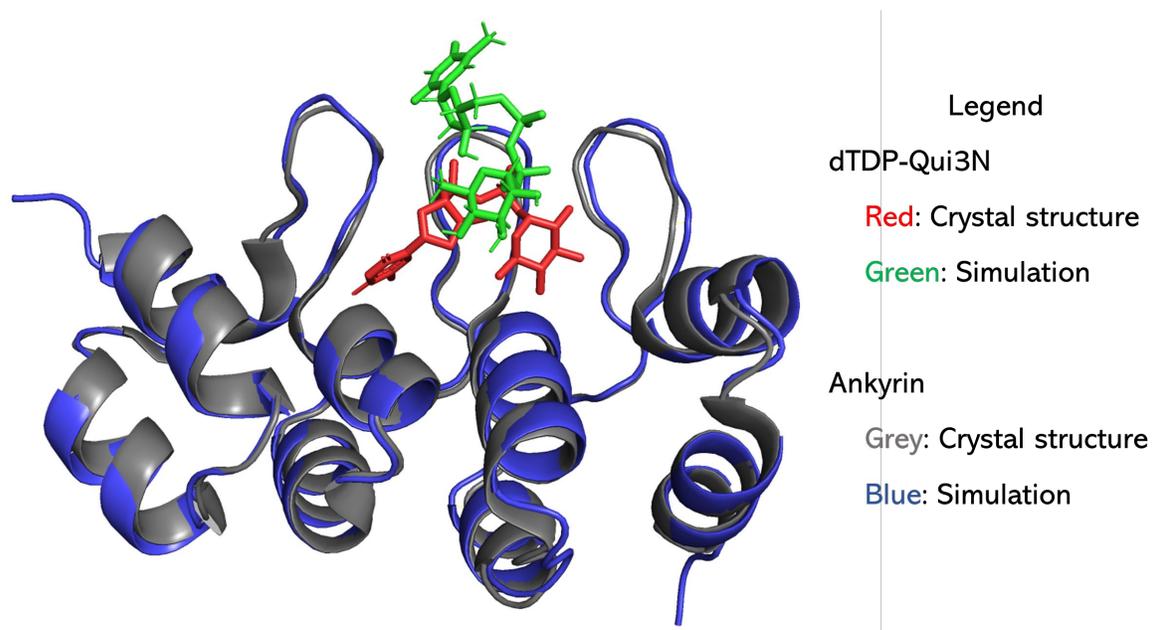


Figure 4.14: Overlay of the simulated and crystal structure interactions between dTDP-Qui3N and the ankyrin domain. The end point of the maximum hotspot ($z = -3.4$, $\theta = 162$) from the hydrogen bond heatmap between the ankyrin proteins (blue) and dTDP-Qui3N (green) is shown. The simulated ankyrin structure (blue) was aligned with the crystal structure (grey), alongside the crystal structure binding site of dTDP-Qui3N (red). The ankyrin domain contains residues 269-397 of PDB 4XCZ.

Chapter 5: Discussion

5.1: Enzymatic Synthesis of SLe^X-PEG₃-Azide

CHAPTER 5: DISCUSSION

Glycans have various physiological functions with implications in numerous diseases that have made them important biomarkers for diagnostic and therapeutic purposes. Due to advances in the field of glycobiology new technologies such as lectin arrays are leading to the rapid discovery of glycan epitopes of interest. The study of these glycans is currently limited by a specific tools that can be used to target glycan structures [29]. Directed evolution methods coupled with display screening methods such as mRNA display allows for the rapid production of binding proteins for a variety of targets. Display methods require malleable starting scaffolds that can tolerate an array of mutations without denaturing, such as affibodies and designed ankyrin repeat proteins (DARPin). Binding protein variants selected from these scaffolds can be further improved by rational design, which can be accomplished using computational methods such as molecular dynamics (MD) simulations. In this project, we developed methods that aim to accelerate the production of glycan binding proteins using directed evolution and computational approaches. We created site saturated affibody and DARPin libraries compatible with mRNA display methods, that can be tested for binding affinity to the tumour associated carbohydrate (TACA) SLe^X-PEG₃-azide. We optimized an enzymatic synthesis protocol for SLe^X-PEG₃-azide, which can be immobilized and used in screening methods for display techniques such as mRNA display. We validated the immobilization approach using a lectin based fluorescent assay. Additionally, we created an MD approach for the identification of protein-ligand interactions, that can complement the sequence data obtained from mRNA display.

5.1 Enzymatic Synthesis of SLe^X-PEG₃-Azide

Sialyl Lewis X (SLe^X) was selected as the target ligand for designing novel glycan binding proteins (GBPs). SLe^X is a tumour associated carbohydrate antigen (TACA) associated with various types of cancers. Earlier work by Haoyu Wu described an enzymatic synthesis protocol for SLe^X-PEG₃-azide, using a three-step enzymatic pathway [13]. The bacterial glycosyltransferases *HpGalT*, *Cst-I*, and *HpFucT*, are used to transfer galactose, sialic acid, and fucose, respectively, from their nucleotide-activated donors to a GlcNAc-PEG₃-azide starting compound. Following this protocol, we noted that the addition of sialic acid was

Chapter 5: Discussion

5.2: SLe^X-PEG₃-Aide Immobilization on Magnetic Beads

incomplete and resulted in a side product with an MS peak at $m/z = 731.29$ (**SFigure 6.1**). The enzyme *HpFucT* used in the reaction is an $\alpha 1,3$ fucosyltransferase, which adds fucose to the GlcNAc residue. If the sialyl addition by Cst-I is incomplete *HpFucT* still fucosylates GlcNAc, resulting in Le^X-PEG₃-azide, which explains the MS peak at m/z 731.29. Le^X-PEG₃-azide is a poor substrate for Cst-I, reducing the yield of the desired product SLe^X-PEG₃-azide.

Altering the reaction protocol to increase the incubation time with the sialyltransferase Cst-I did not increase the MS peak at m/z of 830.32 (not shown); therefore, we aimed to reduce formation of the by-product by enzymatically digesting unreacted intermediates using *SpHex* (*N*-acetylhexosaminidase) and BgaA (galactosidase). *SpHex* and BgaA can degrade GlcNAc-PEG₃-azide and LacNAc-PEG₃-azide, respectively, but cannot degrade the sialylated 3'-SLN-PEG₃-azide as both enzymes act on terminal monosaccharides [150,151]. Therefore, prior to the fucosylation step, we added *SpHex* and BgaA to the reaction mixture as in section 3.2 (**Figure 4.1**). *SpHex* and BgaA degraded the precursors – with the major MS peak of m/z of 830.32 corresponding to the desired product, 3'-SLN-PEG₃-azide (**Figure 4.2**). The reaction mixture containing 3'-SLN-PEG₃-azide was then incubated with *HpFucT*, producing the desired product SLe^X-PEG₃-azide (**Figure 4.3**). From a 1 mL reaction we obtained 3.0 mg of SLe^X-PEG₃-Azide, a 61 % yield, based on the theoretical yield of a complete reaction. Removing the precursors was essential, as all intermediate contain the azide tag, which would lead to interference in the immobilization of SLe^X-PEG₃-azide and would affect downstream selection for binding proteins. The purified SLe^X-PEG₃-azide is intended to be used in the selection of novel GBPs, which requires the ligand to be immobilized.

5.2 SLe^X-PEG₃-Aide Immobilization on Magnetic Beads

Our aim is to create glycan binding proteins (GBPs) for SLe^X using the high throughput selection method mRNA display, which requires methods for the immobilization and the detection of SLe^X. Purified SLe^X-PEG₃-azide was immobilized using click chemistry enabled dibenzocyclooctyne (DBCO) labelled magnetic nanoparticles as in section 3.3.1. Click chemistry allows for the simple and covalent

Chapter 5: Discussion

5.3: Site Saturated DARPin and Affibody Library Design for use in mRNA Display

immobilization of SLe^X-PEG₃-azide and could be expanded to other glycan targets, given that they can be synthesized with an azide group.

The click chemistry reaction is expected to occur rapidly at room temperature, without a catalyst [152]. Proper immobilization of the glycan target is essential for the downstream selection of GBPs; therefore, we developed a fluorescent assay to test whether the SLe^X-PEG₃-azide immobilization was successful (**Figure 4.4A**). We incubated the glycan labelled beads with biotinylated AAL lectin, which binds to the fucose in SLe^X. Streptavidin-HRP was added to bind the biotinylated AAL lectin. Then we added Amplex™ red and H₂O₂. Amplex red reacts with HRP and H₂O₂ to make resorufin, the absorbance of which can be measured. The concentration of immobilized SLe^X-PEG₃-azide is expected to positively correlate with the fluorescence intensity. Testing two different concentrations of SLe^X-PEG₃-azide (2.0 mM, 1.0 mM) and a control (0 mM), the fluorescent intensity positively correlated with the concentration of SLe^X-PEG₃-azide (Two-way ANOVA, $\alpha = 0.05$, $n = 3$, $p < 0.001$; **Figure 4.4B**). It should be noted that there is signal in the negative control containing 0 mM SLe^X, this could likely be reduced by increasing the number and stringency (increased Triton) of washing steps. Future experiments could investigate that; however, the AAL lectin also has a low binding affinity [153], and each wash would reduce the signal we would get from the assay. Future work could improve upon the sensitivity of this assay by altering the washing steps or using different lectins, but for the purposes of this work, we aimed to test whether we could detect immobilized SLe^X.

Since the immobilized SLe^X was detected on the magnetic beads it indicates that our click-chemistry immobilization protocol was successful and could be used to generate carbohydrate labelled beads for selection methods, such as mRNA display. We aim to generate DARPin and affibody libraries for use in mRNA display that can be used to screen for SLe^X binding using the magnetic SLe^X beads generated here.

5.3 Site Saturated DARPin and Affibody Library Design for use in mRNA Display

One factor limiting the study of glycans is a lack of specific tools that can be used to detect them [29]. We aim to accelerate the discovery novel GBPs by mRNA display using site saturated DARPin and affibody

Chapter 5: Discussion

5.3: Site Saturated DARPin and Affibody Library Design for use in mRNA Display

libraries. Glycan-protein interactions tend to be weak, with dissociation constants in the mM range. In nature, the weak binding affinity is usually compensated through multivalent glycan-protein interactions. One binding protein scaffold that allows for multivalency are DARPins, which contain internal repeat units that can be altered to accommodate multiple binding sites [17,18]. Although DARPins have never been screened for their ability to bind glycans, a bacterial ankyrin domain has been characterized with a binding site for a nucleoside sugar [19]. Due to the structural similarities of the ankyrin domain and DARPin scaffolds, DARPins may be adaptable to bind carbohydrates. We designed a modular DARPin library to be assembled without a vector, as mRNA display requires linear DNA libraries. The modules are N-terminal, I1, I2, I3, and C-terminal, with the internal repeats (I1-I3) being interchangeable. The DARPin library is amplified in two steps, first the site saturated modules are amplified, and second, type IIS restriction enzyme cut sites are added to using primers (**Figure 4.6**).

The PCR amplification of the DARPin modules, N, C, I1, I2, and I3, was successful, with the second step resulting in the band sizes of 171, 145, 137, 124, and 141 bp, respectively (**Figure 4.7A**). All modules were then digested and ligated as in section 3.4.2. The ligated products contain 450 fmol of DNA, which is PCR amplified for visualization by agarose gel electrophoresis. However, the PCR product of the digestion and ligation resulted in multiple band sizes, with none matching the expected 595 bp size of the assembled DARPin library (**Figure 4.7B**). It is unclear whether it is the ligation that is failing, or the subsequent amplification. It could be possible that the site saturation within the library is creating new type IIS cut sites, however, this seems unlikely to be the problem as we assembled the library in parts with some success. The N to I1, and I3 to C, fragments can be ligated together successfully, resulting in band sizes of 279 and 257 bp, respectively (**SFigure 6.3**). The specific sequence used for the library may be causing secondary structure formation, in which case future work should aim to create a library with synonymous codons.

In contrast to the DARPin library, the affibody library was synthesized successfully by overlap extension PCR. The library was created by two consecutive PCR reactions, first amplifying the site saturation, and

Chapter 5: Discussion

5.3: Site Saturated DARPin and Affibody Library Design for use in mRNA Display

then adding the T7 promoter and the terminal puromycin linker sequence (**Figure 4.8A**). Amplification of the affibody library was successful, resulting in the expected band size of 259 bp (**Figure 4.8C**).

Our aim is to test the DARPin and affibody libraries using mRNA display, which can be broken down into a four-step approach: library generation, transcription, translation, and screening (**Figure 5.1**). The library generation was previously discussed, here we focus on the transcription. Since the DARPin library generation was not successful we only discuss the results of the transcription and puromycin linkage of the affibody library. The *in vitro* transcription of the affibody library was successful, resulting in 10 μL of 9100.6 ng/ μL mRNA purified from a 30 μL transcription reaction. In mRNA display, the mRNA is linked to a puromycin linker, which creates a covalent bond between the mRNA and the protein during translation. We successfully linked the affibody mRNA library to the puromycin linker, as shown by urea-PAGE (**Figure 4.9**). The puromycin-linked mRNA is higher up on the urea-PAGE, as expected due to its increased molecular weight from the addition of the puromycin linked oligo.

In combination with the directed evolution approach by mRNA display discussed here, we aimed to create computational approaches that could be used to further improve GBPs by rational protein design. mRNA display data will select for sequences of proteins associated with higher binding avidity but does not provide information on the binding site location or types of interactions. In the following section we discuss an *in silico* molecular dynamics (MD) approach that is designed to discover glycan-protein interactions.

Chapter 5: Discussion

5.4: In Silico Study of Glycan-Protein Interactions

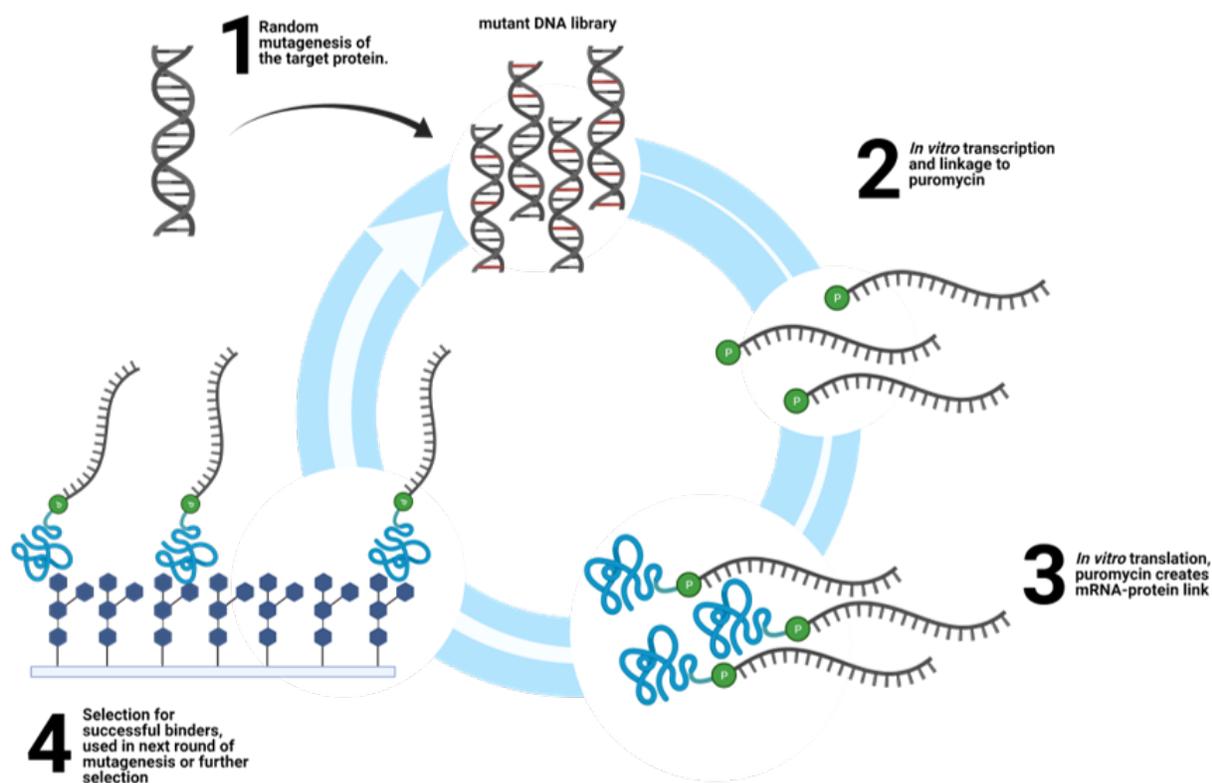


Figure 5.1: Directed evolution by mRNA display. The approach can be broken down into four steps: library synthesis, transcription, translation, and selection. Following *in vitro* transcription of the library, the mRNA is linked to puromycin. During translation puromycin is covalently linked to the nascent peptide. The peptide-puromycin-mRNA library can then be screened for binding against an immobilized target.

5.4 In Silico Study of Glycan-Protein Interactions

We aim to create a computational approach that can complement data obtained from display methods such as mRNA display. The nucleotide sugar dTDP-Qui3N was selected to be used in the *in silico* study of glycan-protein interactions, since it is currently the only known carbohydrate ligand for an ankyrin that binds directly to the carbohydrate structure. This allows us to simulate the ankyrin domain of *Providencia alcalifaciens* N-formyltransferase (hereinafter: ankyrin), which has a binding site for dTDP-Qui3N, as a positive control [19].

We set up MD simulations that probe for interactions between dTDP-Qui3N and three different proteins: ankyrin, DARPin, and MtDARPin (Section 3.8). The MtDARPin was mutated to replicate the dTDP-Qui3N binding site found in the ankyrin domain, as explained in section 3.7. The simulations probe for glycan-protein interactions of the protein at different locations and the interactions over time are depicted

Chapter 5: Discussion

5.4: In Silico Study of Glycan-Protein Interactions

in a heatmap (**Figure 4.13**). The heatmap location are indicated by z and θ values, which correspond to locations on the protein that are being probed for glycan-protein interactions. The dTDP-Qui3N binding site in the ankyrin domain, as represented on the hydrogen bond heatmap, is expected to be at a z and θ value of -3.4 and 168, respectively. This corresponds to the coordinates of the hotspot maximum, indicating that the probing simulation correctly identified the binding site (**Figure 4.13**). When aligning the simulation hotspot structure ($z = -3.4$, $\theta = 168$) of the ankyrin and dTDP-Qui3N with the crystal structure, it further highlights that we were able to correctly identify the dTDP-Qui3N binding site using the MD approach (**Figure 4.14**). Although, there are differences between the exact location of dTDP-Qui3N ligand in the simulation when compared to the crystal structure. This is expected, as the binding simulations are carried out over 8-10 ns; whereas it can take hundreds of nanoseconds to microseconds for small molecules to bind a protein binding site [154]. With this SMD approach we expect that even though we do not have enough time to let the ligand find its proper binding pocket, we can find the general area of the protein binding site in a short timescale. We expect that even if the ligand is not inside the binding pocket, it would have more interactions with the protein closer to the binding site. Longer simulations could be set up in future experiments to test whether dTDP-Qui3N will eventually bind as in the crystal structure.

In comparison with the ankyrin simulations, we did not see as strong glycan-protein interactions in the DARPin and MtDARPin simulations (**Figure 4.13B, C**). The lack of interactions in the DARPin simulations were expected, as the DARPin protein does not have a binding site for dTDP-Qui3N and was chosen as a negative control. Interactions between DARPin and dTDP-Qui3N were still present, but at a lower intensity than in the ankyrin simulations. However, the MtDARPin was expected to have more interactions with dTDP-Qui3N, as it was mutated to replicate the binding site of the ankyrin. This indicates that grafting the binding site from the ankyrin onto the MtDARPin scaffold may be causing other changes within the protein that are impeding dTDP-Qui3N binding. Grafting binding sites of structural similar proteins has been successfully done in the literature [155], but by grafting the entire binding site into the MtDARPin there may be other changes we are introducing to the overall structure. Future *in silico* studies

Chapter 5: Discussion

5.4: In Silico Study of Glycan-Protein Interactions

of MtDARPin variants containing partial binding site grafts could be set up, which may circumvent structural issues caused by grafting all residues of the ankyrin binding site.

We also created scripts for analyzing interactions with the phosphates in dTDP-Qui3N and the proteins, as well as interactions with the sugar rings and aromatic interactions, since those are more important when it comes to carbohydrate protein interactions (**SFigure 6.5**). However, the heatmaps obtained from that did not match the heatmaps obtained by the hydrogen bond scripts. It may be explained by our MDAnalysis script only looking at vector distances between two atoms, without taking the angles between interacting atom groups being measured. For example, in hydrogen bonds, the angle between the donor and acceptor atoms (defined as D–H···A) must be less than 20° for a hydrogen bond to occur (**Figure 5.2**). Since the script that counts the number of interactions with aromatic residues (pi-stacking, CH-pi) does not take bond angle limitations into considerations, it may incorrectly count some interactions simply based on proximity of atoms.

Overall, this SMD probing approach used here is a computationally affordable simulation that was able to correctly identify the dTDP-Qui3N binding site in the ankyrin domain. This is the first case of SMD probing simulations being used to identify glycan-protein interactions. Although MD probing simulations have been used to study and improve protein-drug interactions, the approach described here is less computationally expensive. Typical probing simulations include multiple probe (i.e. ligand) molecules and a map is created based on the time the probe spends at different locations on the protein, combined with the free energy of the probe [28]. In contrast, our approach was able to identify the glycan binding site simply based on glycan-protein hydrogen bonds, over shorter (10 ns) simulations. This approach could be a useful tool for probing glycan protein interactions and could be used for the rational design of glycan binding proteins. Data obtained from directed evolution techniques, such as selected proteins from mRNA display, could be simulated using this SMD probing approach to identify the binding sites. In turn, the data from the SMD probing simulation could be used to create protein chimeras to improve the binding affinity or specificity

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

of GBPs. To verify the results of the *in silico* determined interactions, future work should experimentally assess the binding interactions of dTDP-Qui3N with the ankyrin, DARPin, and MtDARPin proteins.

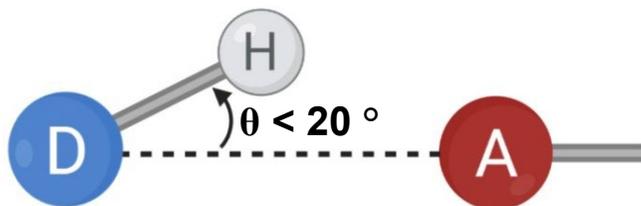


Figure 5.2: Hydrogen bond requirements. The angle formed between atoms D–H···A has to be less than 20 ° for a hydrogen bond to occur.

5.5 Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

The *in silico* approach mentioned in the previous section will need to be validated in the future. For this end we aimed to synthesize dTDP-Qui3N to be used for mutational studies that can test the binding affinities of mutated proteins to dTDP-Qui3N. The enzymatic synthesis of dTDP-Qui3N was previously described [134]. While we were able to produce dTDP-Qui3N, as indicated by the MS peak at m/z 546.09 (**SFigure 6.6**); upon purification attempts, the compound was barely detectable after freeze drying (**Figure 4.11**). We think alternative synthesis protocol could be tried to increase the product yield and stability, as other researcher were able to produce enough dTDP-Qui3N for co-crystallization with proteins [19].

If dTDP-Qui3N can be successfully synthesized in the future, the binding interactions with the ankyrin domain, DARPin, and MtDARPin could be investigated experimentally. To this end we recombinantly expressed and purified DARPin and MtDARPin (**SFigure 6.7**), but the expression of ankyrin was not successful. It should be noted that purified DARPin and MtDARPin fractions contained higher kDa impurities. It has been noted in the literature that DARPins can be stable even above 90 °C [156], hence the impurities may simply be protein that is not fully denatured. Future experiments could test the thermostability of the DARPin and ankyrin proteins purified here.

CHAPTER 6: CONCLUSION AND FUTURE DIRECTIONS

In this study we aimed to address the lack of available glycan binding proteins (GBPs) by developing new methods, experimental and computational, that can accelerate the production of GBPs. To this aim, we

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

developed a protocol for immobilizing glycans on magnetic nanoparticles (MNPs) using click chemistry, enabling the use of glycans for directed evolution methods such as mRNA display. Additionally, we created a fluorescent lectin assay to test the immobilization of glycans on MNPs. We successfully immobilized and detected the tumour associated carbohydrate antigen (TACA) SLe^x; however, these methods are not limited to SLe^x and future work can test this approach for a wide array of glycans. Future work can also screen the mRNA display compatible affibody library created here for binding affinity to SLe^x. We successfully transcribed the affibody library and linked our library to puromycin. The library is now ready for selection against immobilized SLe^x, or other TACAs of interest. In contrast to the affibody library, the designed ankyrin repeat (DARPin) library is not ready for selection by mRNA display. A cost effective, modular DARPin library was designed; however, the modules have not been successfully ligated. We have demonstrated that 450 fmol of linear DNA modules can be ligated together by type IIS digestion and T7 ligation in pairs, but the assembly of all five modules needs to be optimized in future work. Once optimized, the modular DARPin library could be tested for binding ability to SLe^x using mRNA display.

We also developed an *in silico* approach using molecular dynamics (MD) simulations that aims to complement display methods. The MD simulations were able to identify the binding site of the nucleoside sugar dTDP-Qui3N in the ankyrin domain of an N-formyltransferase based on the hydrogen bond interactions. The approach was also tested using a recreation of the dTDP-Qui3N binding site in a mutated DARPin scaffold (MtDARPin); however, the grafted binding site did not show interactions with dTDP-Qui3N *in silico*. This may be a limitation of the MtDARPin design rather than the MD simulations but should be investigated further using mutagenesis studies. Future work could also apply the MD approach to identify the binding sites of protein variants selected using display methods, which could aid in the rational design higher affinity binding proteins.

The MD approach, display library design, and glycan immobilization strategy described here could be used to accelerate the production of novel GBPs, paving the way for the development of new glycan targeting diagnostic and therapeutic tools.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

REFERENCES

1. Warkentin, R.; Kwan, D.H. Resources and Methods for Engineering “Designer” Glycan-Binding Proteins. *Molecules* **2021**, *26*, 380, doi:10.3390/molecules26020380.
2. Flynn, R.A.; Pedram, K.; Malaker, S.A.; Batista, P.J.; Smith, B.A.H.; Johnson, A.G.; George, B.M.; Majzoub, K.; Villalta, P.W.; Carette, J.E.; et al. Small RNAs Are Modified with N-Glycans and Displayed on the Surface of Living Cells. *Cell* **2021**, *184*, 3109-3124.e22, doi:10.1016/j.cell.2021.04.023.
3. Varki, A.; Sharon, N. *Historical Background and Overview*; Cold Spring Harbor Laboratory Press, 2009;
4. Pearce, O.M.T. Cancer Glycan Epitopes: Biosynthesis, Structure and Function. *Glycobiology* **2018**, *28*, 670–696, doi:10.1093/glycob/cwy023.
5. Nardy, A.F.F.R.; Freire-de-Lima, L.; Freire-de-Lima, C.G.; Morrot, A. The Sweet Side of Immune Evasion: Role of Glycans in the Mechanisms of Cancer Progression. *Front Oncol* **2016**, *6*, 54, doi:10.3389/fonc.2016.00054.
6. Maeda, T.; Alexander, C.M.; Friedl, A. Induction of Syndecan-1 Expression in Stromal Fibroblasts Promotes Proliferation of Human Breast Cancer Cells. *Cancer Research* **2004**, *64*, 612–621, doi:10.1158/0008-5472.CAN-03-2439.
7. Du, W.W.; Yang, B.B.; Shatseva, T.A.; Yang, B.L.; Deng, Z.; Shan, S.W.; Lee, D.Y.; Seth, A.; Yee, A.J. Versican G3 Promotes Mouse Mammary Tumor Cell Growth, Migration, and Metastasis by Influencing EGF Receptor Signaling. *PLOS ONE* **2010**, *5*, e13828, doi:10.1371/journal.pone.0013828.
8. Samraj, A.N.; Pearce, O.M.T.; Läubli, H.; Crittenden, A.N.; Bergfeld, A.K.; Banda, K.; Gregg, C.J.; Bingman, A.E.; Secrest, P.; Diaz, S.L.; et al. A Red Meat-Derived Glycan Promotes Inflammation and Cancer Progression. *Proceedings of the National Academy of Sciences* **2015**, *112*, 542–547, doi:10.1073/pnas.1417508112.
9. Kannagi, R.; Izawa, M.; Koike, T.; Miyazaki, K.; Kimura, N. Carbohydrate-Mediated Cell Adhesion in Cancer Metastasis and Angiogenesis. *Cancer Science* **2004**, *95*, 377–384, doi:10.1111/j.1349-7006.2004.tb03219.x.
10. Cohen, E.N.; Fouad, T.M.; Lee, B.-N.; Arun, B.K.; Liu, D.; Tin, S.; Gutierrez Barrera, A.M.; Miura, T.; Kiyokawa, I.; Yamashita, J.; et al. Elevated Serum Levels of Sialyl Lewis X (SLeX) and Inflammatory Mediators in Patients with Breast Cancer. *Breast Cancer Res Treat* **2019**, *176*, 545–556, doi:10.1007/s10549-019-05258-0.
11. Tang, H.; Singh, S.; Partyka, K.; Kletter, D.; Hsueh, P.; Yadav, J.; Ensink, E.; Bern, M.; Hostetter, G.; Hartman, D.; et al. Glycan Motif Profiling Reveals Plasma Sialyl-Lewis X Elevations in Pancreatic Cancers That Are Negative for Sialyl-Lewis A *. *Molecular & Cellular Proteomics* **2015**, *14*, 1323–1333, doi:10.1074/mcp.M114.047837.
12. Nakamori, S.; Kameyama, M.; Imaoka, S.; Furukawa, H.; Ishikawa, O.; Sasaki, Y.; Kabuto, T.; Iwanaga, T.; Matsushita, Y.; Irimura, T. Increased Expression of Sialyl Lewisx Antigen Correlates with Poor Survival in Patients with Colorectal Carcinoma: Clinicopathological and Immunohistochemical Study1. *Cancer Research* **1993**, *53*, 3632–3637.
13. Wu, H. Enzymatic Synthesis of Cell-Surface Modifying Agents to Improve the Efficiency of Stem Cell Therapies. masters, Concordia University, 2021.
14. Laederach, A.; Reilly, P.J. Modeling Protein Recognition of Carbohydrates. *Proteins: Structure, Function, and Bioinformatics* **2005**, *60*, 591–597, doi:10.1002/prot.20545.
15. Mandal, D.K.; Kishore, N.; Brewer, C.F. Thermodynamics of Lectin-Carbohydrate Interactions. Titration Microcalorimetry Measurements of the Binding of N-Linked Carbohydrates and Ovalbumin to Concanavalin A. *Biochemistry* **1994**, *33*, 1149–1156, doi:10.1021/bi00171a014.
16. Cummings, R.D.; Schnaar, R.L.; Esko, J.D.; Driekamer, K.; Taylor, M.E. Principles of Glycan Recognition. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart,

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

- G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., Seeberger, P.H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2015.
17. Wu, Y.; Batyuk, A.; Honegger, A.; Brandl, F.; Mittl, P.R.E.; Plückthun, A. Rigidly Connected Multispecific Artificial Binders with Adjustable Geometries. *Sci Rep* **2017**, *7*, 11217, doi:10.1038/s41598-017-11472-x.
 18. Walser, M.; Rothenberger, S.; Hurdiss, D.L.; Schlegel, A.; Calabro, V.; Fontaine, S.; Villemagne, D.; Paladino, M.; Hospodarsch, T.; Neculcea, A.; et al. Highly Potent Anti-SARS-CoV-2 Multivalent DARPin Therapeutic Candidates 2021, 2020.08.25.256339.
 19. Woodford, C.R.; Thoden, J.B.; Holden, H.M. A New Role for the Ankyrin Repeat Revealed by the Study of the N-Formyltransferase from Providencia Alcalifaciens. *Biochemistry* **2015**, *54*, 631–638, doi:10.1021/bi501539a.
 20. Mangipudi, A. DESIGNING A PLATFORM FOR DEVELOPING NOVEL GLYCAN-BINDING PROTEINS USING YEAST SURFACE DISPLAY. **2021**, doi:10.7298/61gq-8948.
 21. Newton, M.S.; Cabezas-Perusse, Y.; Tong, C.L.; Seelig, B. In Vitro Selection of Peptides and Proteins—Advantages of mRNA Display. *ACS Synth. Biol.* **2020**, *9*, 181–190, doi:10.1021/acssynbio.9b00419.
 22. Cotten, S.W.; Zou, J.; Wang, R.; Huang, B.; Liu, R. mRNA Display-Based Selections Using Synthetic Peptide and Natural Protein Libraries. In *Ribosome Display and Related Technologies: Methods and Protocols*; Douthwaite, J.A., Jackson, R.H., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2012; pp. 287–297 ISBN 978-1-61779-379-0.
 23. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J Chem Phys* **2020**, *153*, 044130, doi:10.1063/5.0014475.
 24. Guvench, O.; Mallajosyula, S.S.; Raman, E.P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T.J.; Jamison, F.W.; MacKerell, A.D. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *J. Chem. Theory Comput.* **2011**, *7*, 3162–3180, doi:10.1021/ct200328p.
 25. Brady, J. Molecular Dynamics Simulations of Carbohydrate Molecules. In *Advances in Biophysical Chemistry*; Elsevier, 1996; Vol. 1, pp. 155–202 ISBN 978-1-55938-978-5.
 26. Kawatkar, S.P.; Kuntz, D.A.; Woods, R.J.; Rose, D.R.; Boons, G.-J. Structural Basis of the Inhibition of Golgi α -Mannosidase II by Mannostatin A and the Role of the Thiomethyl Moiety in Ligand–Protein Interactions. *J. Am. Chem. Soc.* **2006**, *128*, 8310–8319, doi:10.1021/ja061216p.
 27. Wen, X.; Yuan, Y.; Kuntz, D.A.; Rose, D.R.; Pinto, B.M. A Combined STD-NMR/Molecular Modeling Protocol for Predicting the Binding Modes of the Glycosidase Inhibitors Kifunensine and Salacinol to Golgi α -Mannosidase II. *Biochemistry* **2005**, *44*, 6729–6737, doi:10.1021/bi0500426.
 28. Bakan, A.; Nevins, N.; Lakdawala, A.S.; Bahar, I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2435–2447, doi:10.1021/ct300117j.
 29. Glycosciences, N.R.C. (US) C. on A. the I. and I. of G. and *Examples of Outstanding Questions in Glycoscience*; National Academies Press (US), 2012;
 30. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res* **2014**, *42*, D490-495, doi:10.1093/nar/gkt1178.
 31. Balzarini, J. Targeting the Glycans of Gp120: A Novel Approach Aimed at the Achilles Heel of HIV. *The Lancet Infectious Diseases* **2005**, *5*, 726–731, doi:10.1016/S1473-3099(05)70271-1.
 32. Akkouch, O.; Ng, T.B.; Singh, S.S.; Yin, C.; Dan, X.; Chan, Y.S.; Pan, W.; Cheung, R.C.F. Lectins with Anti-HIV Activity: A Review. *Molecules* **2015**, *20*, 648–668, doi:10.3390/molecules20010648.
 33. Jandú, J.J.B.; Moraes Neto, R.N.; Zagmignan, A.; de Sousa, E.M.; Brelaz-de-Castro, M.C.A.; dos Santos Correia, M.T.; da Silva, L.C.N. Targeting the Immune System with Plant Lectins to Combat Microbial Infections. *Front Pharmacol* **2017**, *8*, doi:10.3389/fphar.2017.00671.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

34. Wang, Z.; Park, K.; Comer, F.; Hsieh-Wilson, L.C.; Saudek, C.D.; Hart, G.W. Site-Specific GlcNAcylation of Human Erythrocyte Proteins. *Diabetes* **2009**, *58*, 309–317, doi:10.2337/db08-0994.
35. Ercan, A.; Cui, J.; Hazen, M.M.; Batliwalla, F.; Royle, L.; Rudd, P.M.; Coblyn, J.S.; Shadick, N.; Weinblatt, M.E.; Gregersen, P.; et al. Hypogalactosylation of Serum N-Glycans Fails to Predict Clinical Response to Methotrexate and TNF Inhibition in Rheumatoid Arthritis. *Arthritis Research & Therapy* **2012**, *14*, R43, doi:10.1186/ar3756.
36. Kuno, A.; Miyoshi, E.; Nakayama, J.; Ohyama, C.; Togayachi, A. Glycan Biomarkers for Cancer and Various Disease. In *Glycoscience: Basic Science to Applications: Insights from the Japan Consortium for Glycobiology and Glycotechnology (JCGG)*; Taniguchi, N., Endo, T., Hirabayashi, J., Nishihara, S., Kadomatsu, K., Akiyoshi, K., Aoki-Kinoshita, K.F., Eds.; Springer: Singapore, 2019; pp. 297–309 ISBN 9789811358562.
37. Velkov, V.V.; Medvinsky, A.B.; Sokolov, M.S.; Marchenko, A.I. Will Transgenic Plants Adversely Affect the Environment? *J. Biosci.* **2005**, *30*, 515–548, doi:10.1007/BF02703726.
38. Oliveira, C.; Carvalho, V.; Domingues, L.; Gama, F.M. Recombinant CBM-Fusion Technology — Applications Overview. *Biotechnology Advances* **2015**, *33*, 358–369, doi:10.1016/j.biotechadv.2015.02.006.
39. US6146428A - Enzymatic Treatment of Denim - Google Patents Available online: <https://patents.google.com/patent/US6146428A/en> (accessed on 3 November 2020).
40. Osten, C. von der; Bjornvad, M.E.; Vind, J.; Rasmussen, M.D. Process and Composition for Desizing Cellulosic Fabric with an Enzyme Hybrid 2000.
41. Dang, K.; Zhang, W.; Jiang, S.; Lin, X.; Qian, A. Application of Lectin Microarrays for Biomarker Discovery. *ChemistryOpen* **2020**, *9*, 285–300, doi:10.1002/open.201900326.
42. Barondes, S.H. Bifunctional Properties of Lectins: Lectins Redefined. *Trends in Biochemical Sciences* **1988**, *13*, 480–482, doi:10.1016/0968-0004(88)90235-6.
43. Kocourek, J.; Hořejší, V. Defining a Lectin. *Nature* **1981**, *290*, 188–188, doi:10.1038/290188a0.
44. Feng, Y.; Guo, Y.; Li, Y.; Tao, J.; Ding, L.; Wu, J.; Ju, H. Lectin-Mediated in Situ Rolling Circle Amplification on Exosomes for Probing Cancer-Related Glycan Pattern. *Anal Chim Acta* **2018**, *1039*, 108–115, doi:10.1016/j.aca.2018.07.040.
45. Hashim, O.H.; Jayapalan, J.J.; Lee, C.-S. Lectins: An Effective Tool for Screening of Potential Cancer Biomarkers. *PeerJ* **2017**, *5*, doi:10.7717/peerj.3784.
46. *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., Seeberger, P.H., Eds.; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2015;
47. Bertozzi, C.R.; Kiessling, and L.L. Chemical Glycobiology. *Science* **2001**, *291*, 2357–2364, doi:10.1126/science.1059820.
48. Liener, I.E.; Sharon, N.; Goldstein, I.J. *The Lectins: Properties, Functions, and Applications in Biology and Medicine*; Academic Press, 1986;
49. Hu, D.; Tateno, H.; Hirabayashi, J. Lectin Engineering, a Molecular Evolutionary Approach to Expanding the Lectin Utilities. *Molecules* **2015**, *20*, 7637–7656, doi:10.3390/molecules20057637.
50. Bonnardel, F.; Mariethoz, J.; Salentin, S.; Robin, X.; Schroeder, M.; Perez, S.; Lisacek, F.; Imberty, A. UniLectin3D, a Database of Carbohydrate Binding Proteins with Curated Information on 3D Structures and Interacting Ligands. *Nucleic Acids Res* **2019**, *47*, D1236–D1244, doi:10.1093/nar/gky832.
51. Armenta, S.; Moreno-Mendieta, S.; Sánchez-Cuapio, Z.; Sánchez, S.; Rodríguez-Sanoja, R. Advances in Molecular Engineering of Carbohydrate-Binding Modules. *Proteins* **2017**, *85*, 1602–1617, doi:10.1002/prot.25327.
52. Guillén, D.; Sánchez, S.; Rodríguez-Sanoja, R. Carbohydrate-Binding Domains: Multiplicity of Biological Roles. *Appl Microbiol Biotechnol* **2010**, *85*, 1241–1249, doi:10.1007/s00253-009-2331-y.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

53. Simpson, H.D.; Barras, F. Functional Analysis of the Carbohydrate-Binding Domains of *Erwinia Chrysanthemi* Cel5 (Endoglucanase Z) and an *Escherichia Coli* Putative Chitinase. *Journal of Bacteriology* **1999**, *181*, 4611–4616, doi:10.1128/JB.181.15.4611-4616.1999.
54. Simpson, P.J.; Jamieson, S.J.; Abou-Hachem, M.; Karlsson, E.N.; Gilbert, H.J.; Holst, O.; Williamson, M.P. The Solution Structure of the CBM4-2 Carbohydrate Binding Module from a Thermostable *Rhodothermus Marinus* Xylanase. *Biochemistry* **2002**, *41*, 5712–5719, doi:10.1021/bi012093i.
55. Boraston, A.B.; Creagh, A.L.; Alam, M.M.; Kormos, J.M.; Tomme, P.; Haynes, C.A.; Warren, R.A.; Kilburn, D.G. Binding Specificity and Thermodynamics of a Family 9 Carbohydrate-Binding Module from *Thermotoga Maritima* Xylanase 10A. *Biochemistry* **2001**, *40*, 6240–6247, doi:10.1021/bi0101695.
56. Furtado, G.P.; Lourenzoni, M.R.; Fuzo, C.A.; Fonseca-Maldonado, R.; Guazzaroni, M.-E.; Ribeiro, L.F.; Ward, R.J. Engineering the Affinity of a Family 11 Carbohydrate Binding Module to Improve Binding of Branched over Unbranched Polysaccharides. *International Journal of Biological Macromolecules* **2018**, *120*, 2509–2516, doi:10.1016/j.ijbiomac.2018.09.022.
57. Cicortas Gunnarsson, L.; Nordberg Karlsson, E.; Albrekt, A.-S.; Andersson, M.; Holst, O.; Ohlin, M. A Carbohydrate Binding Module as a Diversity-carrying Scaffold. *Protein Eng Des Sel* **2004**, *17*, 213–221, doi:10.1093/protein/gzh026.
58. Sakata, T.; Takakura, J.; Miyakubo, H.; Osada, Y.; Wada, R.; Takahashi, H.; Yatsunami, R.; Fukui, T.; Nakamura, S. Improvement of Binding Activity of Xylan-Binding Domain by Amino Acid Substitution. *Nucleic Acids Symp Ser (Oxf)* **2006**, *50*, 253–254, doi:10.1093/nass/nrl126.
59. Eyers, P.A.; Murphy, J.M. The Evolving World of Pseudoenzymes: Proteins, Prejudice and Zombies. *BMC Biology* **2016**, *14*, 98, doi:10.1186/s12915-016-0322-x.
60. Schimpl, M.; Rush, C.L.; Betou, M.; Eggleston, I.M.; Recklies, A.D.; van Aalten, D.M.F. Human YKL-39 Is a Pseudo-Chitinase with Retained Chitooligosaccharide-Binding Properties. *Biochem J* **2012**, *446*, 149–157, doi:10.1042/BJ20120377.
61. Lee, C.G.; Da Silva, C.A.; Dela Cruz, C.S.; Ahangari, F.; Ma, B.; Kang, M.-J.; He, C.-H.; Takyar, S.; Elias, J.A. Role of Chitin and Chitinase/Chitinase-Like Proteins in Inflammation, Tissue Remodeling, and Injury. *Annu Rev Physiol* **2011**, *73*, doi:10.1146/annurev-physiol-012110-142250.
62. Lee, S.; Choi, J.; Mohanty, J.; Sousa, L.P.; Tome, F.; Pardon, E.; Steyaert, J.; Lemmon, M.A.; Lax, I.; Schlessinger, J. Structures of β -Klotho Reveal a ‘Zip Code’-like Mechanism for Endocrine FGF Signalling. *Nature* **2018**, *553*, 501–505, doi:10.1038/nature25010.
63. Chen, L.; Xu, Y.; Wong, W.; Thompson, J.K.; Healer, J.; Goddard-Borger, E.D.; Lawrence, M.C.; Cowman, A.F. Structural Basis for Inhibition of Erythrocyte Invasion by Antibodies to Plasmodium Falciparum Protein CyRPA. *eLife* **2017**, *6*, e21347, doi:10.7554/eLife.21347.
64. Favuzza, P.; Guffart, E.; Tamborrini, M.; Scherer, B.; Dreyer, A.M.; Rufer, A.C.; Erny, J.; Hoernschemeyer, J.; Thoma, R.; Schmid, G.; et al. Structure of the Malaria Vaccine Candidate Antigen CyRPA and Its Complex with a Parasite Invasion Inhibitory Antibody. *eLife* **2017**, *6*, e20383, doi:10.7554/eLife.20383.
65. Wong, W.; Huang, R.; Menant, S.; Hong, C.; Sandow, J.J.; Birkinshaw, R.W.; Healer, J.; Hodder, A.N.; Kanjee, U.; Tonkin, C.J.; et al. Structure of Plasmodium Falciparum Rh5–CyRPA–Ripr Invasion Complex. *Nature* **2019**, *565*, 118–121, doi:10.1038/s41586-018-0779-6.
66. Little, D.J.; Li, G.; Ing, C.; DiFrancesco, B.R.; Bamford, N.C.; Robinson, H.; Nitz, M.; Pomès, R.; Howell, P.L. Modification and Periplasmic Translocation of the Biofilm Exopolysaccharide Poly- β -1,6-N-Acetyl-D-Glucosamine. *Proc Natl Acad Sci U S A* **2014**, *111*, 11013–11018, doi:10.1073/pnas.1406388111.
67. Stummeyer, K.; Dickmanns, A.; Mühlenhoff, M.; Gerardy-Schahn, R.; Ficner, R. Crystal Structure of the Polysialic Acid-Degrading Endosialidase of Bacteriophage K1F. *Nature Structural & Molecular Biology* **2005**, *12*, 90–96, doi:10.1038/nsmb874.
68. Jakobsson, E.; Jokilampi, A.; Aalto, J.; Ollikka, P.; Lehtonen, J.V.; Hirvonen, H.; Finne, J. Identification of Amino Acid Residues at the Active Site of Endosialidase That Dissociate the

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

- Polysialic Acid Binding and Cleaving Activities in Escherichia Coli K1 Bacteriophages. *Biochem J* **2007**, *405*, 465–472, doi:10.1042/BJ20070177.
69. Yu, C.-C.; Hill, T.; Kwan, D.H.; Chen, H.-M.; Lin, C.-C.; Wakarchuk, W.; Withers, S.G. A Plate-Based High-Throughput Activity Assay for Polysialyltransferase from *Neisseria Meningitidis*. *Anal. Biochem.* **2014**, *444*, 67–74, doi:10.1016/j.ab.2013.09.030.
 70. Yu, C.-C.; Huang, L.-D.; Kwan, D.H.; Wakarchuk, W.W.; Withers, S.G.; Lin, C.-C. A Glyco-Gold Nanoparticle Based Assay for α -2,8-Polysialyltransferase from *Neisseria Meningitidis*. *Chem. Commun.* **2013**, *49*, 10166–10168, doi:10.1039/C3CC45147J.
 71. Montanier, C.; Money, V.A.; Pires, V.M.R.; Flint, J.E.; Pinheiro, B.A.; Goyal, A.; Prates, J.A.M.; Izumi, A.; Stålbrand, H.; Morland, C.; et al. The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. *PLOS Biology* **2009**, *7*, e1000071, doi:10.1371/journal.pbio.1000071.
 72. Woods, R.J.; Yang, L. Glycan-Specific Analytical Tools 2018.
 73. Arbabi-Ghahroudi, M. Camelid Single-Domain Antibodies: Historical Perspective and Future Outlook. *Front Immunol* **2017**, *8*, doi:10.3389/fimmu.2017.01589.
 74. Nelson, A.L. Antibody Fragments. *MAbs* **2010**, *2*, 77–83.
 75. Ahmad, Z.A.; Yeap, S.K.; Ali, A.M.; Ho, W.Y.; Alitheen, N.B.M.; Hamid, M. ScFv Antibody: Principles and Clinical Application Available online: <https://www.hindawi.com/journals/jir/2012/980250/> (accessed on 2 January 2021).
 76. Holliger, P.; Prospero, T.; Winter, G. “Diabodies”: Small Bivalent and Bispecific Antibody Fragments. *PNAS* **1993**, *90*, 6444–6448, doi:10.1073/pnas.90.14.6444.
 77. Sha, F.; Salzman, G.; Gupta, A.; Koide, S. Monobodies and Other Synthetic Binding Proteins for Expanding Protein Science. *Protein Science* **2017**, *26*, 910–924, doi:https://doi.org/10.1002/pro.3148.
 78. Sterner, E.; Flanagan, N.; Gildersleeve, J.C. Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database. *ACS Chem Biol* **2016**, *11*, 1773–1783, doi:10.1021/acscchembio.6b00244.
 79. Stewart, A.; Liu, Y.; Lai, J.R. A Strategy for Phage Display Selection of Functional Domain-Exchanged Immunoglobulin Scaffolds with High Affinity for Glycan Targets. *J Immunol Methods* **2012**, *376*, 150–155, doi:10.1016/j.jim.2011.12.008.
 80. McCullum, E.O.; Williams, B.A.R.; Zhang, J.; Chaput, J.C. Random Mutagenesis by Error-Prone PCR. In *In Vitro Mutagenesis Protocols: Third Edition*; Braman, J., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; pp. 103–109 ISBN 978-1-60761-652-8.
 81. Coco, W.M.; Levinson, W.E.; Crist, M.J.; Hektor, H.J.; Darzins, A.; Pienkos, P.T.; Squires, C.H.; Monticello, D.J. DNA Shuffling Method for Generating Highly Recombined Genes and Evolved Enzymes. *Nature Biotechnology* **2001**, *19*, 354–359, doi:10.1038/86744.
 82. Muteeb, G.; Sen, R. Random Mutagenesis Using a Mutator Strain. In *In Vitro Mutagenesis Protocols: Third Edition*; Braman, J., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; pp. 411–419 ISBN 978-1-60761-652-8.
 83. Poluri, K.M.; Gulati, K. *Protein Engineering Techniques: Gateways to Synthetic Protein Universe*; SpringerBriefs in Forensic and Medical Bioinformatics; Springer Singapore, 2017; ISBN 978-981-10-2731-4.
 84. Yabe, R.; Suzuki, R.; Kuno, A.; Fujimoto, Z.; Jigami, Y.; Hirabayashi, J. Tailoring a Novel Sialic Acid-Binding Lectin from a Ricin-B Chain-like Galactose-Binding Protein by Natural Evolution-Mimicry. *J Biochem* **2007**, *141*, 389–399, doi:10.1093/jb/mvm043.
 85. Stemmer, W.P.C. Rapid Evolution of a Protein in Vitro by DNA Shuffling. *Nature* **1994**, *370*, 389–391, doi:10.1038/370389a0.
 86. Melzer, S.; Sonnendecker, C.; Föllner, C.; Zimmermann, W. Stepwise Error-Prone PCR and DNA Shuffling Changed the PH Activity Range and Product Specificity of the Cyclodextrin Glucanotransferase from an Alkaliphilic *Bacillus* Sp. *FEBS Open Bio* **2015**, *5*, 528–534, doi:10.1016/j.fob.2015.06.002.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

87. Ihssen, J.; Haas, J.; Kowarik, M.; Wiesli, L.; Wacker, M.; Schwede, T.; Thöny-Meyer, L. Increased Efficiency of *Campylobacter* Jejuni N-Oligosaccharyltransferase PglB by Structure-Guided Engineering. *Open Biol* **2015**, *5*, doi:10.1098/rsob.140227.
88. Mendonça, L.M.F.; Marana, S.R. Single Mutations Outside the Active Site Affect the Substrate Specificity in a β -Glycosidase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011**, *1814*, 1616–1623, doi:10.1016/j.bbapap.2011.08.012.
89. Adam, J.; Pokorná, M.; Sabin, C.; Mitchell, E.P.; Imberty, A.; Wimmerová, M. Engineering of PA-IIL Lectin from *Pseudomonas Aeruginosa* – Unravelling the Role of the Specificity Loop for Sugar Preference. *BMC Struct Biol* **2007**, *7*, 36, doi:10.1186/1472-6807-7-36.
90. Kunstmann, S.; Engström, O.; Wehle, M.; Widmalm, G.; Santer, M.; Barbirz, S. Increasing the Affinity of an O-Antigen Polysaccharide Binding Site in *Shigella Flexneri* Bacteriophage Sf6 Tailspike Protein. *Chemistry – A European Journal* **2020**, *26*, 7263–7273, doi:10.1002/chem.202000495.
91. Allhorn, M.; Olin, A.I.; Nimmerjahn, F.; Collin, M. Human IgG/Fc γ R Interactions Are Modulated by Streptococcal IgG Glycan Hydrolysis. *PLoS One* **2008**, *3*, doi:10.1371/journal.pone.0001413.
92. O'Donohue, M.J.; GeoffKneale, G. Site-Directed and Site-Saturation Mutagenesis Using Oligonucleotide Primers. In *DNA-Protein Interactions: Principles and Protocols*; Geoff Kneale, G., Ed.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 1994; pp. 211–225 ISBN 978-1-59259-517-4.
93. Hutchison, C.A.; Phillips, S.; Edgell, M.H.; Gillam, S.; Jahnke, P.; Smith, M. Mutagenesis at a Specific Position in a DNA Sequence. *J. Biol. Chem.* **1978**, *253*, 6551–6560.
94. Imamura, K.; Takeuchi, H.; Yabe, R.; Tateno, H.; Hirabayashi, J. Engineering of the Glycan-Binding Specificity of *Agrocybe Cylindracea* Galectin towards α (2,3)-Linked Sialic Acid by Saturation Mutagenesis. *J. Biochem.* **2011**, *150*, 545–552, doi:10.1093/jb/mvr094.
95. Yamamoto, K.; Konami, Y.; Osawa, T. A Chimeric Lectin Formed from *Bauhinia Purpurea* Lectin and *Lens Culinaris* Lectin Recognizes a Unique Carbohydrate Structure. *J. Biochem.* **2000**, *127*, 129–135, doi:10.1093/oxfordjournals.jbchem.a022573.
96. Lienemann, M.; Boer, H.; Paananen, A.; Cottaz, S.; Koivula, A. Toward Understanding of Carbohydrate Binding and Substrate Specificity of a Glycosyl Hydrolase 18 Family (GH-18) Chitinase from *Trichoderma Harzianum*. *Glycobiology* **2009**, *19*, 1116–1126, doi:10.1093/glycob/cwp102.
97. Seifert, A.; Pleiss, J. Identification of Selectivity-Determining Residues in Cytochrome P450 Monooxygenases: A Systematic Analysis of the Substrate Recognition Site 5. *Proteins: Structure, Function, and Bioinformatics* **2009**, *74*, 1028–1035, doi:10.1002/prot.22242.
98. Cheriyan, M.; Toone, E.J.; Fierke, C.A. Mutagenesis of the Phosphate-Binding Pocket of KDPG Aldolase Enhances Selectivity for Hydrophobic Substrates. *Protein Science* **2007**, *16*, 2368–2377, doi:10.1110/ps.073042907.
99. Schneider, S.; Gutiérrez, M.; Sandalova, T.; Schneider, G.; Clapés, P.; Sprenger, G.A.; Samland, A.K. Redesigning the Active Site of Transaldolase TalB from *Escherichia Coli*: New Variants with Improved Affinity towards Nonphosphorylated Substrates. *ChemBioChem* **2010**, *11*, 681–690, doi:10.1002/cbic.200900720.
100. Lo, W.-C.; Wang, L.-F.; Liu, Y.-Y.; Dai, T.; Hwang, J.-K.; Lyu, P.-C. CPred: A Web Server for Predicting Viable Circular Permutations in Proteins. *Nucleic Acids Res* **2012**, *40*, W232–W237, doi:10.1093/nar/gks529.
101. Stephen, P.; Tseng, K.-L.; Liu, Y.-N.; Lyu, P.-C. Circular Permutation of the Starch-Binding Domain: Inversion of Ligand Selectivity with Increased Affinity. *Chem. Commun. (Camb.)* **2012**, *48*, 2612–2614, doi:10.1039/c2cc17376j.
102. Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 681–697, doi:10.1038/s41580-019-0163-x.
103. Frei, J.C.; Lai, J.R. Protein and Antibody Engineering by Phage Display. *Methods Enzymol* **2016**, *580*, 45–87, doi:10.1016/bs.mie.2016.05.005.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

104. Peltomaa, R.; Benito-Peña, E.; Barderas, R.; Moreno-Bondi, M.C. Phage Display in the Quest for New Selective Recognition Elements for Biosensors. *ACS Omega* **2019**, *4*, 11569–11580, doi:10.1021/acsomega.9b01206.
105. Chasteen, L.; Ayriess, J.; Pavlik, P.; Bradbury, A.R.M. Eliminating Helper Phage from Phage Display. *Nucleic Acids Res* **2006**, *34*, e145, doi:10.1093/nar/gkl772.
106. Yuasa, N.; Koyama, T.; Fujita-Yamaguchi, Y. Purification and Refolding of Anti-T-Antigen Single Chain Antibodies (ScFvs) Expressed in *Escherichia Coli* as Inclusion Bodies. *BioScience Trends* **2014**, *8*, 24–31, doi:10.5582/bst.8.24.
107. Ng, S.; Tjhung, K.F.; Paschal, B.M.; Noren, C.J.; Derda, R. Chemical Posttranslational Modification of Phage-Displayed Peptides. In *Peptide Libraries: Methods and Protocols*; Derda, R., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2015; pp. 155–172 ISBN 978-1-4939-2020-4.
108. Boder, E.T.; Wittrup, K.D. Yeast Surface Display for Directed Evolution of Protein Expression, Affinity, and Stability. *Meth. Enzymol.* **2000**, *328*, 430–444, doi:10.1016/s0076-6879(00)28410-3.
109. Hong, X.; Ma, M.Z.; Gildersleeve, J.C.; Chowdhury, S.; Barchi, J.J.; Mariuzza, R.A.; Murphy, M.B.; Mao, L.; Pancer, Z. Sugar-Binding Proteins from Fish: Selection of High Affinity “Lambodies” That Recognize Biomedically Relevant Glycans. *ACS Chem. Biol.* **2013**, *8*, 152–160, doi:10.1021/cb300399s.
110. Lu, Z.; Kamat, K.; Johnson, B.P.; Yin, C.C.; Scholler, N.; Abbott, K.L. Generation of a Fully Human ScFv That Binds Tumor-Specific Glycoforms. *Scientific Reports* **2019**, *9*, 5101, doi:10.1038/s41598-019-41567-6.
111. Hanes, J.; Plückthun, A. In Vitro Selection and Evolution of Functional Proteins by Using Ribosome Display. *PNAS* **1997**, *94*, 4937–4942, doi:10.1073/pnas.94.10.4937.
112. Kurz, M.; Gu, K.; Lohse, P.A. Psoralen Photo-Crosslinked mRNA–Puromycin Conjugates: A Novel Template for the Rapid and Facile Preparation of mRNA–Protein Fusions. *Nucleic Acids Res* **2000**, *28*, e83.
113. Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. In Vitro Virus: Bonding of mRNA Bearing Puromycin at the 3'-Terminal End to the C-Terminal End of Its Encoded Protein on the Ribosome in Vitro. *FEBS Lett* **1997**, *414*, 405–408, doi:10.1016/s0014-5793(97)01026-0.
114. Niwa, T.; Kanamori, T.; Ueda, T.; Taguchi, H. Global Analysis of Chaperone Effects Using a Reconstituted Cell-Free Translation System. *Proc Natl Acad Sci U S A* **2012**, *109*, 8937–8942, doi:10.1073/pnas.1201380109.
115. Lieberoth, A.; Splittstoesser, F.; Katagihallimath, N.; Jakovcevski, I.; Loers, G.; Ranscht, B.; Karagogeos, D.; Schachner, M.; Kleene, R. Lewisx and A2,3-Sialyl Glycans and Their Receptors TAG-1, Contactin, and L1 Mediate CD24-Dependent Neurite Outgrowth. *J. Neurosci.* **2009**, *29*, 6677–6690, doi:10.1523/JNEUROSCI.4361-08.2009.
116. Sprung, R.; Nandi, A.; Chen, Y.; Kim, S.C.; Barma, D.; Falck, J.R.; Zhao, Y. Tagging-via-Substrate Strategy for Probing O-GlcNAc Modified Proteins. *J Proteome Res* **2005**, *4*, 950–957, doi:10.1021/pr050033j.
117. Mahal, L.K.; Yarema, K.J.; Bertozzi, C.R. Engineering Chemical Reactivity on Cell Surfaces through Oligosaccharide Biosynthesis. *Science* **1997**, *276*, 1125–1128, doi:10.1126/science.276.5315.1125.
118. Luchansky, S.J.; Argade, S.; Hayes, B.K.; Bertozzi, C.R. Metabolic Functionalization of Recombinant Glycoproteins. *Biochemistry* **2004**, *43*, 12358–12366, doi:10.1021/bi049274f.
119. Bryan, M.C.; Lee, L.V.; Wong, C.-H. High-Throughput Identification of Fucosyltransferase Inhibitors Using Carbohydrate Microarrays. *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 3185–3188, doi:10.1016/j.bmcl.2004.04.001.
120. Winzor, D.J. Determination of Binding Constants by Affinity Chromatography. *Journal of Chromatography A* **2004**, *1037*, 351–367, doi:10.1016/j.chroma.2003.11.092.
121. Kasai, K. Frontal Affinity Chromatography: A Unique Research Tool for Biospecific Interaction That Promotes Glycobiology. *Proc Jpn Acad Ser B Phys Biol Sci* **2014**, *90*, 215–234, doi:10.2183/pjab.90.215.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

122. Schasfoort, R.B.M. Chapter 1: Introduction to Surface Plasmon Resonance. In *Handbook of Surface Plasmon Resonance*; 2017; pp. 1–26.
123. Gutiérrez Gallego, R.; Haseley, S.R.; van Miegem, V.F.L.; Vliegthart, J.F.G.; Kamerling, J.P. Identification of Carbohydrates Binding to Lectins by Using Surface Plasmon Resonance in Combination with HPLC Profiling. *Glycobiology* **2004**, *14*, 373–386, doi:10.1093/glycob/cwh052.
124. Bellapadrona, G.; Tesler, A.B.; Grünstein, D.; Hossain, L.H.; Kikkeri, R.; Seeberger, P.H.; Vaskevich, A.; Rubinstein, I. Optimization of Localized Surface Plasmon Resonance Transducers for Studying Carbohydrate–Protein Interactions. *Anal. Chem.* **2012**, *84*, 232–240, doi:10.1021/ac202363t.
125. Huang, C.-F.; Yao, G.-H.; Liang, R.-P.; Qiu, J.-D. Graphene Oxide and Dextran Capped Gold Nanoparticles Based Surface Plasmon Resonance Sensor for Sensitive Detection of Concanavalin A. *Biosens Bioelectron* **2013**, *50*, 305–310, doi:10.1016/j.bios.2013.07.002.
126. *Handbook of Surface Plasmon Resonance*; 2017; ISBN 978-1-78262-730-2.
127. Gedig, E.T. Chapter 6: Surface Chemistry in SPR Technology. In *Handbook of Surface Plasmon Resonance*; 2017; pp. 171–254.
128. Liu, L.; Prudden, A.R.; Capicciotti, C.J.; Bosman, G.P.; Yang, J.-Y.; Chapla, D.G.; Moremen, K.W.; Boons, G.-J. Streamlining the Chemoenzymatic Synthesis of Complex N -Glycans by a Stop and Go Strategy. *Nature Chemistry* **2019**, *11*, 161–169, doi:10.1038/s41557-018-0188-3.
129. Palcic, M.M. Glycosyltransferases as Biocatalysts. *Current Opinion in Chemical Biology* **2011**, *15*, 226–233, doi:10.1016/j.cbpa.2010.11.022.
130. Shilova, O.N.; Deyev, S.M. DARPins: Promising Scaffolds for Theranostics. *Acta Naturae* **2019**, *11*, 42–53, doi:10.32607/20758251-2019-11-4-42-53.
131. Hofmeister, D.L.; Thoden, J.B.; Holden, H.M. Investigation of a Sugar N-formyltransferase from the Plant Pathogen *Pantoea Ananatis*. *Protein Sci* **2019**, *28*, 707–716, doi:10.1002/pro.3577.
132. Kramer, M.A.; Wetzel, S.K.; Plückthun, A.; Mittl, P.R.E.; Grütter, M.G. Structural Determinants for Improved Stability of Designed Ankyrin Repeat Proteins with a Redesign C-Capping Module. *Journal of Molecular Biology* **2010**, *404*, 381–391, doi:10.1016/j.jmb.2010.09.023.
133. Schilling, J.; Schöppe, J.; Plückthun, A. From DARPins to LoopDARPins: Novel LoopDARPin Design Allows the Selection of Low Picomolar Binders in a Single Round of Ribosome Display. *J Mol Biol* **2014**, *426*, 691–721, doi:10.1016/j.jmb.2013.10.026.
134. Li, Z.Z.; Riegert, A.S.; Goneau, M.-F.; Cunningham, A.M.; Vinogradov, E.; Li, J.; Schoenhofen, I.C.; Thoden, J.B.; Holden, H.M.; Gilbert, M. Characterization of the DTDP-Fuc3N and DTDP-Qui3N Biosynthetic Pathways in *Campylobacter Jejuni* 81116†. *Glycobiology* **2017**, *27*, 358–369, doi:10.1093/glycob/cww136.
135. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876, doi:10.1126/science.abj8754.
136. Calcul Quebec Available online: <https://www.calculquebec.ca/en/> (accessed on 11 June 2022).
137. Homepage | Digital Research Alliance of Canada Available online: <https://alliancecan.ca/en> (accessed on 11 June 2022).
138. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33–38, 27–28, doi:10.1016/0263-7855(96)00018-5.
139. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865, doi:10.1002/jcc.20945.
140. Kim, S.; Lee, J.; Jo, S.; Brooks III, C.L.; Lee, H.S.; Im, W. CHARMM-GUI Ligand Reader and Modeler for CHARMM Force Field Generation of Small Molecules. *Journal of Computational Chemistry* **2017**, *38*, 1879–1886, doi:10.1002/jcc.24829.
141. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616, doi:10.1021/jp973084f.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

142. Brooks, B.R.; Brooks III, C.L.; Mackerell Jr., A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614, doi:10.1002/jcc.21287.
143. Warkentin, R.; Aaram, M. Steered Molecular Dynamics Simulations for Probing Carbohydrate Interactions 2022.
144. Gowers, R.J.; Linke, M.; Barnoud, J.; Reddy, T.J.E.; Melo, M.N.; Seyler, S.L.; Domański, J.; Dotson, D.L.; Buchoux, S.; Kenney, I.M.; et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105, doi:10.25080/Majora-629e541a-00e.
145. Michaud-Agrawal, N.; Denning, E.J.; Woolf, T.B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327, doi:10.1002/jcc.21787.
146. Kolb, H.C.; Finn, M.G.; Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **2001**, *40*, 2004–2021, doi:10.1002/1521-3773(20010601)40:11<2004::aid-anie2004>3.3.co;2-x.
147. Barendt, P.A.; Ng, D.T.W.; McQuade, C.N.; Sarkar, C.A. Streamlined Protocol for mRNA Display. *ACS Comb Sci* **2013**, *15*, 77–81, doi:10.1021/co300135r.
148. Ghosh, T. Artificial Genetically Encoded Peptides and Proteins as Next-Generation Therapeutics: Selection of Ligands for Mycobacterium Tuberculosis UDP-Galactopyranose Mutase as Potential Inhibitors. masters, Concordia University, 2020.
149. Mai, B.K.; Li, M.S. Neuraminidase Inhibitor R-125489--a Promising Drug for Treating Influenza Virus: Steered Molecular Dynamics Approach. *Biochem Biophys Res Commun* **2011**, *410*, 688–691, doi:10.1016/j.bbrc.2011.06.057.
150. Mark, B.L.; Wasney, G.A.; Salo, T.J.S.; Khan, A.R.; Cao, Z.; Robbins, P.W.; James, M.N.G.; Triggs-Raine, B.L. Structural and Functional Characterization of *Streptomyces Plicatus* β -N-Acetylhexosaminidase by Comparative Molecular Modeling and Site-Directed Mutagenesis. *Journal of Biological Chemistry* **1998**, *273*, 19618–19624, doi:10.1074/jbc.273.31.19618.
151. Vian, A.; Carrascosa, A.V.; García, J.L.; Cortés, E. Structure of the β -Galactosidase Gene from *Thermus Sp. Strain T2*: Expression in *Escherichia Coli* and Purification in a Single Step of an Active Fusion Protein. *Appl Environ Microbiol* **1998**, *64*, 2187–2191.
152. Sletten, E.M.; Bertozzi, C.R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew Chem Int Ed Engl* **2009**, *48*, 6974–6998, doi:10.1002/anie.200900942.
153. Olausson, J.; Tibell, L.; Jonsson, B.-H.; Pålsson, P. Detection of a High Affinity Binding Site in Recombinant *Aleuria Aurantia* Lectin. *Glycoconj J* **2008**, *25*, 753, doi:10.1007/s10719-008-9135-7.
154. Deng, H.; Zhadin, N.; Callender, R. Dynamics of Protein Ligand Binding on Multiple Time Scales: NADH Binding to Lactate Dehydrogenase. *Biochemistry* **2001**, *40*, 3767–3773, doi:10.1021/bi0026268.
155. Scheib, U.; Shanmugaratnam, S.; Fariás-Rico, J.A.; Höcker, B. Change in Protein-Ligand Specificity through Binding Pocket Grafting. *Journal of Structural Biology* **2014**, *185*, 186–192, doi:10.1016/j.jsb.2013.06.002.
156. Schilling, J.; Jost, C.; Ilie, I.M.; Schnabl, J.; Buechi, O.; Eapen, R.S.; Truffer, R.; Caflisch, A.; Forrer, P. Thermostable Designed Ankyrin Repeat Proteins (DARPin) as Building Blocks for Innovative Drugs. *Journal of Biological Chemistry* **2022**, *298*, doi:10.1016/j.jbc.2021.101403.

Chapter 6: Conclusion and Future Directions**5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand****APPENDIX****STable 6.1: Oligos used in the 1st PCR for the DARPin library assembly.**

Target	Fwd oligo	Rev oligo	T _A (°C)
DARPin_N_T7_RBS_BbsI	DARPin_N_ultra_fwd	DARPin_N_primer.rev	50 °C
DARPin_I1_BbsI	DARPin_I(1-2)_fwd	DARPin_I(1-3)_rev	58 °C
DARPin_I2_BbsI	DARPin_I(1-2)_fwd	DARPin_I(1-3)_rev	58 °C
DARPin_I3_BbsI	DARPin_I3_Fwd	DARPin_I(1-3)_rev	58 °C
DARPin_C_BbsI	DARPin_C_ultra_fwd	Rev_puromycin	54 °C

STable 6.2: Primers used in the 2nd PCR for the DARPin library assembly.

Target	Fwd primer	Rev primer	T _A (°C)
DARPin_N_T7_RBS_BbsI	T7_long_Fwd	DARPin_N_primer.rev	50 °C
DARPin_I1_BbsI	DARPin_I1_primer_fwd	DARPin_I1_primer_rev	50 °C
DARPin_I2_BbsI	DARPin_I2_primer_fwd	DARPin_I2_primer_rev	50 °C
DARPin_I3_BbsI	DARPin_I3_primer_Fwd	DARPin_I3_primer_rev	50 °C
DARPin_C_BbsI	DARPin_C_primer_Fwd	Rev_puromycin	50 °C

STable 6.3: All oligos and primers used in the project.

Name	Sequence
Affi-NNK13-F129	GTA GAT AAC AAA TTC AAC AAA GAA NNK NNK NNK GCG NNK NNK GAG ATC NNK NNK TTA CCT AAC TTA AAC NNK NNK CAA NNK NNK GCC TTC ATC NNK AGT TTA NNK GAT GAC CCA AGC CAA AGC GCT AAC
Affi-R105	TTT CCG CCC CCC GTC CTA AGA CCC AGA CCC AGA CCC TTT TGG TGC TTG AGC ATC ATT TAG CTT TTT AGC TTC TGC TAA AAG GTT AGC GCT TTG GCT TGG GTC ATC

Chapter 6: Conclusion and Future Directions**5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand**

Affi-T7-F73	TAA TAC GAC TCA CTA TAG GGT TGA ACT TTA AGT AGG AGA TAT ATC CAT GGT AGA TAA CAA ATT CAA CAA AGA A
CGS3-R36	TTT CCG CCC CCC GTC CTA AGA CCC AGA CCC AGA CCC
T7_long_Fwd	GAT CGT TAA TAC GAC TCA CTA TAG GG
DARPin_N_ultra_Fwd	TAA TAC GAC TCA CTA TAG GGT TGA ACT TTA AGT AGG AGA TAT ATC CAT GCG CGG TAG TCA TCA CCA CCA TCA CCA TGG CAG CGA CTT GGG GAA AAA ACT GCT GGA AGC CGC ANN KNN KGG TCA AGA TGA TGA GGT GCG CAT TTT AAT GGC TAA CGG GTC TTC TAG
DARPin_N_primer_Rev	CTA GAA GAC CCG TTA GCC ATT AAA ATG
DARPin_I(1-2)_Fwd	GGA GCA GAT GTT AAC GCT NNK GAT NNK NNK GGT NNK ACC CCG CTT CAT TTG GCG GCA
DARPin_I(1-3)_Rev	TTT CAG CAG CAC CTC GAC AAT CTC CAG GTG GCC NNK NNK TGC CGC CAA ATG AAG CGG GGT
DARPin_I1_primer_Fwd	CGT TAT CTA GAA GAC GCT AAC GGA GCA GAT GTT AAC GCT
DARPin_I1_primer_Rev	CTA GAA GAC CCT GCT CCC GTT TTC AGC AGC ACC TCG AC
DARPin_I2_primer_Fwd	GCT ACT AGA AGA CGG AGC AGA TGT TAA CGC T
DARPin_I2_primer_Rev	CTA GAA GAC TCC GTG TTT CAG CAG CAC CTC GAC
DARPin_I3_Fwd	GGA GCA GAT GTT AAC GCT NNK GCT NNK NNK GGT NNK ACC CCG CTT CAT TTG GCG GCA

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

DARPin_I3_primer_Fwd	GCT TCA CTA GAA GAC AAC ACG GAG CAG ATG TTA ACG CT
DARPin_I3_primer_Rev	CTA GAA GAC CAA CAT CTG CTC CGT GTT TCA GCA GCA CCT CGA C
DARPin_C_primer_Fwd	GCA GTA GAA GAC TGA TGT TAA TGC GCA
DARPin_C_ultra_Fwd	GTA GAA GAC TGA TGT TAA TGC GCA AGA TNN KNN KGG TNN KAC CCC GTT TGA TCT GGC TAT CNN KNN KGG CAA CGA GGA CAT TGC AGA GGT GTT ACA GAA AGC TGC CGG GTC TGG GTC TGG GTC TTA GGA CGG GGG GCG GAA A
Rev_puromycin	TTT CCG CCC CCC GTC CTA A
Puromycin_linker	5'-/5Phos/-CTCCCGCCCCCGTCC-(SPC18) ₅ -CC-(Puromycin)-3'

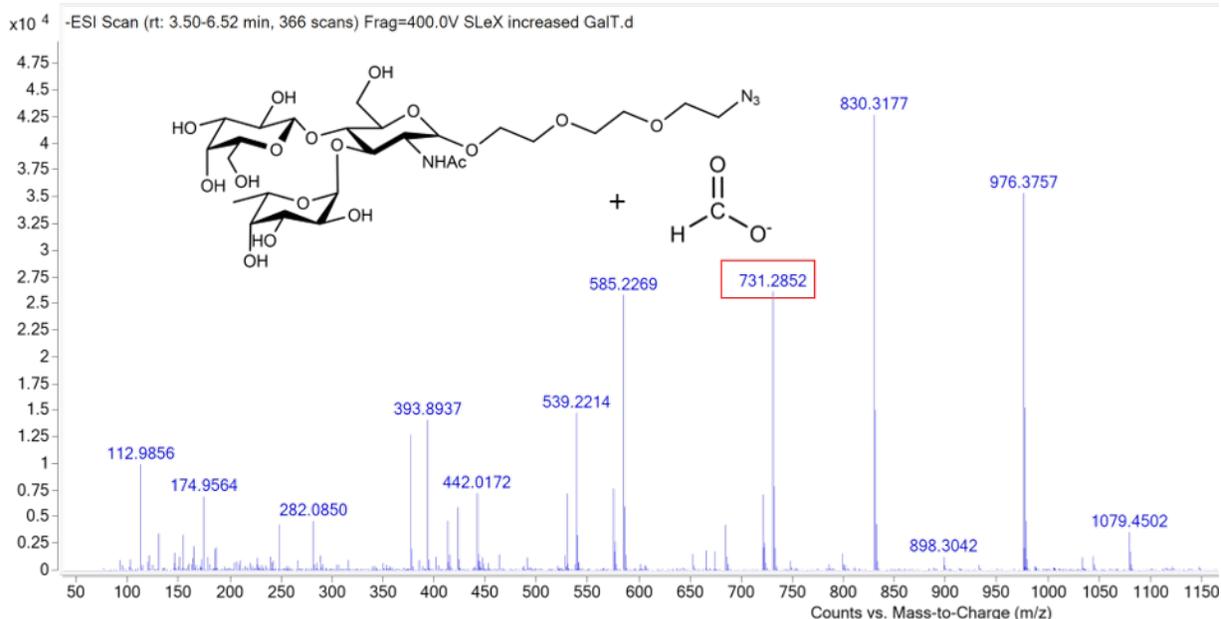


Figure 6.1: Mass spectrometry of wrongly fucosylated intermediate in SLe^X-PEG₃-Azide reaction.
A) Mass spectra following incubation with the glycosyltransferases *HpGalT*, *Cst-I*, and *HpFucT*. Wrongly fucosylated product, Le^X-PEG₃-azide [M + HCOO]⁻, (*m/z* 731.29), is present. The peaks at *m/z* of 830.32 and *m/z* of 976.38 correspond to 3'-SLN-PEG₃-Azide [M - H]⁻ and SLe^X-PEG₃-azide [M - H]⁻, respectively. Samples were analyzed in negative ion mode.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

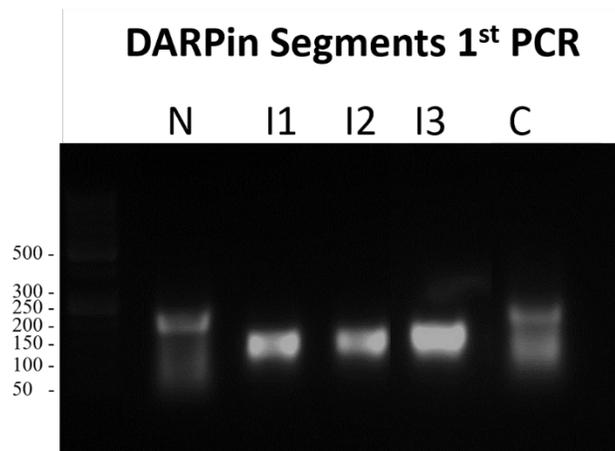


Figure 6.2: DARPin library 1st PCR amplification of DARPin library segments. The expected size of the N, C, I1, I2, and I3, segments are 165, 117, 109, 112, and 142 bp. Note that the N and C terminal fragments contain smears, which may indicate some nonspecific amplification.

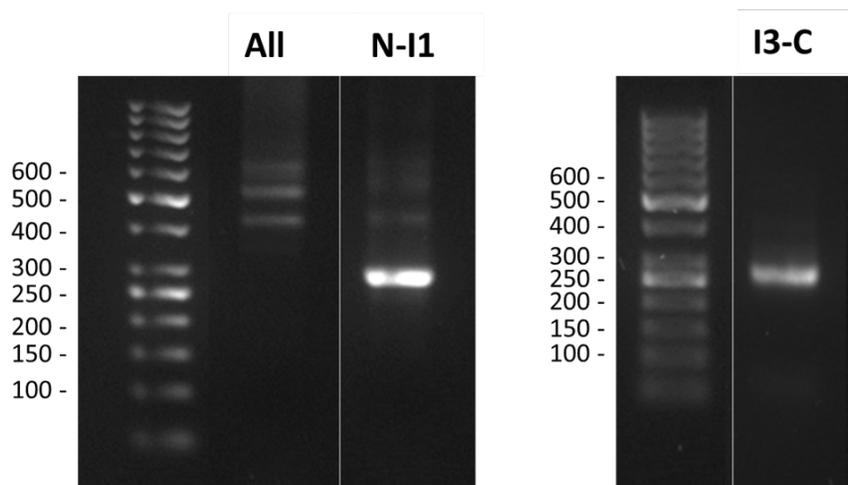


Figure 6.3: DARPin library PCR amplification following ligation of modules. Site saturated DARPin library modules previously PCR amplified using primers containing type IIS restriction enzyme cut sites. The ligation of all fragments (lane: All) is expected to result in a band size of 595 bp. The ligation of N-I1 and I3-C fragments are expected to result in band sizes of 279 and 257 bp, respectively. The modules were digested using BpI and ligated by T7 ligase.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

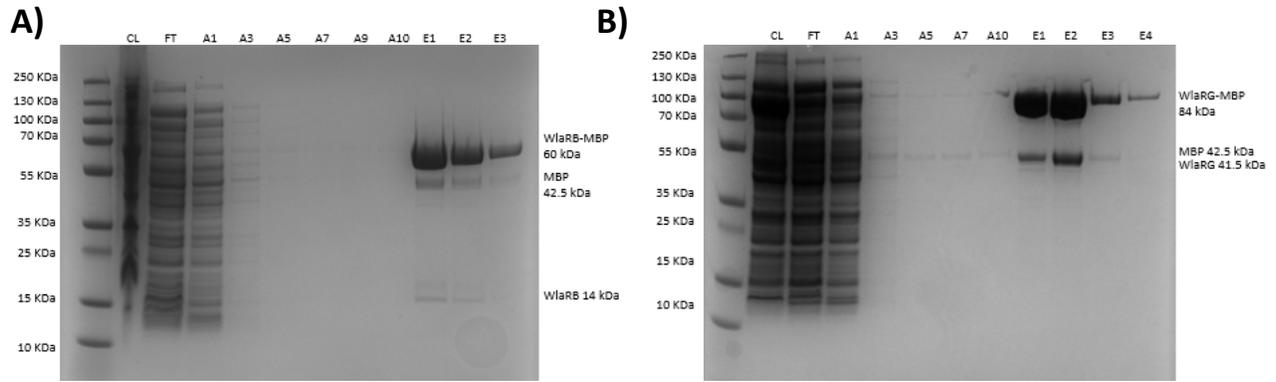
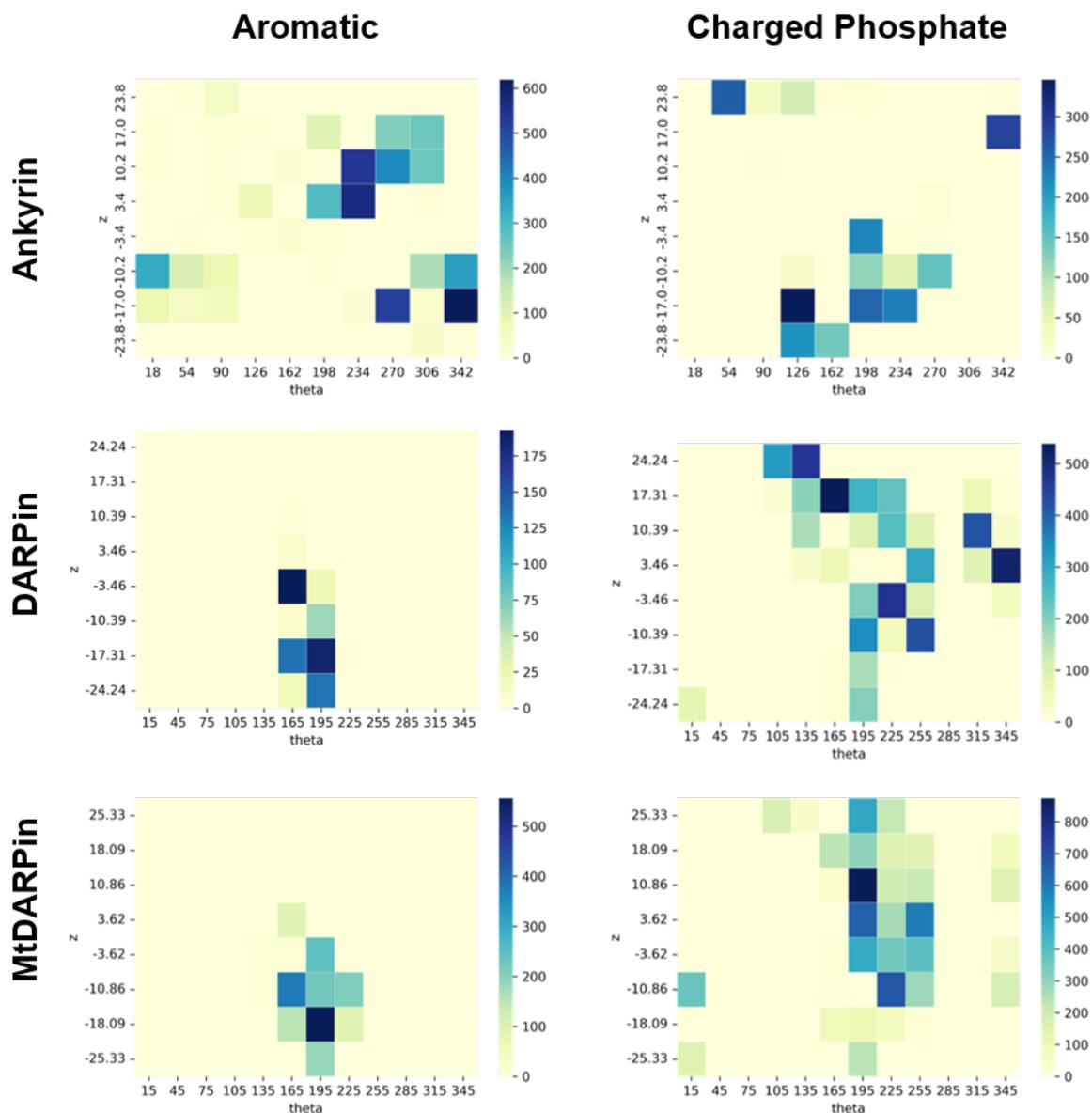


Figure 6.4: SDS-PAGE of the recombinant protein purification of WlaRB and WlaRG. A) SDS-PAGE of WlaRB. The expected molecular weight of WlaRB-MBP is 60 kDa. B) SDS-PAGE of WlaRG. The expected molecular weight of WlaRB-MBP is 84 kDa. Both proteins were MBP tagged for purification using amylose resin affinity columns. Note that the only other bands present in the gel match the expected size of MBP (42.5 kDa), or the proteins without the MBP tag.

Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand



SFigure 6.5: Aromatic and phosphate interaction heatmaps between dTDP-Qui3N and ankyrin, DARPin, or MtDARPin. Aromatic interactions are defined by proximity dTDP-Qui3N ring structures to aromatic amino acids. Charged interactions are based on the proximity of the phosphate oxygen atoms in dTDP-Qui3N to positively charged amino acids in the proteins.

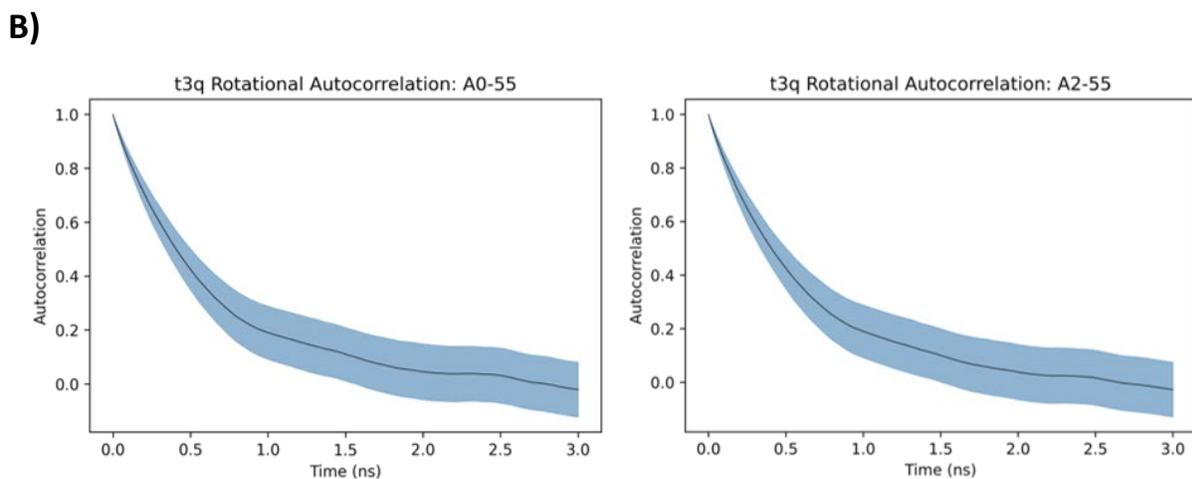
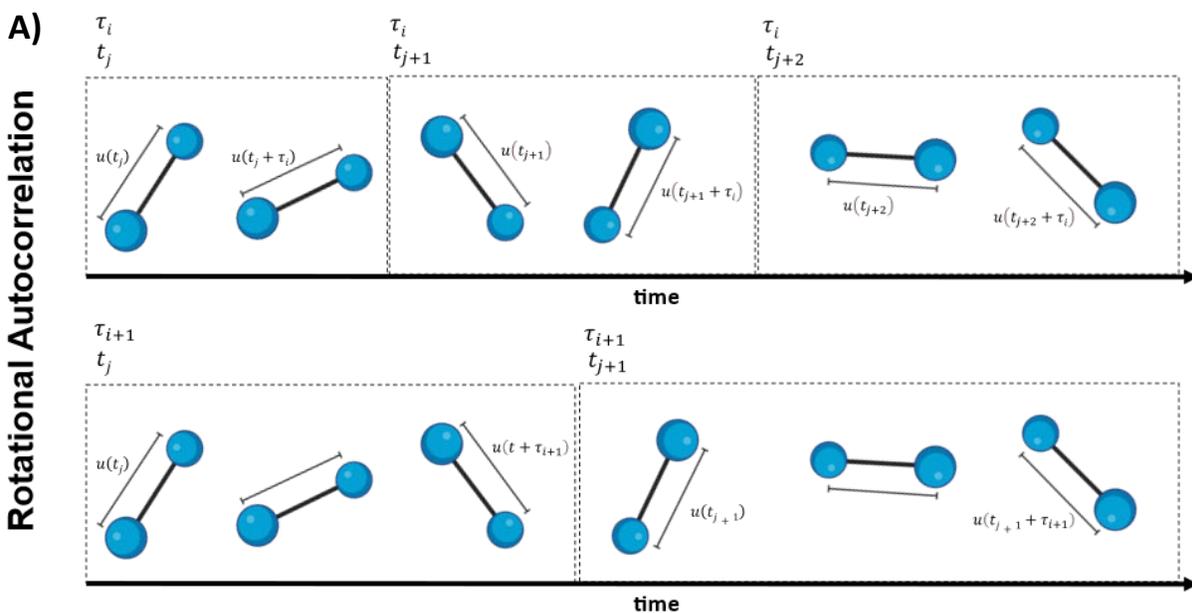
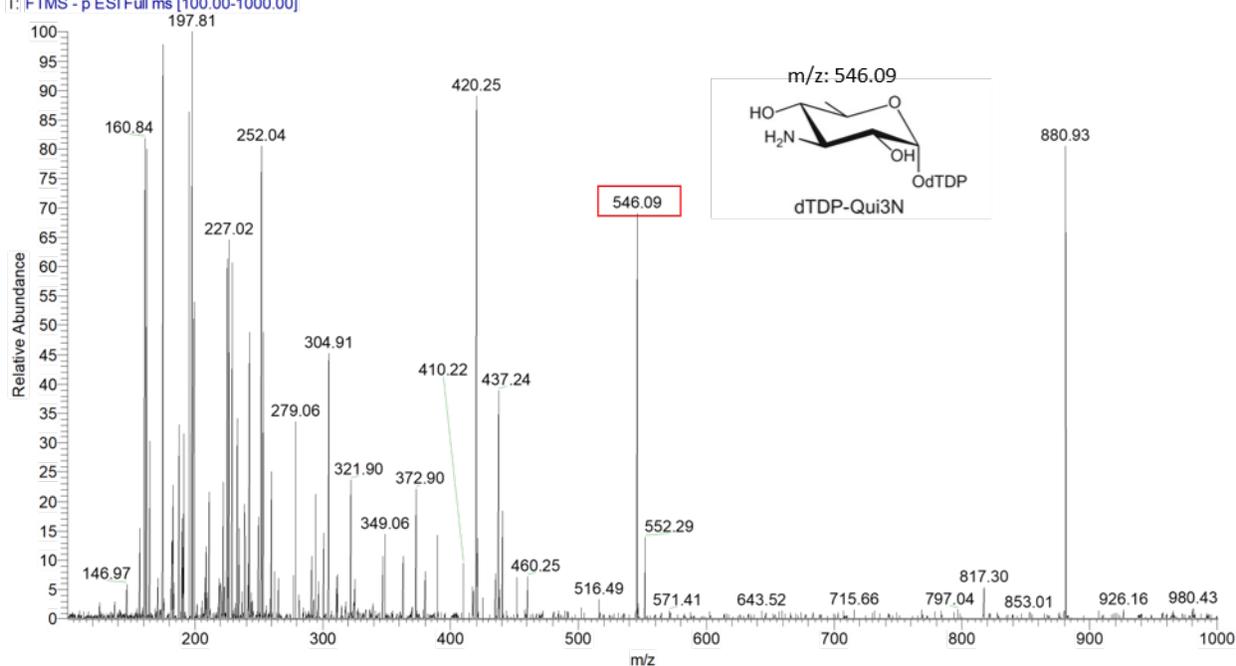


Figure 6.6: Rotational Autocorrelation Approach and Results. **A) Rotational autocorrelation approach.** The rotational autocorrelation finds the time it takes for a unit vector between two atoms (u) to decorrelate from itself. The time window (τ_i) is increased (τ_{i+1}) until an autocorrelation of zero is observed. **B) Rotational autocorrelation of dTDP-Qui3N.** The rotational autocorrelation was measured between atoms 0-55, and 2-55, both of vectors decorrelate between 2-3 seconds. The standard deviation is shown in light blue.

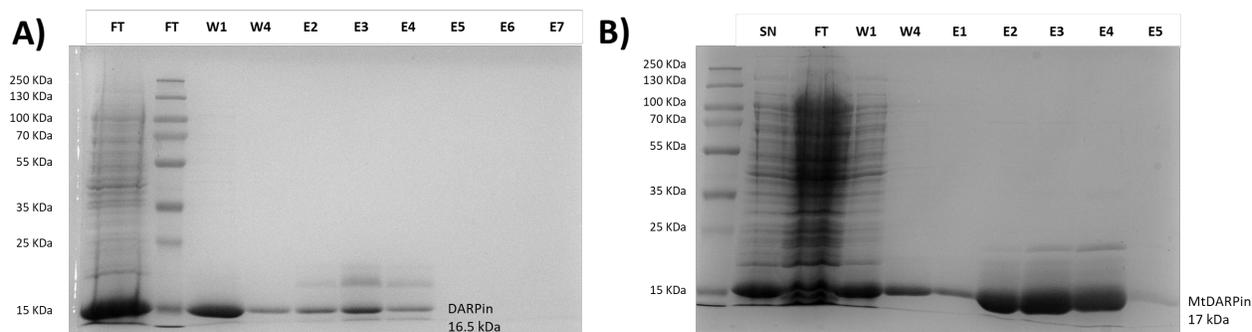
Chapter 6: Conclusion and Future Directions

5.5: Synthesis of Ankyrin, DARPin, MtDARPin and the dTDP-Qui3N Ligand

B8 #6-19 RT: 0.08-0.24 AV: 14 NL: 1.08E4
T: FTMS - p ESI Full ms [100.00-1000.00]



SFigure 6.7: Mass spectra of dTDP-Qui3N prior to freeze drying. The reaction was purified using size exclusion chromatography and lyophilized. The m/z 546.09 corresponds to dTDP-Qui3N [-H]. Sample was analyzed in negative ion mode.



SFigure 6.8: SDS-PAGE of the recombinant protein purification of DARPin and MtDARPin. A) SDS-PAGE of DARPin. The expected molecular weight of DARPin is 16.5 kDa. B) SDS-PAGE of MtDARPin. The expected molecular weight of MtDARPin is 17 kDa. Both proteins are His-tagged for purification using nickel column affinity chromatography. Note that other bands are present.