

LECTENZ: CARBOHYDRATE-RECOGNIZING BIOSENSOR ENGINEERED VIA  
COMPUTATIONALLY-GUIDED DIRECTED EVOLUTION

by

KAUSAR NADIM SAMLI

(Under the Direction of Robert J. Woods)

ABSTRACT

Carbohydrate recognition is an integral part of normal biological processes. It is critical for host-pathogen interactions, biological development, and increasingly important for disease-state biomarker detection. Due to the importance of carbohydrate recognition and variation in host glycosylation, glycans are obvious targets for detection, diagnostic, and therapeutic applications. Not only do glycans serve as important disease biomarkers, they also impact the pharmacological properties of therapeutic biologics. For example, glycan heterogeneity can impact the batch-to-batch consistency, immunogenicity, pharmacokinetics, activity, and biological clearance of recombinant glycoproteins. Given that more than two thirds of therapeutic biologics are glycosylated recombinant proteins, new tools for glycosylation analysis during bioprocess monitoring are required.

Reported here is the development of a novel reagent for detecting the core chitobiose component common to all *N*-linked glycans. Through a combination of computationally guided biocombinatorial library design and *in vitro* directed evolution, the *N*-glycan processing enzyme, PNGase F from *Flavobacterium meningosepticum*, has

been engineered into a catalytically inactive protein with enhanced affinity for the substrates of the wild-type enzyme. The engineering of a lectin-like carbohydrate-recognizing biomolecule from a carbohydrate-processing enzyme (a Lectenz®) was initiated *in silico* to determine optimal carbohydrate-enzyme interactions using molecular dynamics simulations. *In silico* structure/function analyses guided the design of focused biocombinatorial libraries for *in vitro* directed evolution via yeast-displayed selection of Lectenz® candidates. The selected clone, R911, was observed to have a 10x affinity enhancement ( $K_D = 0.26 \mu\text{M}$ ) relative to a non-affinity enhanced control clone (D60A). In addition, enrichment of the *N*-glycan bearing glycoprotein, Ribonuclease B, and *N*-glycopeptides was demonstrated via Lectenz® affinity chromatography. Furthermore, successful enrichment of glycoproteins from the cell extract of a human breast cancer cell line, MCF7, demonstrated the utility of R911 Lectenz® as a capture reagent for the enrichment of glycoproteins from complex mixtures. Molecular modeling of R911 provided insights into mutations critical for affinity and specificity, thus rationalizing experimental observations.

The successful creation of the R911 Lectenz® reagent presents not only a unique solution to the challenge of glycopeptide and glycoprotein sample enrichment, but also demonstrates a novel strategy for engineering glycan-targeting reagents for glycans and glycoconjugates of biological relevance.

INDEX WORDS: Lectenz, Lectin, Glycan, Protein engineering, Directed evolution, Molecular dynamics, MM-GBSA, Binding free energy decomposition, Glycosylation, PNGase F, Yeast display, Glycosensors, Biosensors, Affinity chromatography, Glycoprofiling

LECTENZ: CARBOHYDRATE-RECOGNIZING BIOSENSOR ENGINEERED VIA  
COMPUTATIONALLY-GUIDED DIRECTED EVOLUTION

by

KAUSAR NADIM SAMLI

B.S., The University of Texas at Austin, 2001

M.S., Arizona State University, 2008

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

© 2014

Kausar Nadim Samli

All Rights Reserved

LECTENZ: CARBOHYDRATE-RECOGNIZING BIOSENSOR ENGINEERED VIA  
COMPUTATIONALLY-GUIDED DIRECTED EVOLUTION

by

KAUSAR NADIM SAMLI

Major Professor:	Robert J. Woods
Committee:	James H. Prestegard
	Lance Wells
	Zachary A. Wood

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
May 2014

## DEDICATION

For *merijaan* Farah

--the life we have lived and have yet to live together.

## ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to family, friends, and colleagues whose personal and professional contributions have made this work possible. Foremost among my colleagues is my doctoral advisor, Robert J. Woods, and the members of my advisory committee, Lance Wells, James H. Prestegard, and Zachary A. Wood, who were critical in providing guidance, perspective, and insightful scientific discussion to enhance the significance and quality of this work in addition to expanding my scientific knowledge and expertise. Lori Yang's early contributions to get this project started were instrumental and her tireless efforts at Glycosensors & Diagnostics have been phenomenal. Jodi Hadden, Arunima Singh, Matthew Tessier, Lachele Foley, Keigo Ito, Xiaocong Wang (JArMaLaKX) served as molecular modeling gurus in teaching me the practical applications of these powerful tools. Julie Nelson provided technical resources as Director of the CTEGD FACS facility. Robert Bridger diligently performed LC-MS/MS analysis of critical samples, of which I seemed to possess an endless supply. SPR experiments would not have been possible with the core service overseen by Gerardo Gutierrez-Sanchez. MALDI MS analysis was made possible by the core services overseen by Parastoo Azadi. David Smith and Jaime Heimburg-Molinaro at Emory University made processing of glycan array samples possible. DJ Bernsteel generously cultured and provided MCF7 cells for experiments. Members of the Woods' research group, especially Mark Baine, are some of the most generous colleagues and friends I have had the privilege of working with. I am particular grateful to Al Darvill and Steve

Hajduk for their leadership at the Complex Carbohydrate Research Center and the Dept. of Biochemistry and Molecular Biology respectively, and the support they have extended to me personally in order to pursue opportunities to advance my career as a scientist. In addition, both Steve Hajduk and Lance Well deserve special recognition for their efforts to enhance the BMB graduate program.

Many members of the UGA's leadership team made my exceptional experience as a graduate student at UGA possible. Working with David Lee, Libby Morris, and Maureen Grasso on promoting entrepreneurship at UGA through Global Impact Competitions and *Thinc.* at UGA has been a wonderful experience. In addition, the support, encouragement, and recognition from the Blue Key Honor Society members and advisors provided inspiration through engagement with this singular group of individuals who live to serve.

Several friends and family members saw me through to the end of completing this work. In particular I am grateful to the Hughes and Samli families, for the continuing inspiration and perspective they provide. My fellow graduate students shared in this journey with me and have seen me to the end, and I am deeply grateful for the frequent and much needed respites. Several friends from various seasons of life I found close at hand when needed, including Kathlynn Brown, Mitch Magee, David Dell, and Sarah and Richard Chatfield. Thank you all for your generous and loving friendship.

My beloved wife, Farah, has been my constant companion since the beginning. I could not have wanted for a more giving, loving, and devoted friend in life's journey. I am especially grateful for the recently discovered companionship of Jasper and Enzo, who are always in my corner reminding me that life is to be lived with **joy**.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	v
LIST OF TABLES .....	ix
LIST OF FIGURES .....	xii
CHAPTER	
1 LITERATURE REVIEW AND INTRODUCTION .....	1
Protein Engineering .....	1
Molecular Modeling.....	8
Glycoscience .....	15
PNGase F .....	23
Dissertation Overview .....	32
References.....	35
2 COMPUTATIONALLY-GUIDED DESIGN OF BIOCOMBINATORIAL LIBRARIES .....	46
Computationally Guided Library Design.....	46
Yeast Display Library Construction .....	50
Directed Evolution of PNGase F Clones via Yeast Surface Display.....	55
Methods.....	63
References.....	67

3	EXPERIMENTAL CHARACTERIZATION OF LECTENZ®	
	CANDIDATES .....	69
	Cloning of PNGase F Clones into a Bacterial Expression Vector .....	69
	Expression and Purification of PNGase F Clones .....	72
	Activity and Kinetic Studies .....	80
	Lectenz® Affinity Chromatography .....	85
	Glycan Array Screening.....	94
	Methods.....	97
	References.....	108
4	MOLECULAR DYNAMICS SIMULATIONS OF PNGASE F CLONES.	112
	Structural Models of D60A, R911, and R911 C60A .....	112
	MD Simulations and Binding Free Energy Decomposition of PNGase F Clones with <i>N</i> -Glycan and <i>N</i> -Glycotriptide Ligands .....	129
	MD Simulations and Binding Free Energy Decomposition of PNGase F Clones with a Ser- <i>O</i> -GlcNAc Ligand .....	146
	Methods.....	152
	References.....	154
5	DISCUSSION AND FUTURE PERSPECTIVES.....	156
	References.....	163

## LIST OF TABLES

	Page
Table 1.1: Estimated library diversity of various directed evolution technologies .....	6
Table 1.2: Comparison of enzyme engineering methodologies.....	7
Table 1.3: Comparison of methods used to predict binding free energies.....	13
Table 1.4: Peptide: <i>N</i> -glycanase nomenclature.....	23
Table 2.1: Experimental and theoretical hydrogen bond lengths observed between chitobiose and PNGase F .....	47
Table 2.2: Approximate residue contributions (kcal/mol) to the binding free energy for wtPNGase F bound to substrate, chitobiose .....	49
Table 2.3: Computational alanine scanning of PNGase F bound to chitobiose.....	50
Table 2.4: Comparison of theoretical and estimated synthesized library diversity and coverage of sequence space .....	54
Table 2.5: Sequences of enriched clones from Library 1 and Library 2 selections.....	62
Table 3.1: Physical and chemical properties of PNGase F clones.....	79
Table 3.2: Bovine pancreatic RNase properties.....	80
Table 3.3: Deglycosylation activity of PNGase F clones .....	81
Table 3.4: Surface plasmon resonance kinetic data for PNGase F clones.....	85
Table 3.5: Comparison of MLAC and R911 Lectenz® affinity chromatography with MCF7 cell extract .....	91

Table 3.6: R911 Lectenz® affinity chromatography enrichment of MCF7 cell extract glycoproteins.....	91
Table 3.7: Eluted MCF7 <i>N</i> -glycoproteins identified by LC-MS/MS .....	92
Table 3.8: Eluted MCF7 <i>O</i> -glycoproteins identified by LC-MS/MS .....	94
Table 4.1: Dunbrack library rotamer selection .....	113
Table 4.2: Dynameomics library rotamer selection .....	114
Table 4.3: Structural models of wtPNGase F and clones for rotamer selection .....	119
Table 4.4: Experimental and theoretical hydrogen bond lengths observed between chitobiose and PNGase F .....	120
Table 4.5: Estimated MM-GBSA interaction action energies of rotamer models.....	124
Table 4.6: Rotamer conformations from R911 Dyn MD simulation.....	128
Table 4.7: Frequency of preferred rotamer combinations in R911 Dyn MD simulation	128
Table 4.8: Structural models of wtPNGase F and clones used for MD simulations and MM-GBSA .....	131
Table 4.9: MM-GBSA of 1PNF complexed with $\beta$ -chitobiose .....	132
Table 4.10: MM-GBSA of D60A complexed with $\beta$ -chitobiose.....	132
Table 4.11: MM-GBSA of R911 Dyn complexed with $\beta$ -chitobiose.....	132
Table 4.12: MM-GBSA of R911 C60A Dyn complexed with $\beta$ -chitobiose .....	133
Table 4.13: Estimated MM-GBSA theoretical interaction action energies of models complexed with $\beta$ -chitobiose .....	133
Table 4.14: Structure models of wtPNGase F and clones used for MD simulations & MM-GBSA .....	135
Table 4.15: MM-GBSA of 1PNF (GLH206) complexed with glycotriptide ligand...	140

Table 4.16: MM-GBSA of D60A (GLH206) complexed with glycotriptide ligand ..	140
Table 4.17: MM-GBSA of R911 Dyn complexed with a glycotriptide ligand.....	141
Table 4.18: MM-GBSA of R911 C60A Dyn complexed with glycotriptide ligand...	141
Table 4.19: Estimated MM-GBSA theoretical interaction action energies of models complexed with a glycotriptide ligand.....	142
Table 4.20: Summary of R911 mutation theoretical and experimental characteristics ..	144
Table 4.21: Structural models of wtPNGase F and clones used for MD simulations and MM-GBSA .....	146
Table 4.22: Theoretical hydrogen bond lengths between GlcNAc- $\beta$ -Ser and R911 .....	150
Table 4.23: Estimated MM-GBSA theoretical interaction action energies of models complexed with an <i>O</i> -GlcNAc ligand.....	151
Table 5.1: Proposed functions of PNGase F active site residues.....	162

## LIST OF FIGURES

	Page
Figure 1.1: Directed evolution strategies of protein-protein interactions .....	4
Figure 1.2: A classical mechanics force field equation .....	9
Figure 1.3: The thermodynamic cycle for a receptor-ligand interaction .....	12
Figure 1.4: Representative examples of <i>N</i> -glycan complexity .....	18
Figure 1.5: Chemical diversity of glycans .....	20
Figure 1.6: The PNGase F deglycosylation reaction .....	24
Figure 1.7: Sequence display of PNGase F (PDB ID: 1PNF) .....	26
Figure 1.8: Surface hydrophobicity representation of the PNGase F-chitobiose complex (PDB ID 1PNF).....	27
Figure 1.9: Schematic diagram showing the intermolecular hydrogen bonding contacts between PNGase F, <i>N,N'</i> -diacetylchitobiose and water molecules .....	29
Figure 1.10: Active site hydrogen bond network of the PNGase F:chitobiose complex..	30
Figure 1.11: Schematic representation of the Lectenz® design strategy.....	33
Figure 2.1: RMSD in the $C\alpha$ positions in the PNGase F – chitobiose complex.....	46
Figure 2.2: PNGase F binding pocket.....	48
Figure 2.3: Sequence and restriction map of non-amplified GenScript Library 1.....	51
Figure 2.4: Sequence and restriction map of non-amplified GeneArt Library 2 .....	52
Figure 2.5: PNGase F modified pPNL6 yeast display library plasmid map.....	53

Figure 2.6: Yeast cell-surface display. Representation of Aga2p-PNGase F fusion protein displayed via Aga1p on the yeast cell surface .....	55
Figure 2.7: Flow cytometry of induced yeast library.....	56
Figure 2.8: Ribonuclease B glycoforms.....	57
Figure 2.9: Magnetic-Activated Cell Sorting (MACS) of yeast cells.....	59
Figure 2.10: Fluorescent-Activated Cell Sorting (FACS) of double positive yeast cells.	60
Figure 2.11: Yeast display PNGase F clonal selection and enrichment with GeneArt Library 2.....	61
Figure 2.12: Yeast colony PCR and sequencing primers .....	66
Figure 2.13: Yeast colony PCR program.....	66
Figure 3.1: PCR amplification primers for PNGase F clones selected via yeast display .	69
Figure 3.2: PCR amplification primers for PNGaseF-pOPH6 ompA-PNGase F-His6 sequence .....	70
Figure 3.3: DNA gel of pOPH6 II and PNGase F cloning .....	70
Figure 3.4: PNGase F-pOPH6 II vector map.....	71
Figure 3.5: D60A IMAC elution chromatogram .....	73
Figure 3.6: SDS-PAGE and Western Blot of IMAC purified D60A clone .....	73
Figure 3.7: D60A SEC chromatogram on a Superdex 75 10/300 GL column .....	74
Figure 3.8: SDS-PAGE and Western Blot of SEC purified D60A clone .....	74
Figure 3.9: R911 IMAC elution chromatogram .....	75
Figure 3.10: R911 SEC chromatogram on a Superose 12 10/300 GL column.....	75
Figure 3.11: SDS-PAGE and Western Blot of IMAC and SEC purified R911.....	76
Figure 3.12: RNase B pH scouting .....	83

Figure 3.13: SPR sensograms of wtPNGase F, D60A, R911, and R911 C60A .....	84
Figure 3.14: D60A affinity chromatography with RNase A and RNase B .....	87
Figure 3.15: R911 Lectenz® affinity chromatography of RNase A vs RNase B.....	87
Figure 3.16: R911 Lectenz® affinity chromatography of deglycosylated RNase B .....	88
Figure 3.17: R911 Lectenz® affinity chromatography of tryptic digests of RNase A and RNase B .....	88
Figure 3.18: R911 Lectenz® affinity chromatography of RNase B using free chitobiose for competitive elution .....	89
Figure 3.19: R911 Lectenz® affinity chromatography of MCF7 cell extract using free chitobiose for competitive elution .....	90
Figure 3.20: Glycan array screening of D60A and R911 clones .....	96
Figure 4.1: R911 D57L rotamers overlaid on wtPNGase F D57.....	115
Figure 4.2: R911 D60C rotamer overlaid on wtPNGase F D60.....	116
Figure 4.3: Comparison of R911 I156L rotamers.....	117
Figure 4.4: R911 G192I Dunbrack rotamers overlaid on wtPNGase F G192 .....	118
Figure 4.5: R911 R248W rotamers overlaid on wtPNGase F R248.....	118
Figure 4.6: Experimental and theoretical hydrogen bonds .....	120
Figure 4.7: Interaction energy stability during 100 ns 1PNF MD simulation .....	121
Figure 4.8: Interaction energy stability during 100 ns R911 Dun MD simulation.....	121
Figure 4.9: D57L rotamer histograms of Chi1 and Chi2 dihedral angles.....	125
Figure 4.10: D60C rotamer histogram of Chi1 dihedral angle .....	126
Figure 4.11: E206S rotamer histogram of Chi1 dihedral angles .....	126
Figure 4.12: I156L rotamer histograms of Chi1 and Chi2 dihedral angles .....	126

Figure 4.13: G192I rotamer histograms of Chi1 and Chi2 dihedral angles.....	118
Figure 4.14: R248W rotamer histograms of Chi1 and Chi2 dihedral angles.....	118
Figure 4.15: R911 Dyn MD simulation hydrogen bonds and preferred rotamers .....	129
Figure 4.16: Interaction energy stability during 100 ns 1PNF (GLH206) MD simulation with the glycotriptide ligand .....	136
Figure 4.17: Chair and skew-boat ring conformations of reducing GlcNAc in wtPNGase F complexed with glycotriptide .....	137
Figure 4.18: 1PNF (GLH206) and glycotriptide MD simulation hydrogen bonds at 60 ns time point.....	138
Figure 4.19: 1PNF (GLH206) and glycotriptide MD simulation hydrogen bonds at 84 ns time point.....	139
Figure 4.20: Interaction energy stability during 50 ns 1PNF (GLH206) MD simulation with an <i>O</i> -GlcNAc peptide ligand .....	147
Figure 4.21: Interaction energy stability during 50 ns 1PNF (GLH206) MD simulation with an <i>O</i> -GlcNAc peptide ligand .....	148
Figure 4.22: Interaction energy stability during 50 ns 1PNFD60A (GLH206) MD simulation with an <i>O</i> -GlcNAc peptide ligand .....	149
Figure 4.23: R911 and GlcNAc- $\beta$ -Ser MD simulation at 45 ns time point .....	149

## CHAPTER 1

### LITERATURE REVIEW AND INTRODUCTION

#### **1.1 Protein Engineering**

Recent advances in *in silico* modeling have shifted protein engineering from discovery-based efforts towards hypothesis-driven strategies by providing the intellectual framework to predict and rationalize protein design.<sup>1-7</sup> This transition to semi-rational protein engineering is due to the convergence of several factors, including the rapid growth of structural information in the Protein Data Bank, the advancement of molecular modeling tools, algorithms and processing power, and the development of directed evolution strategies.<sup>8-15</sup> Semi-rational protein engineering strategies now lead the way towards achieving *de novo* protein engineering.<sup>5, 16-18</sup>

##### *1.1.1 Rational Design & Directed Evolution*

The development of site-directed mutagenesis was a key milestone leading to the development of the field of protein engineering. In the early 1980's, structures were starting to be solved by X-ray crystallography, and the principles of enzyme catalysis were well understood by enzymologists.<sup>19</sup> However, a significant barrier to investigating enzyme activity and mechanisms was the inability to introduce precise amino acid mutations into the active sites in order to test hypotheses.<sup>19</sup> Site-directed mutagenesis was first demonstrated in 1978 to make DNA mutations, and used to make defined amino acid substitutions to tyrosyl-transfer RNA synthetase (cysteine to serine) and  $\beta$ -lactamase (serine to cysteine) in 1982.<sup>20-22</sup> The development of a methodology to introduce amino acid mutations of interest into enzymes launched the field of protein engineering. As

increasing numbers of structures were solved with X-ray crystallography, the early days of structure-guided hypothesis-driven approach to investigating enzyme mechanisms launched a rational design approach to protein engineering. Furthermore, as molecular cloning techniques and oligonucleotide synthesis became routine, site-directed mutagenesis became an indispensable tool. While significant advances were made in the understanding of protein chemistry and enzymology, like the quantitation of hydrogen bonding contributions<sup>23</sup>, it also became clear that rationally introduced mutations often negatively impacted enzyme activity and stability. Furthermore, failed attempts to alter substrate specificity led to the realization that local interactions were not solely responsible for recognition and activity, and that amino acids with no direct contact with the substrate played an indirect but critical role in defining active site substrate specificity.<sup>24-26</sup> These efforts underscored the increasing importance of systematically investigating the effect of individual mutations through site-saturation mutagenesis, site-specific mutagenesis with all possible amino acids.<sup>19</sup>

Site-directed mutagenesis, alanine-scanning mutagenesis, and site-saturation mutagenesis became standard tools for protein engineering via rational design. However, there remained a need to investigate the synergistic effects of mutations by introducing multiple mutations in defined combinations. Furthermore, the engineering of proteins with designed properties could not be easily achieved with site-directed mutagenesis, which is effective for studying the role of individual amino acids. Nonetheless, the significance of these techniques is evidenced by the fact that they remain an important part of the ever-growing tool set of rational design methodologies for the field of protein engineering.

*In vitro* selection was first demonstrated in 1964 using chemical mutagenesis to evolve a catabolic pathway in the bacterium *Aerobacter aerogenes*.<sup>27</sup> In 1967, the first *in vitro* selection of an RNA molecule was demonstrated.<sup>28</sup> Shortly after the pioneering work that established the early rational design methodologies for protein engineering, the first application of directed evolution in 1985 demonstrated the selection of randomly mutagenized peptides via phage display.<sup>29</sup> Unlike, rational protein design, directed evolution demonstrated the distinct advantage that prior knowledge of the protein structure or the effects of amino acid mutations were not required. Furthermore, this work set the pathway toward developing other display technologies that would enable the high throughput investigation of mutagenized variants of a protein, significantly advancing understanding of protein-protein interactions and the field of protein engineering.<sup>31</sup>

The development of combinatorial approaches to introduce multiple mutations permitted a rapid way to investigate the effect of multiple random mutations and developed in parallel to rational design methodologies. Relative to site-directed mutagenesis, these random mutagenesis methodologies may seem to take more of a ‘blind’ or ‘irrational’ design approach. However, the uses of sophisticated high throughput screening methodologies with defined selection criteria serve as a way for the knowledge-guided enrichment of clones.

Nature’s process of evolution via natural selection is now routinely mimicked in the laboratory to rapidly evolve biomolecules with desired characteristics under defined selection pressure through the process of directed evolution.<sup>32</sup> Directed evolution provides a mechanism to engineer proteins (and ligands) with enhanced affinity, altered

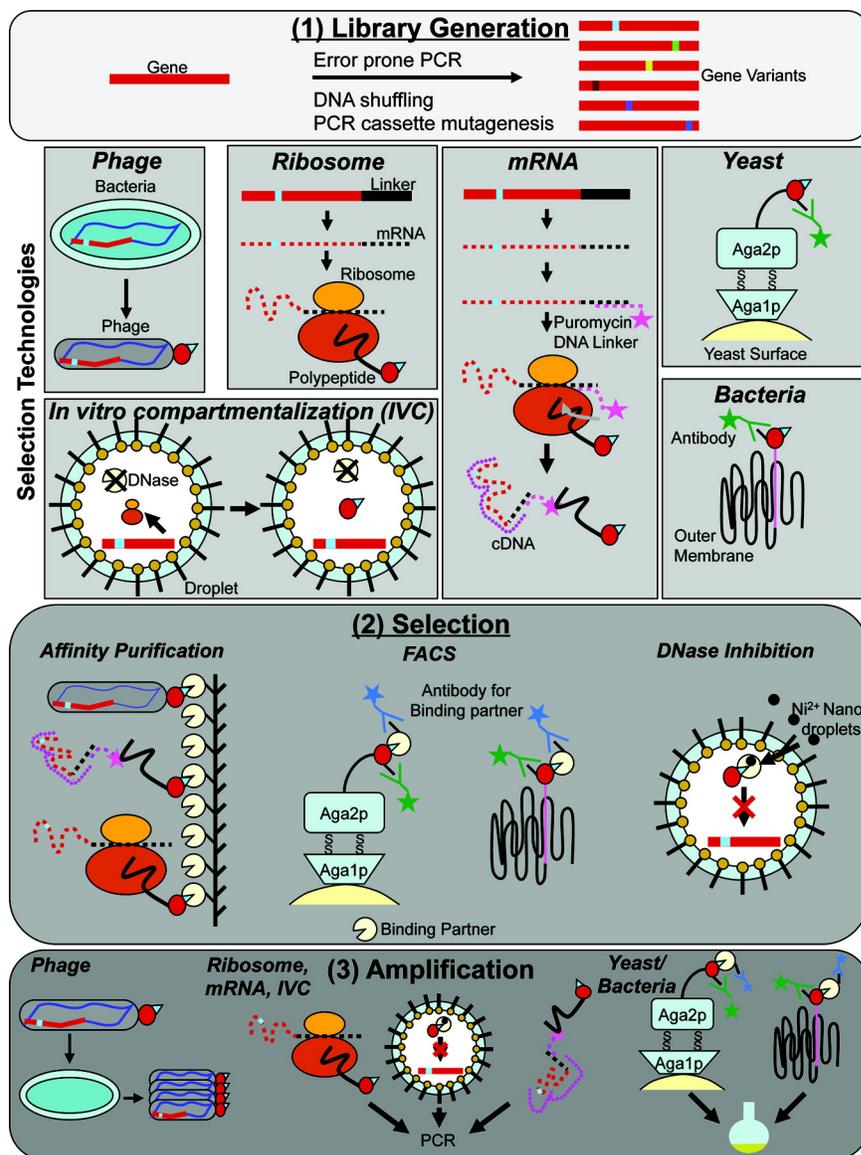


Figure 1.1: Directed evolution strategies of protein-protein interactions. 1) Library generation: the gene of interest is subjected to molecular evolution through either random mutagenesis or through cassette mutagenesis strategies to create molecular diversity. Depending on the display system being used, the library of gene variants are integrated into an expression system (Phage, Ribosome, mRNA, Yeast, Bacterial, or in vitro compartmentalization) for display and selection. 2) Selection: affinity purification is achieved by screening library clones against immobilized targets. Magnetic-activated cell sorting and Fluorescence-activated cell sorting are frequently utilized to rapidly enrich for clones with desired functionality. 3) Amplification: selected clones with relevant characteristics are amplified, for the next round of selection. Library diversity decreases as clonal enrichment is achieved through iterative rounds of selection and amplification. Reprinted with permission from Bonsor, D.A. & Sundberg, E.J. Dissecting protein-protein interactions using directed evolution. *Biochemistry* **50**, 2394-2402. Copyright 2011 American Chemical Society.

specificity of enzymes and receptors, altered catalytic activity, improved thermostability, antibody affinity maturation, therapeutic molecule development, and *de novo* proteins.<sup>3, 7,</sup>

16, 18, 33-46

Library generation, selection, and amplification are the three common steps that define directed evolution methodologies. An overview of various display technologies used for directed evolution are outlined in Figure 1.1.<sup>31</sup> Ideally, a high diversity library of mutagenized protein clones is screened against target molecules, via iterative rounds of selection and amplification.<sup>31</sup> This iterative process allows the selective enrichment of a small population of functionally relevant molecules while the majority of the undesired molecules are successively removed. After several rounds of selection, the enriched pool is sequenced to determine the identities of the functionally relevant clones.<sup>31</sup> Selected clones can then be characterized to evaluate their characteristics. An advantage of this approach is that should selected clones fail to exhibit the desired characteristics, the selection process can be repeated with modified selection parameters. This permits multiple use of the same library for selection of clones with different characteristics.

### *1.1.2 Semi-Rational Design*

The efficiency of directed evolution decreases as the number of mutagenized positions in the protein increases, with a practical limit of  $10^{10}$  clones, or approximately 7 fully randomized sites for some display technologies.<sup>31</sup> The maximum diversity of a library is dependent on several factors including the choice of directed evolution system to be utilized (Table 1.1). A strong incentive therefore exists to rationally select the optimal sites for randomization.

Table 1.1: Estimated library diversity of various directed evolution technologies.

Directed Evolution System	Estimated Library Diversity
Yeast display <sup>47, 48</sup>	$\sim 10^7 - 10^9$
In vitro compartmentalization <sup>31</sup>	$\sim 10^{10}$
Phase display <sup>49</sup>	$\sim 10^6 - 10^{11}$
Bacterial display <sup>50</sup>	$\sim 10^{11}$
mRNA display <sup>31</sup>	$\sim 10^{12} - 10^{14}$
Ribosome display <sup>51</sup>	$\sim 10^{12} - 10^{14}$
SELEX <sup>52</sup>	$\sim 10^{13} - 10^{15}$

Library design and generation is a critical component of all directed evolution strategies. Historically, libraries have relied on random design strategies. However, given the challenges with library diversity, many experimental library generation techniques now incorporate a knowledge-guided library design approach. For library design, the gene encoding the protein of interest is subjected to evolution through the use of various methodologies. Random design strategies include error-prone PCR and recursive PCR.<sup>19, 53</sup> DNA shuffling strategies provide a way of exploring favorable conserved mutations across a protein family or multiple unrelated families. These strategies may be homologous recombination-dependent or -independent and include single-gene shuffling, family shuffling, SHIPREC, ITCHY, and SCRATCHY.<sup>19, 53</sup> In addition, oligonucleotide-directed randomization without codon degeneracy is also frequently used; however, codon degeneracy is being employed more frequently to minimize codon bias by equally representing amino acid distribution in the sequence space covered by the library.<sup>6</sup> Furthermore, several computational modeling tools have been developed for the generation of knowledge-guided libraries.

Experimental and computational semi-rational design strategies rely on structural and/or evolutionary information to guide the design of focused libraries for directed evolution.<sup>19, 30, 32</sup> Employing computational simulations and structural biology tools with

high throughput directed evolution strategies makes feasible the rational design of novel protein libraries focused towards identifying clones with desired functionality (Table 1.2).<sup>8-12, 14, 15</sup>

Table 1.2: Comparison of enzyme engineering methodologies. Reprinted from Current Opinion in Biotechnology 16, Chica, R.A., Doucet, N. & Pelletier, J.N., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, 378-384, Copyright 2005, with permission from Elsevier.<sup>7</sup>

	Rational design	Random mutagenesis	Semi-rational design
High-throughput screening or selection method	Not essential	Essential	Advantageous but not essential
Structural and/or functional information	Both essential	Neither essential	Either is sufficient
Sequence space exploration	Low	Moderate, random	Experimental: moderate, targeted Computational: vast, targeted
Probability of obtaining synergistic mutations	Moderate	Low	High

A key objective of the field of protein engineering is to elucidate the chemical and structural mechanisms of biological interactions. The advent of novel technologies, driven by the increased understanding of biological interactions, has resulted in the development of protein engineering applications focused on designing biomolecules with novel functionality. This represents a subtle but significant shift from early discovery-based efforts towards hypothesis-driven strategies and application oriented objectives, reflecting a natural maturation of the field over the past 30 years. Many of the recent advances in protein engineering have been aided by parallel developments in the field of molecular modeling.

## 1.2 Molecular Modeling

The theoretical methods and computational techniques used to model molecular systems provide insights into the behavior of molecules. Molecular modeling emphasizes the representation and manipulation of three-dimensional molecular structures, and the properties that define those structures.<sup>54</sup> The simulation of molecular dynamics can be achieved with Quantum Mechanics (QM) or Molecular Mechanics (MM). In QM, physical properties of a model are determined to a high degree of accuracy by calculating solutions to the wave function thereby generating a description of electron distribution. MM differs in that the energy of a system is calculated as a function of the nuclear positions only based on Newtonian or classical physics. A clear advantage of MM is that large systems can be readily computed, whereas in QM solutions are time-consuming, and can be impractical for large biomolecules.<sup>54</sup> However, unlike MM, QM can calculate properties that are explicitly dependent on electrons, thus deriving properties based on the electronic distribution in a molecule, permitting investigation of reactions mechanisms.<sup>55</sup>

### *1.2.1 Molecular Mechanics*

Molecular Mechanics relies on classical mechanics to define the boundaries of molecular motion by approximating the potential energy. This approximation is achieved by employing a classical mechanics force field appropriate to the molecular system being examined. Based on Newtonian physics (classical mechanics), the force field is a mathematical model that defines the potential energy of a molecule as a function of the three-dimensional structure. Molecular modeling force fields are developed based on the mathematical description of intra- and inter- molecular forces with the system (Figure 1.2).<sup>54</sup> Intra-molecular forces are defined by three types of bonded interactions: 1) bond stretching/vibration, 2) bond/angle bending, and 3) bond torsion rotation. Inter-molecular

forces are defined by two types of non-bonded interactions: 1) van der Waals (repulsion and attraction) and 2) electrostatics. Bond stretching and bending terms describe the molecular connectivity and internal flexibility, whereas the non-bonded and torsion terms describe the three-dimensional structure and molecular interactions.

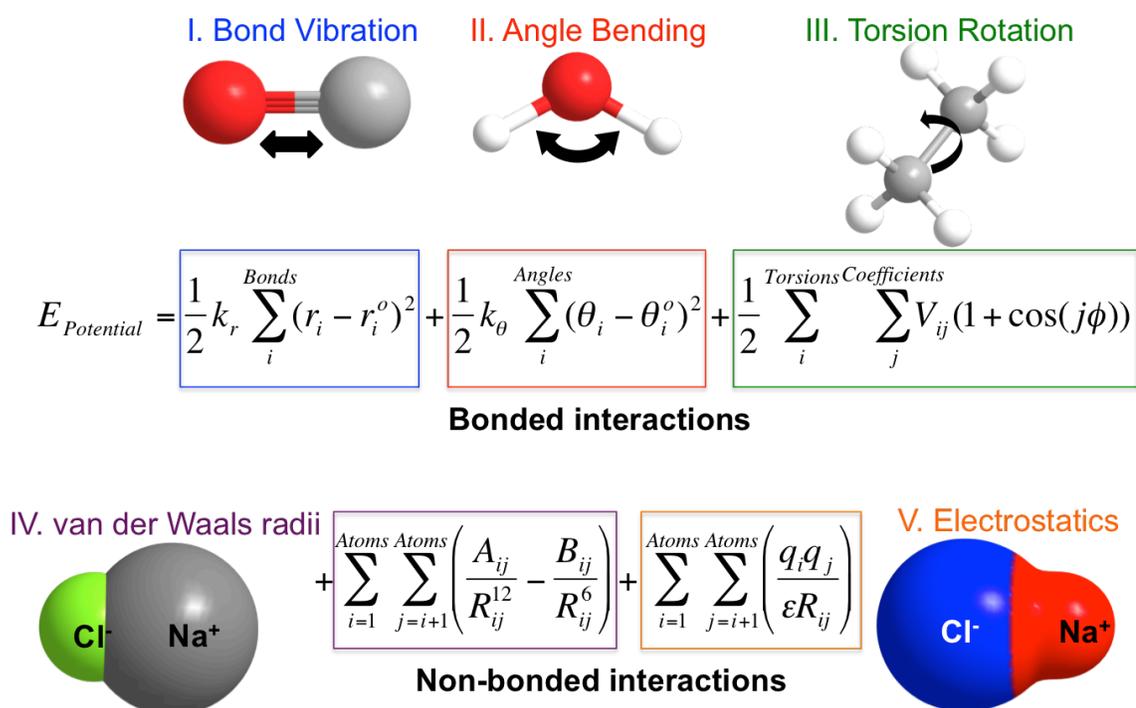


Figure 1.2: A classical mechanics force field equation.

The bond stretching term defines the potential energy contribution due to stretching modeled as a harmonic potential (analogous to Hooke's law) with respect to a reference bond length. Several hundred kcal/mol are required for bond stretching, a high energetic penalty. The angle bending term defines potential energy contribution also as a harmonic potential with respect to a reference angle. Unlike bond stretching, bond angle bending requires relatively minimal energy ( $\sim 0.009 \text{ kcal mol}^{-1} \text{ deg}^{-1}$ ). The torsion terms defines the potential energy as a function of bond rotation between a quartet of atoms. The non-bonded terms are calculated between all pairs of atoms between different molecules or the same molecule (but separated by at least three bonds). The van der

Waals interactions are described by a Lennard-Jones potential defining both the repulsive ( $1/R^{12}$ ) and attractive forces ( $1/R^6$ ) between pairs of atoms. Electrostatic interactions are modeled using a Coulombic potential term in which potential energy is inversely proportional to the distance between pairs of atoms. Several force fields have been developed that are suited for different kinds of interactions (proteins, carbohydrate, nucleic acids, etc.).

### *1.2.2 Molecular Dynamics Simulations*

The time-dependent changes in the conformation of a molecule due to molecular motion (Molecular Dynamics) can be computed using a force field equation over discrete time steps, thus permitting conformational sampling of a biomolecule of interest. The Verlet algorithm, a specific expression of the Taylor expansion series, is used to compute new atomic position based on position and acceleration over sequential time steps. The potential energy of the system is recalculated (using the force field equation) at each time step for all the atoms in the system at their new positions. Thus, the acceleration based on known mass and force can be computed at each time step. Thus, the force field equation, the Verlet algorithm, and Newton's second law are three key calculations required at each time step of a Molecular Dynamics simulation.

Molecular Dynamics (MD) simulations based on Newtonian physics provides a useful way to study the conformational energy landscape accessible to biomolecules.<sup>56, 57</sup> A key requirement for MD simulations is the availability of experimental structural data to provide the initial assignment for biologically relevant atomic coordinates for a biomolecule of interest. Such structural data is frequently obtained from X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) experiments and is made available via the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)). The Protein Data Bank (PDB)

provides a growing repository of structural models that can be used for MD simulations.<sup>15</sup> Biomolecules for which experimental structural data is not available, homology modeling can be utilized to approximate a structural model based on sequence similarity.<sup>58</sup>

A key assumption in generating MD data (whether through QM or MM approaches) is that the force field approximations describing the energy as a function of nuclear coordinates are accurate and transferable to larger systems. The precision and accuracy of force fields are continually improving through parameter development.<sup>59</sup> The parameterization process relies on experimental data to define energy terms and refinement of parameters has led to numerous force fields suitable for different kinds of systems with improved stability and accuracy.<sup>9, 13, 60-64</sup> MD data are now increasingly validated with experimental data, signifying the increasing reliability of the practical use of MD simulations for conformational analysis.<sup>57, 65, 66</sup>

Critical to the adoption of MD simulations (and the development of force fields employed) is the phenomenal increase in computing power. The first reported MD simulation in 1977 was a mere 9.2 ps trajectory of a small protein in vacuum.<sup>67</sup> Today, a 100 ns simulation of a 36 kDa protein in water can be completed in ~3 days using a single nVidia TESLA M2090 GPU. Coupled with improved force fields, refined structural models, and improved understanding of protein-ligand interactions, MD simulations and the estimation of free energies are being increasingly employed for protein engineering and design applications.

### *1.2.3 Binding Free Energy Decomposition*

The estimation of free energies using molecular mechanics force fields and their application to the study of protein-ligand interactions is of great interest. Not only do free energy estimates provide insights into the underlying process of protein-ligand

interactions but also enable studying states of a system not accessibly experimentally.<sup>9, 10,</sup>  
<sup>68-70</sup> This is particularly significant for protein engineering applications, where binding free energy decomposition on a per residue basis can provide insights into the energetic contribution of individual residues towards the binding free energy of a protein-ligand complex.

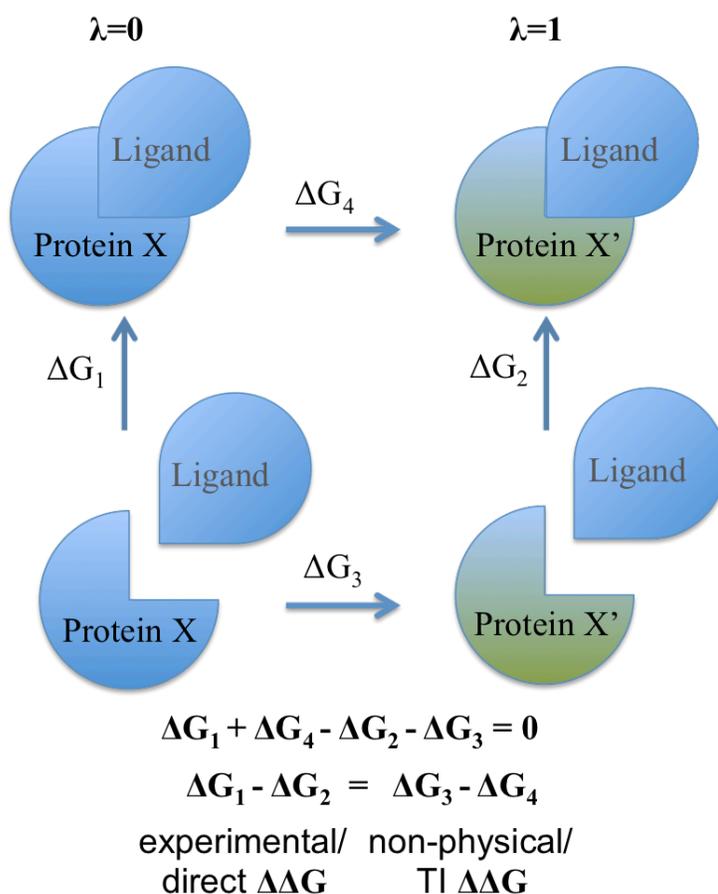


Figure 1.3: The thermodynamic cycle for a receptor-ligand interaction. Depicted is a receptor (Protein X) and its mutagenized clone (Protein X') and their ligand. Direct  $\Delta G$  calculations use the initial free receptor and ligand and the final complex states of the cycle ( $\Delta G_1$ ). TI methods calculate relative the free energy difference receptor-ligand complexes where the protein has been mutagenized ( $\Delta G_4$ ).

The calculation of the free energy of binding of a complex can be determined by using two distinct theoretical methods (Figure 1.3): 1) direct  $\Delta G$  calculations compute the absolute affinity of a ligand for a receptor, and 2) Thermodynamic Integration (TI)

predicts the relative free energy difference between variant ligands for the same receptor (or vice versa). Each method has advantage and disadvantages as indicate in Table 1.3.

Table 1.3: Comparison of methods used to predict binding free energies. DeMarco, M.L. & Woods, R.J. Structural glycobioogy: a game of snakes and ladders. *Glycobiology*, 2008, 18, 426-440, by permission of Oxford University Press.<sup>70</sup>

Comparison	Automated docking	Direct $\Delta G$	Thermodynamic Integration
Computational efficiency	High	Moderate	Low
Ligand set	Can be diverse	Can be diverse	Close structural analogues only
Binding free energy computed	Absolute $\Delta G$	Absolute $\Delta G$	Relative $\Delta G$
Water model	Implicit	Implicit	Explicit
Accuracy dependence	Compounds used for calibration	Water model, force field, and sampling	Sampling and force field

For direct  $\Delta G$  calculations, binding energies can be derived from snap shots from a MD trajectory. The binding energy reaction can be decomposed into individual components representing the initial free states of the ligand and the receptor, and the final receptor-ligand complexed state. The binding reaction can be decomposed into individual components (Equations 1.1 – 1.3). The average molecular mechanical energies ( $E_{MM}$ ), which are predominantly defined by electrostatic (hydrogen-bonds, charge-charge, charge-dipole, dipole-dipole) and dispersive interactions (Van der Waals attractions and repulsions), for conformations selected from the MD trajectory can be used to compute interaction energies. These interaction energies can be adapted to include estimates of entropic contributions from the hydrophobic effect and desolvation energy from Poisson-Boltzman or Generalized Born approximations (MM/PB/GBSA) and/or estimates of the change in conformational and vibrational entropy (Equations 1.2 and 1.3).<sup>8, 12, 71</sup>

Equation 1.1:

$$\Delta\Delta G_{BINDING} = \Delta G_{COMPLEX} - \Delta G_{PROTEIN} - \Delta G_{LIGAND}$$

Equation 1.2:

$$G = E_{MM} + G_{DESOL} - T\Delta S_{CONF/VIB}$$

Equation 1.3:

$$E_{MM} \approx V_{COULOMB} + V_{VDW} + INT_{BOND+ANGLE+DIHEDRAL}$$

Thermodynamic integration can compute free energy differences between two states (e.g.: mutagenized receptors for a ligand, or multiple ligands against a single receptor) by coupling them via a non-physical coordinate called  $\lambda$ .<sup>10, 69, 72-77</sup> A transition from state A ( $\lambda=0$ ) to B ( $\lambda=1$ ) along this reaction coordinate is simulated and any points along the  $\lambda$  coordinate are described by a mixed potential energy function  $V(\lambda)$  as indicated in Equation 1.4.<sup>10</sup>

Equation 1.4: Free energy difference computed by thermodynamic integration.

$$\Delta G_{TI}^0 = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

Free energy difference is computed over the ensemble average of the  $\lambda$  derivative of the mixed potential energy function at discrete values of  $\lambda$ . The integration is carried out over the average of the  $\lambda$  derivative of  $V(\lambda)$  at discrete  $\lambda$  values.<sup>75</sup> MD simulations at these discrete  $\lambda$  values are performed. This permits efficient parallelization and allows additional  $\lambda$  values to be calculated as needed to improve accuracy.

The use of free energy decomposition analysis on a per amino acid basis for identification of target residues for affinity enhancement is relatively new application. Zoete, *et al.*, employed MM-GBSA binding free energy decomposition to identified

important and detrimental residues in a T-cell receptor. Using this information they designed a strategy using sequence modifications to a T cell receptor in order to enhance its affinity for the peptide-MHC complex.<sup>78</sup> Recently, Pierdominici-Sottile, *et al.*, employed free energy calculations to identify mutations required to confer trans-sialidase activity into a *Trypanosoma ranfeli*'s sialidase.<sup>79</sup> These reports demonstrate the potential usefulness of free energy decomposition for protein engineering.

### 1.3 Glycoscience

The systematic study of complex carbohydrate (glycans) and their biological significance has led to glycomics becoming a prominent field of study in the post-genomic era. Historically, glycomics has lagged behind genomics and proteomics; however, the increased understanding of the significance of glycans and glycoconjugates in biological systems has spurred interest, investment, and innovation leading to notable advances in the field of glycomics.<sup>80, 81</sup>

#### 1.3.1 Significance

Carbohydrate recognition is an integral part of biological processes. It is critical for host-pathogen interactions, biological development, and increasingly important for disease-state biomarker detection.<sup>82</sup> Many tumor antigens are glycoproteins or glycolipids, and a variety of carbohydrate epitopes have been identified that are up-regulated in the disease state.<sup>83</sup> Currently approved carbohydrate tumor markers include<sup>84</sup>: Carcinoembryonic Antigen (CEA), a glycoprotein containing 50–80% carbohydrate associated with colorectal, gastric, pancreatic, lung, and breast carcinomas and the developing fetus<sup>85</sup>; Carbohydrate Antigen 19-9 (CA 19-9), or sialylated Lewis A antigen, which is present in a glycolipid found in patients with pancreatic cancer<sup>85</sup>; and Carbohydrate Antigen 15-3 (CA15-3), the most widely used serum marker for breast

cancer, is a glycoprotein fragment derived from mucin protein 1 (MUC1)<sup>86</sup>. Due to the importance of carbohydrate recognition and variation in host glycosylation, glycans are obvious targets for detection, diagnostic, and therapeutic applications.<sup>87-92</sup>

The location of many glycans on the cell surface makes them crucial for cellular interactions and contribution to the control of normal metabolic processes. Glycan structure and abundance are dynamic properties that can be driven the state of cellular processes, resulting in heterogeneity as biological processes are altered between normal and disease states. Furthermore, unlike DNA and protein synthesis, glycan synthesis is a non-template driven enzymatic process managed by many enzymes in a dynamic manner. The complexity in their synthesis may be attributed to the complex roles of glycans in biological processes; however, alterations in glycosylation machinery and activity can result in systemic effects on glycosylated proteins regardless of their level of abundance.

Glycans also impact the pharmacological properties of recombinant therapeutic biologics. Glycan heterogeneity can impact batch-to-batch consistency, immunogenicity, pharmacokinetics, activity and clearance.<sup>93</sup> Unlike the case of proteins and nucleic acids, the sequencing and structural characterization of glycans is a laborious multi-step process, typically requiring sample enrichment, enzyme digestion, and mass-spectrometric analysis, a process which is not amenable to real-time monitoring. Given that more than two thirds of therapeutic biologics are glycosylated recombinant proteins, new tools for glycosylation analysis during bioprocess monitoring are also required.<sup>94</sup>

Despite the significance of glycans, the discovery and routine laboratory analysis of glycans and glycoconjugates is limited by available isolation and analysis techniques,<sup>82</sup> which is not unexpected given the immense diversity of glycan structure.<sup>95</sup> Thus there is

an urgent need for glycan biosensors with defined carbohydrate specificity that can be used to interrogate biological samples to identify abnormal glycosylation states in cancer as well as the production of glycosylated therapeutics biologics.<sup>81</sup>

### 1.3.2 Glycan Biosynthesis and Diversity

The covalent attachment of glycans to nascent proteins is a non-template driven process and requires approximately 1000 gene products, thus the biosynthesis of oligosaccharides requires a significant investment of cellular resources and defects in the cellular machinery required for glycosylation can be fatal.<sup>91, 96-99</sup> The major types of mammalian protein glycosylation are *N*- and *O*-linked glycosylation.

The biosynthesis of *N*-glycan structures occurs on the endoplasmic reticulum membrane and requires over twenty enzymes in humans prior to its *en bloc* co-translational attachment to a nascent protein.<sup>100-104</sup> Synthesis begins with a dolichylpyrophosphate carrier and individual monosaccharides are attached sequentially until a fourteen-saccharide *N*-glycan structure is completed.<sup>102-104</sup> Each different glycosidic linkage requires a unique enzyme. The protein complex, oligosaccharyl transferase, is responsible for the *en bloc* attachment of the fourteen-saccharide *N*-glycan structure to an Asn-X-Ser/Thr sequone (where X can be any amino acid, except Pro) on the nascent peptide chain via an *N*-glycosidic bond to the side chain of the Asn residue.<sup>101, 103</sup> Endoplasmic reticulum chaperones regulate the proper folding of the nascent polypeptide via direct interaction with the *N*-glycan structure prior to the transfer of the high-mannose containing immature glycoprotein to the Golgi.

The biosynthesis of hybrid and complex glycosylation protein Golgi complex where additional enzymes are responsible for further modification and terminal elaboration of the attached *N*-glycan structures as glycoproteins migrates through the *cis*-

*medial*- and *trans*-Golgi processes. Glycosylated proteins with numerous glycoforms are produced in this manner. The non-template driven enzymatic biosynthesis of glycan structures results in significant glycan diversity.

A core five-saccharide *N*-glycan structure (from the original fourteen-saccharide structure) is conserved and increasingly terminally modified in higher eukaryotes, thereby yielding highly diverse *N*-glycosylation (Figure 1.4). Yeast express a high mannose form of *N*-glycosylation.<sup>105</sup> Plants express both high mannose and more complex forms of *N*-glycosylation.<sup>106</sup> Animals have evolved the most complex *N*-glycan structures reflected by the highest diversity of terminal modifications.<sup>99</sup>

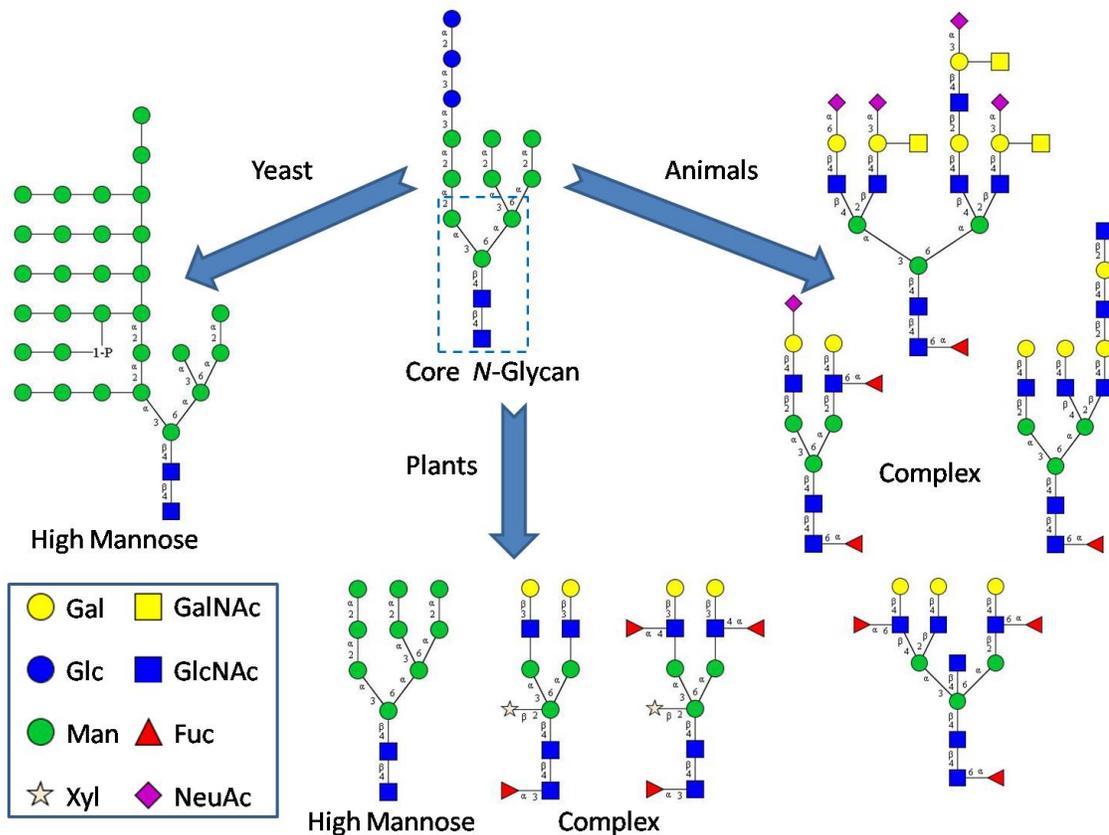


Figure 1.4: Representative examples of *N*-glycan complexity. The fourteen saccharide *N*-glycan structure which is attached *en bloc* to nascent polypeptides prior to modification is shown in the center. The highly conserved five-saccharide *N*-glycan core structure is designated by the dashed box.

*O*-linked glycosylation is defined by covalent attachment of core saccharide(s) to the hydroxyl group of serine and threonine residues.<sup>107, 108</sup> Two major classes of *O*-glycans consist of mucins and proteoglycans. Unlike *N*-glycosylation, which consists of a large core *N*-glycan structure that is trimmed and terminally modified, mucin-type *O*-glycans consist of smaller 8 core structures that lead to significant *O*-glycan diversity. These core structures are similar to the terminal modifications found on *N*-glycans and are enzymatically attached to proteins only in the Golgi complex.

Mucins consist of long polypeptides with repeating Ser/Thr sequences that serve as attachment sites of core *O*-glycan structures. The formation of disulfide-linked oligomers can result in the formation of mucins larger than 1 MDa. Unlike mucins, proteoglycans consist of longer repeating oligosaccharide chains (>100 monosaccharide residues) attached to a polypeptide backbone. Often the oligomer consists of repeating amino derived disaccharide hexoses, which are known predominantly as glycosaminoglycans (GAGs). The 3 types of GAGs, which are differentiated according to the repeating disaccharide unit, are: 1) dermatan sulfate/chondroitin sulfate, 2) heparin sulfate/heparin and 3) keratan sulfate. Proteoglycans are a major component of extracellular matrices and connective tissues. In addition to mucins and proteoglycans, other types of *O*-glycans include  $\alpha$ -linked *O*-fucose,  $\beta$ -linked *O*-xylose,  $\alpha$ -linked *O*-mannose,  $\beta$ -linked *O*-GlcNAc,  $\alpha$ - or  $\beta$ -linked *O*-galactose, and  $\alpha$ - or  $\beta$ -linked *O*-glucose glycans.<sup>108</sup>

The non-template driven process of enzymatic biosynthesis of *N*- and *O*-glycans provides significant diversity to protein structure and function by post-translational modification via glycosylation. Examples of *O*- and *N*-linked glycan chemical and

structural diversity are presented in Figure 1.5.<sup>96</sup> Variations in glycan synthesis provide added complexity in the form of variant glycoforms of each protein. Given that proteins frequently have multiple glycosylation sites and each site can have various glycoforms, deciphering the complexity of glycan biosynthesis and the downstream roles of glycoproteins and their glycoforms is an immense challenge.

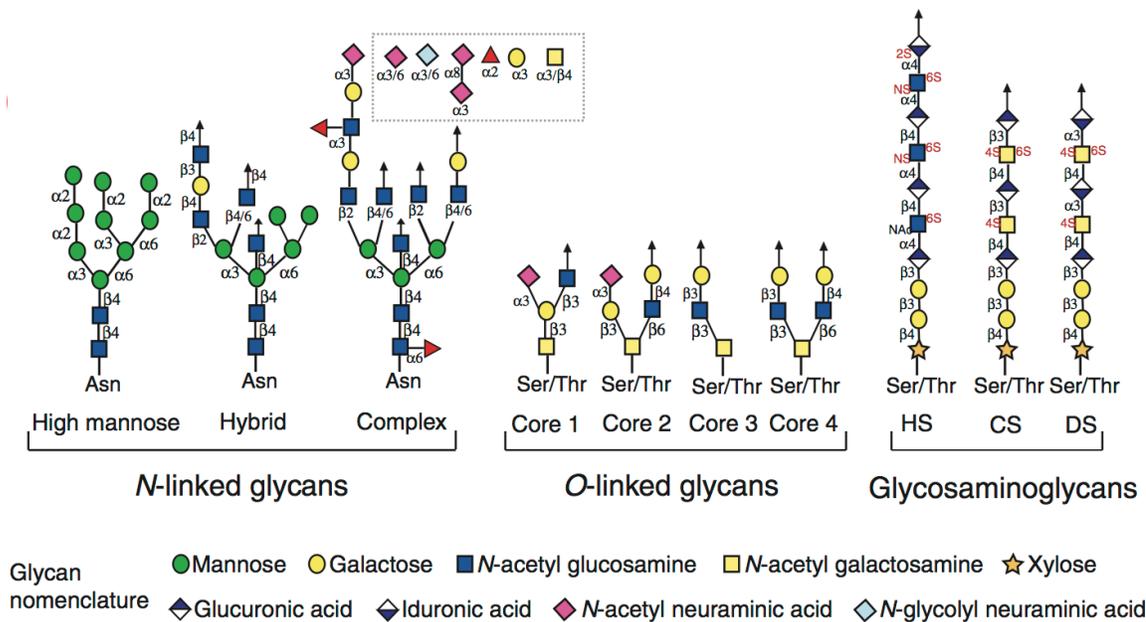


Figure 1.5: Chemical diversity of glycans. Different classes of glycans in the symbol nomenclature developed as a collaborative effort to homogenize glycan representation. Directionality is from nonreducing end at the top to the reducing end at the bottom with the arrows indicating the extension at the nonreducing end. Linkages between monosaccharides contain the anomeric configuration of the monosaccharide ( $\alpha$ , alpha and  $\beta$ , beta) and the oxygen atom in the reducing end monosaccharide to which it is linked to. “/” is used to represent either-or case ( $\beta 3/4$  means  $\beta 3$  or  $\beta 4$ ). In the case of complex N-linked glycans, the common terminal motifs attached to Gal are shown in a dotted box. Abbreviations HS, CS and DS correspond to heparin or heparin sulfate, chondroitin sulfate and dermatan sulfates, respectively. Reprinted by permission from Macmillan Publishers Ltd: Nature Methods, Raman, R., *et al.*, Glycomics: an integrated systems approach to structure-function relationships of glycans. 2, 817-824, copyright 2005.

### 1.3.3 Glycan Recognition

Glycans are recognized by several classes of proteins, including lectins, antibodies, and enzymes. Lectins, glycan-binding proteins (many of which require metal

ions for function), frequently have millimolar to micromolar affinities and increased avidity effects due to multivalent interactions enhances affinity.<sup>80, 109</sup> Although some lectins can discriminate between dissimilar structures, most lectins display remarkably broad specificity, towards similar carbohydrate structures.<sup>110-112</sup> Historically, lectins have been identified from plant or fungal sources, although an increasing number are being identified in animals. Lectin affinity chromatography is the most widely applied glycan, glycopeptide, or glycoprotein isolation technique. However, for whole glycoproteome studies, a limitation of this approach is that it biases glycan detection to a subset of glycoproteins based on the selection of lectin column(s).<sup>80</sup>

It is important to note that the capacity for glycans to be branched, and to display differences among linkage configurations, results in their recognition being highly influenced by both the composition and 3D structure of the glycan.<sup>113</sup> In addition, when the structural similarity of monosaccharides is taken into account, it is common for biologically-unrelated glycans cross react with the same lectin or antibody in a concentration-dependent manner. Thus if enough glycan or protein is present, a weaker, but nevertheless specific, interaction may be detected and potentially misinterpreted.<sup>82, 114, 115</sup> For example, Wheat Germ agglutinin and *Urtica dioica* agglutinin (UDA) are known to recognize both terminal *N*-acetylglucosamine (GlcNAc) and neuraminic acid (Neu5Ac) in the same binding site, by virtue of the fact that these monosaccharides may be oriented in such a way that they present a common 3D binding motif.<sup>116, 117</sup> In addition, UDA recognizes both chitotriose (GlcNAc $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc) associated with fungal cell surfaces, and the mannose (Man)-containing trisaccharide Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc common to *N*-linked glycans.<sup>118, 119</sup> Because such cross-

reactivities are inseparable, they present the core challenge in generating or applying reagents for the characterization of glycan composition. The choice of reagent used for sample enrichment or isolation can therefore bias the outcome of glycomic analyses toward a subset of glycoconjugates based on the binding properties of the lectin or antibody.<sup>80</sup>

Antibodies recognize glycan structures with greater affinity and specificity than lectins; however, they are difficult to generate given that carbohydrates are poor immunogens in general. Thus only a limited selection of anti-carbohydrate antibodies is available and many display cross-reactivity to similar glycan structures.<sup>82, 114</sup>

In contrast to lectins or antibodies, glycan-processing enzymes are often exquisitely selective with regard to substrate structure, reflecting their essential role in glycan processing. Glycosyl hydrolases generally recognize both of the monosaccharide residues comprising the glycosidic linkage, and are often specific for position and configuration of the linkage. For example, the enzymes Endo- $\beta$ -*N*-acetylglucosaminidase H (Endo H) and chitinase, from various sources, are all members of family 18 of the glycohydrolases and share have similar tertiary structures. Despite these similarities, Endo H is exclusively active on the GlcNAc $\beta$ 1-4GlcNAc linkage when present in the *N*-glycan core sequence Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc sequence; it does not hydrolyze the same linkage in chitin.<sup>120</sup> This specificity contrasts with that seen for the lectin UDA. Additionally, many carbohydrate-processing enzymes have non-catalytic carbohydrate-binding modules, which serve to enhance the specificity of enzyme-substrate interactions.<sup>121</sup> Site-directed mutagenesis is often employed to generate inactive mutants, facilitating the characterization of substrate specificity.<sup>120</sup>

An interesting opportunity therefore exists to employ inactive mutants of carbohydrate-processing enzymes as reagents for detecting substrate. Such lectin-like enzyme-derived (Lectenz®) reagents would in principle have the advantage of retaining the inherent specificity of the wild-type enzyme. Indeed examples exist where a single point mutation in an enzyme can lead to a reagent capable of being used in an affinity column to capture specific glycans (polysialic acid) or peptides (anhydrotrypsin).<sup>122, 123</sup> However, as enzymes have evolved for turnover, simple inactive point mutants do not generally have affinities high enough to be practical reagents.

#### 1.4 PNGase F

Peptide:*N*-glycanase (PNGase) enzymes (Table 1.4) are a class of *N*-glycan releasing enzymes that catalyze the cleavage of the amide bond between the asparagine side chain of the polypeptide and the proximal *N*-acetyl- $\beta$ -D-glycosamine (GlcNAc) of the *N*-glycan. The hydrolysis reaction results in the release of the glycan and free ammonia, and conversion of the asparagine to an aspartic acid (Figure 1.6).

Table 1.4: Peptide:*N*-glycanase nomenclature.

Systematic name	<i>N</i> -linked-glycopeptide-( <i>N</i> -acetyl- $\beta$ -D-glycosaminyl)-L-asparagine aminohydrolase
Recommended name	Peptide- <i>N</i> <sup>4</sup> -( <i>N</i> -acetyl- $\beta$ -D-glycosaminyl)asparagine amidase
Synonyms	PNGase, <i>N</i> -oligosaccharide glycopeptidase, Glycopeptidase, Glycoamidase, <i>N</i> -Glycanase
Enzyme Commission #	EC 3.5.1.52

##### 1.4.1 Discovery of PNGase F

The *N*-glycan processing enzyme, Peptide-*N*<sup>4</sup>-(*N*-acetyl- $\beta$ -D-glucosaminyl)asparagine amidase (PNGase F) was identified from the gram negative soil bacterium *Flavobacterium meningosepticum* (formerly known as *Chryseobacterium meningosepticum* and *Elizabethkingia meningosepticum*) by Plummer *et al.* in 1984.<sup>124</sup>

PNGase enzymes have been identified from various species across plants, animals, and fungi; however, PNGase F stands in contrast to these other PNGase enzymes as it is the only confirmed bacterial PNGase enzyme 30 years since its initial discovery.

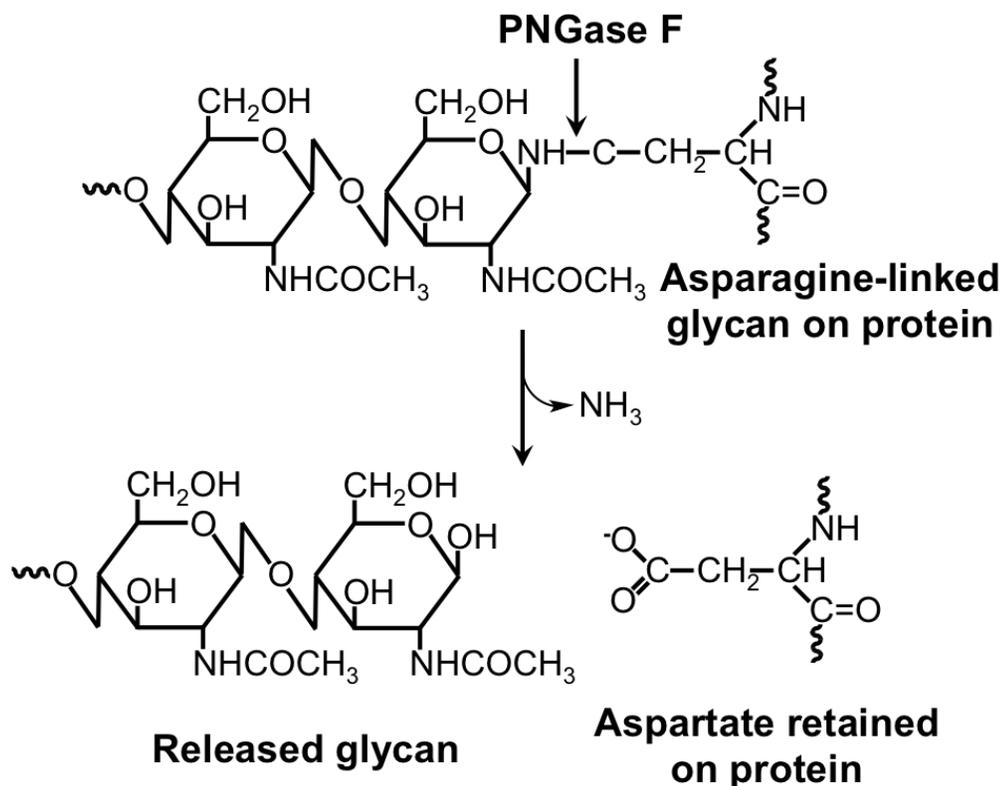


Figure 1.6: The PNGase F deglycosylation reaction. PNGase F catalyzes the release of *N*-linked glycans from the polypeptide backbone by cleaving the *N*-glycosidic bond (amide bond) between the asparagine side chain and the proximal GlcNAc. In addition to the released of free ammonia, the asparagine on the polypeptide protein backbone is converted to an aspartic acid.

Early studies of PNGase F indicated that the enzyme could catalyze the release of all *N*-glycans.<sup>124</sup> However, this was due to enzyme preparations containing a mixture of PNGase F and Endo- $\beta$ -*N*-acetylglucosaminidase F (Endo F) from *F. meningosepticum*.<sup>125</sup> Endo F cleaves the glycosidic bond of the chitobiose moiety where as PNGase F cleaves the amide bond at the glycosylaminy junction.<sup>126</sup> These results confirmed that PNGase F was in fact a peptide: *N*-glycosidase and not an endoglycosidase, resulting in its

reclassification. Additional experiments using fetuin glycopeptides and erythropoietin from Chinese hamster ovary cells indicated that the activity of the enzyme was markedly improved on denatured glycoproteins that had been pretreated with detergents, requiring significantly less enzyme for deglycosylation.<sup>125</sup> However, optimal reaction conditions including buffer composition weren't established until later studies that demonstrated decreased PNGase F activity in the presence of some detergents and metal ions. These studies also confirmed optimal enzymatic activity at pH 8.0 and buffer compositions were optimized to use Tris buffer with no sodium chloride.<sup>126, 127</sup>

Cloning and heterologous expression of PNGase F in *E. coli*, in 1989, allowed high purity preparations for continued study and led to its rapid adoption for total *N*-glycoprotein deglycosylation.<sup>128</sup> However, Tretter, *et al.* demonstrated in 1991 that in contrast to PNGase A, core  $\alpha$ 1,3 fucosylation of the asparagine-linked GlcNAc conferred resistance of a glycopeptide or glycoprotein to PNGase F.<sup>129</sup> Shortly after in 1994, two three-dimensional x-ray crystal structures of PNGase F (PDB IDs 1PNG & 1PGS) were obtained (sans ligand), leading to significant interest in identifying the active site and the hypothesis that the reason  $\alpha$ 1,3 fucosylation confers resistance is likely due to the C3 position of the asparagine-linked GlcNAc being buried into the hydrophobic groove of the binding cleft.<sup>130-132</sup> This hypothesis was confirmed when x-ray crystallography data was obtained for a co-crystallized PNGase F: chitobiose complex.

#### 1.4.2 The X-Ray Crystal Structure of PNGase F

The first structure of PNGase F co-crystallized with the chitobiose ligand, *N,N'*-diacetylchitobiose, (PDB ID 1PNF) was published in 1995 by Kuhn, *et al.* at 2.0Å resolution.<sup>133</sup> The annotated sequence of the crystallized PNGase F enzyme is presented in Figure 1.7. Consistent with the uncomplexed structures (PDB IDs 1PNG & 1PGS),

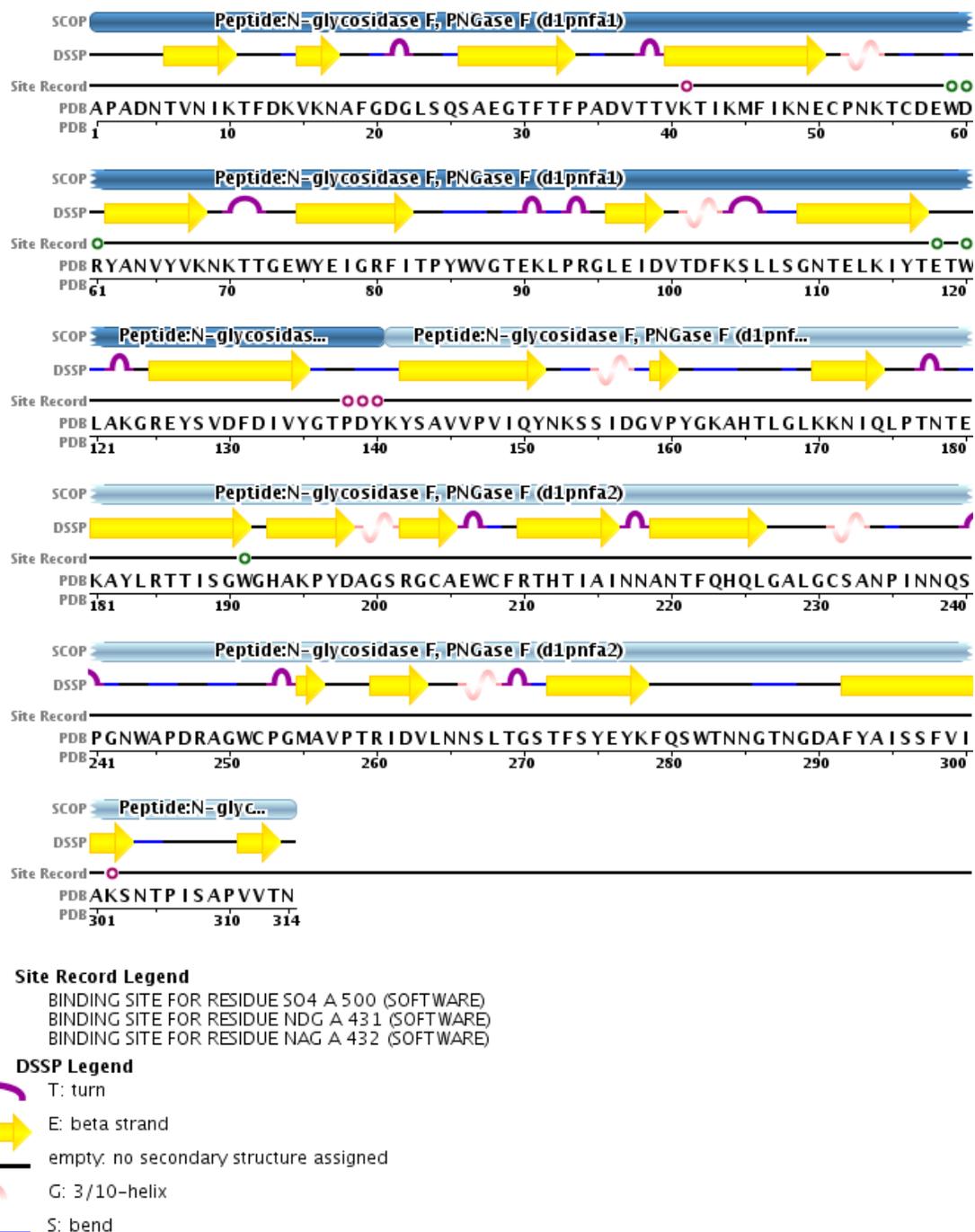


Figure 1.7: Sequence display of PNGase F (PDB ID: 1PNF). The 314 amino acid sequence is depicted and annotated. The two domains are labeled d1pnfa1 & d1pnfa2. Residue 431 corresponds with the reducing GlcNAc and 432 with the 2<sup>nd</sup> GlcNAc of the chitobiose ligand. Three disulphide bonds are located at 51-56, 204-208, and 231-252. Image from the RCSB PDB ([www.rcsb.org](http://www.rcsb.org)) of PDB ID 1PNF (Kuhn, P. et al. (1995) Active Site and Oligosaccharide Recognition Residues of Peptide-*N*<sup>4</sup>-(*N*-acetyl- $\beta$ -D-glucosaminyl)asparagine Amidase F. Journal of Biological Chemistry 270, 29493-29497).

there were no significant changes to the conformation of the complexed PNGase F: chitobiose structure, indicating that the conformation is unaffected by binding of the chitobiose ligand. The folded protein consists of two domains, which are comprised of residues 1-137 and 143-314 respectively. Both domains have eight-stranded antiparallel  $\beta$ -sandwiches that lie adjacent to each other such that the interface runs the full length of the  $\beta$ -sheets with extensive hydrogen bonding contacts. Three possible binding sites had been postulated based on three grooves in the uncomplexed structures.<sup>132</sup> The first groove, a bowl shape, on one face of the molecule, contained residues similar to the active site of L-asparaginases.<sup>132</sup> A shallow S-shaped cleft on the opposite face containing a number of acidic residues and threonine residues was postulated as a second possible

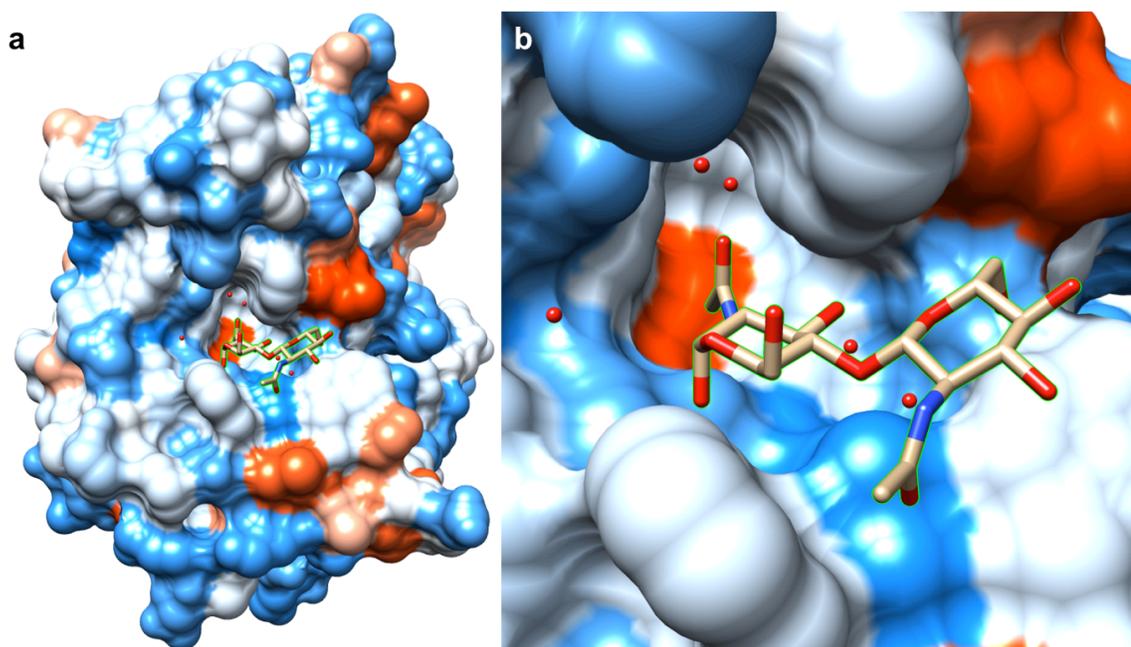


Figure 1.8: Surface hydrophobicity representation of the PNGase F-chitobiose complex (PDB ID 1PNF). a) The co-crystallized complex is shown emphasizing the orientation of the ligand in the binding site located at the interface of the two domains. b) Close up view of the binding cleft shows the orientation of the  $\alpha$ -chitobiose ligand with the *N*-acetyl group of the reducing GlcNAc extended into a deep hydrophobic pocket. Five water molecules are positioned between the protein and chitobiose interface. The *N*-acetyl group of the second GlcNAc is facing the solvent accessible side of the binding cleft.

binding site.<sup>132</sup> A deep cleft at the interface between the two domains at one end of the molecule was postulated as a third binding site. This cleft, containing several acidic residues and serines, possessed the unique attribute of having five tryptophan residues.<sup>132</sup>

The IPNF structural model confirms the deep cleft at the interface of the two domains as the binding cleft for the chitobiose ligand. A surface hydrophobicity representation of the structural model is shown in Figure 1.8a. The orientation of the  $\alpha$ -chitobiose ligand in the binding cleft shows the *N*-acetyl group of the reducing GlcNAc extended into a deep hydrophobic pocket (Figure 1.8b). Five water molecules are positioned between the protein and chitobiose interface. The *N*-acetyl group of the second GlcNAc is facing the solvent accessible side of the binding cleft. The C3 position of the reducing GlcNAc is facing into the binding cleft, confirming that there would be no space for the glycan to fit into this groove if it is 1,3 fucosylated. Unlike the C3 position, the C6 position is pointed outwards towards the solvent exposed side of the cleft, indicating that  $\alpha$ 1,6 fucosylation at this position does not sterically hinder access to the binding cleft.

An extensive network of hydrogen bond interactions is also evidence between the protein and ligand, many of which are facilitated through 5 water molecules positioned in the interface between the protein and the ligand (Wat<sup>75</sup>, Wat<sup>146</sup>, Wat<sup>346</sup>, Wat<sup>348</sup>, Wat<sup>349</sup>). Three of these water molecules (Wat<sup>75</sup>, Wat<sup>146</sup>, Wat<sup>346</sup>) are also present in nearly identical positions in the uncomplexed structures.<sup>130, 132</sup> A total of 10 residues (D60, R61, Y85, E118, W120, S155, G190, W191, E206, R248) are involved in the network of hydrogen bonds with water molecules and the ligand. A schematic diagram showing the intermolecular hydrogen bonding contacts as originally published by Kuhn, *et al.* is

reproduced in Figure 1.9.<sup>133</sup> A three-dimensional representation of this schematic network of hydrogen bonding contacts is presented in Figure 1.10.

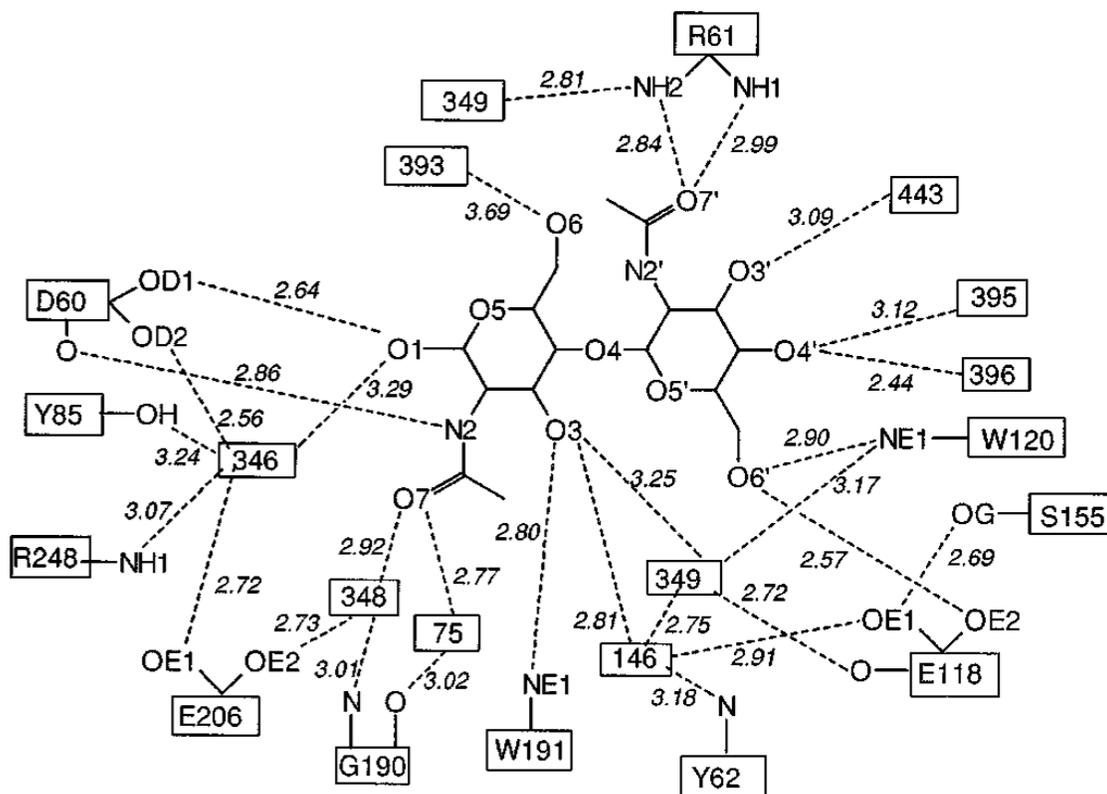


Figure 1.9: Schematic diagram showing the intermolecular hydrogen bonding contacts between PNGase F, *N,N'*-diacetylchitobiose and water molecules. Protein residues are indicated with single-letter amino acid code and sequence number in rectangular boxes, water molecules are indicated by a number, corresponding to their number in the file deposited with the Protein Data Bank. The reducing end GlcNAc residue is on the left. Hydrogen bonding distances, in Å, are shown in *italics*. Note that Wat<sup>349</sup> (349) is present twice, once in contact with O3 and one in Arg-61. This research was originally published in the Journal of Biological Chemistry. Kuhn, P. *et al.* Active Site and Oligosaccharide Recognition Residues of Peptide-*N*<sup>4</sup>-(*N*-acetyl- $\beta$ -D-glucosaminyl)asparagine Amidase F. Journal of Biological Chemistry. 1995; 270: 29493-29497. © the American Society for Biochemistry and Molecular Biology.

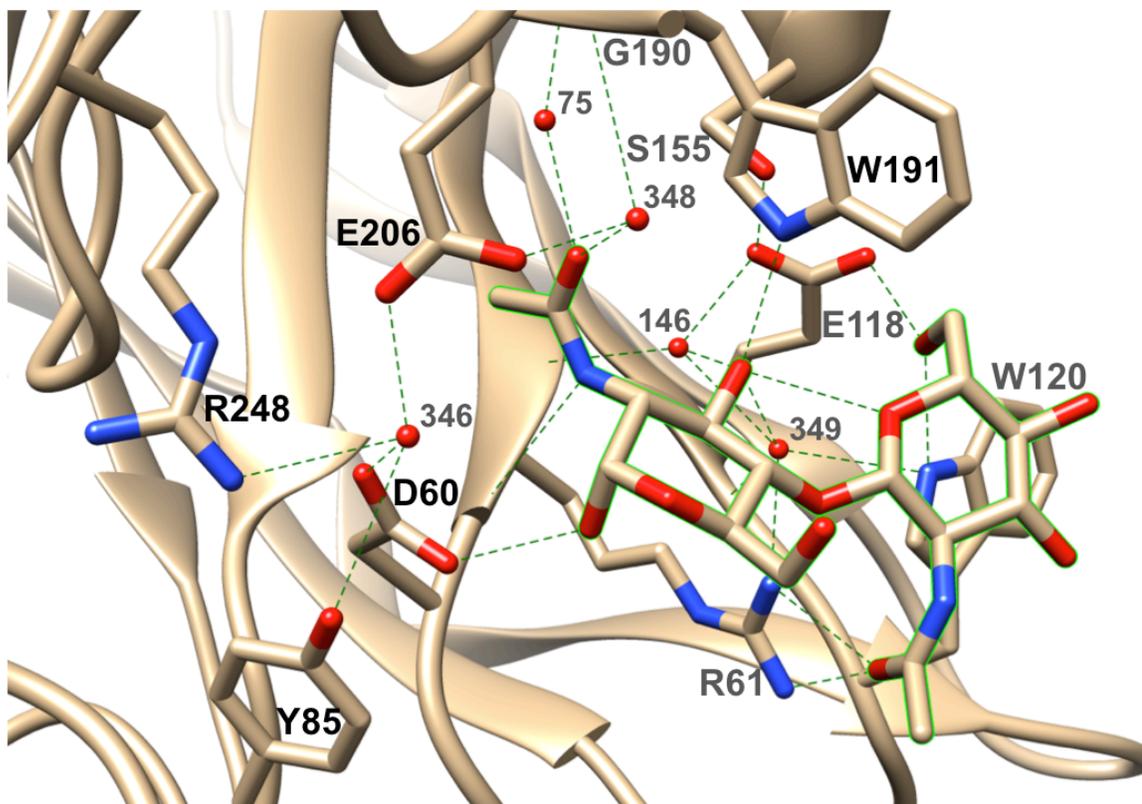


Figure 1.10: Active site hydrogen bond network of the PNGase F:chitobiose complex. The hydrogen bond network in the binding cleft of PNGase F with the  $\alpha$ -chitobiose ligand (green outline) is shown based on experimental x-ray data (PDB ID: 1PNF). Water molecules in the binding cleft between the protein the ligand are depicted as red spheres: Wat<sup>75</sup>, Wat<sup>146</sup>, Wat<sup>346</sup>, Wat<sup>348</sup>, Wat<sup>349</sup>. Amino acids in the binding cleft involved in hydrogen bonds: D60, R61, Y85, E118, W120, S155, G190, W191, E206, R248.

#### 1.4.3 Active Site Residues of PNGase F

Point mutagenesis studies of active site residues in PNGase F have identified D60 as the primary catalytic residue and E206 as likely contributing to stabilization of reaction state intermediates.<sup>133</sup> Based on the position of the chitobiose ligand, D60 and E206 would span both sides of the amine bond that the enzyme would cleave. However, structures of PNGase F complexed with a glycopeptide have not been reposted, thus a mechanism has yet to be confirmed. Mutagenesis studies of E118, which is at the opposite end of the ligand interacting with O6 of the 2<sup>nd</sup> GlcNAc, indicates that E118 is

critical for substrate recognition, a prerequisite for catalytic activity likely mediated by D60 and E206. A model for the reaching mechanism has been proposed facilitated by D60A, E206, and R248, with D60 as the primary catalytic residue.<sup>132, 134</sup> In this model, R248 is postulated to form a hydrogen bond with the carbonyl oxygen of the *N*-glycosidic bond, thus making the Asn-carbonyl carbon more susceptible to nucleophilic attack by a hydroxide ion. This nucleophilic attack would be facilitated by Wat<sup>346</sup> (Wat<sup>422</sup> in PDB ID 1PGS), which is present in both complexed and uncomplexed structures, and is located proximal to D60, E206, and R248. A hydroxide ion could be formed by transfer of a proton from Wat<sup>346</sup> to D60. The Asn-carbonyl carbon would undergo nucleophilic attack from the hydroxide ion, forming the transition state intermediate. D60 would donate its proton to the nitrogen of the amine bond, completing the cleavage of the amide bond. This proposed model would require the pK<sub>a</sub> of D60 being raised from 4.5 to ~8.0, the pH optimum of the reaction. Such a shift in the local environment of the active site could be made feasible by a hydrophobic environment caused by nearby aromatic residues (Y85, W251, W207, and W191) surrounding E206 and D60.

#### *1.4.4 The Significance of PNGase F*

Since the discovery of PNGase F 30 years ago, it has become a standard tool for releasing *N*-linked glycans prior to characterization. PNGase F has the broadest specificity for *N*-glycans bearing glycoproteins, as it recognizes both the chitobiose core as well as the asparagine-linked peptide motif common to *N*-glycan peptide and protein conjugates. Substrate specificity studies have confirmed that the minimum glycan motif required for catalytic activity is the chitobiose core.<sup>135</sup> In addition, the minimal peptide motif recognized is the Asn-X-Ser/Thr glycosylation motif common to all *N*-linked glycans.<sup>135</sup> Interestingly, optimal enzyme activity was observed with the chitobiose-linked

pentapeptide, Tyr-Ile-Asn-Ala-Ser, indicating that the enzyme recognizes residues both upstream and downstream of the glycan-linked asparagine.<sup>135</sup>

Given the broad specificity of PNGase F for the *N*-glycopeptide core common to all *N*-glycan bearing glycoproteins, this enzyme would be an ideal candidate for engineering into an *N*-glycopeptide recognizing reagent. There is currently no single detection reagent that can recognize and enrich *N*-glycopeptides. Such a reagent would be of significant interest to the glycoscience community and engineering it is the focus of this dissertation.

## 1.5 Dissertation Overview

Recent advances in both theoretical and experimental approaches present unique opportunities to advance the field of glycomics. Specifically, employing computational chemistry and structural biology tools with high throughput directed evolution strategies makes feasible the rational *in silico* design of novel protein libraries focused towards identifying clones with desired functionality.<sup>8-15</sup> Computational docking and molecular dynamics have become indispensable tools for investigating the highly complex and flexible nature of protein-glycan interactions.<sup>9, 70</sup> Furthermore, determining binding free energies to evaluate thermodynamic contributions that drive the binding interactions is a powerful computational technique that provides insight into protein-ligand interactions broken down per amino acid that cannot be determined any other way.<sup>8, 10, 12</sup> These computational tools serve to advance understanding of biomolecular interactions and guide the development of biomolecules with novel functionality. Coupling *in silico* structural analysis, molecular dynamics (MD) and binding free energy decomposition strategies with *in vitro* directed evolution will enable knowledge-based protein

engineering that will not only advance both disciplines but also spur the development of novel biomolecules relevant to the field of glycomics.<sup>3, 31, 40</sup>

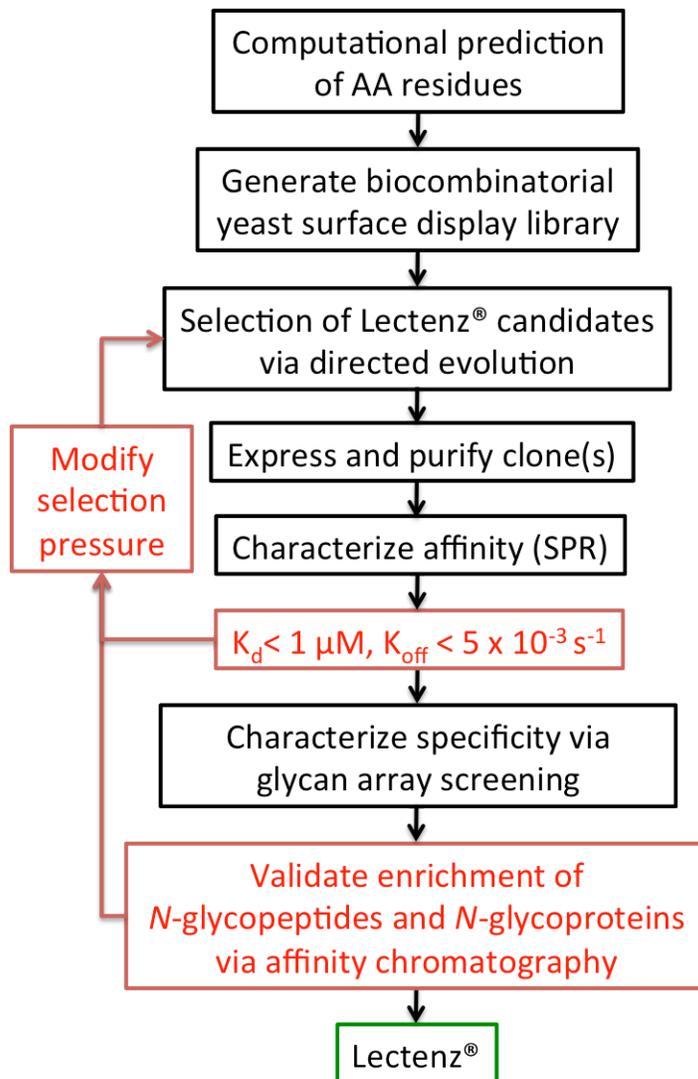


Figure 1.11: Schematic representation of the Lectenz® design strategy. Shown is the integrative strategy utilizing computational methods, knowledge-based library design, selection, and down stream characterization and validation. Red boxes indicate checkpoints where if a selected candidate fails to meet the desired threshold, the selection process can be repeated with modified selection conditions. Once a selected candidate satisfies the specificity and affinity characterization requirements, it is coupled to an affinity matrix to validate affinity chromatography based enrichment of *N*-glycopeptides and *N*-glycoproteins.

Given the advances in *in silico* and *in vitro* protein engineering methodologies and the need for novel glycan detecting reagents, this dissertation describes a novel lectin-like glycan-recognizing biomolecule engineered from a glycan-processing enzyme, which are called Lectenz®. (Lectenz® is a federally registered trademark of Glycosensors & Diagnostics, LLC.) Since glycan-processing enzymes have exquisite specificity for their glycan substrates, they serve as an ideal starting point to generate a catalytically inactive variant for affinity enhancement. Specifically, the *Flavobacterium meningosepticum* *N*-glycan processing enzyme, PNGase F, is engineered into a catalytically-inactive, affinity-enhanced variant for detecting the core glycopeptide component common to all *N*-linked glycans.

A schematic of the Lectenz® design strategy is presented in Figure 1.11. The wild-type PNGase F enzyme, has previously been co-crystallized (PDB ID: 1PNF) with the *N,N'*-diacetylchitobiose disaccharide in the active site at 2.0 Å resolution.<sup>133</sup> This structural model is used to conduct molecular dynamics simulations and binding free energy decomposition analysis to identify critical and tepid amino acid residues proximal to the chitobiose ligand. Critical residues are not selected for saturation mutagenesis, where as tepid residues with weak ligand-binding interaction energies are selected for saturation mutagenesis via directed evolution. Directed evolution is performed using a yeast display system to select for mutagenized PNGase F clones with affinity for the target *N*-glycan bearing glycoprotein, Ribonuclease B (RNase B). The selected Lectenz®, R911, is characterized via surface plasmon resonance for kinetic analysis, glycan array screening for specificity determination, and employed in Lectenz® affinity chromatography for *N*-glycopeptide and *N*-glycoprotein sample enrichment.

The successful creation of a lectin-like reagent from a carbohydrate processing enzyme (a Lectenz®) presents not only a unique solution to the challenge of *N*-glycopeptide and *N*-glycoprotein sample enrichment, but also demonstrates a novel strategy for engineering glycan-targeting reagents for glycans and glycoconjugates of biological relevance.

## 1.6 References

1. Wijma, H.J. et al. Computationally designed libraries for rapid enzyme stabilization. *Protein Eng Des Sel* **27**, 49-58 (2014).
2. Wijma, H.J. & Janssen, D.B. Computational design gains momentum in enzyme catalysis engineering. *Febs J* **280**, 2948-2960 (2013).
3. Feldmeier, K. & Höcker, B. Computational protein design of ligand binding and catalysis. *Current Opinion in Chemical Biology* **17**, 929-933 (2013).
4. Tinberg, C.E. et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212-216 (2013).
5. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-1391 (2008).
6. Patrick, W.M. & Firth, A.E. Strategies and computational tools for improving randomized protein libraries. *Biomolecular Engineering* **22**, 105-112 (2005).
7. Chica, R.A., Doucet, N. & Pelletier, J.N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology* **16**, 378-384 (2005).
8. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling* **51**, 69-82 (2011).
9. Woods, R.J. & Tessier, M.B. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Current Opinion in Structural Biology* **20**, 575-583 (2010).
10. Steinbrecher, T. & Labahn, A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr Med Chem* **17**, 767-785 (2010).
11. D.A. Case, T.A.D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A.

- Roitberg, G. Seabra, I. Kolossvai, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (University of California, San Francisco, 2010).
12. Carrascal, N. & Green, D.F. Energetic decomposition with the generalized-born and Poisson-Boltzmann solvent models: lessons from association of G-protein components. *The journal of physical chemistry. B* **114**, 5096-5116 (2010).
  13. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4**, 435-447 (2008).
  14. Okimoto, N. et al. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *PLoS Comput Biol* **5**, e1000528 (2009).
  15. Berman, H.M. et al. The Protein Data Bank. *Nucl. Acids Res.* **28**, 235-242 (2000).
  16. Karanicolas, J. et al. A De Novo Protein Binding Pair By Computational Design and Directed Evolution. *Mol Cell* **42**, 250-260 (2011).
  17. Lutz, S. Beyond directed evolution--semi-rational protein engineering and design. *Current Opinion in Biotechnology* **21**, 734-743 (2010).
  18. Grove, T.Z., Hands, M. & Regan, L. Creating novel proteins by combining design and selection. *Protein Eng Des Sel* **23**, 449-455 (2010).
  19. Brannigan, J.A. & Wilkinson, A.J. Protein engineering 20 years on. *Nature Reviews. Molecular Cell Biology* **3**, 964-970 (2002).
  20. Winter, G., Fersht, A.R., Wilkinson, A.J., Zoller, M. & Smith, M. Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature* **299**, 756-758 (1982).
  21. Sigal, I.S., Harwood, B.G. & Arentzen, R. Thiol-beta-lactamase: replacement of the active-site serine of RTEM beta-lactamase by a cysteine residue. *Proc Natl Acad Sci U S A* **79**, 7157-7160 (1982).
  22. Hutchison, C.A., 3rd et al. Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* **253**, 6551-6560 (1978).
  23. Leatherbarrow, R.J., Fersht, A.R. & Winter, G. Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc Natl Acad Sci U S A* **82**, 7840-7844 (1985).

24. Graf, L. et al. Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry* **26**, 2616-2623 (1987).
25. Perona, J.J., Hedstrom, L., Rutter, W.J. & Fletterick, R.J. Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* **34**, 1489-1499 (1995).
26. Venekei, I., Szilagyi, L., Graf, L. & Rutter, W.J. Attempts to convert chymotrypsin to trypsin. *FEBS Lett* **379**, 143-147 (1996).
27. Lerner, S.A., Wu, T.T. & Lin, E.C. Evolution of a Catabolic Pathway in Bacteria. *Science* **146**, 1313-1315 (1964).
28. Mills, D.R., Peterson, R.L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci U S A* **58**, 217-224 (1967).
29. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317 (1985).
30. Tobin, M.B., Gustafsson, C. & Huisman, G.W. Directed evolution: the 'rational' basis for 'irrational' design. *Curr Opin Struct Biol* **10**, 421-427 (2000).
31. Bonsor, D.A. & Sundberg, E.J. Dissecting protein-protein interactions using directed evolution. *Biochemistry* **50**, 2394-2402 (2011).
32. Cobb, R.E., Chao, R. & Zhao, H. Directed evolution: Past, present, and future. *AIChE Journal* **59**, 1432-1440 (2013).
33. Socha, R.D. & Tokuriki, N. Modulating protein stability - directed evolution strategies for improved protein function. *Febs J* **280**, 5582-5595 (2013).
34. Stone, J.D., Chervin, A.S., Aggen, D.H. & Kranz, D.M. T cell receptor engineering. *Methods Enzymol* **503**, 189-222 (2012).
35. Shim, J.H., Chen, H.M., Rich, J.R., Goddard-Borger, E.D. & Withers, S.G. Directed evolution of a beta-glycosidase from *Agrobacterium* sp. to enhance its glycosynthase activity toward C3-modified donor sugars. *Protein Eng Des Sel* **25**, 465-472 (2012).
36. Patel, S.C. & Hecht, M.H. Directed evolution of the peroxidase activity of a de novo-designed protein. *Protein Eng Des Sel* **25**, 445-452 (2012).
37. Yip, S.H. et al. Directed evolution combined with rational design increases activity of GpdQ toward a non-physiological substrate and alters the oligomeric structure of the enzyme. *Protein Eng Des Sel* (2011).

38. Jakeman, D.L. & Sadeghi-Khomami, A. A beta-(1,2)-glycosynthase and an attempted selection method for the directed evolution of glycosynthases. *Biochemistry* **50**, 10359-10366 (2011).
39. Cobucci-Ponzano, B., Perugino, G., Rossi, M. & Moracci, M. Engineering the stability and the activity of a glycoside hydrolase. *Protein Eng Des Sel* **24**, 21-26 (2011).
40. Brustad, E.M. & Arnold, F.H. Optimizing non-natural protein function with directed evolution. *Current Opinion in Chemical Biology* **15**, 201-210 (2011).
41. Lopes, A., Schmidt Am Busch, M. & Simonson, T. Computational design of protein-ligand binding: modifying the specificity of asparaginyl-tRNA synthetase. *Journal of computational chemistry* **31**, 1273-1286 (2010).
42. Kittl, R. & Withers, S.G. New approaches to enzymatic glycoside synthesis through directed evolution. *Carbohydrate Research* **345**, 1272-1279 (2010).
43. Yu, L. et al. Phage display screening against a set of targets to establish peptide-based sugar mimetics and molecular docking to predict binding site. *Bioorganic & Medicinal Chemistry* **17**, 4825-4832 (2009).
44. Thompson, S.M. et al. Heparan sulfate phage display antibodies identify distinct epitopes with complex binding characteristics: insights into protein binding specificities. *The Journal of biological chemistry* **284**, 35621-35631 (2009).
45. Hancock, S.M., Rich, J.R., Caines, M.E., Strynadka, N.C. & Withers, S.G. Designer enzymes for glycosphingolipid synthesis by directed evolution. *Nature chemical biology* **5**, 508-514 (2009).
46. Belien, T., Verjans, P., Courtin, C.M. & Delcour, J.A. Phage display based identification of novel stabilizing mutations in glycosyl hydrolase family 11 B. subtilis endoxylanase XynA. *Biochemical and Biophysical Research Communications* **368**, 74-80 (2008).
47. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755-768 (2006).
48. Gera, N., Hussain, M. & Rao, B.M. Protein selection using yeast surface display. *Methods* **60**, 15-26 (2013).
49. Tohidkia, M.R., Barar, J., Asadi, F. & Omidi, Y. Molecular considerations for development of phage antibody libraries. *Journal of drug targeting* **20**, 195-208 (2012).
50. Kenrick, S.A. & Daugherty, P.S. Bacterial display enables efficient and quantitative peptide affinity maturation. *Protein Eng Des Sel* **23**, 9-17 (2010).

51. Dreier, B. & Pluckthun, A. Ribosome display: a technology for selecting and evolving proteins from large libraries. *Methods in molecular biology* **687**, 283-306 (2011).
52. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX--A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24**, 381-403 (2007).
53. Bornscheuer, U. & Kazlauskas, R.J. Survey of protein engineering strategies. *Curr Protoc Protein Sci* **Chapter 26**, Unit26 27 (2011).
54. Leach, A.R. Molecular modelling : principles and applications, Edn. 2nd. (Prentice Hall, Harlow, England ; New York; 2001).
55. Groenhof, G. Introduction to QM/MM simulations. *Methods Mol Biol* **924**, 43-66 (2013).
56. Karplus, M. & McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646-652 (2002).
57. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **102**, 6679-6685 (2005).
58. Baker, D. & Sali, A. Protein Structure Prediction and Structural Genomics. *Science* **294**, 93-96 (2001).
59. Mackerell, A.D., Jr. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* **25**, 1584-1604 (2004).
60. Stortz, C.A., Johnson, G.P., French, A.D. & Csonka, G.I. Comparison of different force fields for the study of disaccharides. *Carbohydrate Research* **344**, 2217-2228 (2009).
61. Fadda, E. & Woods, R.J. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discovery Today* **15**, 596-609 (2010).
62. Genheden, S. & Ryde, U. A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations. *J Comput Chem* **32**, 187-195 (2011).
63. Kirschner, K.N. et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622-655 (2008).
64. Guvench, O. & MacKerell, A.D., Jr. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol Biol* **443**, 63-88 (2008).

65. Benz, R.W., Castro-Roman, F., Tobias, D.J. & White, S.H. Experimental validation of molecular dynamics simulations of lipid bilayers: a new approach. *Biophysical journal* **88**, 805-817 (2005).
66. Showalter, S.A. & Brüschweiler, R. Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field. *Journal of Chemical Theory and Computation* **3**, 961-975 (2007).
67. McCammon, J.A., Gelin, B.R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).
68. de Ruiter, A. & Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Current Opinion in Chemical Biology* **15**, 547-552 (2011).
69. Christ, C.D., Mark, A.E. & van Gunsteren, W.F. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*, NA-NA (2009).
70. DeMarco, M.L. & Woods, R.J. Structural glycobiology: a game of snakes and ladders. *Glycobiology* **18**, 426-440 (2008).
71. Wang, J., Tan, C., Tan, Y.-H., Lu, Q. & Luo, R. Poisson-Boltzmann Solvents in Molecular Dynamics Simulations. *Communications in computational physics* **3**, 22 (2008).
72. Jorgensen, W.L. Efficient Drug Lead Discovery and Optimization. *Accounts of Chemical Research* **42**, 724-733 (2009).
73. Jorgensen, W.L. The many roles of computation in drug discovery. *Science* **303**, 1813-1818 (2004).
74. McCammon, J.A. Theory of biomolecular recognition. *Current Opinion in Structural Biology* **8**, 245-249 (1998).
75. Hummer, G. & Szabo, A. Calculation of free-energy differences from computer simulations of initial and final states. *Journal of Chemical Physics* **105**, 2004 (1996).
76. van Gunsteren, W.F. et al. Computation of Free Energy in Practice: Choice of Approximations and Accuracy Limiting Factors, Vol. 2. (ESCOM, Leiden; 1993).
77. Beveridge, D.L. & DiCapua, F.M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu Rev Biophys Biophys Chem* **18**, 431-492 (1989).
78. Zoete, V., Irving, M.B. & Michielin, O. MM-GBSA binding free energy decomposition and T cell receptor engineering. *Journal of molecular recognition : JMR* **23**, 142-152 (2010).

79. Pierdominici-Sottile, G., Palma, J. & Roitberg, A.E. Free-energy computations identify the mutations required to confer trans-sialidase activity into *Trypanosoma rangeli* sialidase. *Proteins* **82**, 424-435 (2014).
80. Krishnamoorthy, L. & Mahal, L.K. Glycomic analysis: an array of technologies. *ACS chemical biology* **4**, 715-732 (2009).
81. Arnaud, J., Audfray, A. & Imberty, A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem Soc Rev* **42**, 4798-4813 (2013).
82. Kuzmanov, U., Kosanam, H. & Diamandis, E.P. The sweet and sour of serological glycoprotein tumor biomarker quantification. *BMC medicine* **11**, 31 (2013).
83. Hakomori, S. Tumor-associated carbohydrate antigens. *Annu Rev Immunol* **2**, 103-126 (1984).
84. Porcel, J.M. et al. Use of a panel of tumor markers (carcinoembryonic antigen, cancer antigen 125, carbohydrate antigen 15-3, and cytokeratin 19 fragments) in pleural fluid for the differential diagnosis of benign and malignant effusions. *Chest* **126**, 1757-1763 (2004).
85. Goonetilleke, K.S. & Siriwardena, A.K. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur J Surg Oncol* **33**, 266-270 (2007).
86. Taylor-Papadimitriou, J., Burchell, J., Miles, D.W. & Dalziel, M. MUC1 and cancer. *Biochim Biophys Acta* **1455**, 301-313 (1999).
87. Ghazarian, H., Idoni, B. & Oppenheimer, S.B. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta histochemica* **113**, 236-247 (2011).
88. Hart, G.W. & Copeland, R.J. Glycomics hits the big time. *Cell* **143**, 672-676 (2010).
89. Taniguchi, N., Hancock, W., Lubman, D.M. & Rudd, P.M. The Second Golden Age of Glycomics: From Functional Glycomics to Clinical Applications. *Journal of Proteome Research* **8**, 425-426 (2009).
90. An, H.J., Kronewitter, S.R., de Leoz, M.L. & Lebrilla, C.B. Glycomics and disease markers. *Current Opinion in Chemical Biology* **13**, 601-607 (2009).
91. Freeze, H.H. Update and perspectives on congenital disorders of glycosylation. *Glycobiology* **11**, 129R-143R (2001).

92. Haltiwanger, R.S. & Lowe, J.B. Role of glycosylation in development. *Annu Rev Biochem* **73**, 491-537 (2004).
93. Li, H. & d'Anjou, M. Pharmacological significance of glycosylation in therapeutic proteins. *Current Opinion in Biotechnology* **20**, 678-684 (2009).
94. Dance, A. From pond scum to pharmacy shelf. *Nat Med* **16**, 146-149 (2010).
95. Cummings, R.D. The repertoire of glycan determinants in the human glycome. *Molecular BioSystems* **5**, 1087-1104 (2009).
96. Raman, R., Raguram, S., Venkataraman, G., Paulson, J.C. & Sasisekharan, R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods* **2**, 817-824 (2005).
97. Murrell, M.P., Yarema, K.J. & Levchenko, A. The systems biology of glycosylation. *ChemBiochem* **5**, 1334-1347 (2004).
98. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**, 2364-2369 (2001).
99. Drickamer, K. & Taylor, M.E. Evolving views of protein glycosylation. *Trends in Biochemical Sciences* **23**, 321-324 (1998).
100. Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54**, 631-664 (1985).
101. Knauer, R. & Lehle, L. The oligosaccharyltransferase complex from yeast. *Biochim Biophys Acta* **1426**, 259-273 (1999).
102. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1426**, 239-257 (1999).
103. Hashimoto, K. et al. KEGG as a glycome informatics resource. *Glycobiology* **16**, 63R-70R (2006).
104. Weerapana, E. & Imperiali, B. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R-101 (2006).
105. Kukuruzinska, M.A., Bergh, M.L. & Jackson, B.J. Protein glycosylation in yeast. *Annu Rev Biochem* **56**, 915-944 (1987).
106. Rayon, C., Lerouge, P. & Faye, L. The protein N-glycosylation in plants. *J. Exp. Bot.* **49**, 1463-1472 (1998).
107. Taylor, M.E. & Drickamer, K. Introduction to glycobiology, Edn. 2nd. (Oxford University Press, Oxford ; New York; 2006).

108. Cylwik, B., Lipartowska, K., Chrostek, L. & Gruszewska, E. Congenital disorders of glycosylation. Part II. Defects of protein O-glycosylation. *Acta Biochimica Polonica* **60**, 361-368 (2013).
109. Lundquist, J.J. & Toone, E.J. The Cluster Glycoside Effect. *Chemical Reviews* **102**, 555-578 (2002).
110. Debray, H., Decout, D., Strecker, G., Spik, G. & Montreuil, J. Specificity of twelve lectins towards oligosaccharides and glycopeptides related to N-glycosylproteins. *Eur J Biochem* **117**, 41-55 (1981).
111. Liener, I.E., Sharon, N. & Goldstein, I.J. The Lectins : properties, functions, and applications in biology and medicine. (Academic Press, Orlando; 1986).
112. Bertozzi, C.R. & Kiessling, L.L. Chemical glycobiology. *Science (New York, N.Y.)* **291**, 2357-2364 (2001).
113. Meier, S. & Duus, J. Carbohydrate dynamics: Antibody glycans wiggle and jiggle. *Nature chemical biology* **7**, 131-132 (2011).
114. Cunningham, S., Gerlach, J.Q., Kane, M. & Joshi, L. Glyco-biosensors: Recent advances and applications for the detection of free and bound carbohydrates. *Analyst* **135**, 2471-2480 (2010).
115. Manimala, J.C., Roach, T.A., Li, Z. & Gildersleeve, J.C. High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems. *Glycobiology* **17**, 17C-23C (2007).
116. Wright, C.S. 2.2 A resolution structure analysis of two refined N-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes. *J Mol Biol* **215**, 635-651 (1990).
117. Saul, F.A. et al. Crystal structure of *Urtica dioica* agglutinin, a superantigen presented by MHC molecules of class I and class II. *Structure* **8**, 593-603 (2000).
118. Harata, K. & Muraki, M. Crystal structures of *Urtica dioica* agglutinin and its complex with tri-N-acetylchitotriose. *J Mol Biol* **297**, 673-681 (2000).
119. Huang, W., Wang, D., Yamada, M. & Wang, L.X. Chemoenzymatic synthesis and lectin array characterization of a class of N-glycan clusters. *J Am Chem Soc* **131**, 17963-17971 (2009).
120. Rao, V., Cui, T., Guan, C. & Van Roey, P. Mutations of endo-beta-N-acetylglucosaminidase H active site residue Asp130 and Glu132: activities and conformations. *Protein Sci* **8**, 2338-2346 (1999).

121. Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R. Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology & Biotechnology* **85**, 1241-1249 (2010).
122. Jakobsson, E., Schwarzer, D., Jokilammi, A. & Finne, J. Endosialidases: Versatile Tools for the Study of Polysialic Acid. *Topics in current chemistry* (2012).
123. Korecka, L. et al. Bioaffinity magnetic reactor for peptide digestion followed by analysis using bottom-up shotgun proteomics strategy. *J Sep Sci* **31**, 507-515 (2008).
124. Plummer, T.H., Elder, J.H., Alexander, S., Phelan, A.W. & Tarentino, A.L. Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. *Journal of Biological Chemistry* **259**, 10700-10704 (1984).
125. Tarentino, A.L., Gomez, C.M. & Plummer, T.H., Jr. Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24**, 4665-4671 (1985).
126. Mussar, K.J., Murray, G.J., Martin, B.M. & Viswanatha, T. Peptide: N-glycosidase F: studies on the glycoprotein aminoglycan amidase from *Flavobacterium meningosepticum*. *Journal of biochemical and biophysical methods* **20**, 53-68 (1989).
127. Haselbeck, A. & Hosel, W. Studies on the effect of the incubation conditions, various detergents and protein concentration on the enzymatic activity of N-glycosidase F (Glycopeptidase F) and endoglycosidase F. *Topics in Biochemistry* **8**, 1-4 (1988).
128. Lemp, D., Haselbeck, A. & Klebl, F. Molecular cloning and heterologous expression of N-glycosidase F from *Flavobacterium meningosepticum*. *J Biol Chem* **265**, 15606-15610 (1990).
129. Tretter, V., Altmann, F. & MÄRz, L. Peptide-N4-(N-acetyl- $\beta$ -glucosaminy)asparagine amidase F cannot release glycans with fucose attached  $\alpha 1 \rightarrow 3$  to the asparagine-linked N-acetylglucosamine residue. *European Journal of Biochemistry* **199**, 647-652 (1991).
130. Norris, G.E., Stillman, T.J., Anderson, B.F. & Baker, E.N. The three-dimensional structure of PNGase F, a glycosylasparaginase from *Flavobacterium meningosepticum*. *Structure* **2**, 1049-1059 (1994).
131. Norris, G.E., Flaus, A.J., Moore, C.H. & Baker, E.N. Purification and crystallization of the endoglycosidase PNGase F, a peptide:N-glycosidase from *Flavobacterium meningosepticum*. *J Mol Biol* **241**, 624-626 (1994).

132. Kuhn, P., Tarentino, A.L., Plummer, T.H., Jr. & Van Roey, P. Crystal structure of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase F at 2.2-A resolution. *Biochemistry* **33**, 11699-11706 (1994).
133. Kuhn, P. et al. Active Site and Oligosaccharide Recognition Residues of Peptide-N4-(N-acetyl- $\beta$ -D-glucosaminyl)asparagine Amidase F. *Journal of Biological Chemistry* **270**, 29493-29497 (1995).
134. Filitcheva, J. PNGases: A Diverse Family of Enzymes Related by Function Rather Than Catalytic Mechanism, Vol. Ph.D. (Massey University, Palmerston North; 2010).
135. Fan, J.Q. Detailed Studies on Substrate Structure Requirements of Glycoamidases A and F. *Journal of Biological Chemistry* **272**, 27058-27064 (1997).

## CHAPTER 2

### COMPUTATIONALLY-GUIDED DESIGN OF BIOCOMBINATORIAL LIBRARIES

#### 2.1 Computationally Guided Library Design

The wild-type *Flavobacterium meningosepticum* N-glycan processing enzyme, wtPNGase F, has previously been co-crystallized (PDB ID: 1PNF) with a chitobiose disaccharide in the active site at a 2.0 Å resolution.<sup>1</sup> Using this 1PNF x-ray crystal structural model a 5 ns fully solvated MD simulation of the PNGase F – N,N'-diacetylchitobiose (GlcNAc $\beta$ 1-4GlcNAc) complex in water at room temperature and pressure employing the AMBER-GLYCAM protein-carbohydrate force field was performed.<sup>2-4</sup> The root mean squared difference (RMSD) in the positions of the C $\alpha$  atoms, relative to the experimental structure, was determined as a function of the

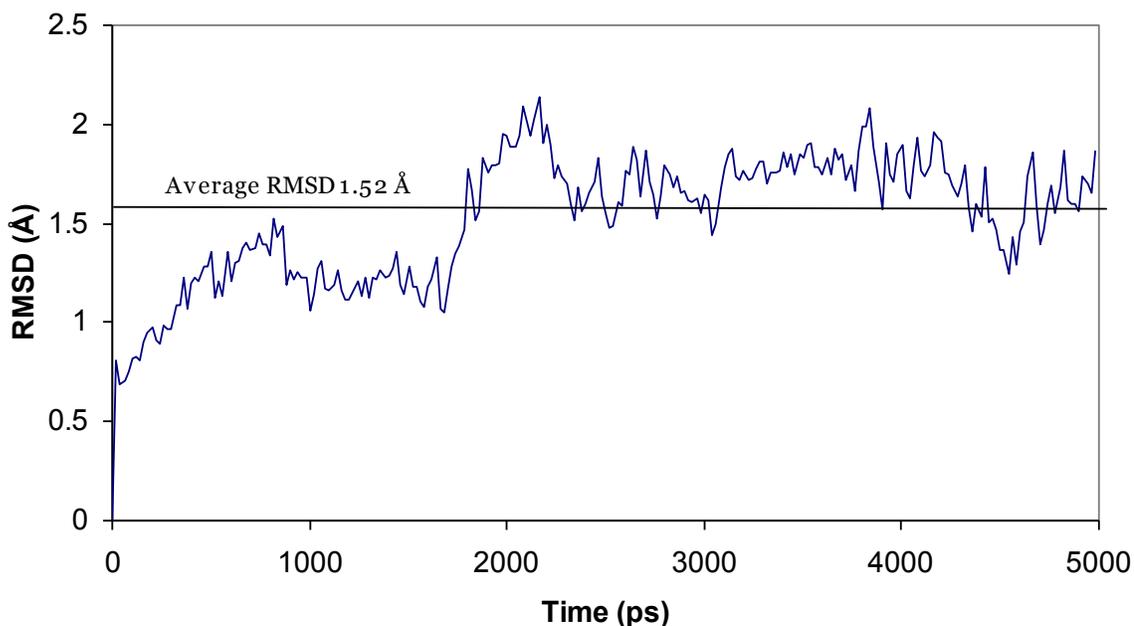


Figure 2.1: RMSD in the C $\alpha$  positions in the PNGase F – chitobiose complex. MD simulation data generated by Woods, *et al.*

simulation time and the relative low 1.5 Å average RMSD (Figure 2.1) indicated that the simulation reproduced the experimental structure. Additionally, the complex maintained experimentally observed hydrogen bond interactions between the disaccharide ligand and the protein (Table 2.1). Given that the simulation of the complex appeared to be stable and consistent with experimental structural data, the interaction energies were then computed. Data from the heating and pre-equilibration period (1 ns) were not included in the subsequent analysis. Using the MM-GBSA protocol as implemented in AMBER, per-residue molecular mechanical (MM) contributions to the binding energy were computed for each of the 314 amino acids in PNGase F over the period of 2-5 ns; the generalized Born (GB) continuum solvent model was employed to estimate desolvation energies.<sup>5</sup> Additionally, MD data were employed in computational alanine scanning. Figure 2.2 depicts residues within 4.5 Å of the ligand.

Table 2.1: Experimental and theoretical hydrogen bond lengths observed between chitobiose and PNGase F. Data generated by Woods, *et al.*

Hydrogen bonds	1PNF X-ray Data (Å) <sup>1</sup>	Average from MD Simulation (Å)
D60-Oδ – GlcNAc316 O1	3.02	2.76 ± 0.1
D60-O – GlcNAc316 NAc	2.97	2.84 ± 0.1
R61-NH – GlcNAc317 OAc	2.84	2.91 ± 0.1
R61-NH – GlcNAc316-O4	2.92	2.90 ± 0.1
R61-NH2 – GlcNAc317 OAc	3.03	2.90 ± 0.1
W120-Nε – GlcNAc317-O6	2.93	2.98 ± 0.1
W191-Nε – GlcNAc316-O3	2.96	3.06 ± 0.1

The estimated interaction energies for residues proximal to the ligand (within 4.5 Å) in addition to any other residues that contributed at least 0.5 kcal/mol to either the total molecular mechanical (sum of van der Waals,  $\Delta E_{VDW}$ , and electrostatic,  $\Delta E_{ELE}$ ) interaction energy ( $\Delta E_{MM}$ ) or the binding free energy are listed in Table 2.2. The per residue binding free energy ( $\Delta G_{BINDING}$ ) was computed as the sum of the molecular

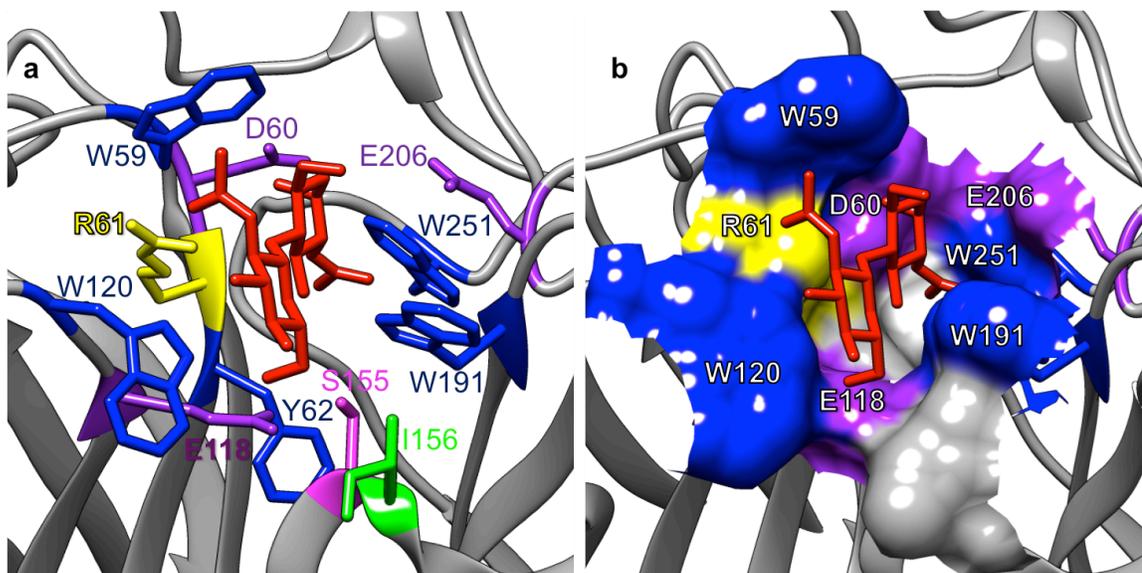


Figure 2.2: PNGase F binding pocket. a) Residues within 4.5 Å of the disaccharide chitobiose ligand (red) in the binding site of PNGase F. b) The solvent accessible surface with critical residues for binding labeled. PDB ID 1PNF. Molecular graphics made with UCSF Chimera package.

mechanical interaction energy ( $\Delta E_{MM}$ ) and the desolvation energy ( $\Delta G_{GB+SA}$ ) of that residue. The per-residue energy analysis enabled the residues proximal to the binding site (Figure 2.2) to be categorized into critical and tepid based on per residue binding free energy.

Critical residues made significantly stabilizing interactions, with the exception of the three residues (D60, E206, & E118) that were indicated to make slightly unfavorable interactions with the substrate. These three residues have each been associated with the catalytic function of PNGase F, which may explain their role in destabilizing the substrate.<sup>1</sup> Based on point mutant studies, D60 has been identified as the primary catalytic residue, whereas E206 and E118 are proposed to help stabilize high-energy reaction intermediates.<sup>1</sup> Consistent with experimental observations of hydrogen bonds and aromatic stacking in the complex, the energy decomposition analysis confirmed that R61, W120, W58, W191, and W251 are critical to ligand binding.<sup>1</sup>

Table 2.2: Approximate residue contributions (kcal/mol) to the binding free energy for wtPNGase F bound to substrate, chitobiose. Residues listed are within 4.5 Å of the ligand or contributed at least 0.5 kcal/mol to either the total molecular mechanical (van der Waals and electrostatics) interaction energy ( $\Delta E_{MM}$ ) or the total binding free energy ( $\Delta G_{BINDING}$ ). Residues required for catalytic activity are indicated in bold.<sup>1</sup> Library columns indicate residues selected for optimization for knowledge-based library design: A=alanine, X=all 20 amino acids, X(-D)=19 amino acids (excluding aspartic acid). MM-GBSA data generated by Woods, *et al.*

Critical Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$	Library 1	Library 2
R61	-1.5	-15.1	-16.7	12.3	-4.4		
W120	-3.1	-2.3	-5.4	1.9	-3.5		
W59	-3.1	-0.2	-3.3	0.3	-3.0		
W191	-1.3	-1.6	-2.9	1.3	-1.6		
W251	-0.7	-0.3	-1.0	0.1	-0.9		
<b>E118</b>	-0.5	-0.1	-0.5	0.6	0.1	X	
<b>D60</b> (nucleophile)	-0.9	-3.9	-4.8	5.2	0.4	A	X(-D)
<b>E206</b>	-0.3	2.1	1.8	-1.1	0.7	X	X
Proximal residues making only weak contributions	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$	Library 1	Library 2
Y62	-0.6	-0.1	-0.6	0.0	-0.6	X	
D57	-0.1	3.0	2.9	-3.5	-0.6	X	X
I156	-0.2	0.1	-0.2	-0.1	-0.3	X	X
S155	-0.3	0.2	-0.1	-0.1	-0.1	X	
R248	-0.1	-1.2	-1.4	1.2	-0.1		X
G192	0.0	0.1	0.0	0.0	0.0	X	X
T119	-0.2	-0.6	-0.7	0.8	0.1		
K123	0.0	-0.5	-0.5	0.6	0.1		
R125	0.0	-0.4	-0.4	0.6	0.1		
<b>Sub-total Interaction Energy</b> <b><math>\Delta G_{BINDING}</math></b>	<b>-12.9</b>	<b>-20.8</b>	<b>-33.7</b>	<b>20.1</b>	<b>-13.6</b>		

Equally important, nine additional residues, proximal to the ligand, were identified that were not making significant energetic contributions to binding. These nine weakly contributing, or tepid, residues represent the best opportunity for affinity

enhancement by utilizing site-saturation mutagenesis libraries for directed evolution. Additionally, relative to wild-type PNGase F, the computational replacement of D60 or E206 with alanine (D60A, E206A) indicated that these mutants should have favorable interaction energies (Table 2.3). In particular, the D60A interaction energy indicates markedly improved substrate affinity relative to wild-type (wt) PNGase F, thus the D60A mutant was selected for expression and further experimental analysis.

Table 2.3: Computational alanine scanning of PNGase F bound to chitobiose. Interaction energies (kcal/mol) for favorable mutants are identified relative to wtPNGase F. Data generated by Woods, *et al.*

Contact Zone Residues	$\Delta\Delta E_{MM}$	$\Delta\Delta G_{GB+SA}$	$\Delta\Delta G_{BINDING}$
D60A	1.8	-4.0	-2.2
E206A	-1.9	1.2	-0.7

## 2.2 Yeast Display Library Construction

Two yeast surface displayed biocombinatorial libraries were designed, which incorporated several computationally predicted residues for optimization as indicated in Table 2.2. Library 1 (GenScript, Piscataway, NJ) was synthesized using NNK codon degeneracy and it incorporated a fixed D60A mutation in all the clones. NNK codon degeneracy reduces the probably of introducing a random stop codon while also minimizing codon bias relative to NNN codon degeneracy.<sup>6</sup> The sequence and sites of mutations for GenScript Library 1 are shown in Figure 2.3. Library 2 (GeneArt AG, Regensburg, Germany) was synthesized using cassette mutagenesis and NNN codon degeneracy and it incorporated a randomized D60 position using 19 amino acids (i.e.: excluding D). The sequence and sites of mutations for GeneArt Library 2 are shown in Figure 2.4.

```

M13.rev
AGGAAACAGCTATGAC ->
AGGAAACAGCTATGACCATGTTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAA
GCGGGCAGTGAGCGGAAGGCCCATGAGGCCAGTTAATTAAGAGGTACCGCTAGCGCT
CCGGCAGATAAATACGGTAAATATTAAAACATTCGACAAAGTAAAAAATGCCTTTGGT
GACGGATTGTCCCAAAGTGCGGAAGGAACCTTTACATTTCCGGCCGATGTAACAGCC
GTAAAAACGATTAAGATGTTTCATTAATAAATGAATGTCCTAATAAACTTGTNNKGAA
TGGGCTCGTNNKGCCAATGTTTATGTAATAAATAAAACAACAGGTGAGTGGTACGAA
ATAGGACGCTTTATTACTCCATATTGGGTGGGAACGGAAAAATTACCTCGTGGACTG
GAAATTGATGTTACAGATTTCAAATCTTTACTATCCGGAAATACAGAACTTAAAATT
TATACGNNKKACATGGCTGGCCAAAGGAAGAGAATACAGTGTAGATTTTCGATATTGTA
TACGGGACACCGGATTATAAATATTCGGCTGTAGTACCTGTAGTTCAGTATAACAAA
TCANNKNNKGACGGAGTCCCTTATGGTAAAGCACATACATTGGCTTTGAAAAAGAAT
ATCCAGTTACCAACAAACACAGAAAAAGCTTATCTTAGAACTACTATTTCCGGATGG
NNKCATGCTAAGCCATATGATGCGGGAAGCAGAGGTTGTGCANNKTGGTGCTTCAGA
ACACACACTATAGCAATAAATAATTTCGAATACTTTCCAGCATCAGCTGGGTGCTTTA
GGATGTTTCAGCAAACCTATCAATAATCAGAGTCCGGGAAATGGACTCCCGACAGA
GCCGGTTGGTGCCCGGGAATGGCAGTTC AACACGTATAGATGTAATAATTCT
TTAATAGGCAGTACTTTTAGTTATGAATATAAATTCCAGAACTGGACAAATAACGGA
ACCAATGGAGATGCTTTTTATGCAATTTCCAGTTTTTGTGATTGCAAAAAGTAATACA
CCTATTAGTGCTCCGGTAGTTACAAACGGATCCGAGCTCATGGCGCGCCTAGGCCTT
GACGGCCTTCCGCCAATTCGCCCTATAGTGAGTCGTATTACGTCGCG
CTCACTGGCCGTCGTTTTACA
<- TGACCGGCAGCAAAATGT
M13.fwd

```

A179C = D60A

N = equimolar A, T, C, or G nucleotide mixture

K = equimolar G or T nucleotide mixture

NheI & BamHI restriction sites

Figure 2.3: Sequence and restriction map of non-amplified GenScript Library 1. The sequence of expressed PNGase F clone is underlined and flanked by NheI (yellow) and BamHI (cyan) restriction sites. A total of eight mutations were engineered into this library construct: D57, D60A, Y62, E118, S155, I156, G192, and E206. The single point A179C nucleotide mutation (red) was introduced to exhibit the D60A amino acid mutation. This mutation is indicated to enhance affinity interactions and while also either inactivating or significantly diminishing the catalytic activity of the enzyme. The seven site-saturation mutagenesis sites were engineered into this library using NNK codon (green) degeneracy where N represents equimolar A, T, C, or G nucleotide mixture and K represents equimolar G or T nucleotide mixture. M13 forward and reverse primer sequences are indicated (blue).

M13.rev  
 AGGAAACAGCTATGAC ->  
AGGAAACAGCTATGACCATGTTAATGCAGCTGGCAGACAGGTTTCCCGACTGGAAA  
 GCGGGCAGTGAGCGGAAGGCCCATGAGGCCAGTTAATTAAGAGGTACC**GCTAGC**GCT  
 CCGGCAGATAATACGGTAAATATTTAAACATTCGACAAAGTAAAAAATGCCTTTGGT  
 GACGGATTGTCCCAAAGTGCGGAAGGAACCTTTACATTTCCGGCCGATGTAACAGCC  
 GTAAAAACGATTAAGATGTTTCATTTAAAAATGAATGTCCTAATAAAACTTGT**NNNGAA**  
 TGG**NNN**CGTTATGCCAATGTTTATGTAAAAAATAAAACAACAGGTGAGTGGTACGAA  
 ATAGGACGCTTTATTACTCCATATTGGGTGGGAACGGAAAAATTACCTCGTGGACTG  
 GAAATTGATGTTACAGATTTCAAATCTTTACTATCCGGAAATACAGAACTTAAAATT  
 TATACGGAGACATGGCTGGCCAAAGGAAGAGAATACAGTGTAGATTTTCGATATTGTA  
 TACGGGACACCGGATTATAAATATTCGGCTGTAGTACCTGTAGTTCAGTATAACAAA  
 TCATCT**NNN**GACGGAGTCCCTTATGGTAAAGCACATACATTGGCTTTGAAAAAGAAT  
 ATCCAGTTACCAACAAACACAGAAAAAGCTTATCTTAGAACTACTATTTCCGGATGG  
**NNN**CATGCTAAGCCATATGATGCGGGAAGCAGAGGTTGTGCA**NNN**TGGTGCCTCAGA  
 ACACACACTATAGCAATAAATAATTTCGAATACTTTCCAGCATCAGCTGGGTGCTTTA  
 GGATGTTTCAGCAAACCCATCAATAATCAGAGTCCGGGAAATTGGACTCCCGAC**NNN**  
 GCCGGTTGGTGCCCGGGAATGGCAGTTC AACACGTATAGATGTAATAATTCT  
 TTAATAGGCAGTACTTTTAGTTATGAATATAAATTCCAGAACTGGACAAATAACGGA  
 ACCAATGGAGATGCTTTTTATGCAATTTCCAGTTTTGTGATTGCAAAAAGTAATACA  
 CCTATTAGTGCTCCGGTAGTTACAAAC**GGATCC**GAGCTCATGGCGCGCCTAGGCCTT  
 GACGGCCTTCCGCCAATTCGCCCTATAGTGAGTCGTATTACGTCCGG  
 CTC**ACTGGCCGTCGTTTTACA**  
 <- **TGACCGGCAGCAAATGT**  
 M13.fwd

**N** = equimolar A, T, C, or G nucleotide mixture  
**N** = nucleotide mixture resulting in all amino acids  
 except aspartic acid (D)  
**NheI** & **BamHI** restriction sites

Figure 2.4: Sequence and restriction map of non-amplified GeneArt Library 2. The sequence of expressed PNGase F clone is underlined and flanked by **NheI** (yellow) and **BamHI** (cyan) restriction sites. A total of six site-saturation mutagenesis sites were engineered into this library construct: D57, D60(-D), I156, G192, E206, and R248. Five of the six site-saturation mutagenesis sites were engineered into this library using **NNN** codon (green) degeneracy where **N** represents equimolar A, T, C, or G nucleotide mixture. For the site, D60 (magenta), a modified nucleotide mixture resulting in all amino acids except aspartic acid was utilized. M13 forward and reverse primer sequences are indicated (blue).

Synthetic degenerate oligonucleotides were constructed with the objective being to include the defined amino acid subsets at the defined position (Figure 2.3 and 2.4).

PCR products were obtained using these oligonucleotide and full-length fragments were

gel purified. Employing a cassette mutagenesis strategy, the full-length products of both libraries were cloned into the pPNL6 vector using the NheI and BamHI restriction sites. The Pacific Northwest National Laboratory provided an aliquot of a yeast cell-surface displayed nonimmune library of human antibody scFv fragments (pPNL6).<sup>7</sup> This library was modified to replace the scFv fragment with the PNGase F enzyme (PNGaseF-pPNL6) as depicted in Figure 2.5 (Dr. Loretta Yang). EBY100 yeast cells were transformed with the PNGase F-pPNL6 libraries for surface display (Figure 2.6).<sup>8</sup> Titration and random sequencing of clones was carried out to assess the quality of the library, the efficiency of transformation, and the percent sequence space covered. A summary of sequence coverage estimates for both libraries is presented in Table 2.4. Library 1 was designed with seven sites for site-saturation mutagenesis. The theoretical diversity of the number of unique clones is  $1.28 \times 10^9$ . However, based on the sequence identity and the total number of transformants the estimated synthesized diversity is only  $2.40 \times 10^6$  clones. This represents sequence coverage of approximately 0.18% indicating

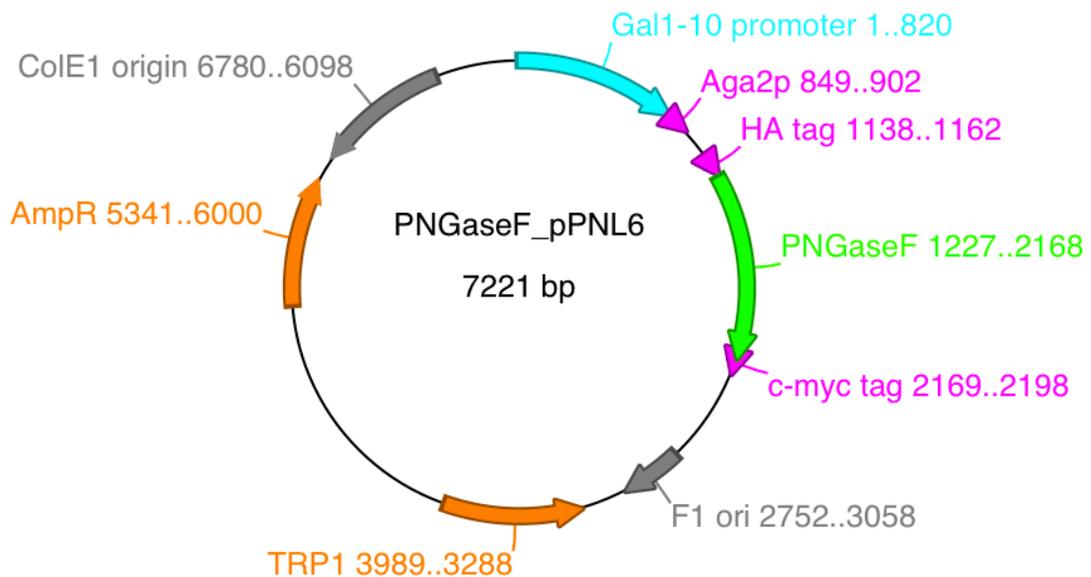


Figure 2.5: PNGase F modified pPNL6 yeast display library plasmid map.

inefficiency both in construction and transformation of Library 1. Library 2 was designed with six sites for site-saturation mutagenesis, representing a theoretical diversity of  $6.08 \times 10^7$  unique clones. The estimated synthesized diversity of Library 2 was determined to be  $1.36 \times 10^7$  clones. Library 2 has sequence coverage of approximately 22.3%, and based on the number of clones represents a 5.7-fold higher synthesized diversity than Library 1.

Table 2.4: Comparison of theoretical and estimated synthesized library diversity and coverage of sequence space.

	Amino Acid Randomization	Theoretical Diversity	Synthesized Diversity	% coverage
Library 1	7 ( $20^7$ )	$1.28 \times 10^9$	$\sim 2.40 \times 10^6$	$\sim 0.18\%$
Library 2	6 ( $20^5 \times 19^1$ )	$6.08 \times 10^7$	$\sim 1.36 \times 10^7$	$\sim 22.3\%$

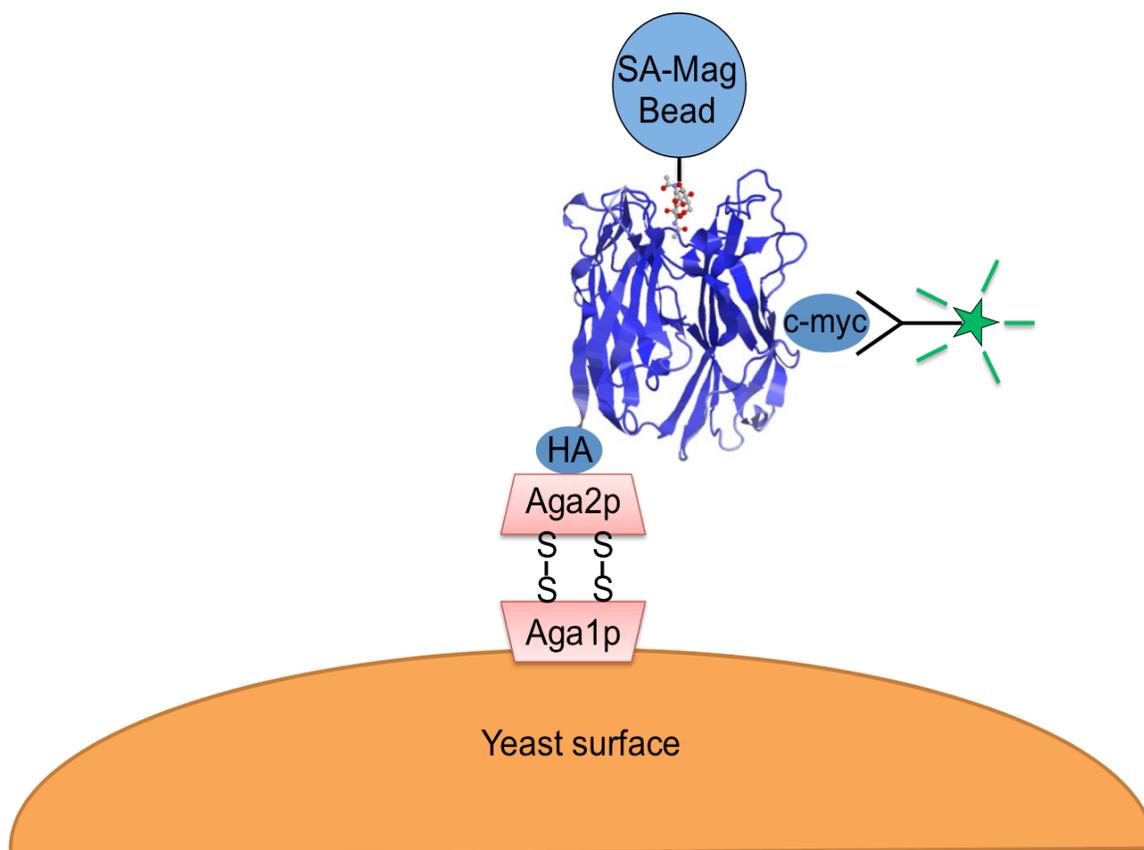


Figure 2.6: Yeast cell-surface display. Representation of Aga2p-PNGase F fusion protein displayed via Aga1p on the yeast cell surface. Selected PNGase F clone(s) interact with the *N*-glycan target. The *N*-glycan target is biotinylated and bound to streptavidin coated magnetic beads. The original pPNL6 construct includes a HA tag between the Aga2p protein and the fused protein, PNGase F in this case. A C-terminal c-myc tag is included and is detected with an anti-c-myc fluorescent antibody by flow cytometry to confirm expression of the full length Aga2p-PNGase F fusion protein on the yeast cell surface prior to each round of selection. Approximately, 50,000 copies of Aga-2p protein are normally displayed on the yeast cell surface.

### 2.3 Directed Evolution of PNGase F Clones via Yeast Surface Display

The constructed yeast-displayed PNGase F libraries were utilized for selecting clones with enhanced affinity for target *N*-glycans. Yeast libraries were grown overnight in selective growth media in a shaking incubator at 30 °C for approximately 24 hours. The expression and display of the Aga2p-PNGase F fusion protein on the yeast cell surface is under a Gal1-10 promoter (Figures 2.5 & 2.6), thus the yeast libraries were induced overnight in galactose containing media in a shaking incubator at 20 °C.

Induction efficiency was determined via flow cytometry using a primary anti-c-myc antibody to detect the C-terminal c-myc tag on the fully expressed Aga-2p-PNGase F fusion protein (Figure 2.7). An induced yeast display library with at least 60% induced clones was used for selection of high affinity clones against *N*-glycan targets.

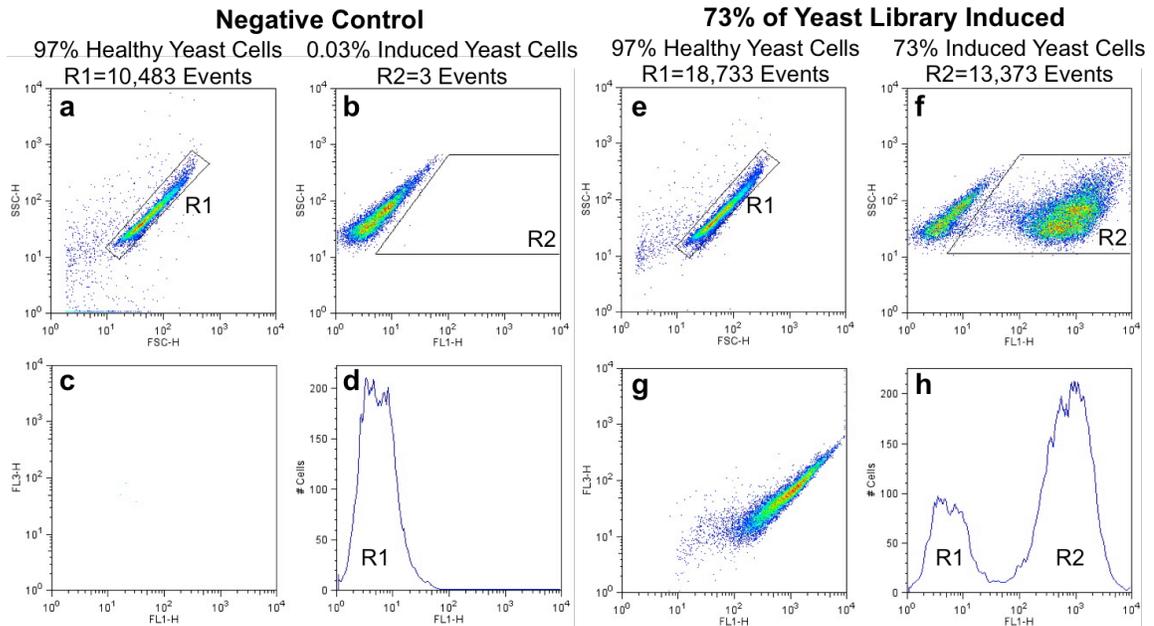


Figure 2.7: Flow cytometry of induced yeast library. Panels a – d: Uninduced yeast library flow cytometry data (negative control). Panels e – h: Induced yeast library flow cytometry data. The healthy yeast cell population was identified by Region 1 (R1) as shown in panels a & e. The sub-population of induced yeast cells within R1 that expressed full length, C-terminal c-myc tagged, Aga2p-PNGase F fusion protein as detected by an anti-c-myc antibody was identified by Region 2 (R2) as shown in panels b & f. R2 sub-population of induced yeast cells are also shown in panels c & g. Histograms of both R1 & R2 populations are shown in panels d & h.

Two *N*-glycan targets were employed in the selection strategy to enrich clones that will retain the broad specificity of wtPNGase F enzyme for *N*-glycan structures. The primary *N*-glycan target was bovine pancreatic Ribonuclease B (RNase B), which contains a single *N*-glycosylation site at asparagine 34 (N34) and has nine high mannose glycoforms (Figure 2.8).<sup>9-11</sup> The reported average molecular weight of RNase B is 15,095 Da derived from the relative abundance of each of the glycosylated species.<sup>12</sup> RNase B

and its non-glycosylated form RNase A, with a reported molecular weight of 13,680 Da, are well-characterized enzymes and frequently used as standards for validating carbohydrate analysis techniques.<sup>11-16</sup> Interestingly, based on comparison of NMR spectra of RNase A and RNase B, the *N*-glycosylation of RNase B has no discernable impact on its structure.<sup>17</sup> However, RNase B exhibits greater stability than RNase A, consistent with observations that glycosylation reduces the denaturing tendency promoted by the preferential hydration of the groups buried in the core of the protein.<sup>14, 15</sup> A secondary *N*-glycan target was Asialofetuin which is created by enzymatically desialylating fetuin with a neuraminidase, and contains less than 0.5% *N*-acetylneurminic acid. Fetuin, isolated from fetal calf serum, is a 48.4 kDa glycoprotein with three *N*-glycosylation sites and five *O*-glycosylation sites and has relatively more complex *N*-glycan structures in comparison to the high mannose structures found on RNase B.<sup>18</sup> The percent weight composition of fetuin is 74% polypeptide, 8.3%hexose, 5.5% hexosamines, and 8.7% sialic acid. Both *N*-glycan target glycoproteins were denatured to make the *N*-glycans fully accessible to the yeast surface displayed PNGase F clones. Furthermore, the denatured glycoproteins were biotinylated in order to present them on Dynabeads®

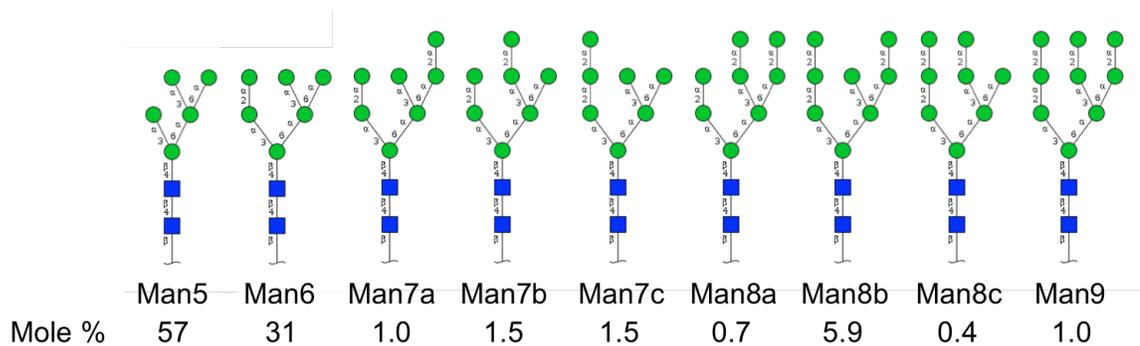


Figure 2.8: Ribonuclease B glycoforms. RNase B has a single *N*-glycosylation site at N34, which can consist of nine glycoforms of Man<sub>5-9</sub>GlcNAc<sub>2</sub>. The mole percentage of these glycoforms is listed below each of the nine glycan structures.

Biotin Binder, streptavidin coated 2.8  $\mu\text{m}$  magnetic beads for selection and for detection with fluorescently labeled streptavidin for FACS.

The selection strategy incorporated two rounds of Magnetic-Activated Cell Sorting (MACS) using streptavidin coated 2.8  $\mu\text{m}$  magnetic beads (Dynabeads® Biotin Binder) followed by a third round of Fluorescence-Activated Cell Sorting (FACS) using denatured RNase B and Asialofetuin as target *N*-glycans (Figures 2.9 and 2.10).<sup>19, 20</sup> The library underwent negative selection against uncoated magnetic beads at the start of first round of selection to remove any bead-binding clones from the library prior to positive selection against *N*-glycan targets of interest.

The set of 2x MACS (Figure 2.9) and 1x FACS (Figure 2.10) rounds of selection were repeated for a total of nine rounds (Figure 2.11a). The target *N*-glycan bearing RNase B was exclusively used for all nine rounds of selection with Library 1. However, both *N*-glycan bearing RNase B and Asialofetuin glycoproteins were concurrently used as targets for Library 2 during parallel rounds of selection. A portion of the amplified library output from round three with RNase B was concurrently selected against Asialofetuin during rounds 4-6. The outputs from both target RNase B and Asialofetuin selections were pooled after round six. As before, a portion of this combined output pool was again concurrently selected against both target RNase B and Asialofetuin during rounds 7-9 in parallel. At the end of the round nine, both RNase B and Asialofetuin selection output pools were recombined once again. Figure 2.10a depicts the progress of the selection process with both RNase B and Asialofetuin with Library 2.

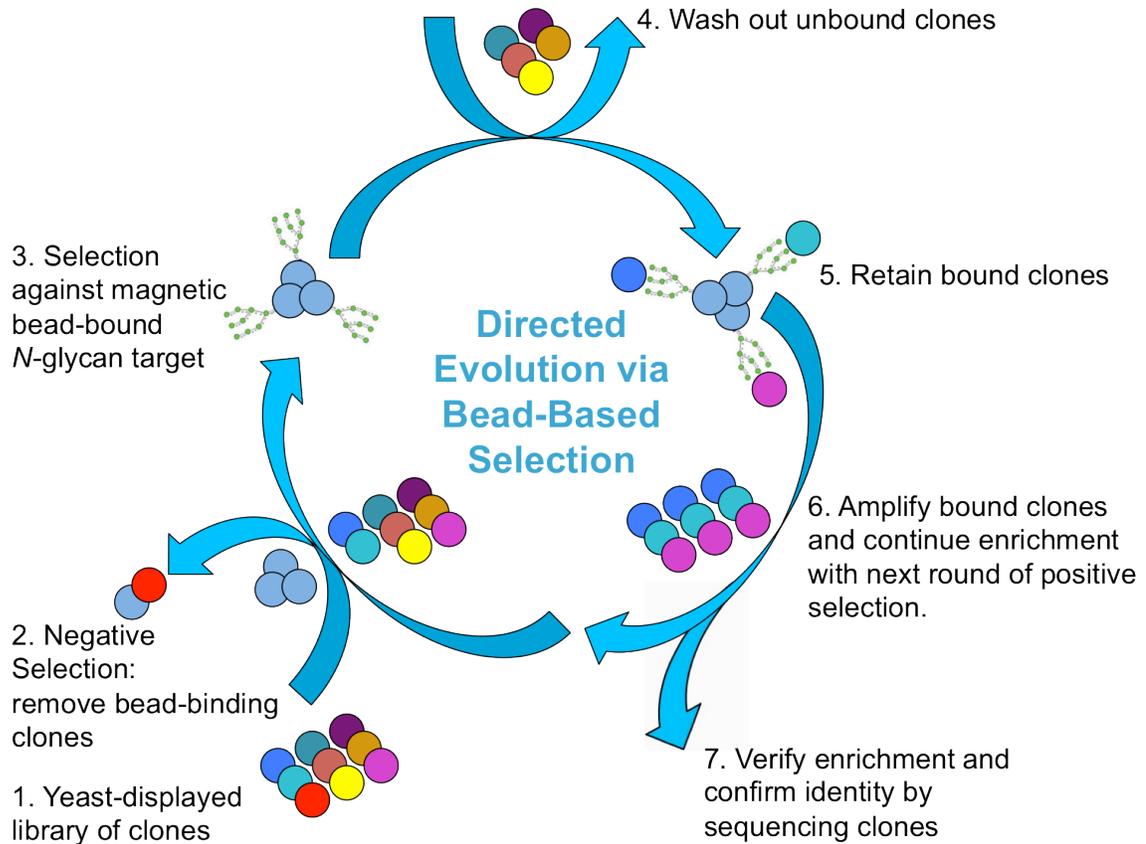


Figure 2.9: Magnetic-Activated Cell Sorting (MACS) of yeast cells. Yeast-displayed PNGase F library selection against *N*-glycan target bound to magnetic beads. Steps 1 – 2: The PNGase F library is subjected to a negative selection against biotin-streptavidin magnetic beads (no dRNase B) at the start of the first round of selection to remove any biotin-streptavidin-magnetic bead-binding clones. Step 3: Biotinylated denatured RNase B is pre-incubated with streptavidin coated magnetic beads (2.8  $\mu$ m diameter) prior to initiating positive selection. Step 4: Unbound yeast clones are washed away. Step 5: Bound yeast clones are retained. Step 6: Bound clones are amplified for the next round of selection. Step 7: Clones are sequenced to monitor enrichment and convergence after each round of FACS (not shown). The converged clone(s) are selected as Lectenz® candidate(s) for downstream characterization.

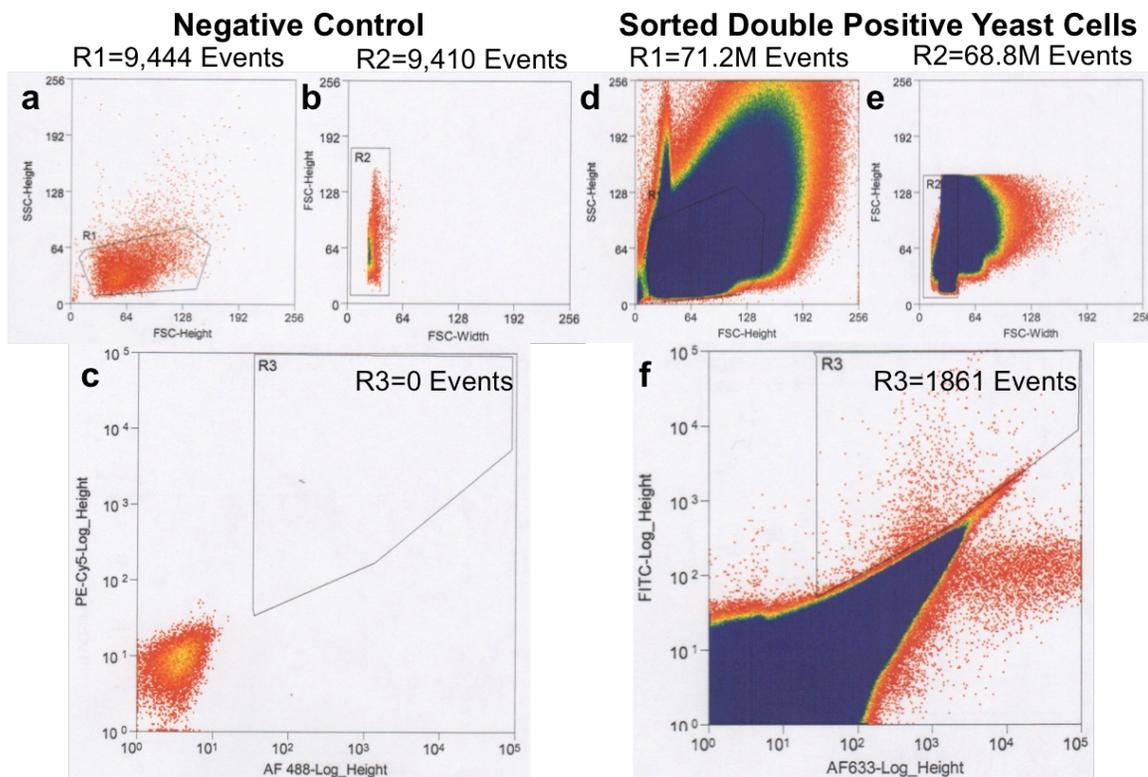


Figure 2.10: Fluorescent-Activated Cell Sorting (FACS) of double positive yeast cells. Panels a – c: Negative control FACS data for uninduced yeast library. Panels d – f: FACS data for induced yeast library incubated with glycan target. Homogenous population of yeast cells was sorted using Region 1 (R1) and Region 2 (R2) as shown in panels a, b, d, and e. Region 3 (R3) sorted double positive events representative of yeast cells identified by the presence of both the *N*-glycan target (FITC signal) and fully expressed PNGase F (AF633 signal) as shown in panels c and f.

During nine rounds of iterative selection and amplification (approximately 50 clones were sequenced at the end of every 3<sup>rd</sup> round) enrichment of clones was observed. The clone with the highest level of enrichment, designated R911, had the following mutations relative to wtPNGase F: D57L, D60C, I156L, G192I, E206S, and R248W. A graphical representation of the prevalence of amino acids at the six computationally selected mutagenesis sites can be viewed in Figure 2.11b.

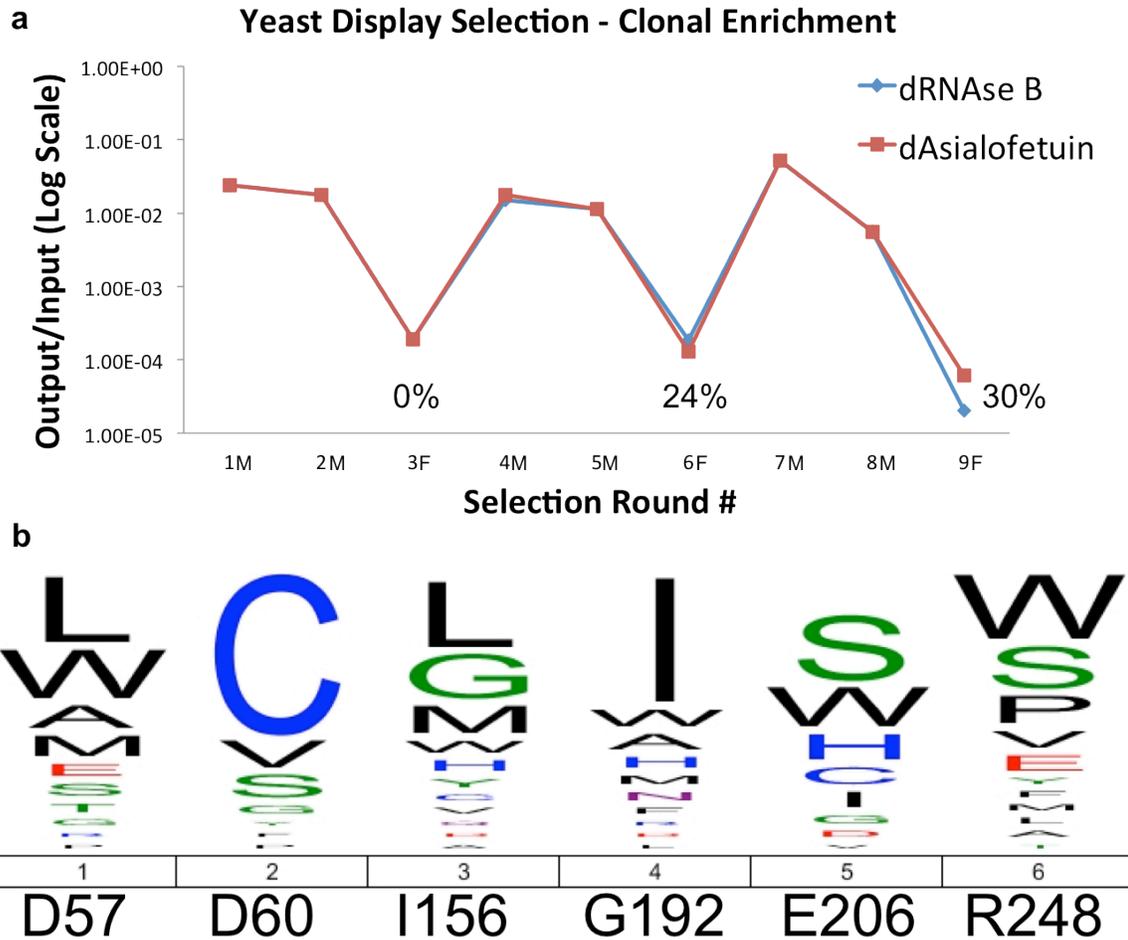


Figure 2.11: Yeast display PNGase F clonal selection and enrichment with GeneArt Library 2. (a) Iterative rounds of yeast display selection, amplification, and percent enrichment of PNGase F clones is shown. An aliquot of each library input sample, negative selection sample, wash sample, and output sample is titered in every round to monitor the progress of the selection and enrichment. Data is displayed as an output/input ratio representing the number of clones recovered from the bead-bound sample after selection relative to the starting number of input clones for that round. MACS based selection was performed on rounds 1M, 2M, 4M, 5M, 7M, and 8M. After every two rounds of MACS, FACS based selection was performed on rounds 3F, 6F, and 9F. Ideally, each round of selection will enrich functionally relevant clones which bind to the target *N*-glycan structure on either denatured RNase B (dRNase B) or denatured Asialofetuin (dAsialofetuin) leading to convergence after several rounds of selection. Enrichment and convergence are monitored with DNA sequencing of ~50 randomly selected clones after every 3<sup>rd</sup> round of panning via FACS. The enrichment of clone R911 relative to all clones sequenced is shown as a percentage at every 3<sup>rd</sup> round of selection. (b) Amino acid ice logo of enriched clone sequences. The wtPNGase F sequence is shown on the bottom. Preferred amino acids at the six randomized positions are shown as a graphical representation. This data is based on ~150 clone sequences obtained from selection rounds 3F, 6F, and 9F. The top most residue in each position is also the sequence of the selected clone R911.

Table 2.5: Sequences of enriched clones from Library 1 and Library 2 selections. Listed are identities of the preferred amino acids at the residues selected for site-saturation mutagenesis. For comparison, the wtPNGase F sequence is included. Clones R617 and R6113 were selected from Library 1 solely against the target *N*-glycan bearing RNase B glycoprotein. Clones R911 and R9113 were selected from Library 3 against both target *N*-glycan bearing RNase B and Asialofetuin glycoproteins. The blue boxes represent sites that were not selected for site-saturation mutagenesis.

wtPNGase F	D57	D60	Y62	E118	S155	I156	G192	E206	R248
R617 (Library 1)	R	A	G	A	D	T	C	S	R
R6113 (Library 1)	C	A	W	A	Q	T	T	R	R
R911 (Library 2)	L	C	Y	E	S	L	I	S	W
R9113 (Library 2)	W	C	Y	E	S	M	I	W	S

Sequence identity of the top two enriched clones selected from both Library 1 and

Library 2 are summarized in Table 2.5. The selection of tryptophan in multiple sites is significant because aromatic side chains are known to interact with the hydrophobic face of monosaccharides.<sup>21</sup> The D60 position in Library 2, which was subjected to site-saturation mutagenesis to 19 amino acids (except D), showed the same D60A mutation in both R9 clones. This may indicate that a cysteine in this position is highly favored for binding interactions. However, the presence of cysteine in all four enriched clones is potentially also a cause for concern as the addition of a single cysteine could potentially disrupt the three pre-existing disulfide bonds at 51-56, 204-208, and 231-252 in PNGase F. Other interesting observations include the preference for E118A and I156T mutations in both the R6 clones from Library 1. Similarly, both the R9 clones from Library 2 show a preference for D60C and G192I mutations. The G192I mutation is significant because a relatively small glycine residue has been replaced with a bulky hydrophobic isoleucine side chain. An increase in the hydrophobicity of the binding pocket may enhance protein-carbohydrate interactions; however, the presence of a bulky side chain could also partially block access to the binding pocket. To investigate the utility of the selected R6 and R9 clones as Lectenz® affinity reagents and characterize their properties, the selected

PNGase F clones were cloned into a bacterial expression vector for expression and purification in an *E.coli*.

## 2.4 Methods

### 2.4.1 Molecular Dynamics and Per-residue Binding Free Energy Decomposition

A 5 ns fully solvated MD simulation of the PNGase F – *N,N'*-diacetylchitobiose (GlcNAc $\beta$ 1-4GlcNAc) complex was performed in water at room temperature and pressure employing the AMBER-GLYCAM protein-carbohydrate force field. The per-residue contributions to the binding energy was computed for each of the 313 amino acids in PNGase F, employing the generalized Born (GB) continuum solvent model as implemented in AMBER.<sup>5</sup> In a typical MM-GB/PB calculation, the free energy is computed for the protein ( $\Delta G_{\text{PROTEIN}}$ ), ligand ( $\Delta G_{\text{LIGAND}}$ ), and complex ( $\Delta G_{\text{COMPLEX}}$ ) for each structural “snapshot” extracted from the MD trajectories. From the 5 ns trajectory, the first 1 ns was discarded and 2000 snapshots were selected (at 2 ps intervals) from the remaining 4 ns for molecular mechanical (MM) binding energy analysis. The binding free energy ( $\Delta G_{\text{BINDING}}$ ) is then computed by subtraction. As shown in Equation 1, averaging over the entire trajectory results in the final average interaction energies ( $\langle \Delta G_{\text{BINDING}} \rangle$ ), where the averaging is over the MD snapshots.

Equation 2.1

$$\langle \Delta G_{\text{BINDING}} \rangle = \langle \Delta G_{\text{COMPLEX}} \rangle - \langle \Delta G_{\text{PROTEIN}} \rangle - \langle \Delta G_{\text{LIGAND}} \rangle$$

The free energies of the components are computed by separating the energies into three categories (Equation 2), namely molecular mechanical ( $\Delta E_{\text{MM}}$ , electrostatic and van der Waals), entropic ( $\Delta S_{\text{MM}}$ ), and solvation ( $\Delta G_{\text{SOLVATION}}$ ).

Equation 2.2:

$$\langle \Delta G \rangle = \langle \Delta E_{\text{MM}} \rangle - T \langle \Delta S_{\text{MM}} \rangle + \langle \Delta G_{\text{SOLVATION}} \rangle$$

#### *2.4.2 Computational Alanine Scanning and Electrostatic Scanning*

Following the single trajectory mutation protocol proposed in the Kollman group and implemented in AMBER, the set of snapshots for the wild type complex was employed for each mutant calculation of the energy terms in Equations 1 and 2.<sup>9</sup> The mutant side chain is truncated, replacing  $C\gamma$  with a hydrogen atom, and setting the  $C\beta$ —H bond length and direction to those of the residue in the wild type  $C\beta$ — $C\gamma$ . The underlying approximations of the single trajectory mutation protocol are that the mutant and the wild type undergo similar conformational changes from the unbound to the bound state, and that local side chain reorganizations are small perturbations relative to the alanine mutation itself.<sup>9</sup> One can run separate trajectories on the wild type and mutant species, however this introduces substantial noise (due to lack of cancellation of internal energy components) and is computationally demanding. Separate simulations would be justified in the case of mutations to larger or charged residues.

In order to probe for the effect of an ionized residue at a particular position, alanine scanning was modified to employ an alanine with a theoretical net positive (Ala+) or negative (Ala-) charge. All atoms in the alanine carried the standard partial charges, while the total charge on the residue was set to +1 or -1 by adjusting the charge on the  $C\beta$  atom.

#### *2.4.3 Synthesis of Yeast-Displayed PNGase F Clones Library*

The Genscript library was synthesized (Genscript, Piscataway, NJ) using NNK codon degeneracy and incorporates mutagenesis sites as indicated under Library 1 in Table 2.1. The GeneArt library was synthesized (Life Technologies, Carlsbad, CA) using cassette mutagenesis and NNN codon degeneracy and it incorporates a randomized D60

position using 19 amino acids (i.e.: excluding D) as indicated under Library 2 in Table 2.1). The synthesized libraries were cloned into the pPNL6 vector (Figure 2.5).

#### *2.4.4 Yeast Display Library Transformation into EBY100*

The cloned libraries in the PNGaseF-pPNL6 vector were transformed into EBY100 yeast cells for surface display (Figure 2.6 and Table 2.4) per the recommended protocol.<sup>8</sup>

#### *2.4.5 Induction of Yeast Display Library*

The yeast library was induced as per the recommended protocol in the Yeast Display scFv Antibody Library User's Manual (Rev: MF031112) ([www.sysbio.org/dataresources/index.stm](http://www.sysbio.org/dataresources/index.stm)) provided by Pacific Northwest National Laboratory (Richland, WA). The EBY100 transformed yeast libraries are induced in galactose containing media to express the surface displayed the Aga2p-PNGase F clones (Figure 2.6) Induction efficiency is determined by flow cytometry to ensure at least 60% of the yeast cells are expressing the C-terminal c-myc tag (Figure 2.7).

#### *2.4.6 Directed Evolution of PNGase F Clones Library via Yeast Surface Display*

The *N*-glycan bearing glycoproteins, RNase B (Sigma R7884) and Asialofetuin (Sigma A4781), were used as selection targets.<sup>9, 18</sup> Both glycoproteins were denatured to ensure maximum exposure of the *N*-glycan and glycopeptide region to the yeast surface displayed PNGase F clones.

The selection strategy incorporates two rounds of magnetic activated cell sorting (MACS) Figure 2.9 followed by a third round of fluorescence activated cell sorting (FACS) Figure 2.10 using a mixture of denatured RNase B and denatured Asialofetuin as target *N*-glycans.<sup>19, 20</sup> The set of 2x MACS and 1x FACS rounds of selection were repeated for a total of nine rounds as described in Figure 11a.

#### 2.4.7 Yeast Colony PCR for Sequencing

Approximately, 50 colonies from every third round of selection were picked and mixed in 20  $\mu$ L of 0.1% SDS in molecular biology grade water (Thermo Scientific SH30538.02) and heated for 5 minutes at 95  $^{\circ}$ C, then stored on ice. 2  $\mu$ L of lysed yeast cell mixture was used to provide template DNA for amplification through polymerase chain reaction (PCR). A PCR master mix was prepared using Taq DNA Polymerase (Life Technologies 10966-034) and dNTP mix (Life Technologies 18427-013) with a final volume of 50  $\mu$ L per reaction, as per the manufacturers recommended protocol. Forward and reverse primers (Figure 2.12) were mixed into the PCR master mix at a final concentration of 0.2  $\mu$ M. PCR was performed with a Mastercycler EP (Eppendorf) with a thermocycle programed as shown in Figure 2.13. PCR product (1163 base pair length)

```
PPNL6For Forward PCR Primer:  
5'-GTACGAGCTAAAAGTACAGTG-3'  
PNL6Rev Reverse PCR Primer:  
5'-TAGATACCCATACGACGTTTC-3'  
ForSeqP2 Forward Sequencing Primer:  
5'-TCTGCAGGCTAGTGGTGGTG-3'
```

Figure 2.13: Yeast colony PCR and sequencing primers.

1. 95  $^{\circ}$ C 5 min
2. 95  $^{\circ}$ C 30 sec
3. 55  $^{\circ}$ C 30 sec
4. 72  $^{\circ}$ C 45 sec
5. Repeat 30x Steps 2 – 4.
6. 72  $^{\circ}$ C 5 min
7. 4  $^{\circ}$ C hold

Figure 2.12: Yeast colony PCR program.

was verified using a 0.7 % Agarose gel and imaged using a Multiimage Light Cabinet (Alpha Innotech, Inc.) and submitted for sequencing to MWG Operon using a forward sequencing primer (Figure 2.12).

## 2.5 References

1. Kuhn, P. et al. Active Site and Oligosaccharide Recognition Residues of Peptide-N4-(N-acetyl- $\beta$ -D-glucosaminyl)asparagine Amidase F. *Journal of Biological Chemistry* **270**, 29493-29497 (1995).
2. Woods, R.J., Dwek, R.A., Edge, C.J. & Fraser-Reid, B. Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM\_93 Parameter Development. *The Journal of Physical Chemistry* **99**, 3832-3846 (1995).
3. Case, D.A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688 (2005).
4. Kirschner, K.N. et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622-655 (2008).
5. Tsui, V. & Case, D.A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **56**, 275-291 (2001).
6. Patrick, W.M. & Firth, A.E. Strategies and computational tools for improving randomized protein libraries. *Biomolecular Engineering* **22**, 105-112 (2005).
7. Miller, K.D., Pefaur, N.B. & Baird, C.L. Construction and screening of antigen targeted immune yeast surface display antibody libraries. *Curr Protoc Cytom* **Chapter 4**, Unit4 7 (2008).
8. Benatuil, L., Perez, J.M., Belk, J. & Hsieh, C.M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* **23**, 155-159 (2010).
9. Morris, T.A., Peterson, A.W. & Tarlov, M.J. Selective binding of RNase B glycoforms by polydopamine-immobilized concanavalin A. *Analytical chemistry* **81**, 5413-5420 (2009).
10. Mega, T., Oku, H. & Hase, S. Characterization of Carbohydrate-Binding Specificity of Concanavalin A by Competitive Binding of Pyridylamino Sugar Chains. *J Biochemistry* **111**, 396-340 (1992).
11. Fu, D., Chen, L. & O'Neill, R.A. A detailed structural characterization of ribonuclease B oligosaccharides by <sup>1</sup>H NMR spectroscopy and mass spectrometry. *Carbohydr Res* **261**, 173-186 (1994).
12. Noble, J.E., Knight, A.E., Reason, A.J., Di Matola, A. & Bailey, M.J. A comparison of protein quantitation assays for biopharmaceutical applications. *Mol Biotechnol* **37**, 99-111 (2007).

13. Prien, J.M., Prater, B.D. & Cockrill, S.L. A multi-method approach toward de novo glycan characterization: a Man-5 case study. *Glycobiology* **20**, 629-647 (2010).
14. Giancola, C. et al. Thermodynamic stability of the two isoforms of bovine seminal ribonuclease. *Biochemistry* **39**, 7964-7972 (2000).
15. Del Vecchio, P., Catanzano, F., de Paola, B. & Barone, G. Thermodynamic Stability of Ribonuclease B. *Journal of Thermal Analysis and Calorimetry* **61**, 363-368 (2000).
16. Rudd, P.M., Scragg, I.G., Coghill, E. & Dwek, R.A. Separation and analysis of the glycoform populations of ribonuclease B using capillary electrophoresis. *Glycoconj J* **9**, 86-91 (1992).
17. Joao, H.C., Scragg, I.G. & Dwek, R.A. Effects of glycosylation on protein conformation and amide proton exchange rates in RNase B. *FEBS Lett* **307**, 343-346 (1992).
18. Zauner, G., Koeleman, C.A., Deelder, A.M. & Wührer, M. Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *J Sep Sci* **33**, 903-910 (2010).
19. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755-768 (2006).
20. Ackerman, M. et al. Highly avid magnetic bead capture: An efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnology Progress* **25**, 774-783 (2009).
21. Asensio, J.L., Ardá, A., Cañada, F.J. & Jiménez-Barbero, J. Carbohydrate–Aromatic Interactions. *Accounts of Chemical Research* **46**, 946-954 (2012).

## CHAPTER 3

### EXPERIMENTAL CHARACTERIZATION OF LECTENZ® CANDIDATES

The four Lectenz® candidates (R617, R6113, R911, and R9113) selected from the computationally-guided yeast-display library selections were expressed in an *E.coli* expression system and purified via Immobilized Metal Affinity Chromatography (IMAC) followed by Size Exclusion Chromatography (SEC) to obtain pure protein. The purified proteins were used to investigate their utility as *N*-glycopeptide affinity reagents using Surface Plasmon Resonance (SPR), glycan array screening, and affinity chromatography.

#### 3.1 Cloning of PNGase F Clones into a Bacterial Expression Vector

The high yield and soluble expression of *Flavobacterium meningosepticum* PNGase F in *E.coli* using the pOPH6 bacterial expression vector was reported by Loo *et al.*<sup>1</sup> The pOPH6 vector incorporates a N-terminal OmpA periplasmic secretion tag to direct PNGase F to the periplasm. The construct also includes a C-terminal histidine tag for IMAC purification of expressed PNGase F. Using this vector, a D60A point mutant was constructed by site-directed mutagenesis (Dr. Loretta Yang).

PNGaseF-pOPH6 Forward Primer:

5' - **CGCAGGCCGGAATTC**AGCTCCGGCAGATAATACc-3'

EcoRI

PNGaseF-pOPH6 Reverse Primer:

5' - **TGGTGATGCGGATCCA**AGTTTGTAACCTACCGGAGCAC-3'

BamHI

Figure 3.1: PCR amplification primers for PNGase F clones selected via yeast-display. The 5' bold sequence matches the PNGaseF-pOPH6 sequence and the 3' end matches the PNGase F sequence in the PNGaseF-pPNL6 yeast display plasmid. The lowercase "c" at the 3' end of the PNGaseF-pOPH6 forward primer is a "G" in the PNGaseF-pOPH6 plasmid. The full length PCR product is flanked with EcoRI and BamHI restriction sites used to ligate the digested product into the PNGaseF-pOPH6 empty vector.



which includes the XbaI restriction site upstream of the OmpA sequence. In addition, the reverse primer was designed to introduce a XhoI restriction site immediate downstream of the stop codon. These restriction sites were used to double digest the PCR product and clone the gene into Xba I and XhoI double-digested pBluescript II KS(-) vector (Figure 3.3). The ligated plasmid containing the OmpA-PNGase F-His6 expression sequence clone was identified as PNGaseF-pOPH6 II and a vector map is provided in Figure 3.4. Five pOPH6 II plasmids were constructed, each containing one of the five PNGase F clones of interest: D60A, R617, R6113, R911, and R9113. Unlike the original pOPH6 vector, expression and purification was successfully achieved using the pOPH6 II vector.

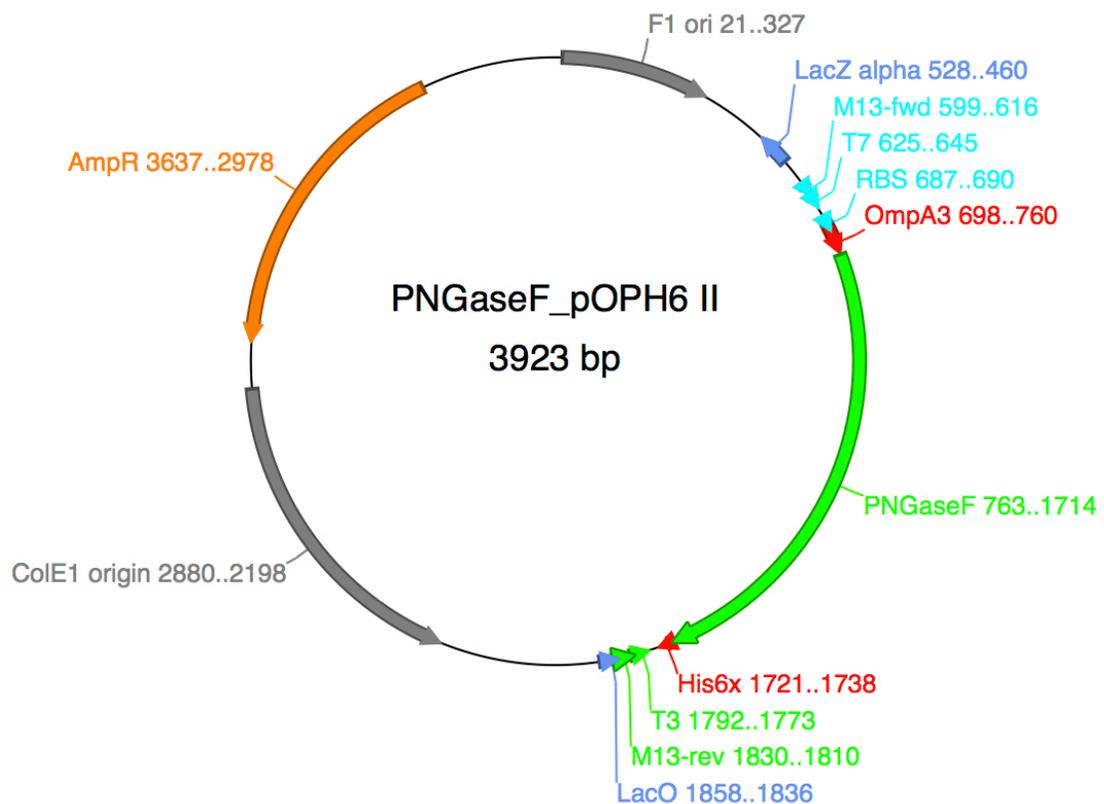


Figure 3.4: PNGase F-pOPH6 II vector map. The *E.coli* expression vector, pOPH6 II, was based on the pBluescript II KS(-) vector and has the OmpA-PNGase F-His6x sequence from pOPH6 for expression. The PNGase F-pOPH6 II expression plasmid is used for the D60A single point mutant as well as the four PNGase F clones selected from the yeast-display library selections: R617, R6113, R911, and R9113.

### 3.2 Expression and Purification of PNGase F Clones

Expression of yeast-display selected PNGase F clones using the original pOPH6 vector was unsuccessful using previously published protocols.<sup>1,2</sup> However, expression and purification was successfully achieved with the PNGase F-pOPH6 II vector using a protocol developed by Filitcheva *et al.*<sup>2</sup> This protocol was adapted to optimize expression and purification of PNGase F clones.

All five PNGase F-pOPH6 II (D60A, R617, R6113, R911, and R9113) plasmids were transformed into *E.coli* BL21-Gold(DE3) competent cells for expression. Expression of the protein of interest is under the control of the isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) inducible T7 promoter. In summary, starter 50 mL LB cultures containing 100  $\mu$ g/mL carbenicillin were inoculated with a single transformed colony selected from a LB-carbenicillin agar plate and grown overnight in a shaking incubator at 37 °C. The culture was expanded to 37 °C pre-warmed 1 L LB with carbenicillin. The temperature was dropped to 22 °C between OD<sub>600</sub> 0.4 - 0.5 and the culture was induced with IPTG and incubation continued for approximately 22 hours. The culture was harvested by collecting the cell pellet and subjecting it to mechanical cell lysis via a French Press. The cell lysate was centrifuged to separate insoluble cell debris from the supernatant containing the periplasmic fraction. This periplasmic fraction was loaded onto an IMAC column and the PNGase F clone eluted over an imidazole gradient. The fractions of the elution peak were pooled and concentrated using a 10 kDa cutoff Vivaspin concentrator and run through size exclusion chromatography for enhanced purity. The PNGase F clone elution peak fractions were pooled, concentrated, and protein yield determined by UV 280 absorbance ( $A_{280}$ ).

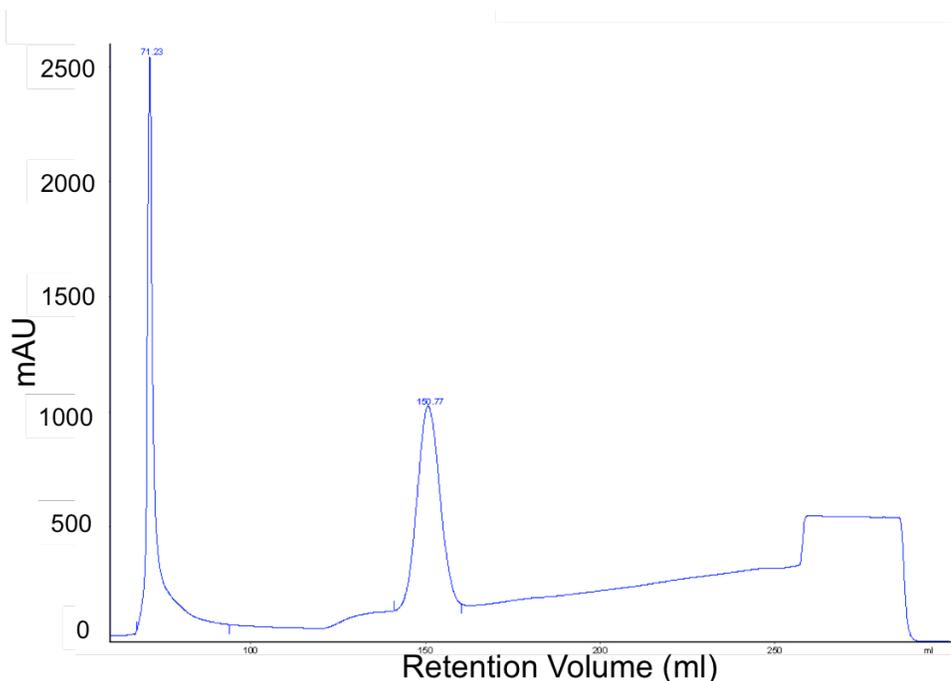


Figure 3.5: D60A IMAC elution chromatogram. Gradient elution profile of D60A. The first sharp peak is a 50 mM imidazole wash (8.3% B). The second peak corresponds with elution of D60A with an absorbance maximum at ~110 mM imidazole (20.5% B).

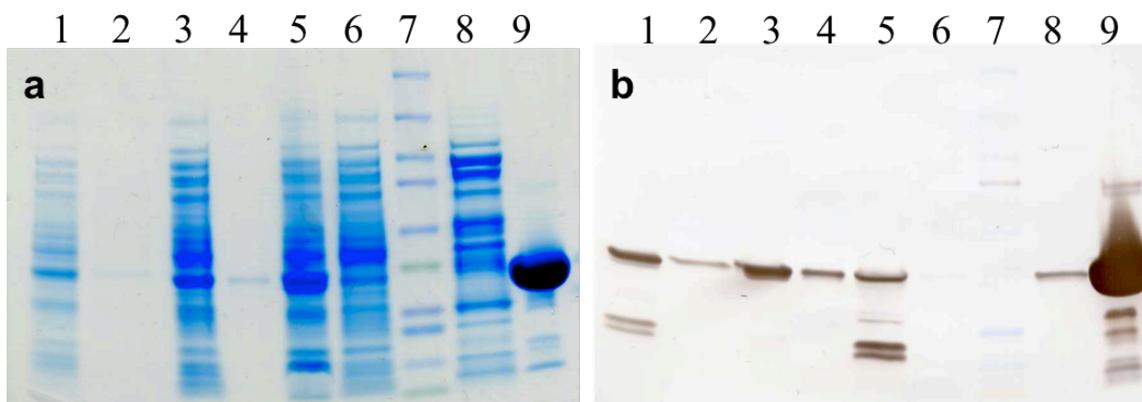


Figure 3.6: SDS-PAGE and Western Blot of IMAC purified D60A clone. a) Coomassie stained denaturing SDS-PAGE of D60A expression and IMAC purification samples. b) Western Blot of duplicate gel of D60A expression and IMAC purification samples. 1:5000 dilution of mouse anti-His6 HRP antibody used and developed with DAB substrate. A 36 kDa band corresponding with expressed D60A is visible across lanes 1-3, 5, and 8-9. Lane 1 = culture. Lane 2= culture supernatant. Lane 3 = soluble periplasmic fraction. Lane 4 = positive control PNGase F (300 ng). Lane 5 = insoluble cell lysate. Lane 6 = loading flow through. Lane 7 = Protein markers: 250 kDa, 150 kDa, 100 kDa, 75 kDa (visible on blot), 50 kDa, 37 kDa (green), 25 kDa, 20 kDa (visible on blot), 15 kDa, 10 kDa (green). Lane 8 = 50 mM imidazole wash. Lane 9 = Pooled elution peak fractions from imidazole gradient (24  $\mu$ g).

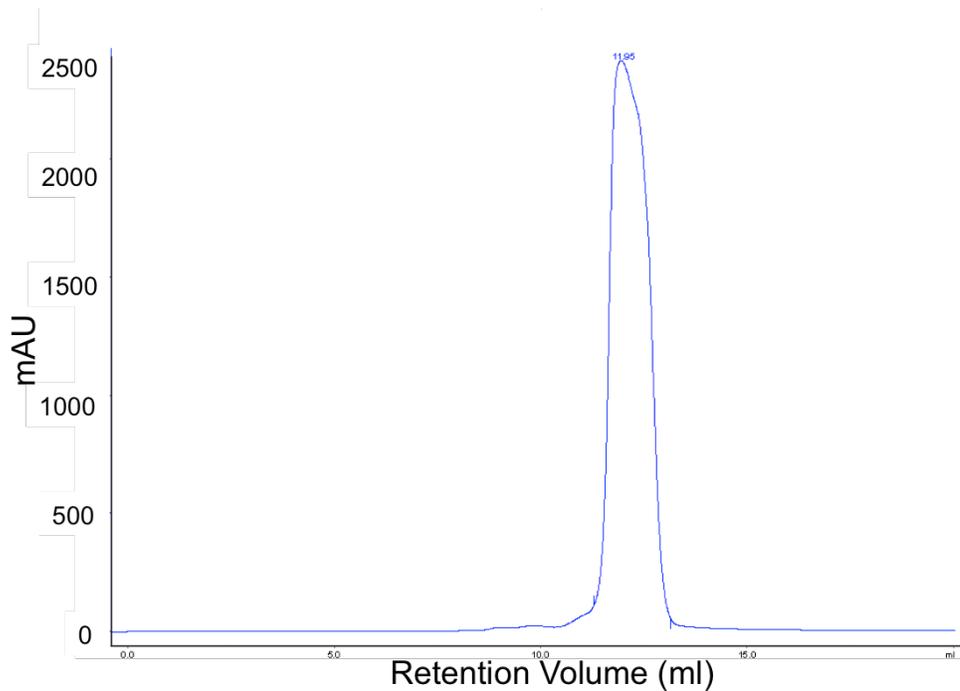


Figure 3.7: D60A SEC chromatogram on a Superdex 75 10/300 GL column. IMAC elution fractions were run through SEC to obtain high purity D60A protein. Both wtPNGase F and D60A eluted at 12 ml retention volume.

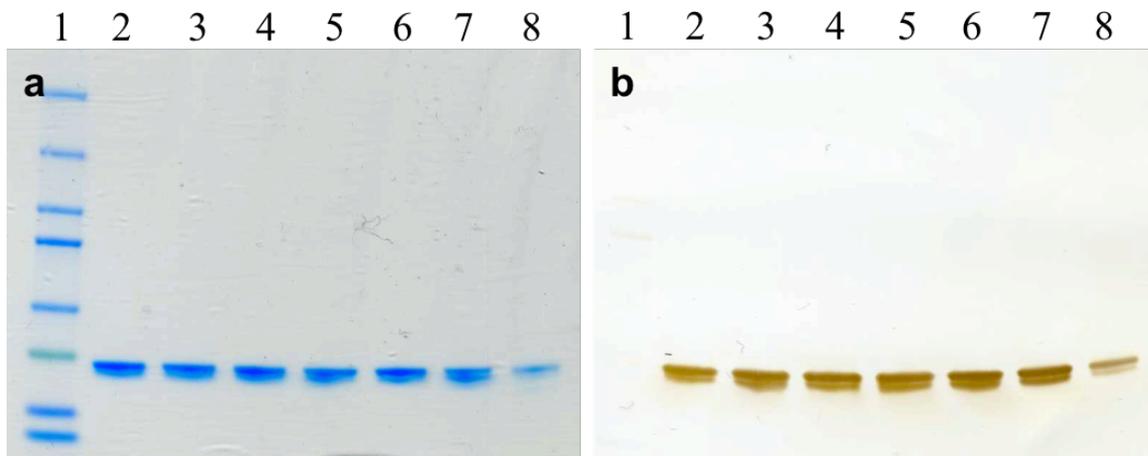


Figure 3.8: SDS-PAGE and Western Blot of SEC purified D60A clone. a) Coomassie stained denaturing SDS-PAGE of D60A SEC elution peak fractions. b) Western Blot of duplicate gel of D60A SEC elution peak fractions. 1:5000 dilution of mouse anti-His6 HRP antibody used and developed with DAB substrate. A 36 kDa band corresponding with expressed D60A is visible across lanes 2-7. Lane 1 = Protein markers (not visible in Western Blot): 250 kDa, 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa (green), 25 kDa, 20 kDa. Lanes 2-7 = D60A elution fractions (1  $\mu$ g each). Lane 8 = Positive control PNGase F (500 ng). The gel bands were purposefully allowed to migrate longer than normal in order to visualize doublet bands around 36 kDa corresponding with OmpA-D60A and D60A without the N-terminal OmpA secretion tag.

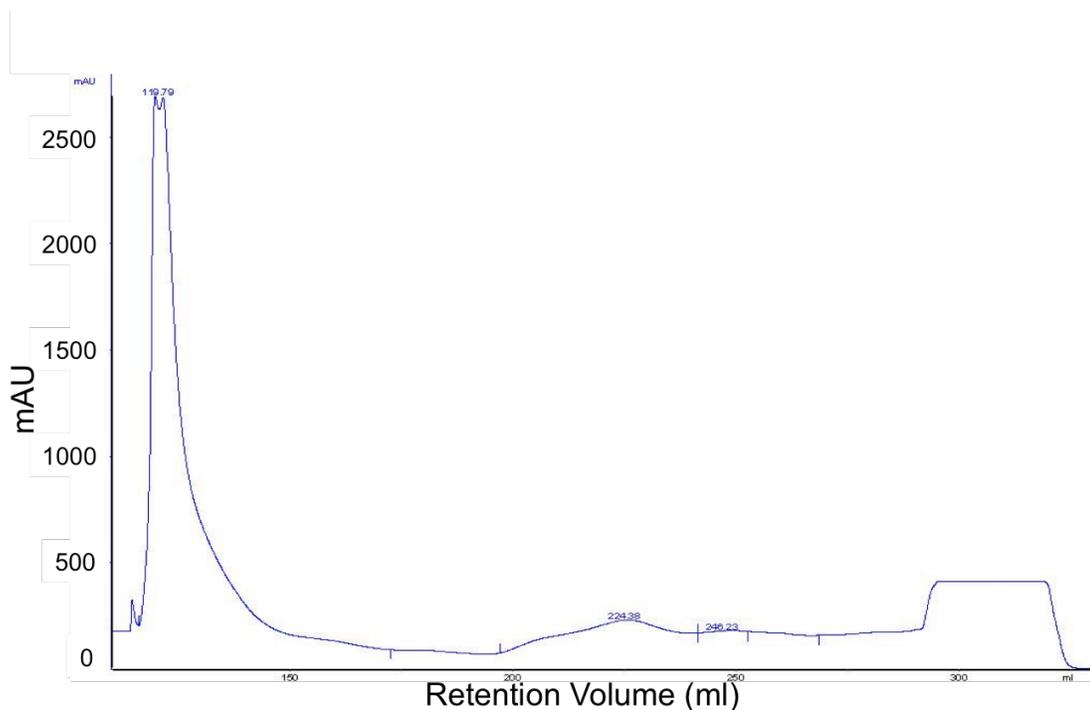


Figure 3.9: R911 IMAC elution chromatogram. Gradient elution profile of R911. The first sharp peak is a 50 mM imidazole wash (8.3% B). The shallow broad peaks correspond with elution of R911 and between 14.5% B - 31% B.

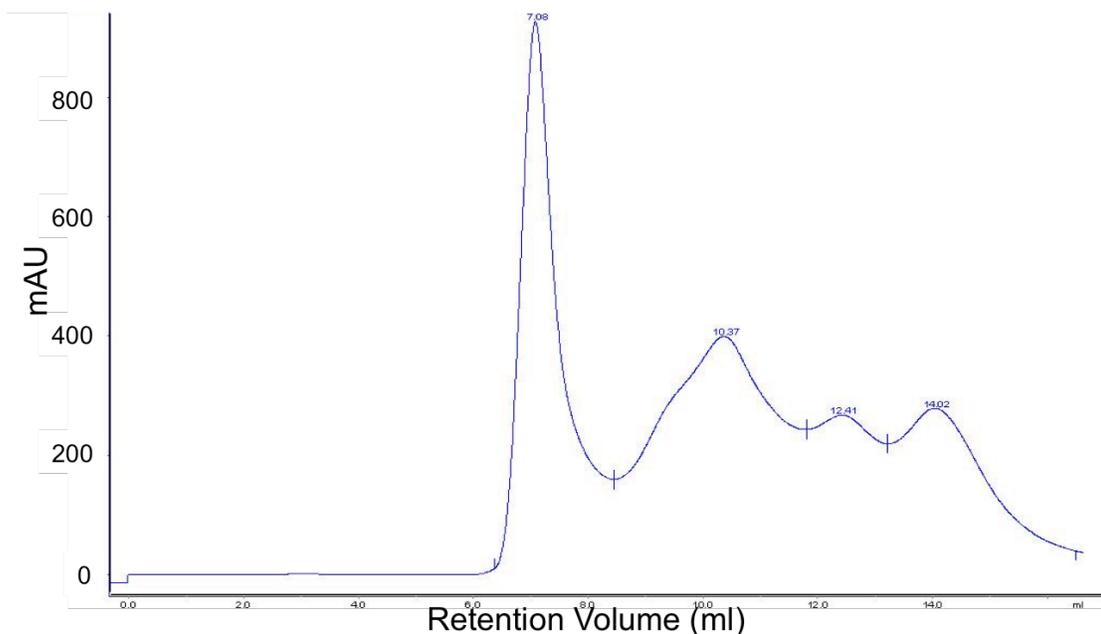


Figure 3.10: R911 SEC chromatogram on a Superose 12 10/300 GL column. IMAC elution fractions were run through SEC. The fourth peak with a peak maximum at 14.02 mL retention volume is consistent with D60A elution on this same column.

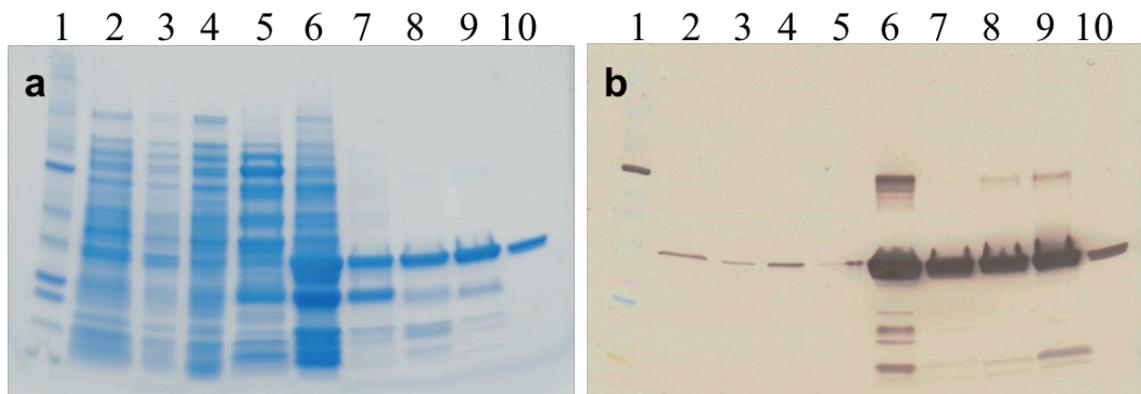


Figure 3.11: SDS-PAGE and Western Blot of IMAC and SEC purified R911. a) Coomassie stained denaturing SDS-PAGE of R911 IMAC and SEC elution fractions. b) Western Blot of duplicate gel of R911 IMAC and SEC elution fractions. 1:5000 dilution of mouse anti-His6 HRP antibody used and developed with DAB substrate. A 36 kDa band corresponding with expressed R911 is visible across lanes 2–9. Lane 1 = Protein markers: 250 kDa, 150 kDa, 100 kDa, 75 kDa (brown on blot), 50 kDa, 37 kDa (green), 25 kDa, 20 kDa, 15 kDa, 10 kDa (green on gel, yellow on blot). Lane 2 = culture. Lane 3 = insoluble cell lysate. Lane 4 = soluble periplasmic fraction. Lane 5 = 50 mM imidazole wash. Lane 6 = pooled IMAC elution fractions #42–63. Lane 7 = SEC fraction #15 corresponding to second SEC elution peak in Figure 3.10 with 10.37 mL retention volume. Lane 8 = SEC fraction #23 corresponding to third SEC elution peak in Figure 3.10 with 12.41 mL retention volume. Lane 9 = SEC pooled fractions #29–37 corresponding to fourth SEC elution peak in Figure 3.10 with 14.02 mL retention volume. Lane 10 = Positive control D60A (1  $\mu$ g)

The IMAC and SEC chromatogram elution profiles of expressed wtPNGase F and the D60A clone were similar. Figure 3.5 shows the IMAC elution chromatogram of D60A. Expression and IMAC purification samples were analyzed by denaturing SDS-PAGE and Western Blot and gel and blot images are shown in Figure 3.6. Similarly, Figures 3.7 and 3.8 show the corresponding SEC elution chromatogram of D60A, and SDS-PAGE and Western Blot images are shown in Figure 3.8. SDS-PAGE gel comparison of the soluble periplasmic fraction sample (Figure 3.6a Lane 3) with the insoluble cell lysate (Figure 3.6a Lane 5) indicates a significantly larger 36 kDa protein band consistent with D60A migration is present in the insoluble cell lysate. However,

only a small portion of this protein band was detected on the Western Blot (Figure 3.6b Lanes 5). Taken together these data suggest that a majority of the expressed D60A protein was in the soluble periplasmic fraction (Figure 3.6 Lane 3) and a minimal amount was in the insoluble cell lysate (Figure 3.6b Lane 5) by Western Blot and that the significantly larger 36 kDa band observed in the coomassie stained gel of the insoluble cell lysate was not D60A (Figure 3.6a Lane 5). Analysis of the loading flow through sample indicates that the his-tagged D60A protein was specifically being retained on the IMAC column as no D60A protein was detected in the Western Blot (Figure 3.6 Lane 6). A significant amount of non-specific proteins were visible in the SDS-PAGE gel (Figure 3.6 Lane 6) of the same loading flow through sample. The 50 mM imidazole wash step removed the majority of non-specific proteins with minimal loss of D60A protein as seen in Figure 3.6 Lane 8. Thus, the IMAC eluted D60A pooled sample shows minimal contamination with non-specific proteins in Figure 3.6 Lane 9, even when the gel and duplicate blot were overloaded with 24  $\mu$ g of total protein. SEC purification of D60A improves the purity even further as neither non-specific protein elution peaks (Figure 3.7) nor protein bands were detected by SDS-PAGE and Western Blot analysis (Figure 3.8) of the individual eluted fractions. Wild-type PNGase F and D60A clones were both successfully expressed and purified with a yield of  $\sim$ 3.0 mg high purity protein from a 1 L expression culture each. The identities of purified PNGase F and D60A were confirmed by MALDI. Additionally, sequence identity of D60A was also confirmed by LC-MS/MS (Rob Bridger).

The IMAC and SEC chromatogram elution profiles of R911 differed significantly from those of wtPNGase F and D60A as shown in Figures 3.9 and 3.10. However,

denaturing SDS-PAGE and Western Blot analysis of R911 expression and purification samples indicate similar results as that of PNGase F and D60A, suggesting that the difference in elution profiles of the R911 may be due to changes in structure of the native R911 relative to native PNGase F and D60A which cannot be distinguished by comparing denaturing gels. A significant difference in the IMAC elution profile of R911 was the elution of broad peak between 14.5% B and 31% B gradient (Figure 3.9) compared to the relatively sharp D60A IMAC elution peak at 20.5% B (Figure 3.5). Similarly, the R911 SEC elution profile shows four distinct elution peaks (Figure 3.10) of which the three latter elution peaks correspond to relatively pure R911 elution samples by denaturing SDS-PAGE and Western Blot analysis (Figure 3.11 Lanes 7, 8, and 9). Given these data, and that size exclusion chromatography separates proteins based on size and shape, it was likely that R911 structural isomers (likely a mixture of folded and misfolded R911) were present which cannot be distinguished by denaturing gel analysis. Furthermore, of these three R911 elution peaks, only the third elution peak, with a peak maximum at 14.02 mL retention volume, was consistent with D60A elution (peak maximum also at ~14 mL) on the same Superose 12 10/300 GL column. This indicates that the third R911 elution peak at 14.02 mL retention volume was the correctly folded R911 isomer. The total R911 protein yield from a 2 L LB culture is ~2.0 mg, corresponding to ~1.31 mg (65%) from SEC elution peak 1, ~0.3 mg (15%) from SEC elution peak 2, and ~0.4 mg (20%) from SEC elution peak 3. With only 20% of the total R911 correctly folded, the effective yield was only ~0.4 mg. Circular dichroism or NMR experiments could help to identify the general structural difference between these hypothesized three folded and misfolded R911 isomers.

Several attempts were made to express and purify R617, R113, and R9113 clones; however, the IMAC elution profile always had a similarly broad but even shallower elution profile relative to R911 with insufficient quantity of protein for detection by Western Blot. Wild-type PNGase F has three disulfide bonds required for proper folding, and since all four yeast-display selected clones introduced a cysteine residue at one of the site for mutagenesis (Table 2.4), it would not be surprising if the addition of an extra cysteine was contributing to the suspected misfolded R911 and the inability to purify his-tagged R617, R6113, and R9113 for study. Thus point mutants R617 C57D, R6113 C192G, R911 C60A, and R9113 C60A were constructed where R617 and R6113 cysteine residues were reverted back to wild-type and R911 and R9113 cysteine residues were mutated to alanine instead of wild-type aspartate given that D60 is required for catalytic activity. Table 3.1 lists the physical and chemical properties of PNGase F clones of interest.

Table 3.1: Physical and chemical properties of PNGase F clones. ExPASy ProtParam calculated properties based on amino acid sequence are reported.<sup>3</sup> Molecular weight, isoelectric point, and extinction coefficients ( $\epsilon$ ) values are listed.

PNGase F Clones (326 Amino Acids)	Molecular Weight	Isoelectric Point	$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ )	$\epsilon$ ( $L \text{ g}^{-1} \text{ cm}^{-1}$ )	$\epsilon$ <sup>1%</sup>
wtPNGase F	36251.6	7.75	73715	2.0334	20.334
D60A	36207.6	8.14	73715	2.0359	20.359
R617	36104.5	8.70	72225	2.0004	20.004
R617 C192G	36058.4	8.74	72225	2.0030	20.030
R6113	36260.7	8.82	77725	2.1435	21.435
R6113 C57D	36272.7	8.74	77725	2.1428	21.428
R911	36281.8	8.36	79215	2.1833	21.833
R911 C60A	36249.7	8.40	79215	2.1853	21.853
R9113	36372.9	8.36	84715	2.3291	23.291
R9113 C60A	36340.8	8.40	84715	2.3311	23.311

The successful expression of cysteine point mutants was confirmed by Western Blot analysis of IMAC purified R617 C57D, R6113 C192G, R911 C60A, and R9113 C60A (data not shown). However, only sufficient amount of R911 C60A could be

produced for experimental requirements, thus only R911 and R911 C60A were investigated further. The cysteine point mutants confirmed that the presence of an extra cysteine is contributing only in part to the altered elution profiles and the proposed structural isoforms of R911. Interestingly, the elution profile of R911 C60A (data not shown) was consistent with that of R911 (Figures 3.9, 3.10, 3.11), indicating that other five mutated residues must also contribute to the altered IMAC and SEC elution profile of R911.

### 3.3 Activity and Kinetic Studies

In order to convert the wtPNGase F enzyme into a high affinity Lectenz® reagent, catalytic activity needed to be abolished while simultaneously enhancing affinity. The PNGase F D60A single point mutant was of particular interest for three reasons: 1) residue D60 is required for catalytic activity based on D60N point mutant studies that demonstrated this mutation made the enzyme catalytically inactive<sup>4</sup>, 2) computational alanine scanning data (Table 2.3) predicted favorable interaction energy for substrate affinity, and 3) given that the D60A single point mutant was not affinity enhanced via directed evolution, it was appropriate to use D60A as a catalytically inactive, non-affinity enhanced control for comparison to the affinity enhanced R911 clone.

Table 3.2: Bovine pancreatic RNase properties. RNase A and RNase B was obtained from Sigma. The reported purities of RNase A and RNase B are 90% and 80% respectively as determined by SDS-PAGE. RNase B is the glycosylated variant of RNase A. The glycosylation site at N34 is reported to have nine glycoforms, thus the reported mass is an average derived from the relative abundance of each of the glycosylated species<sup>a, 5</sup>. Molecular weight, isoelectric point, and extinction coefficients ( $\epsilon$ ) values are listed.

RNase (124 Amino Acids)	Molecular Weight (Da)	Isoelectric Point	$\epsilon$ ( $M^{-1} cm^{-1}$ )	$\epsilon$ ( $L g^{-1} cm^{-1}$ )	$\epsilon$ <sup>1%</sup>
RNase A	13,700	9.6	8,640	0.71	7.1
RNase B	15,095 <sup>a</sup>	-	8,213	0.80	8.0

The enzymatic activity of clones D60A and R911 were both investigated. Properties of the glycosylated substrate, RNase B, and the non-glycosylated version, RNase A, are presented in Table 3.2. A gel shift assay was used to determine *N*-deglycosylation catalytic activity of the clones on denatured RNase B relative to the wtPNGase F enzyme (Table 3.3). The D60A single point mutant has significantly decreased catalytic activity (~13% relative to wtPNGase F) while the R911 clone displayed no detectable catalytic activity in samples from overnight reactions. The deglycosylation of RNase B was further confirmed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) using an AB SCIEX 5800 TOF-TOF. The deglycosylated RNase B product's mass was confirmed to be consistent with RNase A, with the difference being that N34 becomes D34 due to deamination by PNGase F during the *N*-glycan cleavage reaction.<sup>6</sup>

Table 3.3: Deglycosylation activity of PNGase F clones. A gel shift assay was used to determine deglycosylation activity of PNGase F clones relative to wtPNGase F. 50 ng of wtPNGase F, D60A, and R911 each was incubated with 50 µg of denatured RNase B in 50 mM EPPS, pH 8.0 in a 50 µL reaction volume at 37 °C overnight. Samples were analysed on a SDS-PAGE gel and altered migration of deglycosylated RNase B product relative to RNase B was observed. The scanned gel image was analyzed by ImageJ software to quantitate deglycosylated product relative to RNase B substrate.<sup>7</sup> Deglycosylated product confirmed by MALDI TOF-TOF mass spectrometry.

	wtPNGase F	D60A	R911
Deglycosylation Activity on RNase B	100%	13.4%	Not detected

Biomolecular interaction kinetic experiments were conducted on a Biacore 3000 instrument via surface plasmon resonance (SPR). SPR is a phenomenon that occurs when plane-polarized incident light stimulates oscillations of electrons, or the propagation of electromagnetic waves (plasmons), parallel to a metal (conductive)/dielectric interface. Plasmon waves propagate at the interface of the metal and liquid (or air) mediums extending out about 300 nm, and changes at the interface due to the adsorption of

molecules to the surface results in changes in wave propagation causing a shift in the angle of the reflected incident light under conditions of total internal reflection. Total internal reflection is achieved by placing a glass prism placed directly against a gold surface, where plasmons are excited. Due to the high sensitivity of SPR to shifts in mass, it has been adapted for biomolecular interaction measurement.<sup>8</sup> By using a carboxy-modified dextran gold surface, a target ligand of interest can be covalently immobilized onto a dextran-derivitized gold surface using amine coupling.

The evaluation of lectin-carbohydrate interactions by SPR is a well-established technique for kinetic analysis.<sup>9, 10</sup> CM-5 carboxy methyl dextran sensor chips are utilized for amine-coupling of ligands of interest. However, this approach results in a randomized orientation of the ligand and may not be suitable where ligand orientation is particularly critical for interaction with an analyte of interest or the effect of ligand orientation is of specific interest. Ligand orientation can be achieved by using a Ni-NTA derivitized dextran surface to capture histidine-tagged proteins.<sup>11</sup> This approach has a significant drawback that the capture molecule will leech off the surface since they are not covalently linked. Recently, covalent immobilization of histidine-tagged proteins to overcome leeching has been demonstrated.<sup>12</sup> Using microfluids, an analyte of interest is flowed through a flow cell containing the immobilized ligand where interactions between the biomolecules (immobilized ligand and analyte) can occur. Simultaneously, on the opposite side of the sensor surface the degree of change in the angle of the reflected light is proportional to the change in mass.

Denatured RNase B and denatured RNase A, which has the same peptide sequence as RNase B but lacks *N*-glycosylation, were covalently coupled to CM-5 chips using amine-coupling chemistry. Prior to immobilizing the RNase ligands to the carboxy methyl dextran surface on the CM-5 sensor chip a pH scouting experiment indicated an optimal pH of 5.5 for efficient coupling

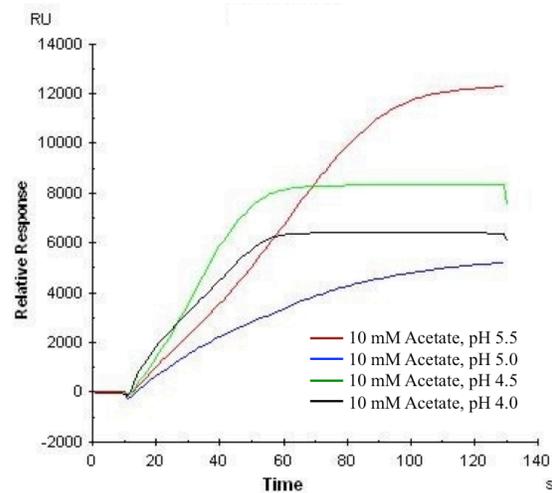


Figure 3.12: RNase B pH scouting. A 10 mM acetate pH 5.5 coupling buffer yielded the most efficient coupling of RNase ligands to the carboxy methyl dextran CM-5 sensor surface chip.

(Figure 3.12). A high-density surface was prepared with sufficient ligand coupling to yield a theoretical  $R_{MAX}$  of  $\sim 3000$  RU. To assess the impact of the D60C mutation in R911 relative to wtPNGase F, the R911 C60A mutant was also evaluated. Steady-state binding kinetics using a bimolecular interaction model was determined using a Biacore 3000 instrument.

The use of SPR to measure binding kinetics between RNase B and yeast PNGase enzyme and a mutant enzyme has been demonstrated.<sup>13</sup> Using a similar strategy, denatured RNase B was immobilized on a CM-5 sensor chip and serially-diluted concentrations of wtPNGase F, D60A, R911, and R911 C60A were passed over the sensor surface, while binding kinetics were measured. A summary of the kinetic results is presented in Table 3.4 and sensograms are shown in Figure 3.13. The wtPNGase F has a  $K_D$  of  $6.4 \mu\text{M}$  and an off-rate ( $k_{off}$ ) of  $0.1 \times 10^{-1} \text{ s}^{-1}$ . Relative to the D60A control clone ( $K_D = 2.7 \mu\text{M}$ ), the selected R911 clone has 10x enhanced affinity ( $K_D = 0.26 \mu\text{M}$ ).

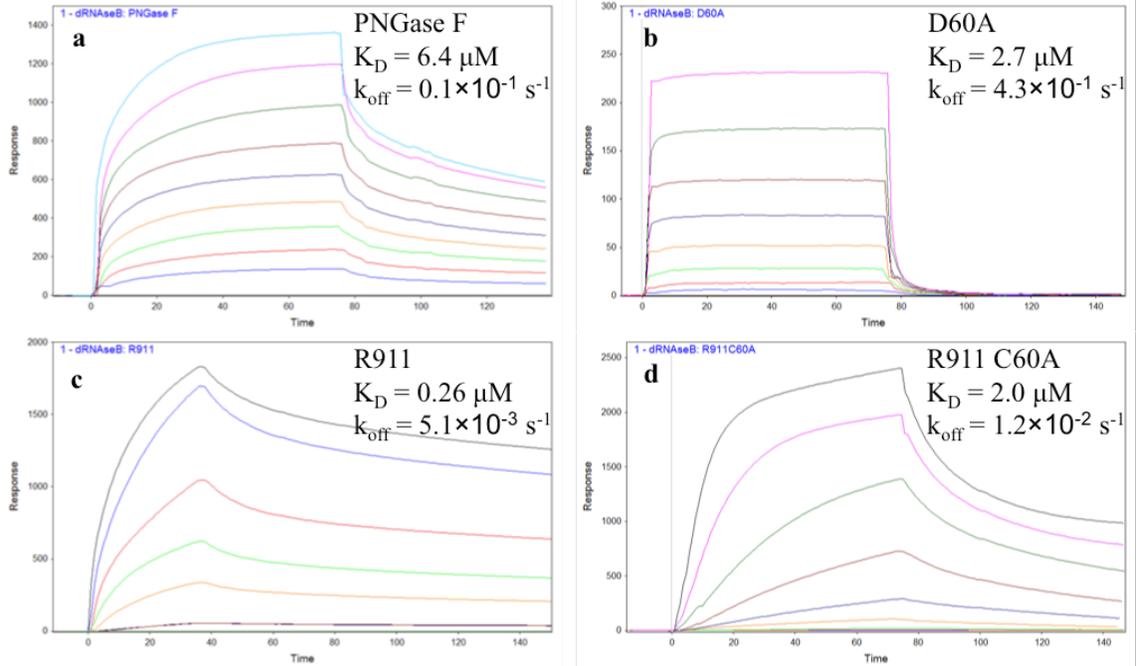


Figure 3.13: SPR sensograms of wtPNGase F, D60A, R911, and R911 C60A. A high-density surface was prepared by immobilizing denatured RNase B to yield a maximum response ( $R_{MAX}$ ) of  $\sim 3200$  RU. a) wtPNGase F: 250 nM – 64  $\mu$ M serial dilutions, b) D60A: serial dilutions 72 nM – 20  $\mu$ M, c) R911: 78 nM – 5  $\mu$ M serial dilutions, and d) R911 C60A: 78 nM – 10  $\mu$ M serial dilutions. The data obtained were analyzed by Scrubber 2.0c.

Furthermore, the selected R911 clone has an 84x decreased off-rate ( $k_{off} = 5.1 \times 10^{-3} \text{ s}^{-1}$ ) relative to the D60A control clone ( $k_{off} = 4.3 \times 10^{-1} \text{ s}^{-1}$ ). The R911 C60A variant clone exhibited only 1.3x enhanced affinity ( $K_D = 2.0 \text{ } \mu\text{M}$ ) and 35x decreased off-rate ( $k_{off} = 1.2 \times 10^{-2} \text{ s}^{-1}$ ) relative to the D60A control clone. Denatured Ribonuclease A (dRNase A) was also utilized as a negative control ligand as it is a non-glycosylated version of RNase B; however, unsurprisingly specific kinetic measurements could not be measured given that RNase A lacks the *N*-glycan moiety recognized by PNGase F. Specifically, a high-density surface with immobilized dRNase A was prepared to yield a theoretical maximum response ( $R_{MAX}$ ) of  $\sim 1800$  RU; however, measured responses with wtPNGase F, D60A, R911, and R911 C60A exceeded  $R_{MAX}$ , indicating that interactions were non-specific.

Table 3.4: Surface plasmon resonance kinetic data for PNGase F clones. Dissociation ( $K_D$ ) and off-rate constants ( $k_{off}$ ) for the interaction between denatured glycoprotein RNase B (dRNase B) with PNGase F mutagenized clones as determined by surface plasmon resonance. Kinetic data determined with dRNase B as immobilized ligand on CM5 sensor chip and PNGase F clones as analytes. Experimental  $\Delta G$  binding interaction energy ( $\Delta G_{BIND-EXP}$ ) was calculated from the  $K_D$ .

PNGase F Clones	$\Delta G_{BIND-EXP}$ (kcal/mol)	$K_D$ (M)	Relative Affinity Enhancement	$k_{off}$ ( $s^{-1}$ )	Relative Off-Rate Enhancement
wtPNGase F	-7.103	$6.4 \times 10^{-6}$	-	$0.1 \times 10^{-1}$	-
D60A	-7.609	$2.7 \times 10^{-6}$	1x	$4.3 \times 10^{-1}$	1x
R911	-8.990	$2.6 \times 10^{-7}$	10x	$5.1 \times 10^{-3}$	84x
R911 C60A	-7.768	$2.0 \times 10^{-6}$	1.3x	$1.2 \times 10^{-2}$	35x

The activity assay and kinetic data indicate that the selected R911 clone is catalytically inactive and has significantly enhanced affinity relative to the non-affinity enhanced PNGase F D60A control clone. In addition, the enhanced off-rate of R911 is significant because a slow off-rate is a key criteria for a useful affinity reagent to enrich target glycans, unlike enzymes, which generally have rapid turn-over to release product. The kinetic analysis of R911 C60A clone provides additional insight into the importance of the cysteine residue at position 60. Both the affinity and the off-rate are negatively impacted by the C60A mutation. This indicates two critical pieces of information: 1) the D60C mutation in R911 is critical for high affinity and 2) that improved affinity also directly impacts the slower off-rate of R911. Based on these results, R911 satisfies the kinetic criteria for a Lectenz® candidate (Figure 1.11). A computational modeling-based analysis of the energetic contributions of R911 mutations relative to wtPNGase F is discussed in Chapter 4.

### 3.4 Lectenz® Affinity Chromatography

Lectin affinity chromatography is the most widely applied technique for the isolation and enrichment of glycans and glycoconjugates.<sup>14</sup> Despite the inherent limitations of current carbohydrate-detection reagents like antibodies and lectins,

numerous affinity-based glycan and glycoconjugate enrichment formats have been developed signifying the critical need for this application. Common enrichment techniques include lectins conjugated to agarose/sepharose packed in centrifugal devices, spin or low-pressure LC columns, and HPLC-compatible matrices that enable high-pressure/high flow rate lectin chromatography, and lectin-modified gold nanoparticles embedded in pipette-tips.<sup>14, 15</sup> Recently, serial lectin affinity chromatography has been employed to enrich glycoproteins of interest from complex samples like sera and cancer cell lysates.<sup>16, 17</sup> However, the choice of reagent used for sample enrichment or isolation can therefore bias the outcome of glycomic analyses toward a subset of glycoconjugates based on the binding properties of the lectin or antibody.<sup>18</sup>

The application of the R911 Lectenz® candidate for enrichment of glycoconjugates in an affinity chromatography format was investigated. Using HiTrap N-hydroxysuccinamide (NHS)-activated HP columns, purified PNGase F D60A and R911 clones were covalently linked to the column matrix to evaluate affinity chromatography based enrichment of *N*-glycopeptides and *N*-glycoproteins. The coupling efficiencies of the clones to the NHS-activated columns consistently ranged between 80% - 87% for all NHS-activated column-coupling reactions. The binding buffer consisted of 10 mM HEPES, 10 mM NaCl, pH 7.4 where as the elution buffer consisted of 10 mM HEPES, 150 mM NaCl, pH 7.4, and a constant flow rate of 0.4 mL/min through out all chromatography runs.

### 3.4.1 Enrichment of RNase B versus RNase A

The control D60A affinity chromatography results indicate no enrichment of *N*-glycosylated RNase B relative to RNase A (Figure 3.14). The flow through peaks at 1.25 mL retention volume indicate that both RNase B and RNase A flowed through the column during initial loading (0 – 5 mL retention volume) and were not retained due to

interactions with D60A. No elution peak was observed for RNase B when running elution buffer (5 – 10 mL retention volume). A small elution peak is visible for RNase A at 7.22 mL retention volume. This may be attributed to impurities in the RNase A sample as it is 90% pure and likely contains some RNase

B as a contaminant. In addition, wtPNGase F is known to recognize both the chitobiose core as well as the peptide glycosylation sequone (Asn – X(-Pro) – Ser/Thr) on the peptide backbone, thus, it is possible that the Asn – Leu – Thr glycosylation sequone on RNase A is being weakly recognized by the D60A single point mutant. Nonetheless, the

relative quantity of the small elution peak observed from the RNase A is minimal.

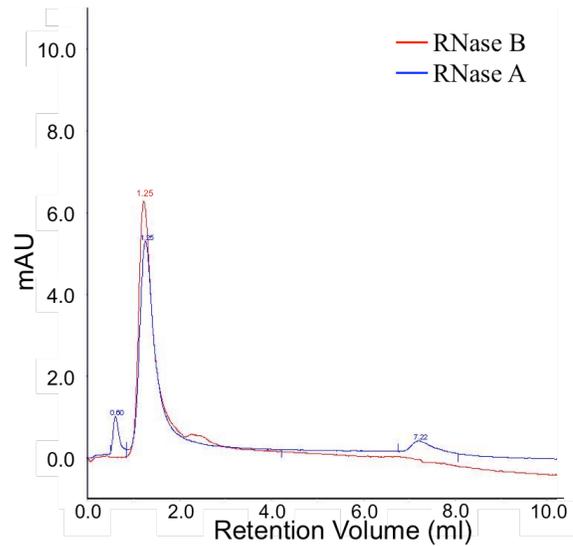


Figure 3.14: D60A affinity chromatography with RNase A and RNase B.

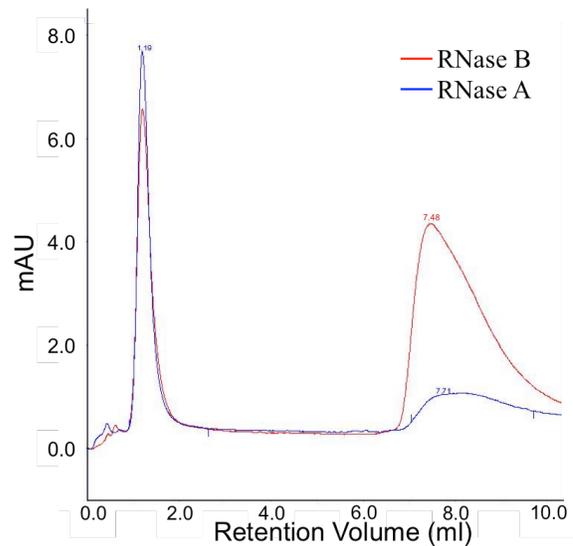


Figure 3.15: R911 Lectenz® affinity chromatography of RNase A vs RNase B.

Unlike D60A, R911 affinity chromatography loading and elution profiles indicate enrichment of *N*-glycosylated RNase B compared to RNase A (Figure 3.15). To confirm specific R911:RNase B glycan interactions, RNase B was deglycosylated with PNGase F. The deglycosylated RNase B was run through the R911 affinity column (Figure 3.16).

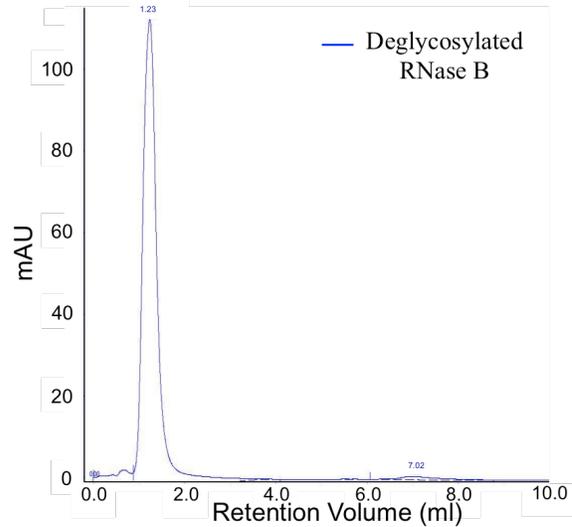


Figure 3.16: R911 Lectenz® affinity chromatography of deglycosylated RNase B.

The chromatogram shows that deglycosylated RNase B was not retained by R911 and flowed through the column. Taken together, these results confirm specific interaction between R911 and RNase B glycans.

### 3.4.2 Affinity Chromatography of RNase A and RNase B Tryptic Digests

To investigate the separation of peptides from *N*-glycopeptides using R911, RNase A and RNase B were digested with trypsin. The tryptic digests were loaded on to the R911 column. RNase A tryptic digest peptides flowed through the column, whereas part of the RNase B tryptic digest sample was retained on the column and eluted with elution buffer (Figure 3.17). The flow

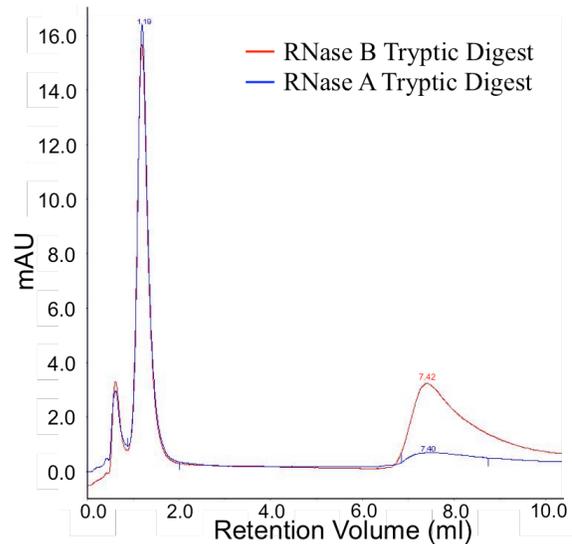


Figure 3.17: R911 Lectenz® affinity chromatography of tryptic digests of RNase A and RNase B.

through and elution samples were analysed by LC-MS/MS which confirmed that some *N*-glycopeptides were enriched.

### 3.4.3 Competitive Elution of RNase B with Free Chitobiose

Competitive elution with chitobiose of R911 bound RNase B was performed to further confirm specific interaction of R911 with the chitobiose core of *N*-glycan structure. RNase B was first loaded onto the R911 affinity matrix and then competitively eluted with free chitobiose in the binding buffer, instead of the standard elution buffer (Figure 3.18). LC-MS/MS analysis of the eluted sample

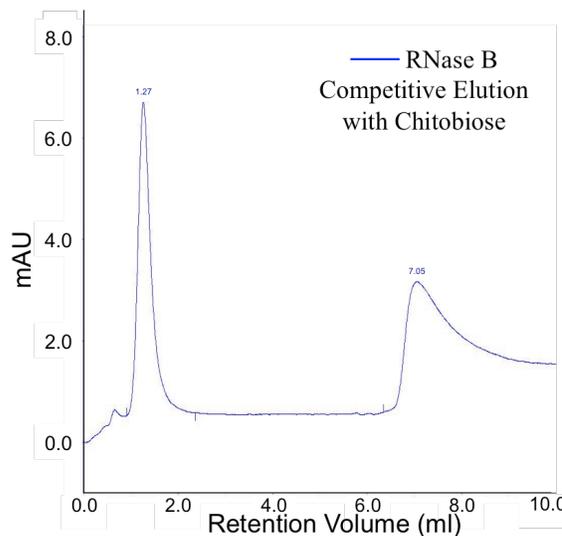


Figure 3.18: R911 Lectenz® affinity chromatography of RNase B using free chitobiose for competitive elution.

confirmed that RNase B was competitively eluted with chitobiose.

### 3.4.4 Enrichment of *N*-glycoproteins from MCF7 Whole Cell Extract using Competitive Elution with Chitobiose

The application of the R911 as an *N*-glycoprotein affinity enrichment reagent was demonstrated using MCF7 whole cell extract. Cell extract (100 µg) was loaded onto a R911 affinity column and then competitively eluted with free chitobiose in the binding buffer, instead of the standard elution buffer. (Figure 3:19). A majority of cell extract proteins flowed through the column corresponding with the observed peak at 1.19 mL retention volume and approximately 6.6 µg of protein was retained on the column and competitively eluted with free chitobiose corresponding with the peak at 6.63 mL retention volume.

The stock MCF7 cell extract sample and the competitively eluted sample were analyzed by LC-MS/MS and proteins were identified with UniProt and UniPep databases.<sup>19, 20</sup> R911 Lectenz® Affinity Chromatography (LAC) results were compared with a reported Multi-Lectin Affinity Chromatography (MLAC) experiment also conducted with MCF7 cell extract using Jac, ConA and WGA

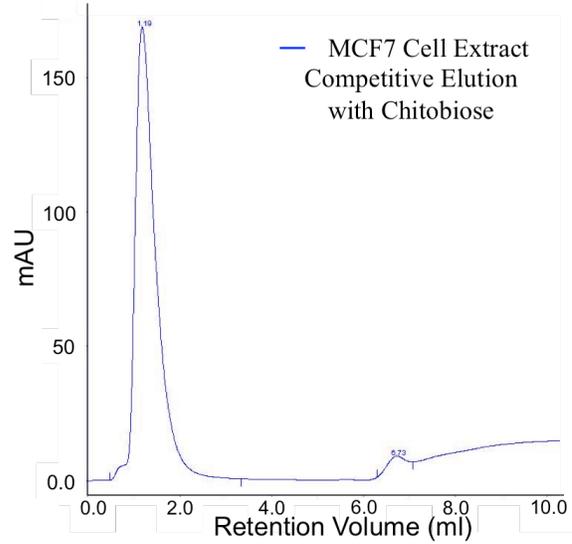


Figure 3.19: R911 Lectenz® affinity chromatography of MCF7 cell extract using free chitobiose for competitive elution.

lectins and are summarized in Table 3.5.<sup>17</sup> MLAC with MCF7 cell extract resulted in the elution of 88 proteins of which 84% are glycoproteins. R911 Lectenz® affinity chromatography resulted in the elution of 73 proteins of which 71.2% are glycoproteins. The glycoproteins eluted by R911 LAC are predominantly different than the glycoproteins eluted by MLAC. Furthermore, 11 glycoproteins identified by MLAC were present in the MCF7 cell extract, but not enriched by R911 LAC. These differences are not surprising given the different specificities of the capture reagents employed for enrichment.

A summary of glycoprotein enrichment by R911 LAC is provided in Table 3.6. Relative to the MCF7 cell extract stock sample, the R911 LAC eluted glycoprotein sample represents a 3.4x glycoprotein enrichment. Furthermore, the eluted glycoproteins consist of 42.5% (31) *N*-glycoproteins and 28.8% (21) *O*-glycoproteins, representing a 2.0x fold *N*-glycoprotein enrichment and 5.2x *O*-glycoprotein enrichment.

Table 3.5: Comparison of MLAC and R911 Lectenz® affinity chromatography with MCF7 cell extract.

Chromatography	Multi-Lectin Affinity Chromatography <sup>17</sup>	Lectenz® Affinity Chromatography
Capture reagent(s)	Jac, Con A, WGA Lectins	R911 Lectenz®
Capture reagent (mg)	3.4 mg total lectin	0.22 mg R911 Lectenz®
MCF7 cell extract (mg)	0.7 mg	0.1 mg
Capture conditions	O/N incubation at 4 °C	0.4 mL/min flow-rate at 4 °C
Eluted glycoproteins	86.5%	71.2%
Differences	11 glycoproteins detected in common that were eluted by MLAC, but not by Lectenz®	Majority of eluted glycoproteins are different than MLAC

The enrichment of *O*-glycoproteins by R911 LAC was unexpected given that

R911 is derived from the *N*-glycan processing enzyme PNGase F and substrate specificity of the enzyme for the *N*-glycopeptide and chitobiose core is well established.<sup>4,</sup>

<sup>21-24</sup> Insight into the enrichment of *O*-glycoproteins is provided from the observation that 76% (16) of the eluted *O*-glycoproteins are *O*-GlcNAcylated, indicating that the common structural motif being recognized of R911 is likely the reducing GlcNAc of both *N*-glycoproteins and *O*-GlcNAcylated glycoproteins.<sup>25, 26</sup> Thus, enrichment of both *N*-glycoproteins and *O*-GlcNAcylated glycoproteins can be achieved by the R911 Lectenz® making it a unique capture reagent which can recognize a common core motif in both *N*-glycoproteins and *O*-GlcNAcylated glycoproteins. The enriched *N*-glycoproteins and *O*-glycoproteins are listed in Tables 3.7 and 3.8 respectively.

Table 3.6: R911 Lectenz® affinity chromatography enrichment of MCF7 cell extract glycoproteins.

	Stock MCF7 Cell Extract Sample	R911 Lectenz® Eluted MCF7 Sample	Enrichment
Total glycoprotein %	26.9%	71.2%	3.4x
<i>N</i> -glycoprotein %	21.2%	42.5%	2.0x
<i>O</i> -glycoprotein %	5.5%	28.8%	5.2x
		(76% are <i>O</i> -GlcNAcylated)	

Table 3.7: Eluted MCF7 *N*-glycoproteins identified by LC-MS/MS.

Accession	Gene Symbol	Cellular Location	<i>N</i> -Glycoprotein	Mass (kDa)
8WZ42-2	TITIN	Golgi, cytoplasm, nucleus	Isoform 2 of Titin	3803.48
P21333-2	FLNA	trans-Golgi, cytoplasm, cytoskeleton	Isoform 2 of Filamin-A	279.83
P19835-2	CEL	Secreted	Isoform Short of Bile salt-activated lipas	71.77
P02768	ALBU	Secreted	Serum Albumin Precursor	69.30
P29401	TKT	extracellular vesicular exosome, nucleus, peroxisome, cytosol	Transketolase	67.82
P04264	KRT1	extracellular space	Keratin, type II cytoskeletal 1	65.98
P35908	KRT2A	Golgi, extracellular space	Keratin, type II cytoskeletal 2 epiderma	65.38
P33527	Z29074	extracellular space	Keratin 9, cytoskeletal, (Cytokeratin 9)	62.07
P48669	K2CF	extracellular vesicular exosome	Keratin, Type II Cytoskeletal 6F (Cytokeratin 6F) (CK 6F) (K6F Keratin)	60.01
P13645	KICJ	extracellular vesicular exosome, cytoplasm	Keratin, Type I Cytoskeletal 10 (Cytokeratin 10) (K10) (CK 10)	59.46
P04745	AMY1A	Secreted	Alpha-amylase 1	57.71
P19013	K2C4	cytoskeleton	Keratin, Type II Cytoskeletal 4 (Cytokeratin 4) (K4) (CK4)	57.21
P01008	SERPINC1	Secreted, extracellular space	Antithrombin-III	52.55
P08729	K2C7	Golgi apparatus, cytoplasm, cytoskeleton	Keratin, Type II Cytoskeletal 7 (Cytokeratin 7) (K7) (CK 7)	51.29
P16233	PNLIP	Secreted	Pancreatic triacylglycerol lipase	51.11
P15086	CBPB1	Secreted	Carboxypeptidase B	47.32
P15085	CBPA1	Secreted	Carboxypeptidase A1	47.09
O60664	PLIN3	Golgi, endosome membrane	Perilipin-3	47.03
P48052	CBPA2	Secreted	Carboxypeptidase A2	46.98
P28799-2	GRN	Secreted	Isoform 2 of Granulins	46.94
Q9H8S1	Q9H8S1	Secreted	cDNA FLJ13286 fis, clone OVARC1001154, highly similar to homo sapiens clone 24720 Epithelin 1 and 2 mRNA	44.08

P55259-2	GP2	Secreted	Isoform Beta of Pancreatic secretory granule membrane major glycoprotein GP2	43.35
P09467	FBP1	extracellular vesicular exosome, cytosol	Fructose-1,6-bisphosphatase 1	36.80
Q9NP79	VTA1	cytoplasm, endosome membrane	Vacuolar protein sorting-associated protein VTA1 homolog	33.84
Q01105	SET	ER, cytoplasm, nucleus	Protein SET	33.45
P09493-3	TPM1	extracellular vesicular exosome, cytoplasm, cytoskeleton	Isoform 3 of Tropomyosin alpha-1 chain	32.84
P06753	TPM3	extracellular vesicular exosome, cytoplasm, cytoskeleton	Tropomyosin alpha-3 chain	32.78
P09093	CEL3A	Secreted	Chymotrypsin-like elastase family member 3A	29.45
P17538	CTRB1	Secreted	Chymotrypsinogen B	27.83
F5H7S3	TPM1	cytoplasm, cytoskeleton	Tropomyosin alpha-1 chain	27.51
P62158	CALM1	exosome vesicles, plasma membrane, cytoplasm, cytoskeleton	Calmodulin	16.81

---

Table 3.8: Eluted MCF7 *O*-glycoproteins identified by LC-MS/MS. Sixteen of the twenty-one *O*-glycoproteins are confirmed to be *O*-GlcNAcylated\*.<sup>25, 26</sup>

Accession	Gene Symbol	Cellular Location	<i>O</i> -Glycoprotein	Mass (kDa)
O60271-7	JIP4	membrane, cytosol, extracellular space	Isoform 7 of C-Jun-amino-terminal kinase-interacting protein 4	54.35
A34720	KRT8	cytoplasm, nucleus,	*Cytokeratin 8 (version 2) - human (P05787)	53.70
P45379-10	TNNT2	cytosol	*Isoform 10 of Troponin T, cardiac muscle	35.56
P35030-2	TRY3	Secreted	Isoform B of Trypsin-3	28.12
P10412	HIST1H1E	nucleus, chromosome	*Histone H1.4	21.83
P55145	MANF	Secreted	Mesencephalic astrocyte-derived neurotrophic factor	20.67
P05451	REG1A	Secreted, extracellular space	*Lithostathine-1-alpha	18.70
A6ND86	CELA2A	Secreted	Chymotrypsin-like elastase family member 2A (P08217)	15.55
Q5TEC6	Q5TEC6	nucleus, chromosome	*Histone H3	15.40
P68431	H31	nucleus, chromosome	*Histone H3.1	15.38
Q71DI3	H32	nucleus, chromosome	*Histone H3.2	15.36
P84243	H33	nucleus, chromosome, extracellular vesicular exosome	*Histone H3.3	15.30
Q6NXT2	H3C	nucleus, chromosome	*Histone H3.3C	15.19
P16104	H2AX	nucleus, chromosome	*Histone H2A.x	15.12
Q0VAF6	SYCN	secretory granule membrane, transport vesicle membrane	Syncollin	14.38
P04908	H2A1B	nucleus, chromosome	*Histone H2A type 1-B/E	14.11
Q16777	H2A2C	nucleus, chromosome	*Histone H2A type 2-C	13.96
P06899	H2B1J	nucleus, chromosome	*Histone H2B type 1-J	13.88
O60814	H2B1K	nucleus, chromosome	*Histone H2B type 1-K	13.86
P0C0S5	H2AZ	nucleus, chromosome	*Histone H2A.Z	13.53
P62805	H4	nucleus, chromosome	*Histone H4	11.34

### 3.5 Glycan Array Screening

The glycan array, developed by the Consortium for Functional Glycomics (CFG), consists of 610 unique mammalian glycans (version 5.1) and has proven to be an invaluable tool in determining the specificity of glycan-binding proteins.<sup>27, 28</sup> A library of

natural and synthetic glycans are modified with an amino linker containing a spacer. The glycans are covalently linked to NHS-activated glass surface via the amino-modified spacer linker. Each glycan is printed in replicates of six on the array. The surface immobilized glycans do not include the peptide glycosylation sequone (Asn – X(-Pro) – Ser/Thr). The lack of peptide sequone is a deviation from the normal biological context of glycan interactions. For many carbohydrate-recognizing proteins, which recognize terminal glycan structures, this is not a significant issue (e.g.: terminal sialic acid recognizing lectins). However, this is a significant issue for those carbohydrate-processing enzymes which recognize glycan structures in the context of the protein on which the glycan is displayed or being transferred to as in the case of various transferases. Given that the wtPNGase F enzyme is known to recognize the glycopeptide consisting of the sequone and the asparagine-linked chitobiose core the lack of the peptide sequone on the immobilized glycan is a limitation.

PNGase F D60A and R911 clones were submitted to the CFG's Protein-Glycan Interaction Core (formerly Core H) for glycan array screening. The purified proteins were labeled with DyLight 488 and dye:protein labeling ratios were determined to be 2.1:1 for D60A and 8.2:1 for R911. The labeled proteins were incubated on the arrays at a final concentration of 200 µg/mL in buffer consisting of 10 mM HEPES, 10 mM NaCl, pH 7.4 containing 0.1% BSA. After incubation, the array was washed in the same buffer without 0.1% BSA. The dried arrays were scanned on a microarray scanner and signal intensities for individual glycan features/spots were quantitated.

Figure 3.20 shows a side-by-side comparison of glycan array screening results for D60A and R911 clones. In summary, high signal intensity, indicating binding interactions

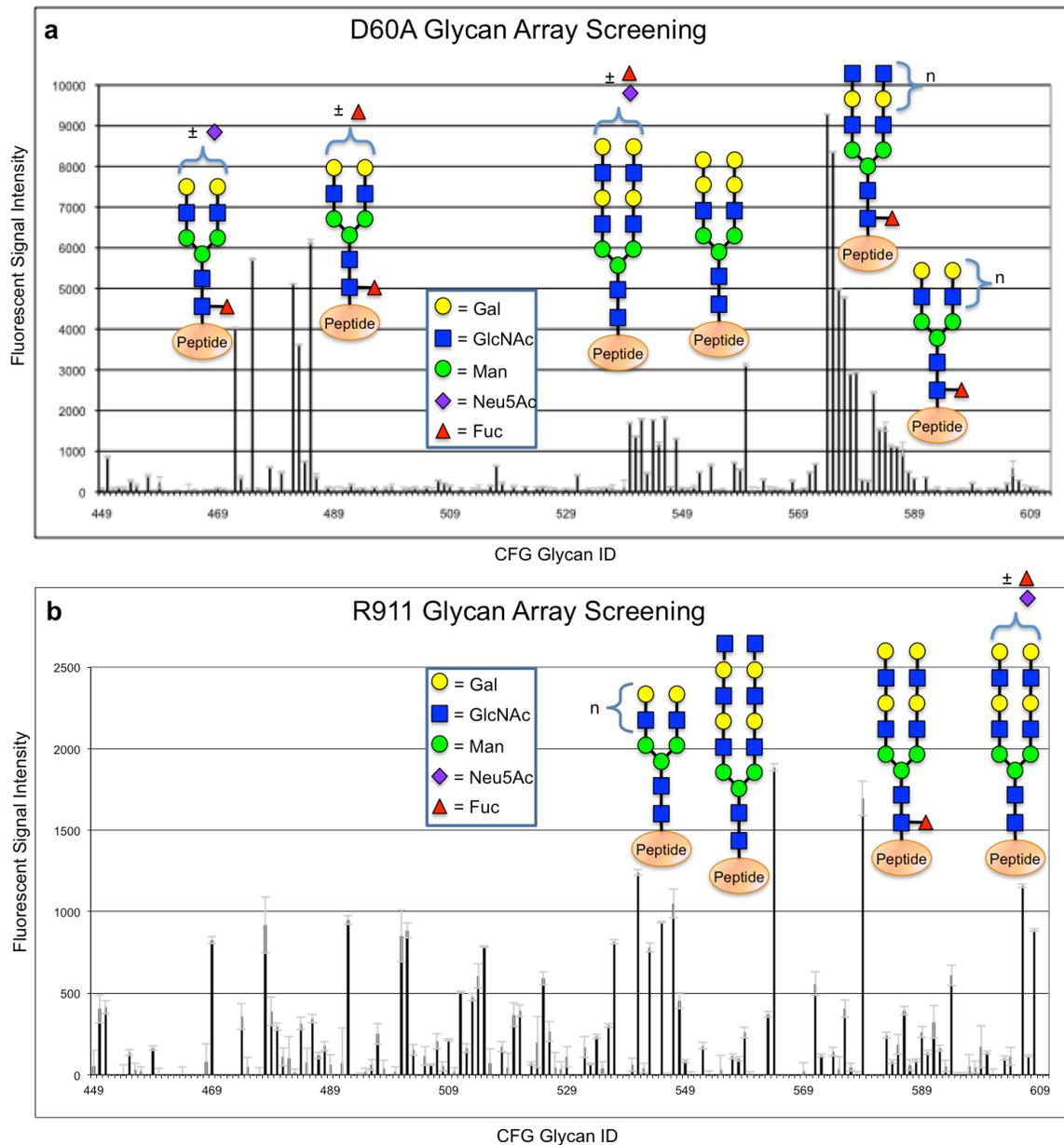


Figure 3.20: Glycan array screening of D60A and R911 clones.

with surface-immobilized *N*-glycans (with and without  $\alpha$ 1,6 core fucosylation), consistent with the reported specificity of the wtPNGase F enzyme was observed for labeled D60A. Furthermore, a lack of signal intensity for D60A interactions with  $\alpha$ 1,3 core fucosylated glycans was also observed. This is consistent with wtPNGase F's inability to release glycans with  $\alpha$ 1,3 linked fucose to the asparagine linked *N*-acetylglucosamine, whereas an  $\alpha$ 1,6 core fucosylated *N*-glycans can be released.

In the case of the labeled R911 clone, a noticeably reduced, but above background, signal intensity for similar surface-immobilized *N*-glycan structures was observed for labeled R911. The lower signal to noise ratio for the R911 data set may be due the possibility that the specificity of the R911 clone has been altered due to the selected mutations. However, this is inconsistent with both SPR affinity data and R911 Lectenz® affinity chromatography results. Thus, the more likely cause for the low signal to noise ratio and the seemingly diminished specificity relative to D60A is the high dye:R911 protein labeling ratio of 8.2:1. A high dye-labeling ratio can damage the binding site due to the higher probability that a dye molecule will react with an available amine group in the binding pocket. The ideal labeling ration is 2:1, as was the case with D60A. To obtain more robust glycan array-based specificity results the R911 glycan array screening will need to be repeated with a lower dye to protein labeling ration. Unfortunately, due to the limited availability of purified R911 as a result of low expression yield, and the need to characterize R911 using multiple techniques, a repeat experiment could not be immediately preformed.

### **3.6 Methods**

#### *3.6.1 Expression of PNGase F Clones*

PNGase F-pOPH6 II (D60A, R617, R6113, R911, and R9113) plasmids were transformed into *E.coli* BL21-Gold(DE3) competent cells obtained from Agilent Technologies (230132) for expression. For each clone, a single colony picked from a Luria Bertani (LB) agar plate (100 µg/ml Carbenicillin) was cultured in 50 mL LB media containing 100 µg/ml Carbenicillin in a shaker (250-300RPM) over night at 37 °C. The following day the culture was expanded into 37 °C pre-warmed 1 L LB media with 100 µg/ml Carbenicillin. Between OD<sub>600</sub> of 0.4 - 0.5 the temperature was dropped from 37 °C

to 22 °C and induced with 1 mM IPTG and the culture was induced overnight (~20 hours). The cell pellet was harvested at 4500 x g (30 minutes) using Avanti JA10 rotor at 4°C. The R911 culture yielded approximately an 8 g cell pellet from a 1 L LB culture. The cell pellet was resuspended in ice cold 20 mL IMAC binding buffer (0.1M EPPS, 0.5M NaCl, 0.01M Imidazole, pH 8.50). An EDTA-free protease inhibitor tablet from Roche (05892791001) was dissolved into 1 mL binding buffer or molecular grade water and mixed into the resuspended cell pellet. Cells were subjected to mechanical lysis three times using a French press at 6,000 psi. The cell lysate was centrifuged at 30,000 x g (45 minutes) in an Avanti JA17 rotor at 4 °C to separate insoluble cell debris from the supernatant containing the periplasmic fraction. The supernatant was collected and filtered using a 0.8 µm filter for every 5 mL of supernatant.

### *3.6.2 Immobilized Metal Affinity Chromatography of PNGase F Clones*

The filtered periplasmic fraction was loaded onto an IMAC column and the PNGase F clone eluted over an imidazole gradient using an AKTA Purifier UPC 10. IMAC Binding buffer (A) consisted of 0.1M EPPS, 0.5M NaCl, 0.01M Imidazole, pH8.50 and the IMAC Elution buffer (B) consisted of 0.1M EPPS, 0.5M NaCl, 0.5M Imidazole, pH8.50. A GE Healthcare HisTrap HP column (17-5247-01) was washed, charged with Ni<sup>2+</sup>, and equilibrated using the manufacturers recommended protocol. A programmed method (Unicorn 5.1) was used for all purification runs. In summary, the nickel-charged HisTrap column was equilibrated in 5 CV binding buffer at 3.5 mL/min flow rate. The periplasmic fraction (~20ml) was loaded into the column at a flow rate of 2 mL/min using a P-960 sample pump. The loaded column was washed with 9 CV of binding buffer (100% A) at a 2 mL/min flow rate. Non-specifically bound proteins were eluted with a 10 CV step elution of 8.3% B (equivalent to 50 mM Imidazole) at a 2

mL/min flow rate. A 43% B gradient elution over 18 CV at a flow rate of 2 mL/min was used to elute the Ni-bound histidine-tagged PNGase F clone and 2.5 mL fractions of eluted protein were collected. The column was wash with 100% B step elution for 8 CV at a flow rate of 2 mL/min followed by re-equilibration with 100% A over 5 CV at a flow rate of 3.5 mL/min. Using a Vivaspin 20 (10kDa cutoff) concentrator, the eluted protein containing fractions were pooled and concentrated down to ~250  $\mu$ L final volume for additional purification via SEC.

### *3.6.3 Size Exclusion Chromatography of PNGase F Clones*

Approximately, 250  $\mu$ L of the concentrated IMAC sample was loaded into a 500  $\mu$ L injection loop for SEC purification using either a SuperDex 75 10/300 GL or a Superose 12 10/300 GL column. The Superose 12 column provided enhanced purification of R911 relative to the SuperDex 75 column. As before, an automated method (Unicorn 5.1) was used for the purification run on the AKTA Purifier UPC 10. The column was equilibrated with 1.5 CV running buffer (50mM EPPS, pH 8.00) at a flow rate of 0.4 mL/min. The sample was injected into the column by flushing the 0.5 mL loop with 2.5 mL running buffer and then the flow rate dropped to 0.2 mL/min. Fraction collection (0.5 mL) was initiated at 6.75 mL retention volume. The fractions corresponding to the elution peak were pooled and concentrated using a 10 kDa cutoff Vivaspin 20 concentrator and protein yield determined by UV 280 absorbance ( $A_{280}$ ). Typical yield after SEC purification of the D60A control clone was ~3.0 mg from a 0.5 L LB culture. In comparison, typical yield for R911 and R911 C60A clones was ~0.3 mg from a 2 L LB culture.

### 3.6.4 SDS-PAGE & Coomassie Staining

Bio-Rad 4-20% TGX gels (456-1093 and 456-1094) and recommended buffers were used for SDS-PAGE of protein samples. All samples were denatured using the manufacturers recommended Laemmli sample buffer recipe at 6x stock concentration containing  $\beta$ -Mercaptoethanol and incubated at 95 °C for 5 minutes prior to loading on the gel. The gel was run for 35 minutes at 200 V (150 mA max). Gels were Coomassie stained using Life Technologies' SimplyBlue SafeStain (LC6060) per the recommended rapid microwave staining and destaining procedure.

### 3.6.5 Western Blot

The following buffers, reagents, and solutions were used for Western Blot with modified manufacturer protocols:

1. 10x Transfer buffer (1 L): 250 mM Tris (30.28 g/L), 1.92 M Glycine (144.1 g/L), 0.05% SDS (5 g/L), pH adjusted to 8.3 by HCl.
2. 10x TBS (1 L): 1.4 M NaCl (81.82 g/L), 250 mM Tris base (30.28 g/L), pH adjusted to 7.4 by HCl.
3. 10x BLOTTO (100 mL): 10% Non-fat dry milk (10 g) from Bio-Rad (170-6404XTU), 90% NANOpure water (90 mL).
4. 1x Blocking buffer (500 mL): 0.5 mL (500  $\mu$ L) Tween 20, 50 mL 10x BLOTTO, 50 mL 10x TBS, 400 mL NANOpure water.
5. 5% Blocking buffer (100 mL for two membranes): 5% non-fat dry milk (4 g), 96 mL of 1x Blocking buffer.
6. 1x TBS (100 mL): 10 mL 10x TBS, 90 mL NANOpure water,
7. 1x Transfer buffer with methanol (1 L) pH 8.3: 100 mL of 10x Transfer buffer, 150 mL of methanol, 750 mL of NANOpure water.

8. PVDF Hybond-P membrane from GE Healthcare (RPN303F).
9. Horseradish peroxidase (HRP) conjugated mouse anti-histidine antibody from Alpha Diagnostic (HISP12-HRP).
10. Thermo Scientific Pierce Metal Enhanced DAB substrate solution (34065).

Protein transfer using TGX gels and PVDF membrane was conducted at 100 V (350 mA max) for 30 minutes with the transfer apparatus kept on ice. A magnetic stir bar was used to circulate the transfer buffer during the transfer process. Post-transfer, the membrane was washed 3x with 20 mL NANOpure water and then blocked in 20 mL 5% blocking buffer for 45 minutes on a shaking platform following by incubation with the anti-histidine HRP conjugated antibody (1:5000) in 10 mL of 1x blocking buffer overnight at 4 °C on a shaker. The following day, the membrane was washed with 20 mL 1x blocking buffer for 5 minutes each followed by 3x washes with 20 mL of 1x TBS also for 5 minutes each on a shaker. The membrane was rinsed with 20 mL NANOpure water before addition of the DAB substrate for development. Depending on the amount of protein loaded on the gel and transferred to the membrane, the membrane was allowed to develop between 1 – 10 minutes. The membrane was rinsed one final time with 20 mL NANOpure water and then dried before scanning using a standard desktop scanner.

### *3.6.6 Deglycosylation Activity of PNGase F Clones*

A gel shift assay was used to determine deglycosylation activity of PNGase F clones relative to wtPNGase F. 50 ng of wtPNGase F, D60A, and R911 each was incubated with 50 µg of denatured RNase B in 50 mM EPPS, pH 8.0 in a 50 µL reaction volume at 37 °C overnight. Samples were analysed on a SDS-PAGE gel and altered migration of deglycosylated RNase B product relative to RNase B was observed. The scanned gel image was analyzed by ImageJ software to quantitate deglycosylated product

relative to RNase B substrate (Supplementary Table 4).<sup>7</sup> Deglycosylated product confirmed by MALDI TOF-TOF mass spectrometry.

### *3.6.7 Protein Denaturation*

5 mg, 0.113  $\mu$ mol asialofetuin (44,189 g/mol), purchased from Sigma (A4781-50MG) was dissolved in 1 ml 0.1 M Tris-HCl (pH 8.0) containing 6 M guanidine HCl (95.53 g/mol) (573 mg in 1 mL) and reduced by the addition of 28 mg, 182  $\mu$ mol DTT (154.25 g/mol) for 1 h at 55°C, followed by addition of 128 mg, 692  $\mu$ mol iodoacetamide (184.96 g/mol) for 30 minutes at room temperature. 0.5 mL of the mixture was desalted with Thermo Scientific Pierce D-Salt Polyacrylamide Desalting Columns, collecting 0.5 mL fractions after the void volume of 1.75 mL.

### *3.6.8 MALDI Mass Spectrometry*

Mass spectrometry was performed using an ABI 5800 MALDI TOF-TOF High Resolutions Mass Spectrometer. Sinapinic Acid matrix was prepared by re-suspending ~10mg Sinapinic Acid in 1ml volume of 30% acetonitrile (ACN) and 0.3% TFA. Matrix to protein sample was mixed at a ratio of 30:1 resulting in 4 pmols of total protein. The sample spotted (1  $\mu$ L) a MALDI plate and air dried prior to loading the plate in to the ABI 5800.

### *3.6.9 LC-MS/MS D60A Sequence Identification*

The protein sample was prepared by adding 8  $\mu$ L of 40 mM  $\text{NH}_4\text{HCO}_3$  to 10  $\mu$ L of D60A sample (10  $\mu$ g) for a total volume of 18  $\mu$ L. The sample was reduced with 2  $\mu$ L of 1 M DTT for one hour at 56 °C and carboxyamidomethylated with 20  $\mu$ L of 55mM iodoacetamide in the dark for 45 minutes. Trypsin (20  $\mu$ g) was reconstituted with 80  $\mu$ L of 40 mM  $\text{NH}_4\text{HCO}_3$  and 10  $\mu$ L (2.5  $\mu$ g) was added to the sample to digest proteins overnight at 37 °C. After digestion, the peptides were acidified with 5  $\mu$ L of 1%

trifluoroacetic acid (TFA). Desalting was performed with a C18 spin column, and the sample was dried down in a vacuum centrifuge. The peptides were re-suspended with 19  $\mu$ L of mobile phase A (0.1% formic acid in water) and 1  $\mu$ L of mobile phase B (80% acetonitrile and 0.1% formic acid in water). The samples were loaded onto a nanospray tapered capillary column/emitter self-packed with C18 reverse-phase resin via a nitrogen pressure bomb for 10 minutes at 1000psi for each run. Each run consisted of a 160 minute gradient of increasing mobile phase B at a flow rate of approximately 200 nL per minute. For the initial protein identification run a LC-MS/MS analysis was performed on a Finnigan LTQ-XL equipped with a nanoelectrospray ion source. An instrument method was used to collect a full MS spectra and generate MS/MS spectra for the 8 most intense peaks using collision-induced dissociation (38% normalized collision energy) with dynamic exclusion set for 30 second intervals. The resulting data was searched against an *E.coli* database with D60A sequence, as well as a target only database, using a Sequest Algorithm. Sequest parameters were altered to search for modifications allowing for oxidation of methionine and alkylation of cysteine. Peptide mass tolerance was set at 1000 ppm and fragment ion tolerance was set at 1 Dalton. Results were filtered at a false discovery rate (FDR) of 1%.

#### *3.6.10 Surface Plasmon Resonance*

The ligands, denatured RNase B, native RNase B, and denatured RNase A, which has the same peptide sequence as RNase B but lacks *N*-glycosylation, were covalently coupled to CM-5 chips using amine-coupling chemistry. Optimal coupling conditions were determined by pH scouting of acetate buffers as per Biacore's recommended protocol (Figure 3.13). A high-density surface area was prepared with sufficient ligand coupling to achieve a calculated  $R_{MAX}$  of 3000 RU. For ligand immobilization, the

coupling buffer consisting of 10 mM Acetate buffer, pH 5.5 was used. The PNGase F clones used as analytes were D60A, R911, and R911 C60A in a serial dilution concentration range starting from 10  $\mu$ M down to 72.5 nM. The running buffer consisted of 10 mM HEPES, 10 mM NaCl, pH 7.4. Steady-state binding kinetics using a bimolecular interaction model were determined using Scrubber 2.0c (Table 3.4 and Figure 3.14).

### *3.6.11 Glycan Array Screening*

The D60A and R911 clones were submitted to the Consortium for Functional Glycomics' Protein-Glycan Interaction Core (formerly Core H) for glycan array screening.<sup>29</sup> Purified D60A was labeled with DyLight 488 and dye:protein labeling ratio was determined to be 2.1:1. Purified R911 was similarly labeled and the dye:protein labeling ratio was determined to be 8.2:1. The clones were incubated on the arrays at a final concentration of 200  $\mu$ g/mL in buffer consisting of 10 mM HEPES, 10 mM NaCl, pH 7.4 containing 0.1% BSA. After a 1 hour incubation, the array was washed in the same buffer without 0.1% BSA four times. Slides are dried under a stream on nitrogen and processed using the standard glycan array data acquisition and analysis protocol. After slides have been dried following the last wash, they are placed in the PerkinElmer ScanArray scanner and data is obtained for each wavelength used for detection (DyLight 488). The PMT setting used is 70% and the laser power used is 90%. After saving, the images are opened in Imogene software and a grid is used to align the spots on the slide using the biotin control spots. Once aligned, the amount of binding to each spot is quantified. The data is analyzed using Microsoft Excel, where the highest and lowest spot of the 6 replicates is removed, and the average of the 4 remaining spots is displayed graphically and in a table along with appropriate statistics

### 3.6.12 Lectenz® Affinity Chromatography

Using 1 mL HiTrap NHS-activated HP columns manufactured by GE Healthcare (17-0716-01) purified PNGase F D60A and R911 clones were covalently linked to the column matrix to evaluate affinity chromatography based enrichment of *N*-glycopeptides and *N*-glycoproteins. Using the manufacturers recommended protocol, coupling efficiencies of PNGase F clones to the NHS-activated columns consistently ranged between 80% - 87% for all NHS-activated column-coupling reactions. The standard binding buffer consisted of 10 mM HEPES, 10 mM NaCl, pH 7.4 where as the standard elution buffer consisted of 10 mM HEPES, 150 mM NaCl, pH 7.4. For competitive elution experiments, the elution buffer consisted of 10 mM HEPES, 10 mM NaCl, 235.6  $\mu$ M (100  $\mu$ g/mL) chitobiose, pH 7.4. Chitobiose was obtained from Sigma (D1523-10). The regeneration buffer consisted of 10 mM HEPES, 500 mM NaCl, pH 7.4. An AKTA Purifier UPC 10 (GE Healthcare) was used for all chromatography experiments configured with a 100  $\mu$ L sample injection loop, 1 mL HiTrap NHS-activated columns, UV<sub>280</sub> detection. For all chromatography runs, the flow rate of 0.4 mL/min was constant. The column was equilibrated with 10 mL or 10 column volumes (CV) of binding buffer, followed by 100  $\mu$ L injection of sample. The sample was allowed to flow through the column using binding buffer over 5 CV to wash out unbound sample. The bound sample was eluted with elution buffer over 5 CV. During binding and elution 0.5 mL fractions were collected. The column was regenerated with 5 CV regeneration buffer and re-equilibrated in 5 CV binding buffer.

### 3.6.13 MCF7 Cell Extract Preparation

Human breast cancer MCF7 cells (generously provided by D.J. Bernstein) were cultured in DMEM media supplemented with 10% fetal bovine serum. Cells were

passed and harvested using trypsin-free cell release. Approximately,  $2.3 \times 10^7$  cells were harvested and washed 3x with 10 mL of phosphate buffered saline at 4 °C by centrifugation at 1000 x g for 5 minutes. The cell pellet was resuspended in 1.5 mL of filter sterilized cell lysis buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% v/v Nonidet P-40) with EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany) and incubated on ice for 30 minutes.<sup>17</sup> The cells were ultrasonicated (Misonix Ultrasonic Liquid Processor Model S-4000) at intervals of 15 seconds for a total of 2 minutes, with a 15 second pause between treatments (30% amplitude). The lysed cells were centrifuged at 17,000 x g for 1 hour in an Eppendorf 5430R at 4 °C. The supernatant containing the MCF7 cell extract was stored at -80 °C in 50 µL aliquots. Protein concentration of the MCF7 cell extract was determined to be 10.67 mg/mL using a Thermo Scientific Pierce BCA protein assay kit (23277). An aliquoted MCF7 cell extract stock was thawed on ice and diluted to 1 mg/mL using 10 mM HEPES, 10 mM NaCl, at pH 7.4. Using a 100 µL sample loop, 100 µg of 1 mg/mL MCF7 cell extract was injected into the R911 Lectenz® affinity column for glycoprotein enrichment.

#### *3.6.14 LC-MS/MS Protein Identification of R911 Lectenz® Affinity Chromatography Eluted Samples*

MCF7 cell extract and proteins eluted from the R911 Lectenz® affinity column were reduced, alkylated and digested with sequence grade trypsin (Promega) using a standard in-solution digest protocol.<sup>30</sup> The samples were acidified with 1% trifluoroacetic acid and desalting was performed using C18 spin columns (Silica C18, The Nest Group, Inc.). Peptides were dried down and resuspended with 39 µL of buffer A (0.1% formic acid) and 1 µL of buffer B (80% Acetonitrile and 0.1% formic acid). The samples were spun through a 0.2 µm filter (Nanosep, Pall Corp) before being loaded into an

autosampler tube and racked into an Ultimate 3000 LC System (Thermo Scientific – Dionex).

LC-MS/MS analysis was performed on an Orbitrap Fusion Tribrid (Thermo Scientific) utilizing a nanospray ionization source. For each sample, 10  $\mu$ L was injected and separated via a 180-minute gradient of increasing buffer B at a flow rate of approximately 200 nL per minute. An instrument method was used to collect full mass spectrum every three seconds and continuously trap and fragment the most intense ions with 38% collision-induced dissociation (CID) and record the resulting MS/MS spectra. Dynamic exclusion was utilized to exclude precursors ions from selection process for 60 seconds following a second selection within a 10 second window.

All MS/MS spectra were searched against a UniProt human database utilizing the SEAQUEST algorithm (Proteome Discoverer 1.4, Thermo Scientific). The SEAQUEST parameters were set to allow for tryptic peptides with a maximum of two internal missed cleavages. Mass tolerances were set to 20 ppm for precursors ions and 0.5 Da for fragment ions. Dynamic mass increases were allowed to account for oxidation of methionine and alkylation of cysteine residues. The spectra were also searched against a random database generated by reversing the human database to determine the false-discovery rate (FDR) of identification. ProteoIQ utilized all SEAQUEST search result files and databases to filter peptide matches and attain accurate protein identifications.<sup>31</sup> Peptides passing a 20% FDR were considered for protein identification and only proteins surviving a 2% FDR were reported.

### 3.6.15 Identification of MCF7 Cell Extract Glycoproteins using UniPep and UniProt Databases

UniProt verified protein identification lists generated from LC-MS/MS analysis of the stock MCF7 sample and the R911 Lectenz® affinity chromatography eluted MCF7 sample were processed through the UniPep database to identify proteins with experimentally confirmed *N*-glycopeptides.<sup>20</sup> In addition, potential glycoproteins with *N*-linked glycosites were also identified via UniPep based on the presence of the *N*-glycosylation sequone (Asn-X-Ser/Thr). A final list of *N*- and *O*- glycoproteins included only those proteins that were confirmed as glycoproteins by UniPep, UniProt, and literature reports.<sup>19, 20, 25, 26, 32</sup> In addition, proteins that were predicted to have an *N*-glycosylation site were only included in the final list if UniProt subcellular localization descriptions were consistent with those expected for glycosylated proteins (Golgi, secreted, vesicular exosome, extracellular space, and histones).

### 3.7 References

1. Loo, T., Patchett, M.L., Norris, G.E. & Lott, J.S. Using Secretion to Solve a Solubility Problem: High-Yield Expression in *Escherichia coli* and Purification of the Bacterial Glycoamidase PNGase F. *Protein Expression and Purification* **24**, 90-98 (2002).
2. Filitcheva, J. PNGases: A Diverse Family of Enzymes Related by Function Rather Than Catalytic Mechanism, Vol. Ph.D. (Massey University, Palmerston North; 2010).
3. Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. in *The Proteomics Protocols Handbook* (ed. J.M. Walker) 571-607 (Copyright Humana Press, 2005).
4. Kuhn, P. et al. Active Site and Oligosaccharide Recognition Residues of Peptide-N4-(N-acetyl-β-D-glucosaminyl)asparagine Amidase F. *Journal of Biological Chemistry* **270**, 29493-29497 (1995).
5. Noble, J.E., Knight, A.E., Reason, A.J., Di Matola, A. & Bailey, M.J. A comparison of protein quantitation assays for biopharmaceutical applications. *Mol Biotechnol* **37**, 99-111 (2007).

6. Kuhn, P., Tarentino, A.L., Plummer, T.H., Jr. & Van Roey, P. Crystal structure of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase F at 2.2-Å resolution. *Biochemistry* **33**, 11699-11706 (1994).
7. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* **9**, 671-675 (2012).
8. Cooper, M.A. Optical biosensors in drug discovery. *Nat Rev Drug Discov* **1**, 515-528 (2002).
9. Schlick, K.H. & Cloninger, M.J. Inhibition binding studies of glycodendrimer-lectin interactions using surface plasmon resonance. *Tetrahedron* **66**, 5305-5310 (2010).
10. Haseley, S.R., Kamerling, J.P. & Vliegthart, J.F.G. Unravelling carbohydrate interactions with Biosensors using surface plasmon resonance (SPR) detection. *Host-Guest Chemistry* **218**, 93-114 (2002).
11. Nieba, L. et al. BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Analytical Biochemistry* **252**, 217-228 (1997).
12. Willard, F.S. & Siderovski, D.P. Covalent immobilization of histidine-tagged proteins for surface plasmon resonance. *Analytical Biochemistry* **353**, 147-149 (2006).
13. Wang, S. et al. N-Terminal Deletion of Peptide:N-Glycanase Results in Enhanced Deglycosylation Activity. *PLoS ONE* **4**, e8335 (2009).
14. Ongay, S., Boichenko, A., Govorukhina, N. & Bischoff, R. Glycopeptide enrichment and separation for protein glycosylation analysis. *J Sep Sci* **35**, 2341-2372 (2012).
15. Alwael, H. et al. Pipette-tip selective extraction of glycoproteins with lectin modified gold nano-particles on a polymer monolithic phase. *Analyst* **136**, 2619-2628 (2011).
16. Jung, K. & Cho, W. Serial affinity chromatography as a selection tool in glycoproteomics. *Anal Chem* **85**, 7125-7132 (2013).
17. Lee, L.Y. et al. An optimized approach for enrichment of glycoproteins from cell culture lysates using native multi-lectin affinity chromatography. *J Sep Sci* **35**, 2445-2452 (2012).
18. Krishnamoorthy, L. & Mahal, L.K. Glycomic analysis: an array of technologies. *ACS chemical biology* **4**, 715-732 (2009).
19. Consortium, T.U. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research* **42**, D191-D198 (2014).

20. Zhang, H. et al. UniPep--a database for human N-linked glycosites: a resource for biomarker discovery. *Genome biology* **7**, R73 (2006).
21. Fan, J.Q. Detailed Studies on Substrate Structure Requirements of Glycoamidases A and F. *Journal of Biological Chemistry* **272**, 27058-27064 (1997).
22. Tretter, V., Altmann, F. & MÄRz, L. Peptide-N4-(N-acetyl- $\beta$ -glucosaminyl)asparagine amidase F cannot release glycans with fucose attached  $\alpha 1 \rightarrow 3$  to the asparagine-linked N-acetylglucosamine residue. *European Journal of Biochemistry* **199**, 647-652 (1991).
23. Mussar, K.J., Murray, G.J., Martin, B.M. & Viswanatha, T. Peptide: N-glycosidase F: studies on the glycoprotein aminoglycan amidase from *Flavobacterium meningosepticum*. *Journal of biochemical and biophysical methods* **20**, 53-68 (1989).
24. Tarentino, A.L., Gomez, C.M. & Plummer, T.H., Jr. Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24**, 4665-4671 (1985).
25. Zhang, S., Roche, K., Nasheuer, H.P. & Lowndes, N.F. Modification of histones by sugar beta-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. *J Biol Chem* **286**, 37483-37495 (2011).
26. Ahmad, W. et al. Human linker histones: interplay between phosphorylation and O-beta-GlcNAc to mediate chromatin structural modifications. *Cell division* **6**, 15 (2011).
27. Taylor, M.E. & Drickamer, K. Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands. *Glycobiology* **19**, 1155-1162 (2009).
28. Adams, G.B. & Scadden, D.T. The hematopoietic stem cell in its place. *Nat Immunol* **7**, 333-337 (2006).
29. Heimburg-Molinaro, J., Song, X., Smith, D.F. & Cummings, R.D. Preparation and analysis of glycan microarrays. *Curr Protoc Protein Sci* **Chapter 12**, Unit12 10 (2011).
30. Lim, J.M. et al. Defining the regulated secreted proteome of rodent adipocytes upon the induction of insulin resistance. *J Proteome Res* **7**, 1251-1263 (2008).
31. Weatherly, D.B. et al. A Heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. *Mol Cell Proteomics* **4**, 762-772 (2005).

32. Takashima, S. Glycosylation and secretion of human  $\alpha$ -amylases. *Advances in Biological Chemistry* **02**, 10-19 (2012).

## CHAPTER 4

### MOLECULAR DYNAMICS SIMULATIONS OF PNGASE F CLONES

Conformational analysis of PNGase F clones D60A, R911, and R911 C60A relative to the wtPNGase F enzyme was investigated via MD Simulations, from which per-residue interaction energies were computed. The x-ray crystallography-based structural model of the wtPNGase F enzyme with the *N,N'*-diacetylchitobiose disaccharide in the active site at 2.0 Å resolution, has previously been reported (PDB ID: 1PNF).<sup>1</sup> The 1PNF model was used to construct mutagenized models of D60A, R911, and R911 C60A clones. In addition, the co-crystallized *N,N'*-diacetylchitobiose disaccharide ligand in the binding pocket of wtPNGase F served as a guide to position modified *N*-glycan structures into the binding pocket of the constructed models of PNGase F clones.

#### **4.1 Structural Models of D60A, R911, and R911 C60A**

##### *4.1.1 Rotamer Selection for Building Models*

The wtPNGase F model, 1PNF, was used as a template to build models of PNGase clones D60A, R911, and R911 C60A. Two rotamer libraries were used to select side chain rotamers for R911 mutations. The x-ray crystallography-based, backbone dependent, Dunbrack library was used to select rotamers and build models R911 Dun and R911 C60A Dun.<sup>2</sup> In addition, the MD-based, backbone independent, Dynaomics library was also used to select rotamers and build models R911 Dyn and R911 C60A Dyn.<sup>3</sup> The highest probability rotamers with the least amount of steric clashes were selected. Dunbrack rotamers evaluated and selected for building R911 Dun and R911 C60A Dun models are listed in Table 4.1. Similarly, Dynaomics rotamers for R911

Dyn and R911 C60A Dyn are listed in Table 4.2. Rotamers for D57L, D60C, I156L, G192I, and R248W mutations are shown in the context of the neighboring residues. The rotamers for E206S are not shown as no clashes were predicted. MD simulations and free energy decomposition were computed to evaluate which rotamers best approximated experimental interaction energies. The rotamer models with the best approximations are used for all subsequent computational studies.

Table 4.1: Dunbrack library rotamer selection. Selected rotamers are indicated in green. Rotamers with the highest probability and lowest number of steric clashes were selected for R911 Dun and R911 C60A Dun models (green).

R911 Rotamer	Rotamer Probability (%)	Average Angles		Clashes #	Selected Dunbrack Rotamer Clash Description and Orientation
		Chi1	Chi2		
1. D57L	64.4334	-173	61.4	4	
2. D57L	28.6281	-64.6	175.9	7	
3. D57L	4.2279	-165.4	171.3	1	Clash w/ T119; similar orientation as D57
1. D60C	74.2327	179.8		1	
2. D60C	25.3616	-62.4		3	
3. D60C	0.4057	63.4		0	similar orientation as D60
1. I156L	93.2665	-63.1	175.6	5	
2. I156L	4.0482	-87.2	49.8	0	
1. G192I	39.1625	62.1	170.9	21	
2. G192I	34.202	-65.1	169.6	17	
3. G192I	10.5124	-169.1	167.9	17	
4. G192I	8.6498	-59.8	-60	16	
5. G192I	4.5635	-166.8	66.7	14	Clash with C204, D290, W191, H193
1. E206S	41.8529	-65.9		0	
2. E206S	31.6039	65.4		0	
3. E206S	26.5433	179.2		0	same orientation as E206
1. R248W	40.4783	-63.5	100.4	9	Clash with E206, W207
2. R248W	29.7865	-70.3	0.4	22	
3. R248W	10.8433	-59.8	-87.2	7	Clash with P253

Table 4.2: Dynameomics library rotamer selection. Selected rotamers are indicated in green. Rotamers with the highest probability and lowest number of steric clashes were selected for R911 Dyn and R911 C60A Dyn models (green).

R911 Rotamer	Rotamer Probability (%)	Average Angles		Steric Clashes #	Selected Dynameomic Rotamer Clash Description and Orientation
		Chi1	Chi2		
1. D57L	65.1945	292	168.5	9	
2. D57L	24.054	183	66.6	4	Clashes: R61, L121; same orientation as D57
3. D57L	4.5273	275.1	79.8	9	
1. D60C	56.2557	298.5		3	
2. D60C	28.6878	183.7		3	
3. D60C	15.0565	54.8		0	same orientation as D60
1. I156L	65.1945	292	168.5	4	Clashes: F292, GlcNAc
2. I156L	24.054	183	66.6	7	
3. I156L	4.5273	275.1	79.8	1	
1. G192I	42.5492	53.3	167.1	19	
2. G192I	28.2902	300.9	176.6	18	
3. G192I	22.6888	301.9	299	15	
4. G192I	3.7004	187.2	165.7	16	
5. G192I	1.4194	184.1	68.3	13	Clashes: C204, D290, A291, W191
1. E206S	73.0609	310.9		0	
2. E206S	24.8453	40.7		0	
3. E206S	2.0938	189.4		2	Clashes: W248
1. R248W	28.2145	294.7	98.5	7	Clashes: W207
2. R248W	16.4889	291.4	348.1	20	
3. R248W	13.8124	181.8	259.2	24	

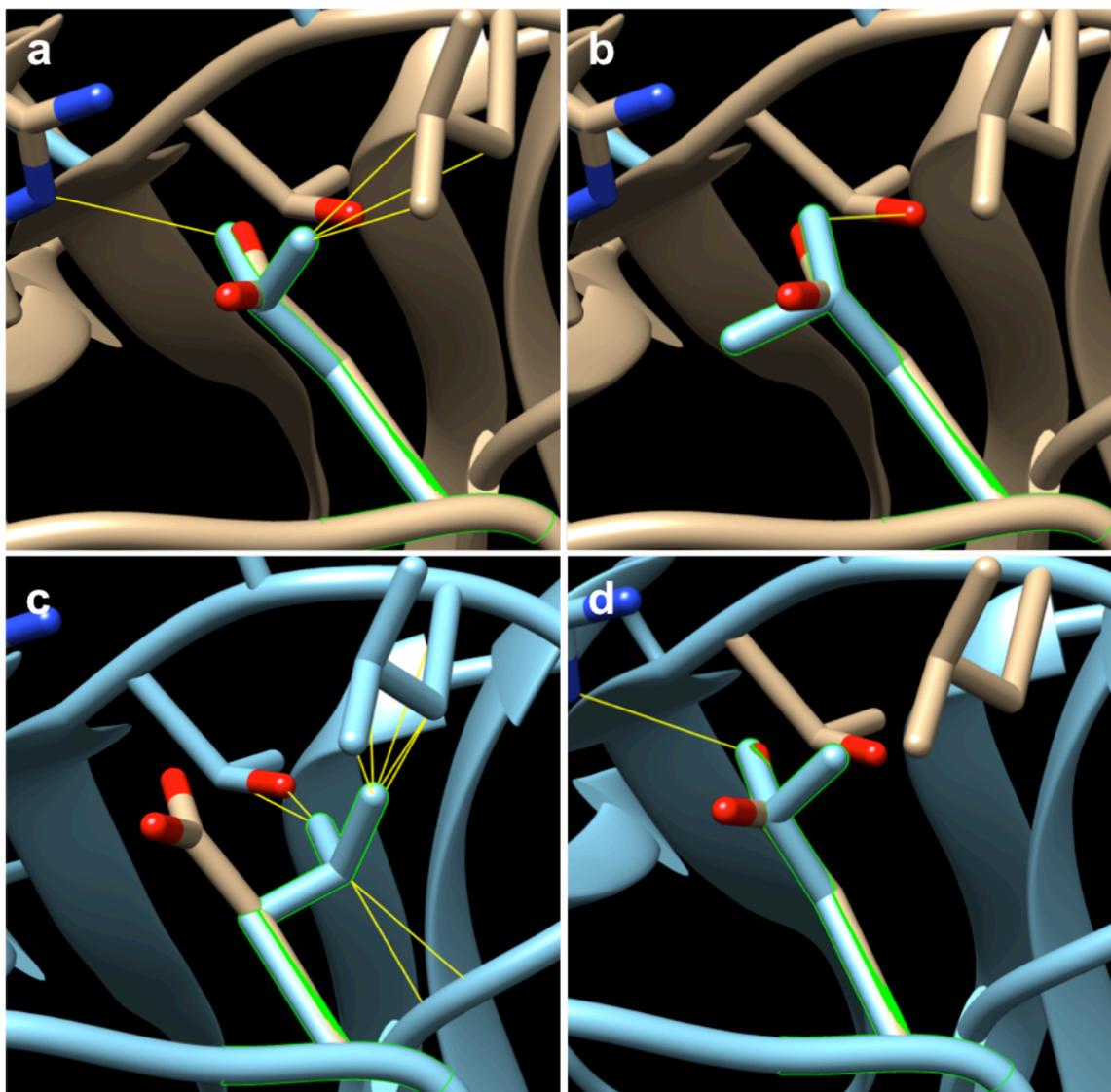


Figure 4.1: R911 D57L rotamers overlaid on wtPNGase F D57. wtPNGase F is colored brown. R911 is colored blue with the rotamer outlined in green. Yellow lines indicate steric clashes with neighboring amino acids. The Dunbrack rotamer depicted in **b** and the Dynameomics rotamer depicted in **d** were selected for modeling based on minimal clashes. a) Dunbrack rotamer 1 has 4 clashes (probability = 64.4334%, Chi1 = -173, Chi2 = 61.4), b) Selected Dunbrack rotamer 2 has 1 clash (probability = 4.2279%, Chi1 = -165.4, Chi2 = 171.3), c) D57L Dynameomics rotamer 1 has 9 clashes (probability = 65.1945%, Chi1 = 292, Chi2 = 168.5), d) selected Dynameomics rotamer 2 has 1 clash (probability = 4.2279%, Chi1 = -165.4, Chi2 = 171.3).

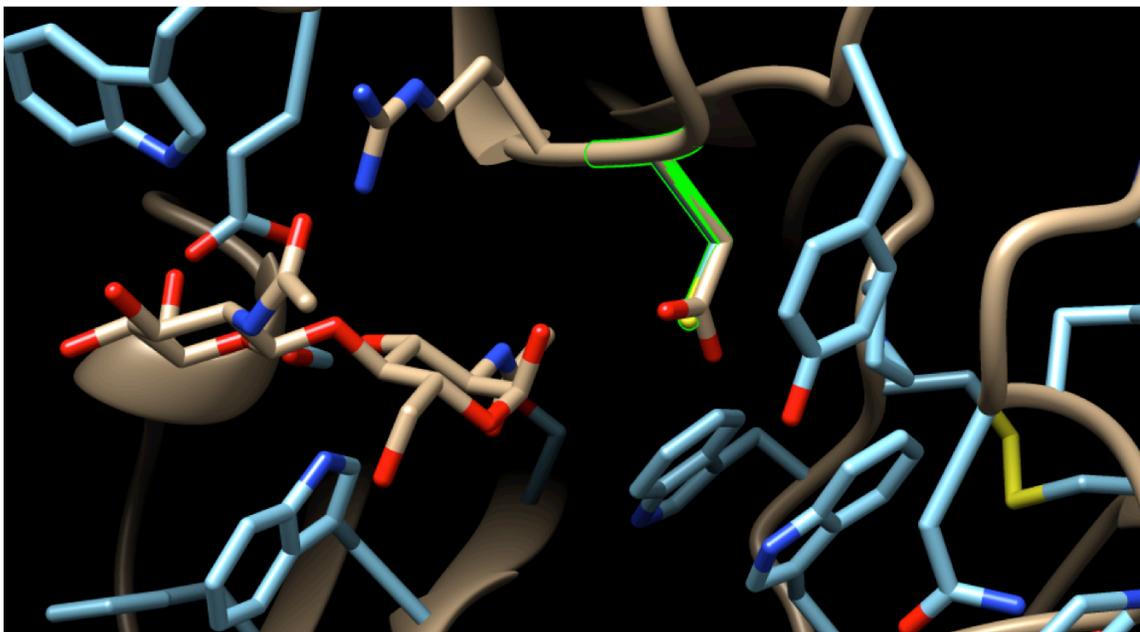


Figure 4.2: R911 D60C rotamer overlaid on wtPNGase F D60. wtPNGase F is colored brown. R911 is colored blue with the rotamer outlined in green. Yellow lines indicate steric clashes with neighboring amino acids. Selected Dunbrack rotamer 3 is depicted (probability = 0.4057%, Chi1 = 63.4) having the same orientation as wtPNGase F D60. The chitobiose ligand is visible to the left of the rotamer. The selected Dymameomics rotamer 3 (probability = 15.0565%, Chi1 = 54.8) also has the same orientation as wtPNGase F D60 (not shown).

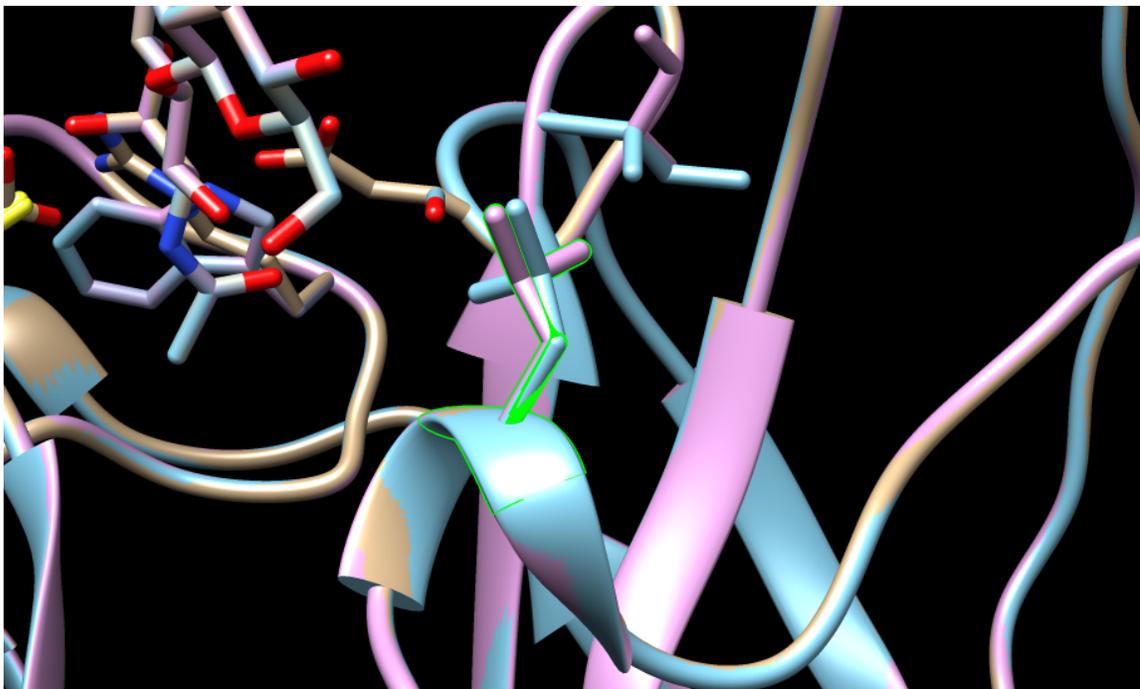


Figure 4.3: Comparison of R911 I156L rotamers. wtPNGase F is colored brown, R911 Dynameomics model is colored pink, and R911 Dunbrack model is colored blue. The selected Dunbrack and Dynameomics I156L rotamers are shown overlaid. The wtPNGase F I156 is not shown. The selected Dunbrack rotamer 2 (blue) has no clashes (probability = 4.0482%, Chi1 = -87.2, Chi2 = 49.8). The selected Dynameomics rotamer 1 (pink) has 4 clashes (65.1945%, Chi1 = 292, Chi2 = 168.5).

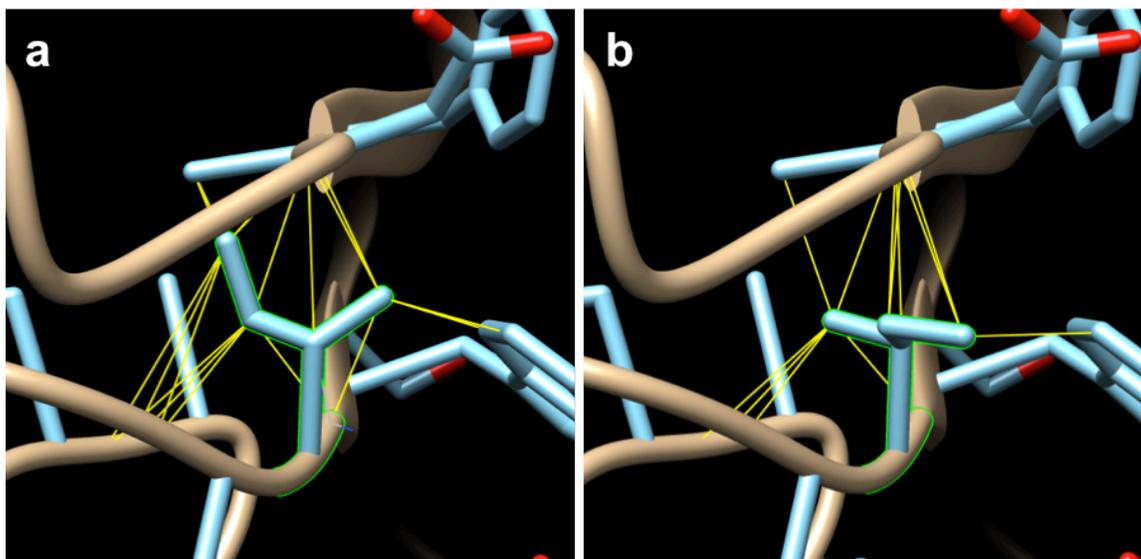


Figure 4.5: R911 G192I Dunbrack rotamers overlaid on wtPNGase F G192. wtPNGase F is colored brown. R911 is colored blue with the rotamer outlined in green. Yellow lines indicate steric clashes with neighboring amino acids. The Dunbrack rotamer depicted in **b** was selected for modeling based on minimal clashes and is similarly oriented to the selected Dymameomics rotamer (not shown). a) Dunbrack rotamer 1 has 21 clashes (probability = 39.1625%, Chi1 = 62.1, Chi2 = 170.9), b) Selected Dunbrack rotamer 2 has 14 clashes (4.5635%, Chi1 = -166.8, Chi2 = 66.7).

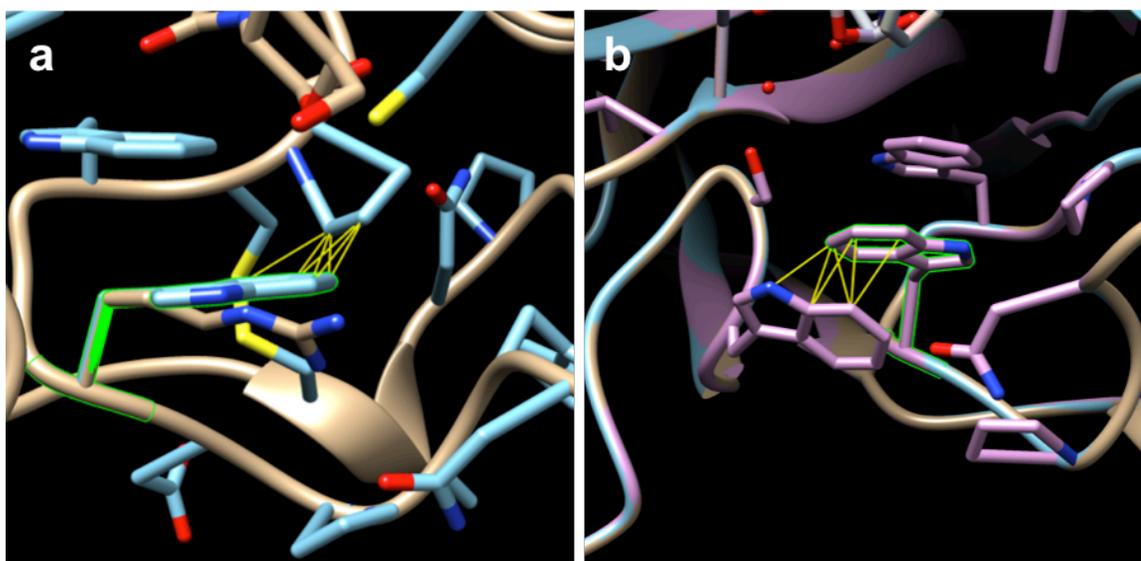


Figure 4.4: R911 R248W rotamers overlaid on wtPNGase F R248. wtPNGase F is colored brown. R911 is colored blue with the rotamer outlined in green. Yellow lines indicate steric clashes with neighboring amino acids. a) Selected Dunbrack rotamer 3 has 7 clashes (probability = 10.8433%, Chi1 = -59.8, Chi2 = -87.2) with P253 is depicted overlaid on wtPNGase F R248. b) Selected Dymameomics rotamer 1 has 7 clashes (probability = 28.2145%, Chi1 = 294.7, Chi2 = 98.5) with W207.

#### 4.1.2 MD Simulations for Validating Rotamers

MD simulations (100 ns) of the wtPNGase F, D60A, and all 4 rotamer models were performed with the *N,N'*-diacetylchitobiose disaccharide (GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH). The root mean squared difference (RMSD) in the positions of C $\alpha$  atoms, relative to the wtPNGase F experimental structure (1PNF), was determined as a function of the simulation time. The average RMSD for each of the 6 models over the course of the 100 ns simulation were low, stable, and ranged between 1.2 Å and 1.3 Å, indicating structural equilibration. A list of 6 structural models used for rotamer analysis and the average RMSD value for each MD simulation is listed in Table 4.3.

Table 4.3: Structural models of wtPNGase F and clones for rotamer selection. 100 ns MD simulations were run for each model. The 1PNF x-ray structure with the co-crystallized  $\alpha$ -chitobiose ligand represents wtPNGase F. The remaining 5 models were constructed from the 1PNF reference structure. The average RMSD value for each simulation is listed.

Structural Model	Ligand	Average RMSD (Å)
1PNF	GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH	1.2404
D60A	GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH	1.2369
R911 Dun	GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH	1.2908
R911 C60A Dun	GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH	1.3090
R911 Dyn	GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH	1.3377
R911 C60A Dyn	GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH	1.3061

In addition to confirming the stability of the simulations, reproducibility of experimentally observed hydrogen bond lengths in 1PNF X-ray data was confirmed. The theoretical hydrogen bonds lengths between protein and the the *N,N'*-diacetylchitobiose disaccharide (GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH) ligand in the 1PNF (wtPNGase F) MD simulation are compared to the experimentally determined hydrogen bond lengths in Table 4.4 and depicted in Figure 4.6. The MD simulation of the 1PNF model accurately reproduced experimental hydrogen bond lengths. Since the other models are derived from 1PNF and consistent RMSD values across all models indicated structural stability, it was

assumed that the remainder of the models were structurally valid as there was on experimental hydrogen bond length data for the R911 and R911 C60A clones.

Table 4.4: Experimental and theoretical hydrogen bond lengths observed between chitobiose and PNGase F. \*The 316 O1 atom of the anomeric hydroxyl in the 1PNF x-ray crystal model is renumbered to 315 O1 in the 1PNF MD simulation.

Hydrogen bonds	1PNF X-ray Data (Å) <sup>1</sup>	Average from 1PNF MD Simulation (Å)
D60-O $\delta$ – GlcNAc316 O1*	2.64	2.64 $\pm$ 0.11
D60-O – GlcNAc316 NAc	2.86	3.03 $\pm$ 0.17
R61-NH – GlcNAc317 OAc	2.84	2.85 $\pm$ 0.14
R61-NH – GlcNAc316-O4	2.81	2.97 $\pm$ 0.16
R61-NH <sub>2</sub> – GlcNAc317 OAc	2.99	3.11 $\pm$ 0.21
E118-O $\epsilon$ – GlcNAc317-O6	2.57	2.75 $\pm$ 0.18
W120-N $\epsilon$ – GlcNAc317-O6	2.90	3.10 $\pm$ 0.19
W191-N $\epsilon$ – GlcNAc316-O3	2.80	3.00 $\pm$ 0.15

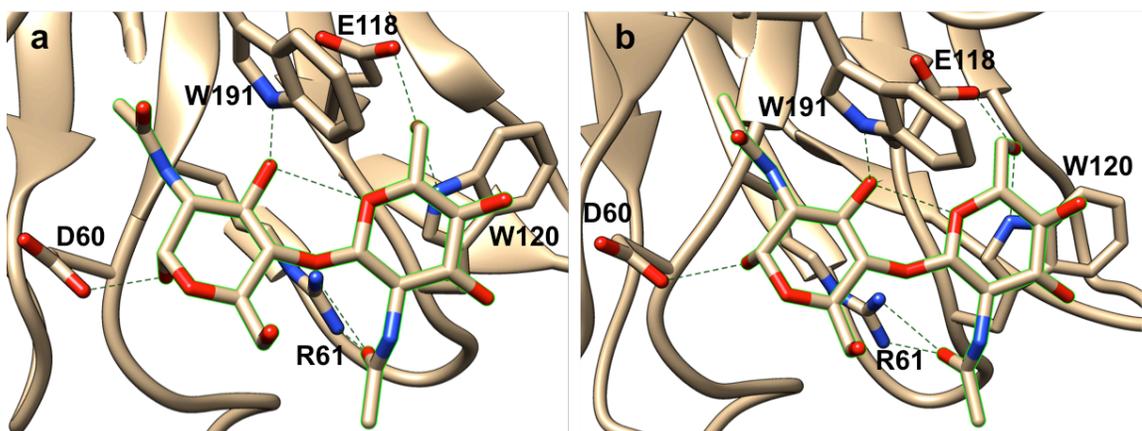


Figure 4.6: Experimental and theoretical hydrogen bonds. Seven hydrogen bonds between wtPNGase F residues D60, R61, E118, W120 and W191 and the chitobiose ligand (green outline) are depicted by the dashed lines. a) Hydrogen bonds reported in the experimental 1PNF x-ray crystallography data. b) Theoretical hydrogen bonds computed from 1PNF MD simulation data.

#### 4.1.3 Energy Convergence and MM-GBSA of Rotamer Models

After confirming that the MD simulation of the complex was stable and consistent with experimental structural data, interaction energy was computed at a 1 ns interval over the duration of the 100 ns MD trajectory. Figure 4.7 shows stable interaction energy over trajectory time indicating energy convergence for the 1PNF MD simulation during the

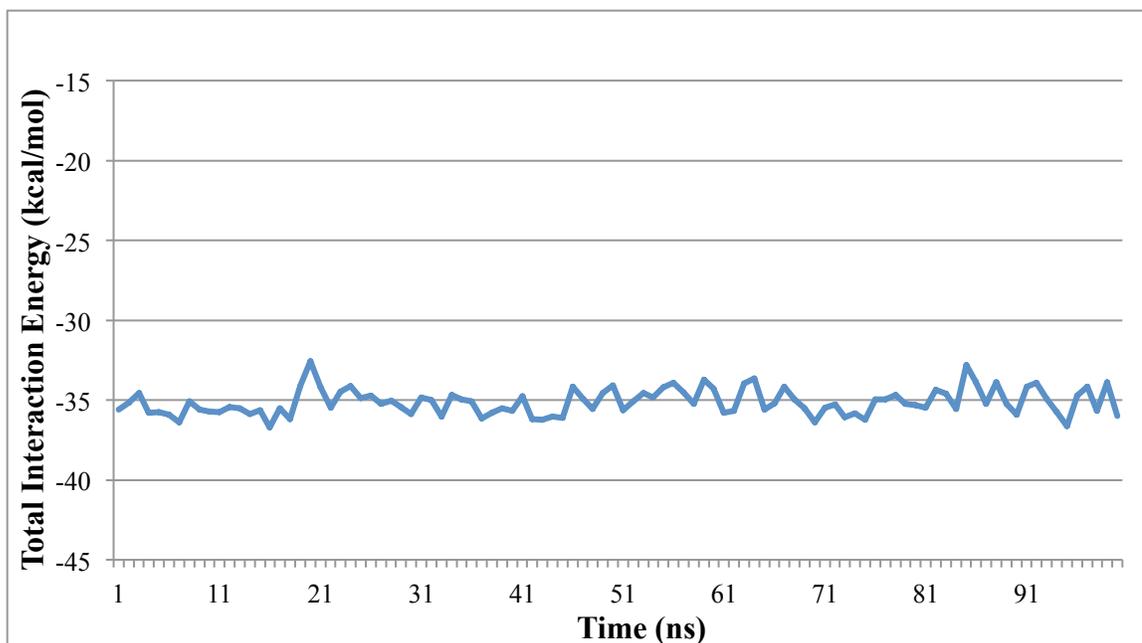


Figure 4.7: Interaction energy stability during 100 ns 1PNF MD simulation.

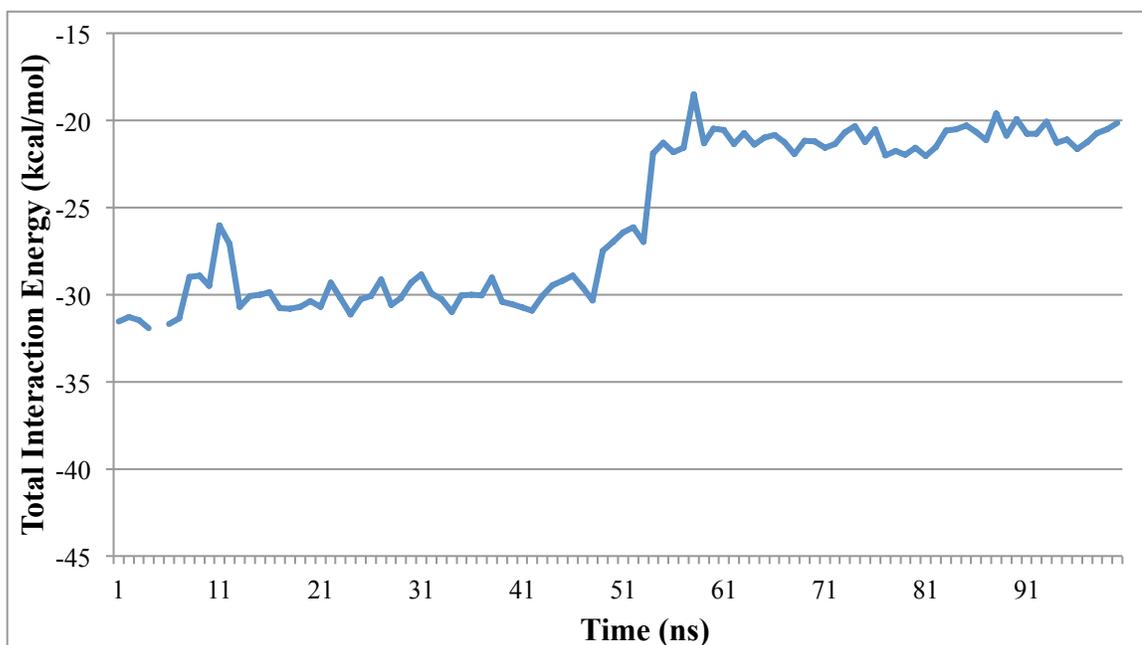


Figure 4.8: Interaction energy stability during 100 ns R911 Dun MD simulation. Interaction energy does not stabilize until after 55 ns into the simulation. Energy data at 5 ns could not be obtained due to loss of data caused by hardware failure of computing node.

100 ns MD simulation. In comparison, Figure 4.8 shows lack of energy convergence during the first 54 ns of the R911 Dun MD simulation. Unlike the R911 Dun MD

simulation, the R911 Dyn simulation had stable interaction energy through out the 100 ns simulation. Analysis of the R911 Dun conformations before and after the 10 kcal/mol interaction energy transition at 54 ns indicated that the conformations during the latter half of the simulation is similar to the R911 Dyn MD simulation. This is specifically observed in the orientation of the R911 D57L mutation that destabilizes R61 hydrogen bonds with the *N*-acetyl group of the second GlcNAc. The R911 Dynameomics model adopts this conformation during energy minimization & equilibration and remains stable in this conformation throughout the 100 ns production run. However the R911 Dunbrack model only adopts the altered R61 orientation in the latter half of the simulation. This would indicate that the Dunbrack model is taking longer to adopt the altered R61 orientation relative to the R911 dynameomics model. Thus, the selection of the latter half of the R911 Dun MD simulation for molecular mechanics generalized Born surface area (MM-GBSA) analysis is rationalized in part by the altered conformation of the R61 which is consistent in both R911 Dunbrack and R911 Dynameomics trajectories. This also demonstrates the importance of longer simulation runs employed in this study to accommodate the sampling requirements to reach energy convergence. Energy convergence analysis of the other MD simulations with the chitobiose (GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH) ligand indicated all simulations had reached structural convergence after the first 60 ns. Therefore, MD production data from the converged portion of the trajectory (last 40 ns) was subjected to MM-GBSA energy analysis.

The binding energy was decomposed into contributions from direct electrostatic interactions, polar and non-polar desolvation and van der Waals contacts, employing the MM-GBSA method.<sup>4</sup> The MM-GBSA procedure yielded total interaction energy of -35.1

kcal/mol for wtPNGase F (1PNF). This value overestimates the experimental binding free energy of -7.1 kcal/mol for wtPNGase F (Table 3.4), which is a typical characteristic of MM-GBSA calculations that omit entropic penalties associated with ligand binding.<sup>4</sup> Entropic effects, arising from changes in conformational flexibility can be estimated, but may require very long MD simulations in order to achieve convergence.<sup>5</sup> However, it may be anticipated that entropic effects arising from reduction in the flexibility of protein side chains will be most significant for those residues that interact strongly with the ligand, and least significant for the tepid or cold residues. For these reasons, the entropic contributions were not computed. Furthermore, conserved water molecules are not included in these MM-GBSA energy estimations and the lack of conserved waters may yield inaccurate estimated energies for E206S, D60C, and R248W, sites that are known to interact with conserved water molecules in the wild-type PNGase F complex experimental X-ray data.<sup>1</sup> In addition, due to approximations made in estimating the decomposed per residue contributions, computed binding energies may have relatively high error, thus making quantitative assessment impermissible. Therefore, qualitative analysis of MM-GBSA data is appropriate.<sup>6</sup>

Since 1PNF represents the structure of the wtPNGase F enzyme, the D60A single point mutant, which was used as a non-affinity optimized experimental control, was similarly used as a control structural model for comparison against the R911 and R911 C60A Dunbrack and Dynameomics rotamer models. The estimated per residue  $\Delta G_{\text{BINDING}}$  (kcal/mol) energies of the mutagenized residues from these 6 MD simulations with the chitobiose (GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH) ligand are presented in Table 4.5.

Table 4.5: Estimated MM-GBSA interaction action energies of rotamer models. Sub-total  $\Delta G_{\text{BINDING}}$  (kcal/mol) of mutagenized residues were compared across 1PNF (wtPNGase F), D60A, and all 4 rotamer models complexed with the chitobiose (GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH) ligand. Sub-total  $\Delta\Delta G_{\text{BINDING}}$  (kcal/mol) energies relative to the D60A control clone indicated that the Dynameomics rotamer models of R911 and R911 C60A best approximated experimental binding free energy ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) trends (green), unlike the Dunbrack models (red).

Residue	1PNF wtPNGase F	D60A	R911 Dun	R911 C60A Dun	R911 Dyn	R911 C60A Dyn
D57/-/L	-0.3 $\pm$ 0.7	-0.3 $\pm$ 0.7	-0.1 $\pm$ 0.0	-0.1 $\pm$ 0.0	-0.1 $\pm$ 0.0	0.0 $\pm$ 0.0
D60/A/C/A	-1.8 $\pm$ 1.7	-2.1 $\pm$ 1.2	-1.5 $\pm$ 0.5	-0.8 $\pm$ 1.1	-4.5 $\pm$ 0.5	-0.1 $\pm$ 0.2
I156/-/L	-1.1 $\pm$ 0.2	-1.0 $\pm$ 0.3	-0.2 $\pm$ 0.1	-0.9 $\pm$ 0.2	-0.1 $\pm$ 0.0	-0.1 $\pm$ 0.1
G192/-/I	0.0 $\pm$ 0.1	0.0 $\pm$ 0.1	-0.9 $\pm$ 0.3	-0.9 $\pm$ 0.3	-0.8 $\pm$ 0.2	-2.9 $\pm$ 0.5
E206/-/S	0.7 $\pm$ 0.9	0.7 $\pm$ 0.9	-0.1 $\pm$ 0.3	-0.1 $\pm$ 0.3	-0.1 $\pm$ 0.0	-1.1 $\pm$ 0.9
R248/-/W	-0.1 $\pm$ 0.5	-0.1 $\pm$ 0.5	-0.1 $\pm$ 0.1	0.2 $\pm$ 0.1	-0.6 $\pm$ 0.1	-0.5 $\pm$ 0.2
Sub-total $\Delta G_{\text{BINDING}}$	-2.7 $\pm$ 1.0	-2.8 $\pm$ 0.8	-2.9 $\pm$ 0.3	-2.6 $\pm$ 0.5	-6.1 $\pm$ 0.2	-4.6 $\pm$ 0.5
Sub-total $\Delta\Delta G_{\text{BINDING}}$	-0.0 $\pm$ 1.2	-	-0.2 $\pm$ 0.8	0.2 $\pm$ 0.9	-3.4 $\pm$ 0.8	-1.9 $\pm$ 0.9
$\Delta\Delta G_{\text{BIND-EXP}}$	0.5	-	-1.9	-0.7	-1.9	-0.7

The total  $\Delta G_{\text{BINDING}}$  for 1PNF (-2.7 kcal/mol) and D60A (-2.8 kcal/mol) are

similar, and indicate that the D60A mutation is slightly energetically favourable, consistent with previous analysis (Table 2.2). The experimental  $\Delta\Delta G_{\text{BIND-EXP}}$  of 1PNF relative to the D60A model confirms that the D60A mutation is energetically favourable by -0.5 kcal/mol. Comparison of the Dunbrack rotamer models, R911 Dun and R911 C60A Dun, relative to D60A indicates that estimated  $\Delta\Delta G_{\text{BINDING}}$  interaction energies are not significantly different than 1PNF. The results from the Dunbrack models are inconsistent with experimental data. However, comparison of the Dynameomics rotamer models, R911 Dyn and R911 C60A Dyn, relative to D60A confirms interaction energy trends consistent with experimental data. Specifically, the R911 C60A mutation relative to R911 has relatively unfavorable interaction energy, but not worse than 1PNF. This data also supports the importance of the D60C mutation in R911 for affinity enhancement relative to 1PNF and D60A. Based on the reproducibility of the experimental interaction

energies, rotamer conformations of the mutagenized residues were determined from the MD simulation. Furthermore, the Dynameomic rotamer models, R911 Dyn and R911 C60A Dyn, were selected for MD simulations using additional glycan and glycotriptide ligands.

#### 4.1.4 Rotamer Conformations from R911 Dyn MD Simulation

Rotamer dihedral angles of mutagenized residues from the R911 Dyn MD simulation were extracted from the energetically converged portion of the trajectory (last 40 ns). Dihedral angle frequency histograms were plotted to identify preferred dihedral angles of all 6 mutagenized residues (Figures 4.9 – 4.14). The preferred rotamer conformations were identified and are listed in Table 4.6. Four rotamers had multiple preferred conformations (D57L Chi2, E206S Chi1, G192I Chi1, and G192I Chi2). Therefore, the most preferred combination of rotamers were identified based on frequency of occurrence in extracted frames from the last 40 ns of the converged trajectory (Table 4.7). It was assumed that the most frequent set of rotamer combinations represents the most favored orientation for ligand interaction. A snapshot from the

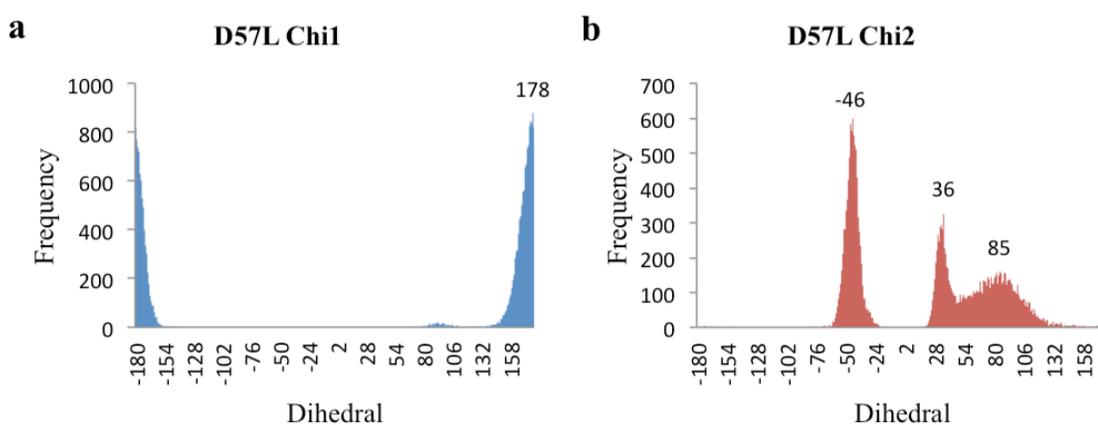


Figure 4.9: D57L rotamer histograms of Chi1 and Chi2 dihedral angles.

trajectory depicting the most favored set of rotamer conformations from the R911 Dyn MD simulation are shown in Figure 4.15.

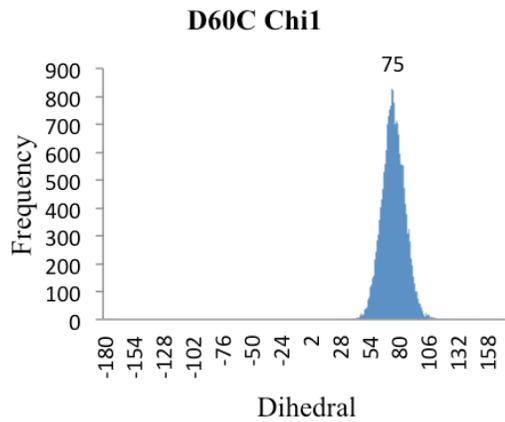


Figure 4.10: D60C rotamer histogram of Chi1 dihedral angle.

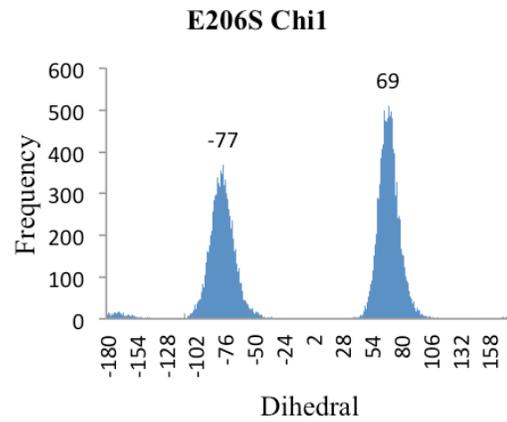


Figure 4.11: E206S rotamer histogram of Chi1 dihedral angles.

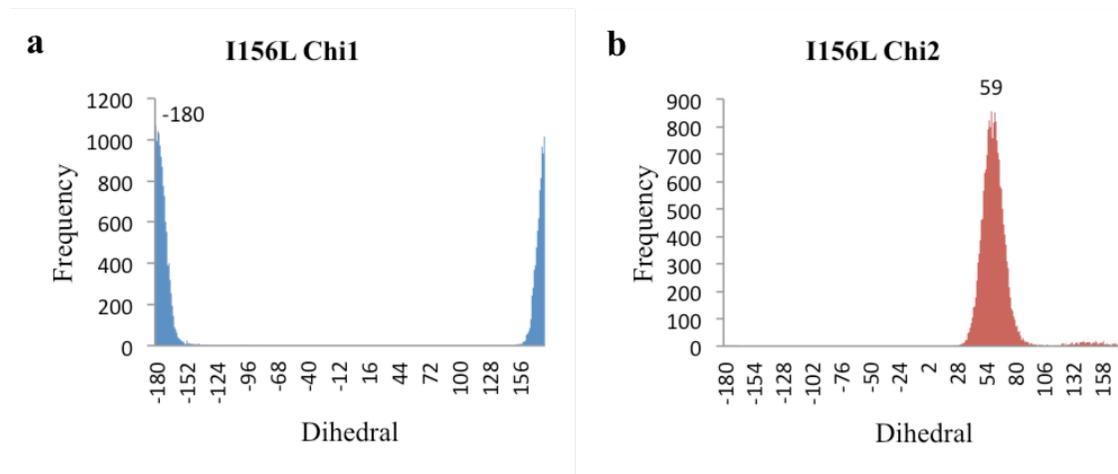


Figure 4.12: I156L rotamer histograms of Chi1 and Chi2 dihedral angles.

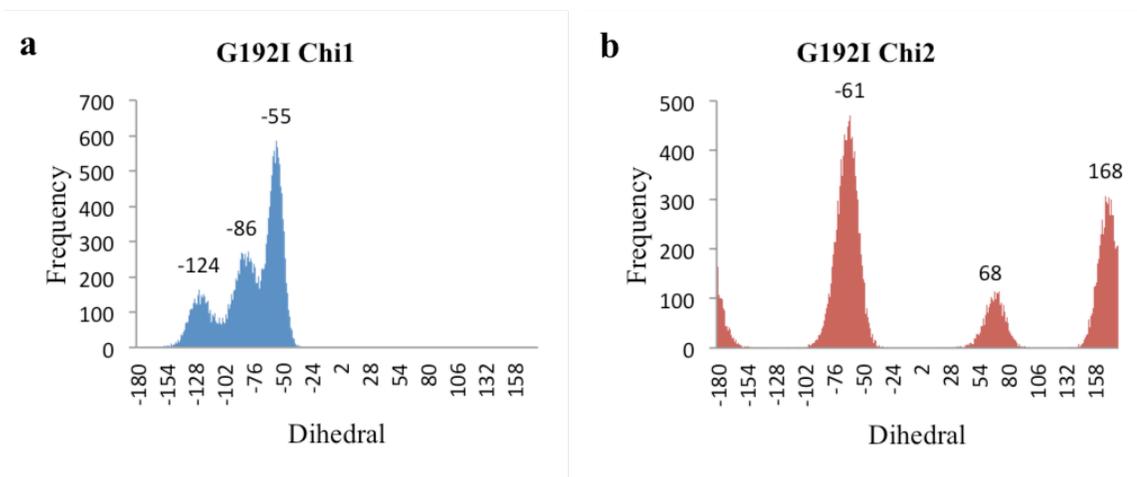


Figure 4.13: G192I rotamer histograms of Chi1 and Chi2 dihedral angles.

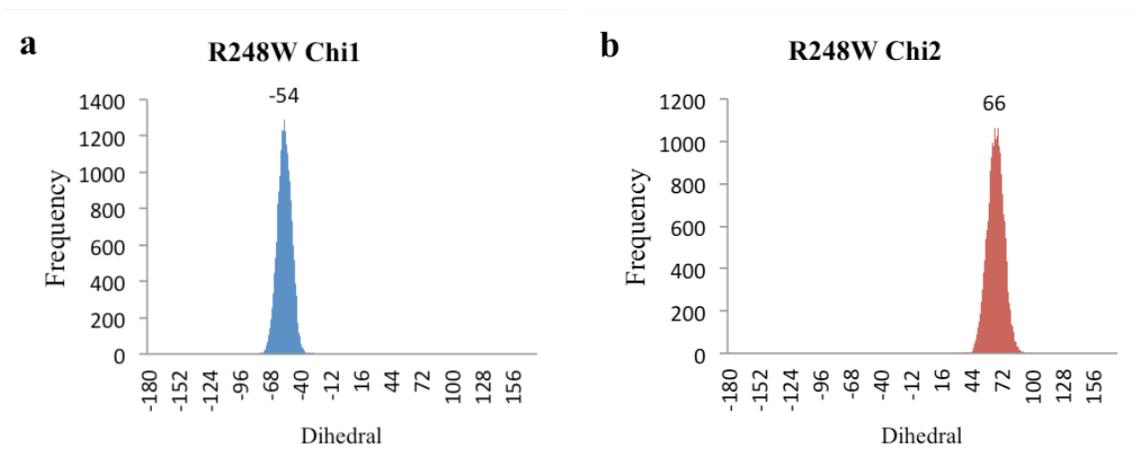


Figure 4.14: R248W rotamer histograms of Chi1 and Chi2 dihedral angles.

Table 4.6: Rotamer conformations from R911 Dyn MD simulation. Preferred rotamer dihedral angles are listed for all 6 mutagenized residues (green). The selected Dymeomics rotamer (Dyn) representing the initial conformation at the start of the MD simulation are include for comparison.

R911 Residue	Rotamer Source	Conformation		Probability (%)	Average Angles		Clashes
		Chi1	Chi2		Chi1	Chi2	
D57L	Dyn	t	g+	24.05	183.0	66.6	4
D57L	MD	t	g-	41.80	178.0	-46.0	
D57L	MD	t	g+	22.59	178.0	36.0	
D57L	MD	t	g+	35.60	178.0	85.0	
D60C	Dyn	g+		15.05	54.8		0
D60C	MD	g+		99.00	75.0		
I156L	Dyn	g-	t	65.19	292.0	168.5	4
I156L	MD	t	g+	99.00	-180.0	59.0	
G192I	Dyn	t	g+	1.41	184.1	68.3	13
G192I	MD	g-	g-	84.06	-55.0	-61.0	
G192I	MD	g-	g-	15.94	-86.0	-61.0	
E206S	Dyn	g-		73.06	310.9		0
E206S	MD	g+		54.65	69.0		
E206S	MD	g-		45.35	-77.0		
R248W	Dyn	g-	g-	28.21	294.7	98.5	7
R248W	MD	g-	g-	99.00	-54.0	66.0	

Table 4.7: Frequency of preferred rotamer combinations in R911 Dyn MD simulation. The frequency of preferred set of rotamer dihedral angles is listed. The most frequent set of rotamer conformations is indicated in green and depicted in Figure 4.

Frequency	D57L Chi2	G192I Chi1	G192I Chi2	E206S Chi1	Probability %
1570	-46	-55	-61	-77	22.52
1477	85	-55	-61	69	21.19
949	-46	-55	-61	69	13.61
868	36	-55	-61	69	12.45
565	85	-55	-61	77	8.11
431	36	-55	-61	-77	6.18
303	-46	-86	-61	-77	4.35
287	85	-86	-61	69	4.12
153	85	-86	-61	77	2.19
139	36	-86	-61	-77	1.99
137	36	-86	-61	69	1.97
92	-46	-86	-61	69	1.32

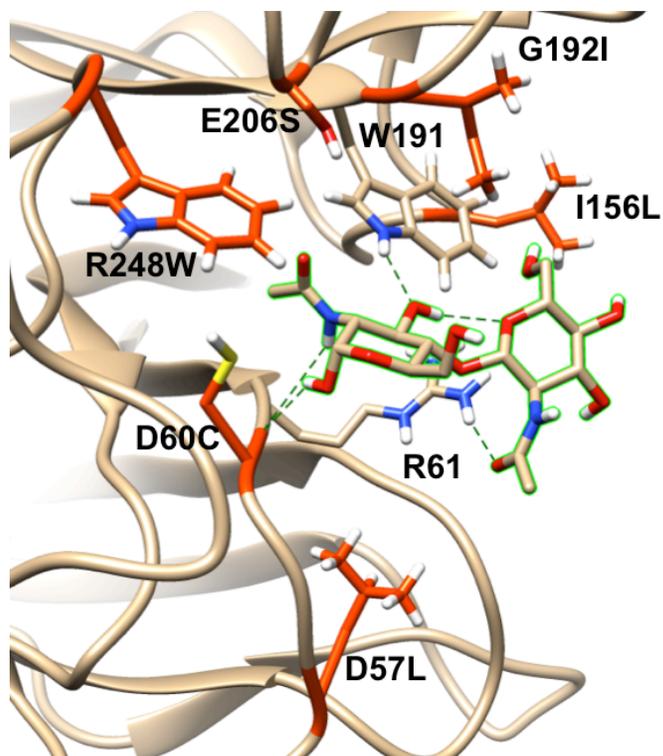


Figure 4.15: R911 Dyn MD simulation hydrogen bonds and preferred rotamers. Five theoretical hydrogen bonds between R911 residues D60C, R61 and W191 and the chitobiose ligand (green outline) are depicted by dashed lines. Rotamers of R911 mutations (orange) are depicted in the most frequent orientation extracted from the simulation trajectory at 73 ns.

#### 4.2 MD Simulations and Binding Free Energy Decomposition (MM-GBSA) of PNGase F Clones with *N*-Glycan and *N*-Glycotriptide Ligands

Using the previously validated R911 and R911 C60A Dynameomics structural models as well as 1PNF and D60A models, 100 ns MD simulations were conducted with modified ligands. One set of four simulations was conducted with the  $\beta$ -chitobiose ligand (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ OH) given that the attachment of the oligosaccharide moiety to N $_8$  of the asparine is in the  $\beta$ -configuration. The 1PNF structure of the wtPNGase F enzyme co-crystalized with  $\alpha$ -chitobiose ligand, even though in solution an equilibrium

state containing a mixture of both  $\alpha$ - and  $\beta$ - anomeric configurations of the O<sub>1</sub> hydroxyl group on the reducing end exists.<sup>1</sup> A second set of four simulations was conducted with the asparagine-linked glycotriptide motif, GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-X[-P]-Thr, recognized by wtPNGase F. Substrate requirement studies with PNGase F have confirmed that this is the essential motif required for optimal catalytic activity.<sup>7</sup> As RNase B was used as the *N*-glycan bearing glycoprotein target for experimental studies, the RNase B peptide sequence was used for the essential glycotriptide ligand (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr). In addition, residue E206 of the 1PNF and D60A models was modified to reflect the protonated state of the carboxyl group (GLH206). Optimal catalytic activity for the wtPNFase F is reported around pH 8.0 - 8.5, thus protonation of glutamate (pK<sub>a</sub> ~4.1) would normally be unlikely. However, point mutant studies of the neighboring hydrophobic tryptophan residues 207 and 251 indicate that a hydrophobic environment is critical for catalytic activity, and that this hydrophobic environment would raise the pK<sub>a</sub> of E206, potentially to ~8.5.<sup>8,9</sup>

#### *4.2.1 Analysis of PNGase F clones Complexed with the $\beta$ -chitobiose Ligand*

Stable simulation trajectories of all models were confirmed by RMSD analysis, followed by energy convergence analysis to confirm stable interaction energies. As with the prior rotamer model studies, data from the converged portion of the trajectories was subjected to MM-GBSA energy analysis. Table 4.8 summarizes the models of the PNGase F clones complexed with the  $\beta$ -chitobiose ligand used for MD simulations, the calculated average RMSD, and the estimated relative binding energies.

Estimated total theoretical binding free energy (Total  $\Delta G_{\text{BINDING}}$ ) of all clones overestimate the experimental binding free energy ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) due to the omission of entropic penalties as previously discussed.<sup>4,5</sup> Relative  $\Delta\Delta G_{\text{BINDING}}$  interaction energies of

the 6 mutagenized residues for the  $\beta$ -chitobiose ligand simulations reproduced the experimental observed interaction energy trends. The wtPNGase F enzyme has marginally less favorable total interaction energy (0.1 kcal/mol) relative to D60A whereas both R911 (-1.8 kcal/mol) and R911 C60A (-0.5 kcal/mol) have favorable interaction energies relative to D60A, with R911 being more favorable than R911 C60A.

Table 4.8: Structural models of wtPNGase F and clones used for MD simulations and MM-GBSA. Structural models with the  $\beta$ -chitobiose ligand in the binding pocket were constructed to conduct 100 ns MD simulations. Simulation stability was confirmed by analyzing RMSD over simulation time and the average RMSD values are listed. Estimated total theoretical binding energy for all 314 amino acids (Total  $\Delta G_{\text{BINDING}}$ ) of each clone are listed (kcal/mol). Estimated theoretical binding interaction energy (Sub-total  $\Delta\Delta G_{\text{BINDING}}$ ) comprised only of the 6 mutagenized residues of all clones relative to D60A are listed (kcal/mol). For comparison, experimental binding free energy ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) of all clones relative to D60A are also listed (kcal/mol).

Structural Model	Ligand	RMSD (Å)	Total $\Delta G_{\text{BINDING}}$	Sub-total $\Delta\Delta G_{\text{BINDING}}$	$\Delta\Delta G_{\text{BIND-EXP}}$
1PNF	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ OH	1.2440	-39.1	$0.1 \pm 1.3$	0.5
D60A	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ OH	1.2674	-31.5	-	-
R911 Dyn	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ OH	1.2944	-17.9	$-1.8 \pm 1.1$	-1.9
R911 C60A Dyn	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ OH	1.2890	-18.3	$-0.5 \pm 1.0$	-0.7

MM-GBSA energy analysis and per-residue contributions were computed to determine over the converged portion of the  $\beta$ -chitobiose ligand trajectories (last 40 ns) were used for  $\Delta\Delta G_{\text{BINDING}}$  relative to D60A. Decomposed energy contributions consist of the total molecular mechanical interaction energy ( $\Delta E_{\text{MM}}$ ), comprised of the sum of van der Waals ( $\Delta E_{\text{VDW}}$ ) and electrostatics ( $\Delta E_{\text{ELE}}$ ) components. The total binding free energy ( $\Delta G_{\text{BINDING}}$ ) is comprised of the generalized Born approximation of polar and non-polar solvation components ( $\Delta G_{\text{GB+SA}}$ ) and  $\Delta E_{\text{MM}}$ . The interaction energies (of the 6 residues mutagenized in wtPNGase F) for all 4 models (1PNF, D60A, R911 Dyn, R911 C60A Dyn) complexed with  $\beta$ -chitobiose are summarized in Tables 4.9 – 4.12.

Table 4.9: MM-GBSA of 1PNF complexed with  $\beta$ -chitobiose. Estimated per residue contributions to the binding free energy (kcal/mol) for wtPNGase F complexed with  $\beta$ -chitobiose are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones. Residues required for catalytic activity are indicated in bold.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57	-0.1	1.9	1.8	-2.2	$-0.4 \pm 0.5$
<b>D60</b> (nucleophile)	-0.8	-14.7	-15.6	13.5	$-2.1 \pm 1.4$
I156	-0.9	0.2	-0.7	-0.4	$-1.1 \pm 0.2$
G192	-0.1	-0.2	-0.3	0.2	$0.0 \pm 0.1$
<b>E206</b>	-0.6	5.3	4.7	-3.7	$1.0 \pm 0.9$
R248	-0.1	-0.7	-0.8	0.7	$-0.1 \pm 0.4$
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-2.7</b>	<b>-8.2</b>	<b>-10.9</b>	<b>8.2</b>	<b><math>-2.7 \pm 0.8</math></b>

Table 4.10: MM-GBSA of D60A complexed with  $\beta$ -chitobiose. Estimated per residue contributions to the binding free energy (kcal/mol) for D60A complexed with  $\beta$ -chitobiose are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57	-0.1	2.3	2.2	-2.5	$-0.3 \pm 0.5$
D60A	-1.0	-2.5	-3.5	1.7	$-1.8 \pm 0.4$
I156	-0.9	0.2	-0.7	-0.4	$-1.1 \pm 0.2$
G192	-0.1	0.0	-0.1	0.1	$0.0 \pm 0.1$
E206	-0.6	2.6	1.9	-1.2	$0.7 \pm 2.0$
R248	-0.1	-1.9	-2.1	1.9	$-0.2 \pm 0.5$
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-3.0</b>	<b>0.7</b>	<b>-2.3</b>	<b>-0.2</b>	<b><math>-2.7 \pm 1.0</math></b>

Table 4.11: MM-GBSA of R911 Dyn complexed with  $\beta$ -chitobiose. Estimated per residue contributions to the binding free energy (kcal/mol) for R911 complexed with  $\beta$ -chitobiose are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57L	-0.1	-0.1	-0.1	0.1	$-0.1 \pm 0.1$
D60C	-0.2	0.2	0.0	-0.2	$-0.2 \pm 0.2$
I156L	-0.3	0.2	-0.1	-0.2	$-0.3 \pm 0.2$
G192I	-1.8	-2.3	-4.1	1.0	$-3.1 \pm 0.6$
E206S	-0.5	-0.9	-1.4	1.0	$-0.4 \pm 0.4$
R248W	-0.4	-0.1	-0.6	0.2	$-0.4 \pm 0.2$
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-3.4</b>	<b>-2.9</b>	<b>-6.3</b>	<b>1.8</b>	<b><math>-4.5 \pm 0.4</math></b>

Table 4.12: MM-GBSA of R911 C60A Dyn complexed with  $\beta$ -chitobiose. Estimated per residue contributions to the binding free energy (kcal/mol) for R911 C60A complexed with  $\beta$ -chitobiose are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57L	-0.1	-0.1	-0.2	0.1	-0.1 $\pm$ 0.0
D60A	-1.0	-2.5	-3.4	1.9	-1.6 $\pm$ 0.2
I156L	-0.1	0.0	-0.1	0.0	-0.1 $\pm$ 0.0
G192I	-0.9	-0.2	-1.1	0.1	-1.0 $\pm$ 0.3
E206S	-0.2	-0.6	-0.7	0.6	-0.1 $\pm$ 0.0
R248W	-0.6	0.3	-0.3	0.0	-0.3 $\pm$ 0.1
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-2.9</b>	<b>-3.0</b>	<b>-5.9</b>	<b>2.7</b>	<b>-3.2 <math>\pm</math> 0.2</b>

Analysis of the decomposed energy indicates that both G192 and E206 in wtPNGase F and the D60A clone have unfavorable  $\Delta G_{BINDING}$  interaction energy, consistent with the 5 ns MD simulation energy decomposition results used to identify tepid and hot residues for library design (Table 2.1). The yeast-display selected mutations G192I and E206S contribute favorably towards the estimated interaction energy of R911 and R911 C60A. In addition, the D57L, I156L, and R248W mutations are also estimated to have slightly favorable interaction energies.

Table 4.13: Estimated MM-GBSA theoretical interaction action energies of models complexed with  $\beta$ -chitobiose. Sub-total  $\Delta G_{BINDING}$  (kcal/mol) of mutagenized residues were compared across 1PNF (wtPNGase F), D60A, R911 Dyn, and R911 C60A Dyn. Sub-total  $\Delta\Delta G_{BINDING}$  (kcal/mol) energies relative to the D60A control clone are indicated. For comparison, experimentally determined binding interaction energy ( $\Delta\Delta G_{BIND-EXP}$ ) of clones is also listed relative to D60A.

Residue	1PNF (GLH206)	D60A (GLH206)	R911 Dyn	R911 C60A Dyn
D57/-/L	-0.4 $\pm$ 0.5	-0.3 $\pm$ 0.5	-0.1 $\pm$ 0.1	-0.1 $\pm$ 0.0
D60/A/C/A	-2.1 $\pm$ 1.4	-1.8 $\pm$ 0.4	-0.2 $\pm$ 0.2	-1.6 $\pm$ 0.2
I156/-/L	-1.1 $\pm$ 0.2	-1.1 $\pm$ 0.2	-0.3 $\pm$ 0.2	-0.1 $\pm$ 0.0
G192/-/I	0.0 $\pm$ 0.1	0.0 $\pm$ 0.1	-3.1 $\pm$ 0.6	-1.0 $\pm$ 0.3
E206/-/S	1.0 $\pm$ 0.9	0.7 $\pm$ 2.0	-0.4 $\pm$ 0.4	-0.1 $\pm$ 0.0
R248/-/W	-0.1 $\pm$ 0.4	-0.2 $\pm$ 0.5	-0.4 $\pm$ 0.2	-0.3 $\pm$ 0.1
Sub-total $\Delta G_{BINDING}$	-2.7 $\pm$ 0.8	-2.7 $\pm$ 1.0	-4.5 $\pm$ 0.4	-3.2 $\pm$ 0.2
Sub-total $\Delta\Delta G_{BINDING}$	0.1 $\pm$ 1.3	-	-1.8 $\pm$ 1.1	-0.5 $\pm$ 1.0
$\Delta\Delta G_{BIND-EXP}$	0.5	-	-1.9	-0.7

A comparison of the theoretical  $\Delta G_{\text{BINDING}}$  interaction energy of all clones with the  $\beta$ -chitobiose ligand is provided in Table 4.13. Experimental binding energies of the clones relative to D60A ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) is also included. Of particular importance is D60, the residue required for catalytic activity. Both D60 and D60A make significant favorable interactions, where as D60C, is estimated to have a significantly less favorable contribution of 0.2 kcal/mol. This theoretical data is contrary to experimental data on R911 and the R911 C60A clones, which indicates that the D60C mutation makes a -1.2 kcal/mol greater contribution ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) towards the overall R911 binding energy relative to R911 C60A (Table 3.4). This discrepancy is likely due to the experimental data being generated using denatured RNase B as the *N*-glycan bearing target ligand, not  $\beta$ -chitobiose as in the MD simulation. This is significant as reports of point mutant studies using D60N, D60E, and D60C, all indicate that D60 is required for catalytic activity.<sup>1, 8</sup> These reports are consistent with the significantly decreased catalytic activity observed for the D60A point mutant (Table 3.3). The critical role of D60 is also supported by 1PNF crystal structure data indicating D60 directly interacts with the anomeric hydroxyl of the reducing GlcNAc of the bound chitobiose ligand.<sup>1</sup> In the case of a glycopeptide, the anomeric hydroxyl would be replaced with the glycosidic bond that the enzyme hydrolyzes. Furthermore, substrate specificity requirement studies with wtPNGase F enzyme demonstrate that the enzyme recognizes both the asparagine-linked carbohydrate moiety as well as the peptide portion consisting of the N-X-T glycosylation sequone.<sup>7</sup> Therefore, MD simulations with the  $\beta$ -chitobiose ligand can neither sufficiently simulate experimental interactions D60 has at the site of the glycosidic linkage of *N*-

glycans, nor interactions of the wtPNGase enzyme with the peptide sequone common to *N*-linked glycans.

#### 4.2.2 Analysis of PNGase F clones Complexed with a Glycotriptide Ligand

MD simulations with the clones complexed with the common *N*-linked glycotriptide (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) were conducted to more accurately model binding interactions and estimate interaction energies. In order to neutralize terminal charges ( $NH_3^+$  for the N-terminal and  $COO^-$  for the C-terminal), the peptide portion of the glycotriptide ligand was modeled with the N-terminal ACE [ $-C(=O)-CH_3$ ] protecting group and the C-terminal NME [ $-C(=O)-NH-CH_3$ ] protecting group as defined in xleap, a component of AMBER Tools 13.<sup>10</sup> Table 4.14 summarizes the models of the PNGase F clones complexed with a glycotriptide ligand used for MD simulations, the calculated average RMSD, and the estimated relative binding energies.

Table 4.14: Structure models of wtPNGase F and clones used for MD simulations & MM-GBSA. Structural models with a glycotriptide ligand in the binding pocket were constructed to conduct 100 ns MD simulations. Simulation stability was confirmed by analyzing RMSD over simulation time and the average RMSD values are listed. Estimated total theoretical binding energy for all 314 amino acids (Total  $\Delta G_{\text{BINDING}}$ ) of each clone are listed (kcal/mol). Estimated theoretical binding interaction energy (Sub-total  $\Delta\Delta G_{\text{BINDING}}$ ) comprised only of the 6 mutagenized residues of all clones relative to D60A are listed (kcal/mol). For comparison, experimental binding free energy ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) of all clones relative to D60A are also listed (kcal/mol).

Structural Model	Ligand	RMSD (Å)	Total $\Delta G_{\text{BINDING}}$	Sub-total $\Delta\Delta G_{\text{BINDING}}$	$\Delta\Delta G_{\text{BIND-EXP}}$
1PNF (GLH206)	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr	1.2238	-44.0	$1.6 \pm 1.2$	0.5
D60A (GLH206)	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr	1.1916	-48.4	-	-
R911 Dyn	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr	1.2705	-54.4	$-1.4 \pm 0.8$	-1.9
R911 C60A Dyn	GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr	1.2618	-51.3	$-1.9 \pm 0.8$	-0.7

The D60A (GLH206), R911 Dyn, and R911 C60A Dyn glycotriptide complexed models all had stable interaction energies through out the simulation.

However, the 1PNF (GLH206) trajectory showed interaction energy fluctuations during the simulation (Figure 4.16).

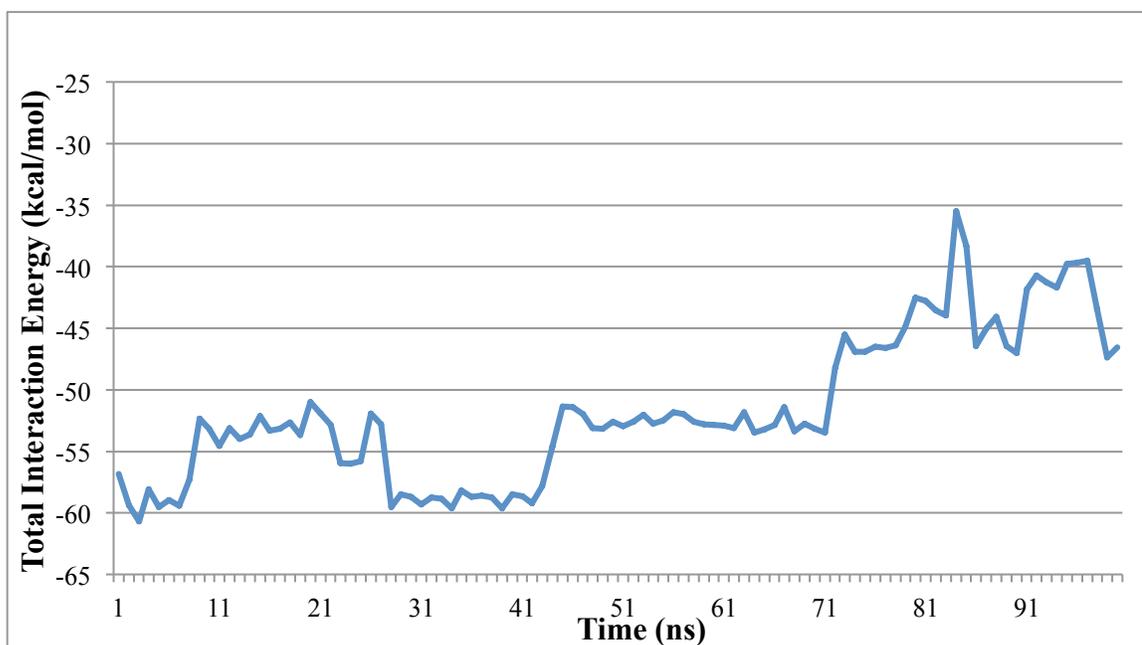


Figure 4.16: Interaction energy stability during 100 ns 1PNF (GLH206) MD simulation with the glycotriptide ligand. The average interaction energy during 51 – 70 ns is approximately -52.7 kcal/mol and during 80 – 100ns it is -44.0 kcal/mol.

Visualization of the 1PNF (GLH206) glycotriptide trajectory indicated a conformational change between 71 – 74 ns, consistent with interaction energy fluctuations. During the first 70 ns of the simulation the reducing GlcNAc had a normal  ${}^4C_1$  chair conformation (Figures 4.17a & 4.18) with the Asn-Leu-Thr tripeptide portion relatively stable. The dynamic motion of the Leu and Thr residues is relatively unrestricted but stable, whereas the Asn residue is relatively constrained since the attached chitobiose is held in the binding pocket during the entire simulation. The *N*-

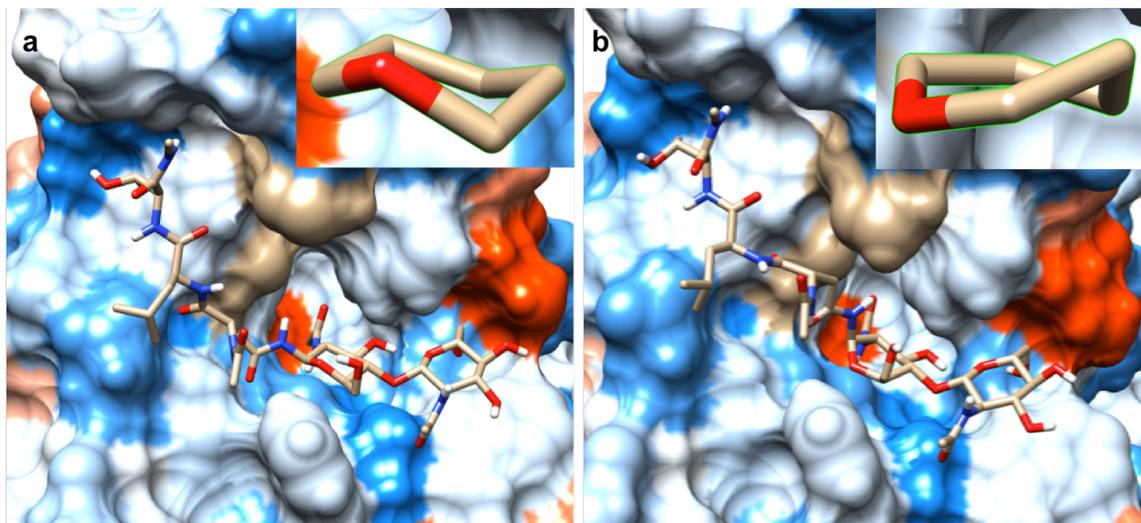


Figure 4.17: Chair and skew-boat ring conformations of reducing GlcNAc in wtPNGase F complexed with glycotriptide. Surface hydrophobicity depicted of wtPNGase F with glycotriptide with in binding pocket. a)  ${}^4C_1$  chair conformation of reducing GlcNAc observed during first  $\sim 70$  ns of MD simulation (60 ns snapshot). Inset shows the  ${}^4C_1$  conformation with ring atoms only. b) Skew-boat conformation during the last 26 ns of the simulation (86 ns snapshot). Inset shows the skew-boat conformation with ring atoms only.

acetyl group of the reducing GlcNAc was extended into a hydrophobic pocket consistent with 1PNF x-ray crystallography data.<sup>1</sup> However, between 71 – 74 ns, the peptide backbone of the Asn residue rotated towards the protein face aligning across a groove extending diagonally upwards from the chitobiose binding pocket (Figure 4.17b). Simulated hydrogen bonds observed between D60-O – GlcNAc316 NAc and Y85-OH – N316-O $\delta$  (chitobiose-linked asparagine) are lost during this conformational change (Figure 4.18). Thus bringing the ligand backbone oxygen atom (N316-O) of chitobiose-linked asparagine into proximity of W207 in the extended groove. This orientation results in the formation of a simulated W207-N $\epsilon$  – N316-O hydrogen bond (Figure 4.19). This conformation change strains the asparagine side-chain glycosidic bond to shift towards a more axial orientation, resulting in the previously observed  ${}^4C_1$  chair conformation of the reducing GlcNAc to shift into a skew-boat conformation (Figures 4.17b and 4.19). The

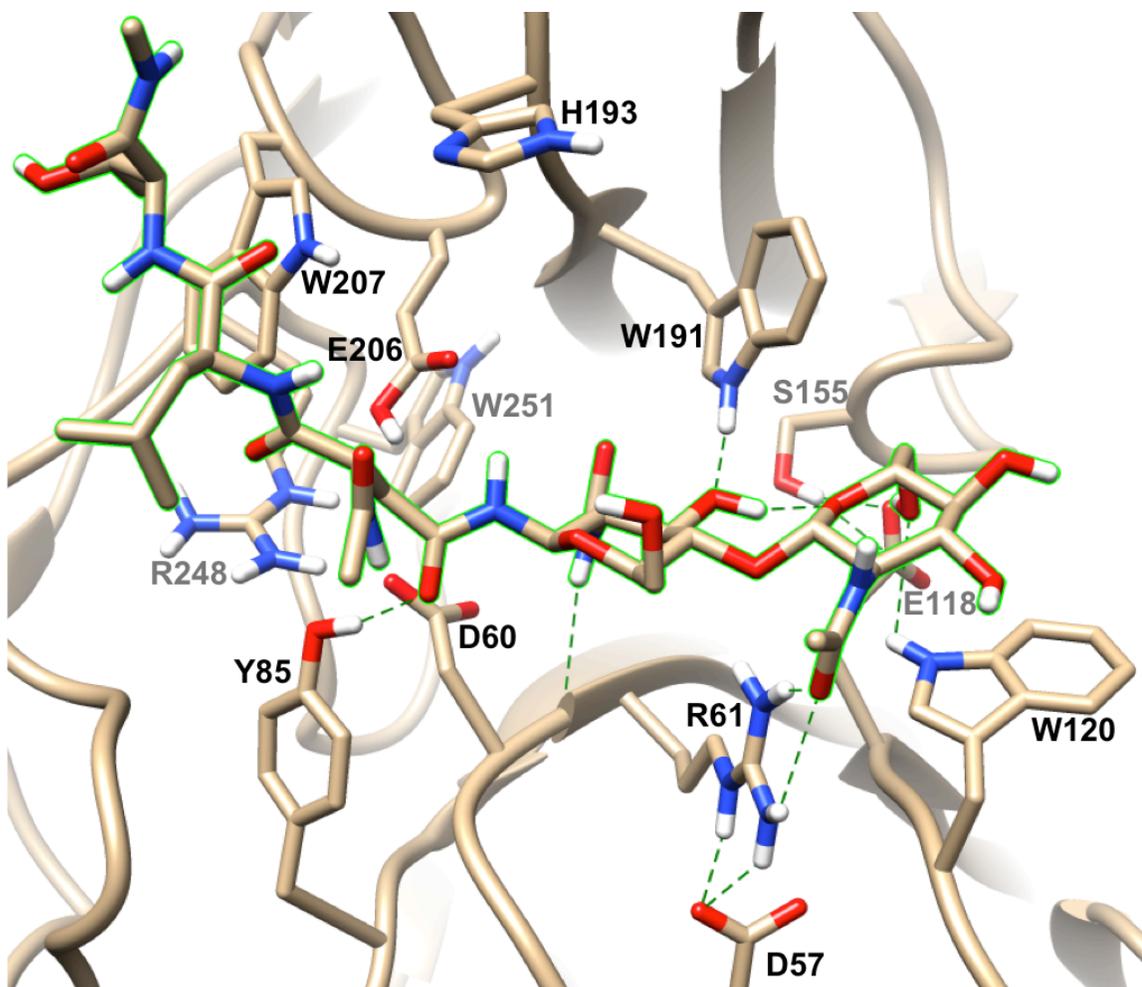


Figure 4.18: 1PNF (GLH206) and glycotriptide MD simulation hydrogen bonds at 60 ns time point. Hydrogen bonds are depicted as dashed green lines. The glycotriptide ligand is outlined in green. Amino acid residues critical for catalytic activity (D60 and E206), substrate recognition, and stabilizing interactions are depicted. The simulated hydrogen bonds D60-O – GlcNAc316 NAc and Y85-OH – N316-O $\delta$  (chitobiose-linked asparagine) are depicted towards the center. The reducing GlcNAc is in a  ${}^4C_1$  conformation.

observed skew-boat conformation is similar to the Michaelis complex in glycosidic mechanisms.<sup>11, 12</sup> The simulated W207-N $\epsilon$  – N316-O hydrogen bond and the skew-boat conformation of the reducing GlcNAc persisted for the remainder of the simulation (26 ns).

The skew-boat conformation change of the reducing GlcNAc was only observed with the 1PNF (GLH206) complexed with the glycotriptide ligand and none of the

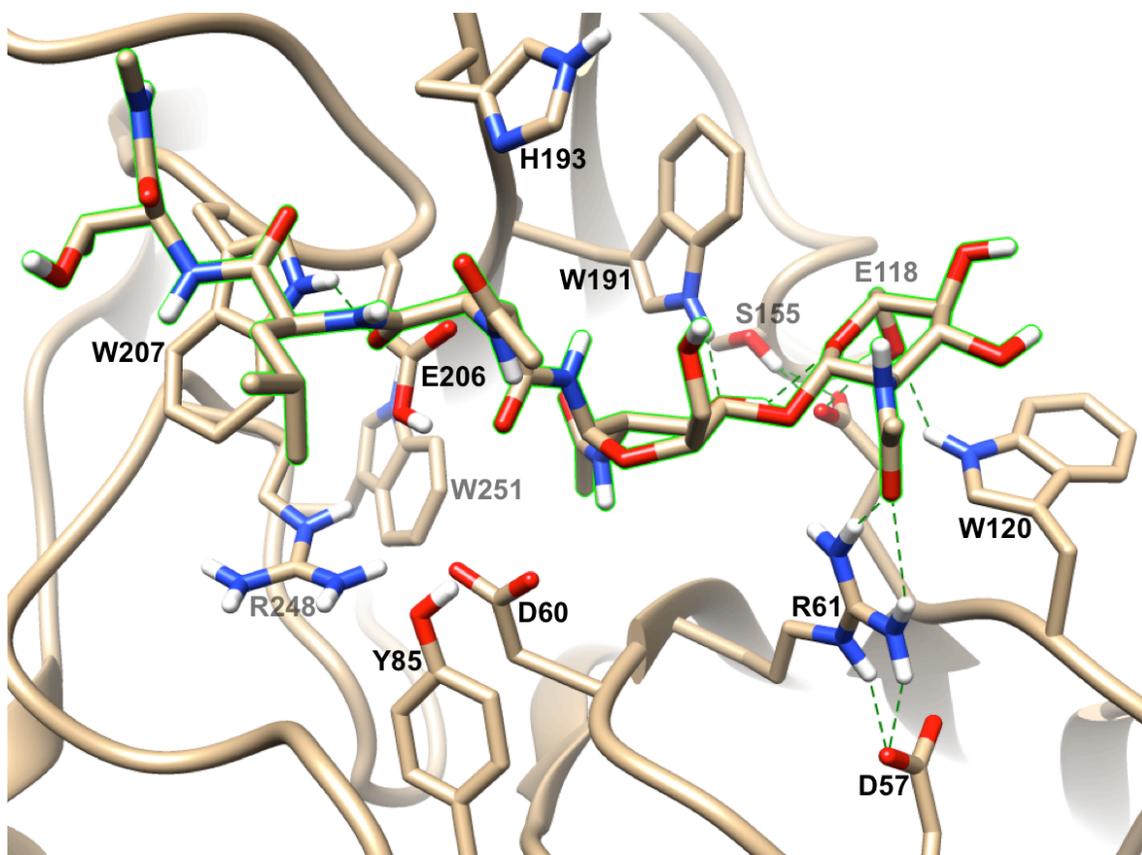


Figure 4.19: 1PNF (GLH206) and glycotriptide MD simulation hydrogen bonds at 84 ns time point. Hydrogen bonds are depicted as dashed green lines. The glycotriptide ligand is outlined in green. Amino acid residues critical for catalytic activity (D60 and E206), substrate recognition, and stabilizing interactions are depicted. The simulated W207-N $\epsilon$  – N316-O (chitobiose-linked asparagine) is depicted towards the top left. The reducing GlcNAc is in a skew-boat conformation.

other 1PNF ligand complexes. Similarly, no similar conformational changes were observed in the D60A (GLH206), R911 Dyn, and R911 C60A Dyn trajectories with the glycotriptide ligand, consistent with stable interaction energies. The observed skew-boat conformation indicates a unique interaction between the glycotriptide and the wtPNGase F enzyme. For this reason, the final 20 ns of the 1PNF (GLH206) trajectory as well as the other three simulations were used for MM-GBSA analysis.

The interaction energies (of the 6 residues mutagenized in wtPNGase F) for all 4 models (1PNF (GLH206), D60A (GLH206), R911 Dyn, R911 C60A Dyn) complexed with the glycotriptide are summarized in Tables 4.15 – 4.18.

Table 4.15: MM-GBSA of 1PNF (GLH206) complexed with glycotriptide ligand. Estimated per residue contributions to the binding free energy (kcal/mol) for wtPNGase F complexed with the glycotriptide (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones. Residues required for catalytic activity are indicated in bold.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57	-0.1	1.8	1.6	-1.9	$-0.3 \pm 0.7$
<b>D60</b> (nucleophile)	-2.1	-1.8	-3.9	4.7	$0.9 \pm 1.5$
I156	-0.8	0.2	-0.6	-0.4	$-0.9 \pm 0.2$
G192	-0.3	0.7	0.4	-0.5	$-0.1 \pm 0.3$
<b>E206</b>	-1.6	-1.2	-2.7	2.0	$-0.7 \pm 0.5$
R248	-0.3	-3.1	-3.5	3.3	$-0.2 \pm 1.2$
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-5.2</b>	<b>-3.3</b>	<b>-8.6</b>	<b>7.2</b>	<b><math>-1.4 \pm 0.9</math></b>

Table 4.16: MM-GBSA of D60A (GLH206) complexed with glycotriptide ligand. Estimated per residue contributions to the binding free energy (kcal/mol) for D60A (GLH206) complexed with the glycotriptide (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57	-0.1	1.9	1.8	-2.1	$-0.3 \pm 0.5$
D60A	-1.0	-2.8	-3.8	2.3	$-1.5 \pm 0.4$
I156	-0.9	0.0	-0.8	-0.2	$-1.0 \pm 0.2$
G192	-0.2	0.3	0.1	-0.2	$-0.1 \pm 0.1$
E206	-1.3	0.3	-1.0	1.1	$0.1 \pm 1.1$
R248	-0.6	-5.9	-6.5	6.3	$-0.2 \pm 1.2$
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-4.1</b>	<b>-6.1</b>	<b>-10.2</b>	<b>7.2</b>	<b><math>-3.0 \pm 0.8</math></b>

Table 4.17: MM-GBSA of R911 Dyn complexed with a glycotriptide ligand. Estimated per residue contributions to the binding free energy (kcal/mol) for R911 Dyn complexed with the glycotriptide (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57L	-0.1	0.0	-0.2	0.1	-0.1 $\pm$ 0.1
D60C	-1.6	-2.2	-3.8	2.0	-1.8 $\pm$ 0.7
I156L	-1.0	-0.1	-1.1	0.1	-1.0 $\pm$ 0.2
G192I	-0.7	0.0	-0.6	0.0	-0.7 $\pm$ 0.2
E206S	-0.2	-0.2	-0.4	0.4	0.0 $\pm$ 0.1
R248W	-1.1	-0.1	-1.2	0.4	-0.8 $\pm$ 0.2
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-4.6</b>	<b>-2.7</b>	<b>-7.3</b>	<b>2.9</b>	<b>-4.5 <math>\pm</math> 0.4</b>

Table 4.18: MM-GBSA of R911 C60A Dyn complexed with glycotriptide ligand. Estimated per residue contributions to the binding free energy (kcal/mol) for R911 C60A Dyn complexed with the glycotriptide (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) are shown. The interaction energy consists of only the 6 residues selected for mutagenesis via directed evolution for the R911 clones.

Key Contact Zone Residues	$\Delta E_{VDW}$	$\Delta E_{ELE}$	$\Delta E_{MM}$	$\Delta G_{GB+SA}$	$\Delta G_{BINDING}$
D57L	-0.1	0.0	-0.1	0.0	-0.1 $\pm$ 0.0
D60A	-1.2	-2.8	-4.0	2.3	-1.7 $\pm$ 0.3
I156L	-1.0	-0.1	-1.1	0.0	-1.1 $\pm$ 0.1
G192I	-0.9	0.2	-0.7	-0.2	-0.9 $\pm$ 0.1
E206S	-0.4	-0.3	-0.7	0.6	-0.1 $\pm$ 0.1
R248W	-1.3	-0.4	-1.7	0.6	-1.1 $\pm$ 0.2
<b>Sub-total Interaction Energy <math>\Delta G_{BINDING}</math></b>	<b>-4.9</b>	<b>-3.4</b>	<b>-8.3</b>	<b>3.4</b>	<b>-4.9 <math>\pm</math> 0.2</b>

A comparison of the theoretical  $\Delta G_{BINDING}$  interaction energy of all clones with the glycotriptide ligand is provided in Table 4.19. Interaction energy analysis estimates of 1PNF (GLH206) and D60A (GLH206) indicate that the D60A mutation has favorable interaction energies relative to the wtPNGase F. This is primarily due to more favorable solvation energy ( $\Delta G_{GB+SA}$ ) contributions for D60A relative to D60. These data are consistent with the computational alanine scanning results (Table 2.2). Relative to the wtPNGase F enzyme, the D60A single point mutant results in more favorable total interaction energy (of the 6 residues indicated) by -1.6 kcal/mol ( $\Delta \Delta G_{BINDING}$ ) relative to 1PNF (GLH206). This is supported by several other data: 1) the stable interaction energy

of D60A relative to 1PNF over the course of the simulation, 2) the lack of glycopeptide conformational change as observed with the 1PNF (GLH206) simulation data, and 3) the experimental binding energy of D60A is -0.5 kcal/mol more favorable than the wtPNGase F enzyme.

Table 4.19: Estimated MM-GBSA theoretical interaction action energies of models complexed with a glycotriptide ligand. Sub-total  $\Delta G_{\text{BINDING}}$  (kcal/mol) of mutagenized residues were compared across 1PNF (GLH206), D60A (GLH206), R911 Dyn, and R911 C60A Dyn complexed with (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr). Sub-total  $\Delta\Delta G_{\text{BINDING}}$  (kcal/mol) energies relative to the D60A control clone are indicated. For comparison, experimentally determined binding interaction energy ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) of clones is also listed relative to D60A.

Residue	1PNF (GLH206)	D60A (GLH206)	R911 Dyn	R911 C60A Dyn
D57/-/L	-0.3 $\pm$ 0.7	-0.3 $\pm$ 0.5	-0.1 $\pm$ 0.1	-0.1 $\pm$ 0.0
D60/A/C/A	0.9 $\pm$ 1.5	-1.5 $\pm$ 0.4	-1.8 $\pm$ 0.7	-1.7 $\pm$ 0.3
I156/-/L	-0.9 $\pm$ 0.2	-1.0 $\pm$ 0.2	-1.0 $\pm$ 0.2	-1.1 $\pm$ 0.1
G192/-/I	-0.1 $\pm$ 0.3	-0.1 $\pm$ 0.1	-0.7 $\pm$ 0.2	-0.9 $\pm$ 0.1
E206/-/S	-0.7 $\pm$ 0.5	0.1 $\pm$ 1.1	0.0 $\pm$ 0.1	-0.1 $\pm$ 0.1
R248/-/W	-0.2 $\pm$ 1.2	-0.2 $\pm$ 1.2	-0.8 $\pm$ 0.2	-1.1 $\pm$ 0.2
Sub-total $\Delta G_{\text{BINDING}}$	-1.4 $\pm$ 0.9	-3.0 $\pm$ 0.8	-4.5 $\pm$ 0.4	-4.9 $\pm$ 0.2
Sub-total $\Delta\Delta G_{\text{BINDING}}$	1.6 $\pm$ 1.2	-	-1.4 $\pm$ 0.8	-1.9 $\pm$ 0.8
$\Delta\Delta G_{\text{BIND-EXP}}$	0.5	-	-1.9	-0.7
$\beta$ -Chitobiose	-16.6 $\pm$ 1.2	-14.4 $\pm$ 1.1	-16.8 $\pm$ 1.5	-16.5 $\pm$ 1.3
Reducing GlcNAc	-10.6 $\pm$ 1.3	-7.9 $\pm$ 1.1	-11.7 $\pm$ 1.7	-11.5 $\pm$ 1.3
Terminal GlcNAc	-6.0 $\pm$ 1.1	-6.5 $\pm$ 1.1	-5.1 $\pm$ 1.2	-5.1 $\pm$ 1.2
Peptide	-4.9 $\pm$ 1.3	-6.5 $\pm$ 1.6	-4.3 $\pm$ 0.9	-4.0 $\pm$ 0.9

The R911 Dyn data indicates that the D60C mutation is -0.3 kcal/mol more favorable than D60A (GLH206). However, the  $\Delta\Delta G_{\text{BINDING}}$  relative to D60A (GLH206) indicates that R911 is -1.4 kcal/mol more favorable. The main favorable contributions are coming from G192I and R248W, increasing the hydrophobicity of the binding pocket. On the other hand D57L and E206S are estimated to make almost no favorable interaction energy contribution. This result might suggest that reverting these two residues back to wild-type may be more favorable. In the case of D57, this observation may be supported by simulation data that indicates that D57 is involved in stabilizing hydrogen bonds with

R61, retaining R61 in a favorable orientation to make direct substrate recognizing hydrogen bonds with the 2<sup>nd</sup> GlcNAc (Figures 4.18 and 4.19). Similarly, E206 interaction data from the 1PNF x-ray crystal model, shows that it is involved in hydrogen bonds with conserved water molecules (Wat<sup>346</sup> and Wat<sup>348</sup>) in the binding pocket and is not directly involved in substrate recognition (Figures 1.9 & 1.10).<sup>1</sup> Both Wat<sup>346</sup> and Wat<sup>348</sup> make direct hydrogen bonds with the reducing GlcNAc. Furthermore, the E206S mutation in R911 is a change from an acidic residue to a polar residue, which is favorable for protein-carbohydrate interactions. However, as these MD simulations were not conducted with conserved water molecules in the binding pocket, the estimated theoretical interaction energies for E206S are likely different than the estimates obtained in this simulation.

The  $\Delta\Delta G_{\text{BINDING}}$  relative interaction energy estimation of R911 (-1.4 kcal/mol) was less favorable than R911 C60A (-1.9 kcal/mol), relative to the D60A (GLH206) model. This is in contrast to experimental data ( $\Delta\Delta G_{\text{BIND-EXP}}$ ). However, these theoretical estimations are within calculated error, indicating the difference is not statistically significant. Given that the decomposition calculations approximate energetic contributions, it is not unusual to obtain MM-GBSA estimations with relatively high error, thus making qualitative assessment of data appropriate.

The energetic contribution of the  $\beta$ -Chitobiose and tripeptide moieties of the glycotriptide ligand were also determined from MM-GBSA analysis (Table 4.19). The majority of the favorable interactions are between the protein and the carbohydrate portion (-14 kcal/mol – -16 kcal/mol) of the glycotriptide relative to the peptide (-4 kcal/mol – -6 kcal/mol). Experimental data from 1PNF x-ray models shows a network of hydrogen bonds between the residues in the binding pocket and the chitobiose ligand

(Figures 1.9 & 1.10).<sup>1</sup> Substrate specificity studies of PNGase F indicate that catalytic activity of PNGase F with a glycotriptide substrate (Chitobiose-Asn-Ala-Thr) is 83%, and with a glycodipeptide substrate (Chitobiose-Asn-Ala) activity is 1.8%, where 100% activity is obtained with a pentapeptide substrate (Chitobiose-Try-Ile-Asn-Ala-Ser).<sup>7</sup> Thus, substrate specificity studies indicate that the peptide portion is critical. As these simulations utilized a glycotriptide ligand, based on the previously mentioned experimental data, it may be expected that more conclusive interaction energy results could be obtained with a glycopentapeptide ligand.

Table 4.20: Summary of R911 mutation theoretical and experimental characteristics. Characteristics of R911 mutations relative to wtPNGase F are listed. Estimated theoretical interaction energies (kcal/mol) relative to D60A (GLH206) complexed with a glycotriptide (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) are included. For comparison, experimentally determined binding interaction energy ( $\Delta\Delta G_{\text{BIND-EXP}}$ ) of clones is also listed relative to D60A.

R911 mutations relative to wtPNGase F	Amino Acid Mutation Characteristics	$\Delta\Delta G_{\text{BINDING}}$
D57L	Acidic to hydrophobic	$-0.1 \pm 0.1$
D60C	Acidic to polar	$-1.8 \pm 0.7$
I156L	Hydrophobicity preserved	$-1.0 \pm 0.2$
G192I	Non-polar to hydrophobic	$-0.7 \pm 0.2$
E206S	Acidic to polar	$0.0 \pm 0.1$
R248W	Basic to hydrophobic	$-0.8 \pm 0.2$
Sub-total $\Delta\Delta G_{\text{BINDING}}$		$-1.4 \pm 0.8$
$\Delta\Delta G_{\text{BIND-EXP}}$		$-1.9$

A summary of the R911 mutations, characteristics, and relative binding

interaction energies are provided in Table 4.20. MM-GBSA data indicates that 4 (D60C, I156L, G192I, & R248W) of the 6 mutagenized residues have favorable interaction energies relative to wtPNGase F ( $\Delta\Delta G_{\text{BINDING}}$ ), with D57L making weak contributions.

The MD trajectory of 1PNF (GLH206) indicates that D57 is involved in hydrogen bonds with R61 throughout the duration of the simulation and visualized in Figures 4.18 & 4.19. This interaction was not originally reported<sup>1</sup> (Figure 1.9) and re-analysis of 1PNF

experimental data confirms D57 hydrogen bonding with R61. This hydrogen bond is critical for holding R61 in place underneath the chitobiose ligand as it is involved in hydrogen bonds with the solvent exposed side of the *N*-acetyl group of the 2<sup>nd</sup> GlcNAc, keeping the 2<sup>nd</sup> half of the chitobiose ligand in place (Figure 1.10). R61 also hydrogen bonds with Wat<sup>349</sup>, which facilitating part of the larger hydrogen bond network on the protein interface side of the chitobiose ligand (Figure 1.10). These data indicate a previously unreported critical substrate-stabilizing role for D57 in wtPNGase F. Thus, mutation of this residue to D57L in R911 may be detrimental to substrate recognition. This is supported by MD trajectory data of R911 with the chitobiose ligand that shows the 2<sup>nd</sup> GlcNAc swinging outward from the binding cleft and adopting a more solvent exposed position. R61, no longer being held in place by D57 hydrogen bonds due to the D57L mutation, moves back into the binding cleft and facilitates hydrogen bonds with the protein facing side of the 2<sup>nd</sup> GlcNAc. These observations are also supported by the MM-GBSA interaction energy estimation for the R911 D57L mutation making negligible favorable contributions towards the interaction energy ( $-0.1 \pm 0.1$  kcal/mol). Thus reverting D57L back to wild-type can reasonably be expected to enhance substrate recognition and affinity.

In the case of R911 E206S, wild-type E206 experimental data indicates hydrogen-bonding interaction with conserved water molecules, which were not accounted for in the theoretical energy estimations. This may have contributed to the neutral interaction energy that was estimated ( $0.0 \pm 0.1$  kcal/mol) for the E206S mutation. Over all, the theoretical sub-total  $\Delta\Delta G_{\text{BINDING}}$  interaction energies for R911 ( $-1.4 \pm 0.8$  kcal/mol) reproduced the experimentally determined value ( $-1.9$  kcal/mol).

### 4.3 MD Simulations and Binding Free Energy Decomposition of PNGase F Clones with a Ser-*O*-GlcNAc Ligand

The enrichment of *O*-GlcNAcylated glycoproteins by R911 Lectenz® affinity chromatography was unexpected (Section 3.4.4) given that R911 is derived from the *N*-glycan processing enzyme PNGase F, which has defined substrate specificity for the core *N*-glycopeptide.<sup>1, 7, 13-15</sup> Structural models of 1PNF, D60A, and R911 were constructed with the common *O*-GlcNAc motif (GlcNAc- $\beta$ -Ser) in the binding pocket and utilized for 50 ns MD simulations and MM-GBSA analysis (Table 4.21). In order to neutralize terminal charges ( $NH_3^+$  for the N-terminal and  $COO^-$  for the C-terminal) the serine residue of the GlcNAc- $\beta$ -Ser ligand was modeled with the N-terminal ACE [ $-C(=O)-CH_3$ ] protecting group and the C-terminal NME [ $-C(=O)-NH-CH_3$ ] protecting group as defined in xleap, a component of AMBER Tools 13.<sup>10</sup>

Table 4.21: Structural models of wtPNGase F and clones used for MD simulations and MM-GBSA. Structural models with a Ser-*O*-GlcNAc ligand in the binding pocket were constructed to conduct 50 ns MD simulations. Simulation stability was confirmed by analyzing RMSD over simulation time and the average RMSD values are listed. Estimated total theoretical binding energy for all 314 amino acids (Total  $\Delta G_{\text{BINDING}}$ ) of each clone are listed (kcal/mol).

Structural Model	Ligand	RMSD (Å)	Total $\Delta G_{\text{BINDING}}$
1PNF (GLH206)	GlcNAc- $\beta$ -Ser	1.2584	-47.2
D60A (GLH206)	GlcNAc- $\beta$ -Ser	1.2090	-46.0
R911 Dyn	GlcNAc- $\beta$ -Ser	1.3029	-42.6

Structural equilibrium was confirmed by the low average RMSD computed during the course of the 50 ns MD simulation, consistent with previous models used in this study. Interaction energies were computed at 1 ns intervals over the duration of the 50 ns MD trajectory for each model (Figures 4.20 – 4.22). Unsurprisingly, the interaction energy for the 1PNF (GLH206) complex with the *O*-GlcNAc glycopeptide remained unstable during the entire 50 ns trajectory (Figure 4.20). The observed instability is likely

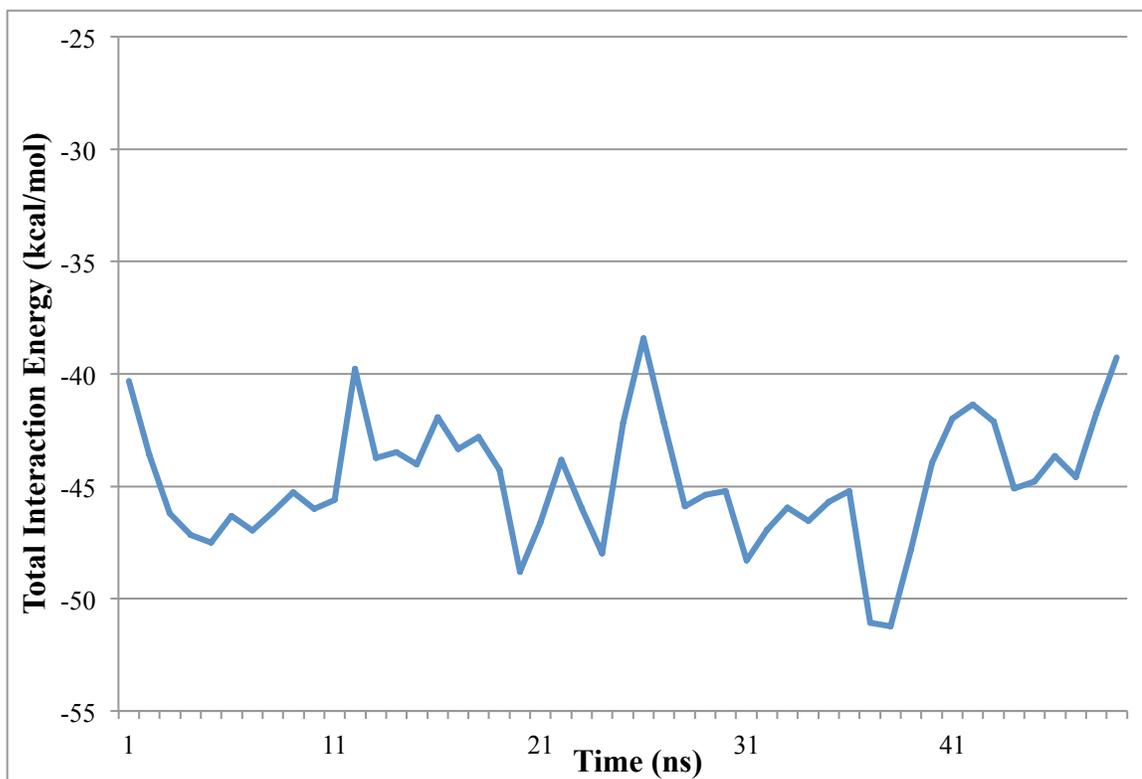


Figure 4.20: Interaction energy stability during 50 ns 1PNF (GLH206) MD simulation with an *O*-GlcNAc peptide ligand. The portion of the trajectory between 29 ns – 39 ns) was selected for MM-GBSA analysis.

a confirmation of experimental data that indicate wtPNGase F specificity for *N*-

glycopeptides. However, it is important to note that the ligand remains in the binding site during the trajectory facilitated by the extension of the *N*-acetyl group into the same hydrophobic pocket as observed in the wtPNGase F experimental structure (Figure 1.8b), indicating that the common *N*-Acetyl group on both the wild-type chitobiose and the *O*-GlcNAc ligands are important for recognition. In addition, a majority of the instability appears to come from the ACE-Ser-NME peptide portion of the *O*-GlcNAc glycopeptide ligand based on the rapid conformation changes visually observed during the 50 ns trajectory, proximal to residue D60. Unlike the 1PNF (GLH206) model, the D60A (GLH206) model had stable interaction energy throughout the 50 ns MD trajectory (Figure 4.21), indicating that the D60 residue is responsible for the interaction energy

instability observed during the 1PNF (GLH206) MD simulation (Figure 4.20) as mutation of this residue to D60A resulted in stabilized interaction energy.

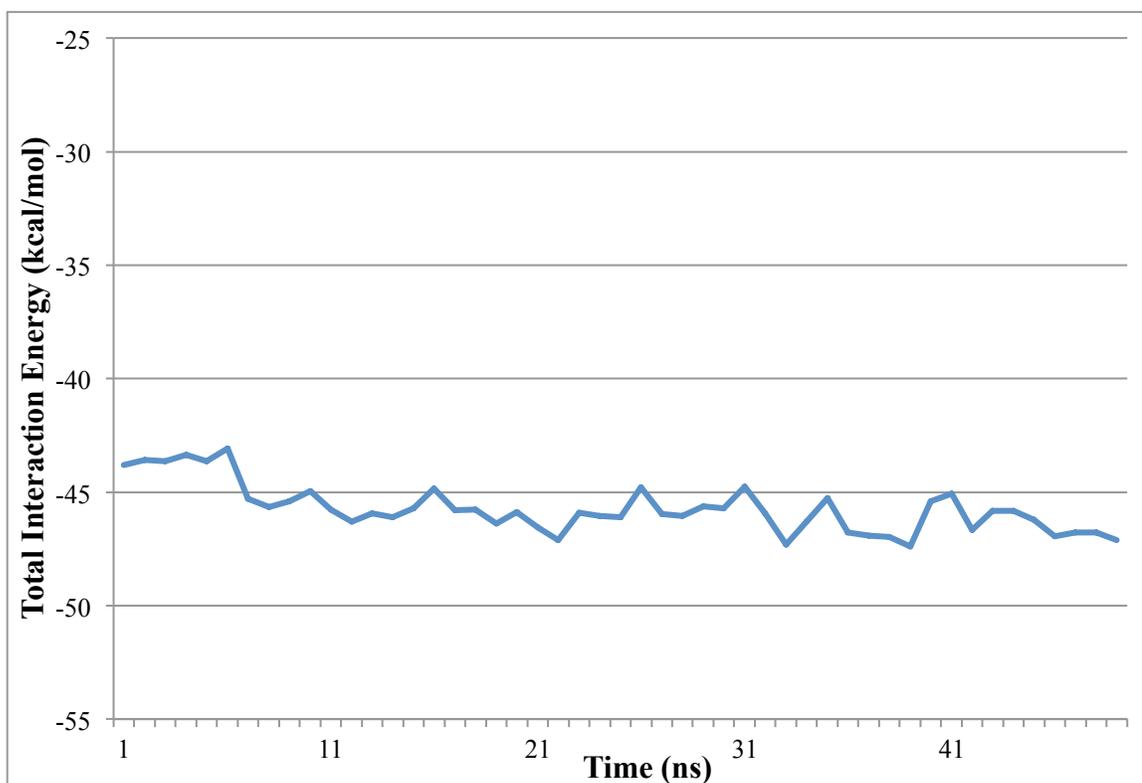


Figure 4.21: Interaction energy stability during 50 ns 1PNFD60A (GLH206) MD simulation with an *O*-GlcNAc peptide ligand. The last portion of the trajectory (39 ns – 49 ns) was selected for MM-GBSA analysis.

The R911 Dyn model of the complex had regions of varied stable interaction energies most notable between 17 ns – 30 ns and again between 39 ns – 50 ns as shown in Figure 4.22. A key difference between these two regions is the ligand confirmation in the earlier time points was altered due to the lack of hydrogen bond interaction between R61 and GlcNAc-O4. Once the R61 and GlcNAc-O4 hydrogen bond is formed starting at 39 ns, the complex adopts a more stable conformation as evidence by the favourable interaction energy during the last 11 ns of MD simulation and by visual analysis of the trajectory (Figure 4.23). A list of the theoretical hydrogen bonds between the R911 and

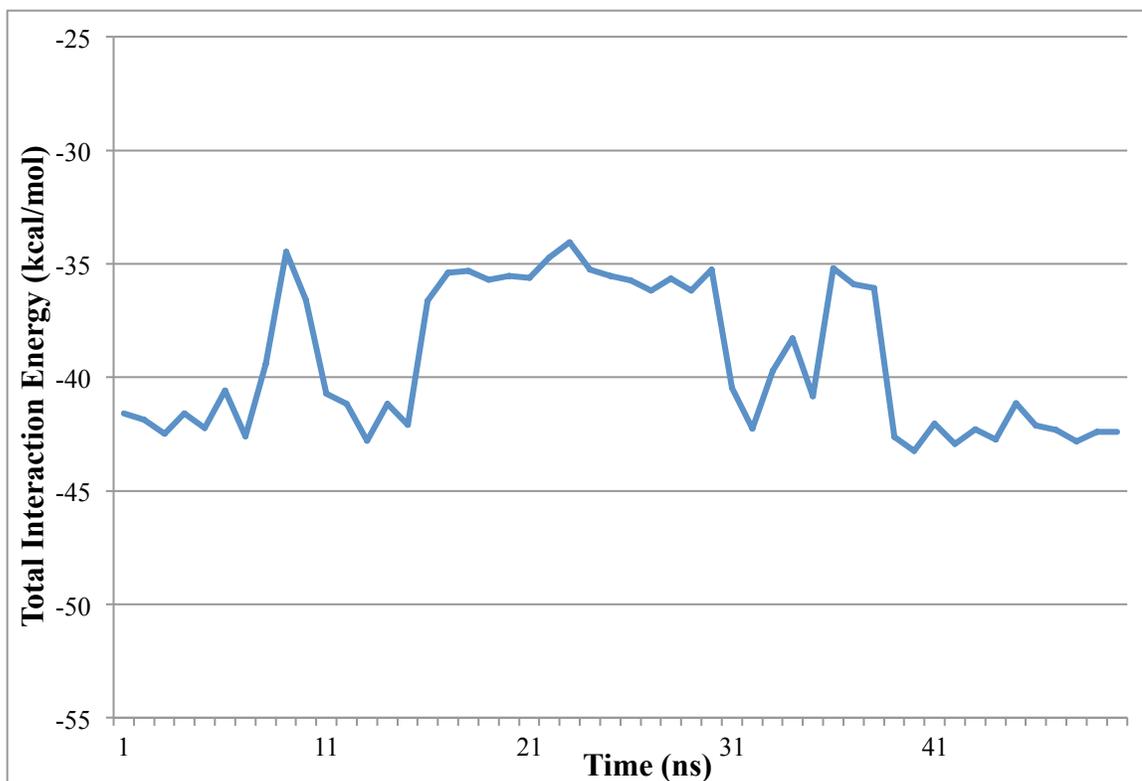


Figure 4.22: Interaction energy stability during 50 ns R911 Dyn MD simulation with an *O*-GlcNAc peptide ligand. The last portion of the trajectory (39 ns – 49 ns) was selected for MM-GBSA analysis.

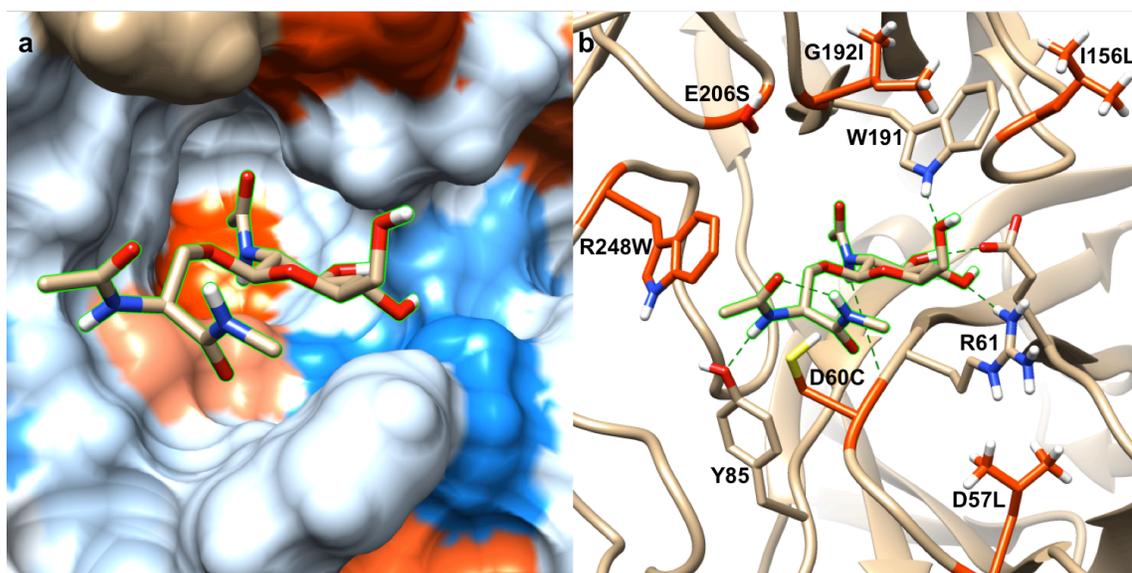


Figure 4.23: R911 and GlcNAc- $\beta$ -Ser MD simulation at 45 ns time point. The GlcNAc- $\beta$ -Ser ligand is outlined in green. a) Surface hydrophobicity representation of the R911-GlcNAc- $\beta$ -Ser complex depicting the ligand in the binding pocket with the *N*-Acetyl group extended into the deep hydrophobic groove. b) Theoretical hydrogen bonds are depicted as dashed green lines. R911 mutagenized residues are depicted in orange.

*O*-GlcNAc ligand is provided in Table 4.22 and depicted in Figure 4.23b. Due to these observations, the 10 ns trajectory between time points 39 ns – 49 ns of the R911 Dyn and D60A (GLH206) trajectories were selected for MM-GBSA analysis. In the case of the 1PNF (GLH206) complex, the most stable region of the trajectory was selected for MM-GBSA analysis (29 ns – 39 ns). However, even this region of the trajectory is relatively unstable and the MM-GBSA data should be considered less than optimal. A longer 100 ns MD simulation did not result in the 1PNF (GLH206) – *O*-GlcNAc complex adopting a more energetically stable conformation. Similarly, no significant changes in conformation were observed for D60A (GLH206) or R911 Dyn models complexed with *O*-GlcNAc when the MD simulation was extended to 100 ns.

Table 4.22: Theoretical hydrogen bond lengths between GlcNAc- $\beta$ -Ser and R911.

Hydrogen bonds	Average from R911 Ser- <i>O</i> -GlcNAc MD Simulation (Å)
D60-O – GlcNAc316-H2N	3.09 $\pm$ 0.22
R61-HH11 – GlcNAc318-O4	2.84 $\pm$ 0.11
E118-O $\epsilon$ – GlcNAc318-H3O	2.69 $\pm$ 0.12
W191-NH – GlcNAc318-O6	3.02 $\pm$ 0.20
Y85-OH – S316-H	3.06 $\pm$ 0.16

The estimated binding free energies for the models with the GlcNAc- $\beta$ -Ser complex are presented in Table 4.23. As observed with the interaction energy and trajectory visualization, the MM-GBSA data also indicates the relative unfavourable interaction energy of D60 in the 1PNF (GLH206) complex ( $-0.5 \pm 1.6$  kcal/mol) relative to the D60A (GLH206) complex ( $-1.6 \pm 0.4$  kcal/mol). This is also supported by the larger estimated error computed for the D60 residue in the 1PNF (GLH206) complex. The estimate interaction energy of the D60C mutation in the R911 Dyn complex is relatively favourable ( $-3.2 \pm 0.7$ ). Unlike the D60C mutation, the E206S mutation is

noticeably unfavorable; however, this may in part be due to an inaccurate under estimation as the conserved water molecule that is observed to interact with this site is not included in the MM-GBSA per residue estimation. Nonetheless, it is conceivable that reverting E206S to wild-type may also enhance ligand recognition specific for GlcNAc- $\beta$ -Ser given the noticeably favorable interaction energy estimated in the D60A (GLH206) for E206 ( $-1.6 \pm 0.7$  kcal/mol).

Table 4.23: Estimated MM-GBSA theoretical interaction action energies of models complexed with an *O*-GlcNAc ligand. Sub-total  $\Delta G_{\text{BINDING}}$  (kcal/mol) of mutagenized residues were compared across 1PNF (GLH206), D60A (GLH206), and R911 Dyn complexed with (GlcNAc- $\beta$ -Ser). Sub-total  $\Delta\Delta G_{\text{BINDING}}$  (kcal/mol) energies relative to the D60A control clone are indicated.

Residue	1PNF (GLH206)	D60A (GLH206)	R911 Dyn
D57/-/L	$-0.3 \pm 0.8$	$-0.4 \pm 0.5$	$-0.1 \pm 0.0$
D60/A/C	$-0.5 \pm 1.6$	$-1.6 \pm 0.4$	$-3.2 \pm 0.7$
I156/-/L	$-0.1 \pm 0.1$	$-0.2 \pm 0.0$	$-0.1 \pm 0.0$
G192/-/I	$-0.1 \pm 0.2$	$-0.1 \pm 0.1$	$-0.4 \pm 0.2$
E206/-/S	$-1.7 \pm 1.1$	$-1.6 \pm 0.7$	$-0.1 \pm 0.2$
R248/-/W	$-0.1 \pm 1.2$	$-2.3 \pm 0.8$	$-0.9 \pm 0.2$
Sub-total $\Delta G_{\text{BINDING}}$	$-3.0 \pm 1.1$	$-6.2 \pm 0.6$	$-4.9 \pm 0.4$
Sub-total $\Delta\Delta G_{\text{BINDING}}$	$3.2 \pm 1.2$	-	$1.3 \pm 0.7$
$\beta$ -GlcNAc	$-17.6 \pm 1.8$	$-14.7 \pm 1.6$	$-16.4 \pm 1.5$
Serine	$-4.9 \pm 1.2$	$-5.7 \pm 0.8$	$-2.9 \pm 0.6$

The I156 site is estimated to make minimal favorable interaction contributions. This is unsurprising as this site is more critical for interaction with the second GlcNAc of the wild-type chitobiose ligand, which is absent in GlcNAc- $\beta$ -Ser ligand. However, the protein loop region of the I156 site may be important for modification via extension to improve specificity toward the GlcNAc- $\beta$ -Ser ligand by blocking access of the 2<sup>nd</sup> GlcNAc in chitobiose to the binding site. The D57L mutation is also making negligible favorable interactions and destabilizing R61 hydrogen bond interactions as observed previously in the R911 chitobiose complex. As discussed previously (Section 4.2.2),

experimental and modeling data indicate that R61 is important for substrate recognition mediated by hydrogen bonds with the *N*-Acetyl group of the 2<sup>nd</sup> GlcNAc residue of the chitobiose. In the R911 Dyn complex with *O*-GlcNAc glycopeptide (GlcNAc- $\beta$ -Ser), R61 hydrogen bonds with GlcNAc-O4 towards the last portion of the trajectory. These observations indicate that reverting D57L to wild-type would likely enhance substrate recognition as well as specificity.

Modeling data with the *O*-GlcNAc glycopeptide (GlcNAc- $\beta$ -Ser) provide a rationalization for the experimentally observed enrichment of *O*-GlcNAcylated glycoproteins. Furthermore, specificity towards *O*-GlcNAcylated glycoproteins may be enhanced by reverting both E206S and D57L to wild-type and extending the loop region of I156L to block larger chitobiose ligands from the binding pocket. In addition, the modus of *O*-GlcNAc recognition by R911 appears to be driven by highly favorable interaction with the reducing GlcNAc ( $-16.4 \pm 1.5$  kcal/mol) relative to the serine residue ( $-2.9 \pm 0.6$  kcal/mol), consistent with observations of favorable interactions with the reducing GlcNAc in the wild-type chitobiose ligand interactions as reported in Table 4.19 and indicated in experimental data of the wtPNGase F complex.<sup>1</sup>

## 4.4 Methods

### 4.4.1 Building D60A and R911 Structures from 1PNF

The 1PNF x-ray structural model was used as the base model from which all other mutagenized PNGase F models were constructed using USCF Chimera v1.8.1.<sup>1,16</sup> Dunbrack and Dynameomics rotamer libraries were utilized to selected preferred rotamers for modeling and editing into the model using USCF Chimera's rotamer selection and torsion angle tools.<sup>2,3</sup> Six models of PNGase F with the  $\alpha$ -chitobiose ligand (GlcNAc $\beta$ 1-4GlcNAc- $\alpha$ OH) were constructed as listed in Table 4.3. Four models of

PNGase F with the  $\beta$ -chitobiose ligand (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ OH) were constructed as listed in Table 4.8. Four models of PNGase F with the glycotriptide ligand (GlcNAc $\beta$ 1-4GlcNAc- $\beta$ -Asn-Leu-Thr) were constructed as listed in Table 4.14. Three models with GlcNAc- $\beta$ -Ser ligand were constructed as listed in Table 4.21. In order to neutralize terminal charges ( $NH_3^+$  for the N-terminal and  $COO^-$  for the C-terminal) the peptide portion of all glycopeptide ligands were modeled with the N-terminal ACE [ $-C(=O)-CH_3$ ] protecting group and the C-terminal NME [ $-C(=O)-NH-CH_3$ ] protecting group as defined in xleap, a component of AMBER Tools 13.<sup>10</sup>

#### 4.4.2 MD Simulations

A 100 ns fully solvated MD simulation of the PNGase F – ligand complex was performed in water at room temperature and pressure (NPT) employing the AMBER-GLYCAM protein-carbohydrate force field. The system was minimized with implicit solvent (5000 steps) using a system restraint mask (protein  $C\alpha$  and ligand ring atoms) to permit all modeled rotamers into energy-minimized conformations. Using tleap the system was explicated solvated with the TIP3P water model. The explicitly solvated system was then energy minimized (2000 steps, NVT) using a system restraint mask. A 30 ps heating step was performed (NVT) also with a system restrain mask, followed by a 1 ns equilibration (NPT) with a ligand restraint only (ligand ring atoms). The 100 ns production run was performed (50,000,000 steps, NPT) and data was saved at ever 0.002 ps, corresponding to 500 frames saved per ns of data. Trajectory analysis was performing using tleap, ptraj, and cpptraj as implemented in AMBER Tools13.<sup>10, 17</sup> Data was visualized using USCF Chimera 1.8.1.<sup>16</sup>

#### 4.4.3 Binding Free Energy Decomposition

The per-residue contributions to the binding energy was computed for each of the 314 amino acids in PNGase F, employing the generalized Born (GB) continuum solvent model as implemented in AMBER as previously described in section 2.4.1 over the energy converged portion of the trajectory.<sup>18</sup>

#### 4.5 References

1. Kuhn, P. et al. Active Site and Oligosaccharide Recognition Residues of Peptide-N4-(N-acetyl- $\beta$ -D-glucosaminyl)asparagine Amidase F. *Journal of Biological Chemistry* **270**, 29493-29497 (1995).
2. Dunbrack, R.L., Jr. Rotamer libraries in the 21st century. *Current Opinion in Structural Biology* **12**, 431-440 (2002).
3. Scouras, A.D. & Daggett, V. The dynamomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Science* **20**, 341-352 (2011).
4. Woods, R.J. & Tessier, M.B. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Current Opinion in Structural Biology* **20**, 575-583 (2010).
5. Genheden, S. & Ryde, U. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys. Chem. Chem. Phys.* **14**, 8662-8677 (2012).
6. Hadden, J.A., Tessier, M.B., Fadda, E. & Woods, R.J. Calculating binding free energies for protein-carbohydrate complexes. (2012).
7. Fan, J.Q. Detailed Studies on Substrate Structure Requirements of Glycoamidases A and F. *Journal of Biological Chemistry* **272**, 27058-27064 (1997).
8. Filitcheva, J. PNGases: A Diverse Family of Enzymes Related by Function Rather Than Catalytic Mechanism, Vol. Ph.D. (Massey University, Palmerston North; 2010).
9. Isom, D.G., Castaneda, C.A., Cannon, B.R., Velu, P.D. & Garcia-Moreno, E.B. Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci U S A* **107**, 16096-16100 (2010).
10. D.A. Case, T.A.D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X.

Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (University of California, San Francisco, 2012).

11. Rye, C.S. & Withers, S.G. Glycosidase mechanisms. *Curr Opin Chem Biol* **4**, 573-580 (2000).
12. Kozmon, S. & Tvaroška, I. Catalytic Mechanism of Glycosyltransferases: Hybrid Quantum Mechanical/Molecular Mechanical Study of the Inverting N-Acetylglucosaminyltransferase I. *Journal of the American Chemical Society* **128**, 16921-16927 (2006).
13. Tretter, V., Altmann, F. & MÄRz, L. Peptide-N4-(N-acetyl- $\beta$ -glucosaminyl)asparagine amidase F cannot release glycans with fucose attached  $\alpha 1 \rightarrow 3$  to the asparagine-linked N-acetylglucosamine residue. *European Journal of Biochemistry* **199**, 647-652 (1991).
14. Mussar, K.J., Murray, G.J., Martin, B.M. & Viswanatha, T. Peptide: N-glycosidase F: studies on the glycoprotein aminoglycan amidase from *Flavobacterium meningosepticum*. *Journal of biochemical and biophysical methods* **20**, 53-68 (1989).
15. Tarentino, A.L., Gomez, C.M. & Plummer, T.H., Jr. Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24**, 4665-4671 (1985).
16. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612 (2004).
17. Roe, D.R. & Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **9**, 3084-3095 (2013).
18. Tsui, V. & Case, D.A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **56**, 275-291 (2001).

## CHAPTER 5

### DISCUSSION AND FUTURE PERSPECTIVES

For efficient generation and screening of biocombinatorial libraries, it is important to limit the library to approximately  $10^9$  clones<sup>1</sup>, which corresponds to 7 randomized positions ( $20^7 = 1 \times 10^9$  clones). When it is not immediately known which residues should be changed, library design can be difficult. This is where input from computational simulations can aid in identifying the appropriate amino acids and thereby focusing the library design. The benefits of computational guidance, particularly in terms of the reduction in the number of potential clones, has been noted.<sup>2</sup> As observed in a recent review by Barakat and Love<sup>3</sup>, computational algorithms blended with *in vivo* screens are leading towards greater and more rapid success in the field of protein design.

Here the computationally-focused yeast display library generated by GeneArt (Library 2) had a diversity of  $\sim 1.36 \times 10^7$  clones, representing an estimated sequence coverage of  $\sim 22\%$  of the theoretical diversity (Table 2.4). Selection via MACS prior to FACS served to ensure that the library was sufficiently enriched prior to using FACS as a stringent selection pressure for the practical sorting of functionally relevant clones (Figures 2.9 and 2.11a). The library was selected against a mixture of representative *N*-glycan targets on RNase B and Asialofetuin to enrich for clones that retained the cognate specificity of the PNGase F enzyme.

An examination of the R911 protein sequence indicated enrichment in residues that are commonly found in protein-glycan interactions. The hydrophobic face of

carbohydrates frequently participates in stacking interactions with aromatic amino acids, which are estimated to contribute 1.5 kcal/mol.<sup>4</sup> It was notable then that selection led to the introduction of a Trp at position 248 that is estimated to favorably contribute  $-0.8 \pm 0.2$  kcal/mol based on MM-GBSA analysis. In addition, several other mutations increased the overall hydrophobicity relative to the wt sequence (D57L, G192I, D60C and E206S) (Table 4.20).

The R248W mutation is of particular interest not only because of its known importance in facilitating carbohydrate-aromatic interactions, but also because of R248's proposed role in the catalytic mechanism of wtPNGase F (making the Asn-carbonyl atom more susceptible to nucleophilic attack) and interaction with Wat<sup>346</sup> (Figures 1.9 & 1.10) in the catalytic site.<sup>5,6</sup> The R248A point mutant has 0.1% catalytic activity relative to the wtPNGase F enzyme.<sup>5</sup> R911's lack of catalytic activity (Table 3.3) may be attributed in part to the R248W mutation. Thus the R248W mutation may not only be enhancing affinity but also contributing to the catalytic inactivation of R911.

Wild-type E206 and D60 span the glycosidic linkage between asparagine and the reducing GlcNAc, are known to participate in hydrogen bonding interactions with conserved water molecules (Wat<sup>346</sup> and Wat<sup>348</sup>) in the x-ray crystal structure 1PNF (Figure 1.9 & 1.10), and contribute to catalytic activity.<sup>6</sup> Thus, it is important to note that the polar mutations of the catalytic residues (D60C and E206S) in R911 are likely also contributing to catalytic inactivity, but potentially preserving the hydrogen bond network that is critical for substrate recognition. MM-GBSA energy interaction analysis with conserved water molecules may provide additional insight into E206S and D60C roles in R911.

MD analysis of the wtPNGase F complexed with the glycotriptide indicated that D57 is important for stabilizing R61 through hydrogen bond interactions, thereby facilitating substrate recognition between R61 and the 2<sup>nd</sup> GlcNAc. This type of stabilizing interaction has been reported between S155 and E118 residues in the 1PNF x-ray crystal structure (Figures 1.9 & 1.10).<sup>6</sup> Like R61, E118 directly interacts with conserved water molecules in the binding cleft as well as the 2<sup>nd</sup> GlcNAc. The orientation of E118 is stabilized via hydrogen bond interactions with S155. Reverting D57L to wild-type in R911 may improve substrate recognition via R61 and enhance affinity.

Free energy decomposition analysis offers a powerful tool to investigate the per residue interaction energy, for which there is no equivalent experimental method. The total interaction energy of -44.0 kcal/mol was computed for the 1PNF (GLH206) MD simulation with the glycotriptide ligand (Table 4.14). This value overestimates the experimental binding free energy of -7.103 kcal/mol for wtPNGase F (Table 3.4), which is a typical feature of MM-GBSA calculations that omit entropic penalties associated with ligand binding.<sup>7</sup> Entropic effects, arising from changes in conformational flexibility can be estimated, but may require very long MD simulations in order to achieve convergence.<sup>8</sup> However, it may be anticipated that entropic effects arising from reduction in the flexibility of protein side chains will be most significant for those residues that interact strongly with the ligand, and least significant for the tepid or cold residues. For these reasons, the entropic contributions were not computed.

Unlike wtPNGase F and D60A, the expression and purification of R911 and R911 C60A clones resulted in a low yield (~150 µg/L). IMAC and SEC elution profiles of these clones differed from wtPNGase F and D60A. Furthermore, Western Blot analysis

of multiple R911 SEC elution peaks suggests the presence of structural isoforms of R911 clones, some of which may be mis-folded R911 clones. Protein mutational tolerance, the risk of multiple mutations decreasing protein stability, is a common issue with protein library design, which can be compensated for by the use of appropriate selection parameters especially when selecting for enhanced enzyme thermostability or activity.<sup>9</sup> However, for affinity enhancement, these challenges persist and recent efforts to minimize destabilizing mutations has led to the development of protein folding algorithms to pre-screen sequence space for stabilizing effects.<sup>9</sup>

SPR kinetic data demonstrates that R911 has sub-micromolar affinity ( $K_D = 0.26 \mu\text{M}$ ) for the *N*-glycan bearing glycoprotein RNase B, a 10x affinity enhancement relative to the non-affinity optimized D60A control. R911 also has an 84x decreased off-rate ( $k_{\text{off}} = 5.1 \times 10^{-3} \text{ s}^{-1}$ ). Where as R911 C60A exhibits relatively lower affinity and decreased off-rate, indicating that D60C R911 mutation makes a critical contribution to the binding interactions, which are further enhances by overall synergistic effects from other mutations. Importantly, the kinetic data satisfies the Lectenz® selection threshold for enhanced affinity and decreased off-rate relative as depicted in the design strategy (Figure 1.9).

The application of the R911 Lectenz® in affinity chromatography demonstrated enrichment of the *N*-glycan bearing glycoprotein RNase B as well as of *N*-glycopeptides derived from RNase B. Furthermore, the lack of enrichment of deglycosylated RNase B and the competitive elution with chitobiose, demonstrates that R911 recognizes the common chitobiose glycopeptide core of *N*-glycan structures. This is consistent with the observed specificity of the wtPNGase F enzyme and the D60A glycan array screening

results. Nonetheless, the glycan specificity of R911 will be further investigated by glycan array screening.

The application of R911 Lectenz® Affinity Chromatography (LAC) for the enrichment of native glycoproteins from MCF7 cell extract by competitive elution resulting in the 3.4x enrichment of both *N*-glycoproteins and *O*-GlcNAcylated *O*-glycoproteins that share a common reducing GlcNAc recognized by R911. This is significant as the R911 Lectenz® is the only known reagent that recognizes both the common chitobiose core of *N*-glycans and *O*-glycoproteins containing the common core *O*-GlcNAcylation motif, making possible the enrichment of two major classes of glycoproteins using a single affinity reagent. Furthermore, in comparison to Multi-Lectin Affinity Chromatography (MLAC) with Jac, ConA, and WGA lectins, R911 LAC resulted in the enrichment of glycoproteins not enriched by MLAC.<sup>10</sup> The difference in the glycoprotein enrichment profiles is not surprising given the different specificities of the capture reagents employed as glycan detection is biased by the type of lectin employed in affinity chromatograph.<sup>11</sup> Not surprisingly, some non-glycoproteins were also identified in the eluted sample from R911 LAC. Another weakness of sample enrichment by affinity chromatography is false positives that results from proteins being captured by non-specific protein-protein interactions other than direct affinity selection of a targeting ligand.<sup>12</sup>

This first-of-its-kind application of biocombinatorial library design based on free energy decomposition for the engineering of a carbohydrate processing enzyme into a catalytically inactive, high affinity capture reagent generally confirms the Lectenz® design strategy and highlights the challenges associated with protein engineering. These

studies indicate that the R911 Lectenz® can be further enhanced by selective mutagenesis to create two additional Lectenz® reagents, one specific for *N*-glycopeptides and *N*-glycoproteins, and a second Lectenz® reagent specific for *O*-GlcNAcylated glycoproteins and glycopeptides. An important next step would be to revert the D57L mutation to wild-type as this is likely the most effective way to enhance substrate specificity for *N*-glycoproteins. An *O*-GlcNAc specific Lectenz® could be engineering by reverting both E206S and D57L to wild-type and extending the loop region of I156L to block larger chitobiose ligands from the binding pocket.

The research presented here also lays the groundwork for the development of 2<sup>nd</sup> generation biocombinatorial libraries for the exploration of alternative sequence spaces for Lectenz® generation. Based on literature reports and the data reported here, Table 5.1 lists the proposed roles of the critical residues identified in the binding cleft of wtPNGase F. This list represents an enhancement of understanding the substrate recognition by PNGase F to guide development of additional Lectenz® candidates. Another critical factor that would advance development would be generation of experimental structural data of a glycotriptide or glycopentapeptide complexed with PNGase F. A complex with the D60A single point mutant developed in this study, which has significantly diminished catalytic activity, would be an equally useful structure to use for Lectenz® engineering. However, the lack of results in obtaining such data over the past 20 years is an indication of the challenges of obtaining experimental structural data.

Table 5.1: Proposed functions of PNGase F active site residues. Residues that impact catalytic activity based on point mutant studies and are proposed as part of the catalytic mechanism are indicated in bold.

IPNF Residue	Proposed Function	Interactions	Contact w/ AA or Ligand
D57	Stabilizing	H-bond w/ R61 (MD)	R61
W59	Impacts catalytic activity <sup>5</sup>	Hydrophobic environment	D60
<b>D60</b>	Catalytic mechanism <sup>5, 6</sup>	H-bond w/ Wat <sup>346</sup> & ligand	1 <sup>st</sup> GlcNAc
R61	Substrate binding/ recognition <sup>5, 6</sup>	H-bond w/ ligand	2 <sup>nd</sup> GlcNAc
Y62	Stabilizing <sup>6, 13</sup>	H-bond w/ Wat <sup>146</sup> w/ N152	N152
I82	Impacts catalytic activity <sup>5</sup>	Hydrophobic environment	D60
Y85	Stabilizing <sup>6</sup>	H-bond w/ Wat <sup>346</sup>	1 <sup>st</sup> GlcNAc
E118	Substrate binding/ recognition <sup>6</sup>	H-bond w/ Wat <sup>349</sup> & ligand	2 <sup>nd</sup> GlcNAc
W120	Substrate binding/ recognition <sup>5</sup>	H-bond w/ Wat <sup>349</sup> & ligand, and potential hydrophobic interaction predicted w/ 1 <sup>st</sup> mannose	2 <sup>nd</sup> GlcNAc & 1 <sup>st</sup> mannose?
S155	Stabilizing <sup>6</sup>	H-bond w/ E118	E118
I156	Stabilizing	Potential hydrophobic interaction predicted w/ 1 <sup>st</sup> mannose (MD)	1 <sup>st</sup> mannose?
W191	Substrate binding/ recognition <sup>6</sup>	H-bond w/ ligand	1 <sup>st</sup> GlcNAc
G190	Stabilizing <sup>6</sup>	H-bond w/ Wat <sup>75</sup> & Wat <sup>348</sup>	1 <sup>st</sup> GlcNAc
H193	Substrate binding/ recognition <sup>5</sup>	-	-
<b>E206</b>	Catalytic mechanism <sup>5, 6</sup>	H-bond w/ Wat <sup>346</sup> & Wat <sup>348</sup>	1 <sup>st</sup> GlcNAc
W207	Impacts catalytic activity <sup>5</sup>	Hydrophobic environment and H-bond with Asn-O (MD)	E206 & Asn-O
<b>R248</b>	Potentially involved in catalytic mechanism <sup>5</sup>	Electrostatic and H-bond w/ Wat <sup>346</sup>	1 <sup>st</sup> GlcNAc
W251	Impacts catalytic activity <sup>5</sup>	Hydrophobic environment	E206

The R911 Lectenz® has been selected using the computationally-guided design of a yeast-surfaced displayed PNGase F biocombinatorial library. The R911 Lectenz® is a

novel pan-specific reagent for detecting the core glycopeptide component common to all *N*-linked glycans and core *O*-GlcNAcylated glycoproteins. This application of the Lectenz® design strategy presents opportunities to engineer additional Lectenz® reagents from carbohydrate-processing enzymes with glycan specificity and enhanced affinity. Lectenz® reagents will thus complement the use of existing carbohydrate-recognizing lectins and antibodies and can be employed in sample enrichment applications like affinity chromatography. The utility of Lectenz® reagents in other applications like glycan detection arrays, FACS and Multiplexed Suspension Arrays, immunohistochemistry, and bioprocess monitoring will be investigated further.

## 5.1 References

1. Bonsor, D.A. & Sundberg, E.J. Dissecting protein-protein interactions using directed evolution. *Biochemistry* **50**, 2394-2402 (2011).
2. Voigt, C.A., Mayo, S.L., Arnold, F.H. & Wang, Z.G. Computationally focusing the directed evolution of proteins. *J Cell Biochem Suppl* **37**, 58-63 (2001).
3. Barakat, N. & Love, J. Molecular Diversity in Engineered Protein Libraries. *Curr Opin Chem Biol* **11**, 335-341 (2007).
4. Asensio, J.L., Ardá, A., Cañada, F.J. & Jiménez-Barbero, J. Carbohydrate–Aromatic Interactions. *Accounts of Chemical Research* **46**, 946-954 (2012).
5. Filitcheva, J. PNGases: A Diverse Family of Enzymes Related by Function Rather Than Catalytic Mechanism, Vol. Ph.D. (Massey University, Palmerston North; 2010).
6. Kuhn, P. et al. Active Site and Oligosaccharide Recognition Residues of Peptide-N4-(N-acetyl- $\beta$ -D-glucosaminyl)asparagine Amidase F. *Journal of Biological Chemistry* **270**, 29493-29497 (1995).
7. Woods, R.J. & Tessier, M.B. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Current Opinion in Structural Biology* **20**, 575-583 (2010).
8. Genheden, S. & Ryde, U. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys. Chem. Chem. Phys.* **14**, 8662-8677 (2012).

9. Socha, R.D. & Tokuriki, N. Modulating protein stability - directed evolution strategies for improved protein function. *Febs J* **280**, 5582-5595 (2013).
10. Lee, L.Y. et al. An optimized approach for enrichment of glycoproteins from cell culture lysates using native multi-lectin affinity chromatography. *J Sep Sci* **35**, 2445-2452 (2012).
11. Krishnamoorthy, L. & Mahal, L.K. Glycomic analysis: an array of technologies. *ACS chemical biology* **4**, 715-732 (2009).
12. Jung, K. & Cho, W. Serial affinity chromatography as a selection tool in glycoproteomics. *Anal Chem* **85**, 7125-7132 (2013).
13. Kuhn, P., Tarentino, A.L., Plummer, T.H., Jr. & Van Roey, P. Crystal structure of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase F at 2.2-A resolution. *Biochemistry* **33**, 11699-11706 (1994).