

TOOLS AND METHODS FOR BIOCHEMICAL AND BIOPHYSICAL ANALYSIS OF O-GLCNAC-MODIFIED PROTEINS

by

Jesús Emanuel Serrano Negrón

M.Sc., University of Puerto Rico, 2017

B.Sc., University of Puerto Rico, 2015

**THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

in the

Department of Molecular Biology and Biochemistry

Faculty of Science

© Jesús Emanuel Serrano Negrón 2025

SIMON FRASER UNIVERSITY

Spring 2025

Copyright in this work is held by the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Declaration of Committee

Name: **Jesús Emanuel Serrano Negrón**

Degree: **Doctor of Philosophy**

Title: **Tools and methods for biochemical and biophysical analysis of O-GlcNAc-modified proteins**

Committee:

Chair: **Sharon M. Gorski**
Professor, Molecular Biology and Biochemistry

David J. Vocadlo
Supervisor
Professor, Chemistry, and Department of Molecular Biology and Biochemistry

Esther M. Verheyen
Committee Member
Professor, Molecular Biochemistry and Molecular Biology

Michel R. Leroux
Committee Member
Professor, Molecular Biochemistry and Molecular Biology

Mani Larijani
Internal Examiner
Professor, Molecular Biochemistry and Molecular Biology

Warren W. Wakarchuk
External Examiner
Professor, Faculty of Science, Biological Sciences
University of Alberta

Abstract

O-GlcNAc is a pivotal post-translational modification involved in critical cellular processes such as transcription, signal transduction, and stress responses. Detecting and characterizing O-GlcNAc modifications are challenging due to their low abundance and labile nature. This thesis addresses these challenges through three interconnected studies. Firstly, a single-chain variable fragment (scFv) derived from the RL2 monoclonal antibody was engineered, demonstrating specificity and sensitivity for O-GlcNAc-modified proteins in various immunodetection assays. Secondly, a novel chemoenzymatic enrichment technique was developed, using β 4GalT1 coupled with a galactose hydrolase identified through metagenomic screening, preserving the glycan moiety and enabling efficient enrichment of O-GlcNAc-modified proteins for subsequent analysis. Lastly, biophysical characterization of O-GlcNAcylated proteins was conducted using circular dichroism (CD) spectroscopy and intrinsic tryptophan fluorescence to assess alterations in secondary structure and overall protein folding in the presence and absence of O-GlcNAc. Results indicated that O-GlcNAc modification enhanced protein stability during thermal denaturation but had minimal impact during equilibrium chaotropic denaturation. Stability effects were observed again in stopped-flow experiments under chaotropic conditions. These findings expand the methodological framework for studying O-GlcNAc modifications, providing novel insights into their biophysical properties and functional significance. This research paves the way for future investigations aimed at exploring the therapeutic potential of targeting O-GlcNAc modifications in disease pathology, offering a comprehensive approach to elucidating the role of O-GlcNAc in protein stability and function. By enabling precise detection, efficient enrichment, and detailed characterization of O-GlcNAc-modified proteins, the methodologies developed here provide a foundation for identifying disease-specific O-GlcNAcylation patterns and their impact on pathological processes. These insights could guide the development of targeted therapies or biomarkers for conditions where O-GlcNAc dysregulation plays a critical role.

Keywords: O-GlcNAc; antibody engineering; chemoenzymatic enrichment;
protein stability; biophysics

Dedication

To those I love the most, Erick, *Dexter* and Hugo.

Acknowledgements

I am profoundly grateful to Prof. David J. Vocadlo for his guidance and support throughout my PhD journey. His wisdom and approachability encouraged me to push the boundaries of my research while providing a strong foundation of expertise. His dedication to his students' success and his constructive feedback have been essential in shaping my work. I also thank my committee members, Prof. Esther Verheyen and Prof. Michel Leroux, for their invaluable feedback and support.

Dr. Dustin T. King has been an exceptional mentor. His expertise in protein biochemistry and biophysics shaped my understanding of these fields. His patient explanations and countless hours of discussion were pivotal. His kindness and encouragement, especially during moments of self-doubt, kept me going. His respect and support made me feel valued and capable even when facing tough challenges.

I also acknowledge the wonderful people in the lab who made my time here memorable: Zarina Madden, Nursah Ertunc, Ben Tiet, Benny Wu, John Lee, Sha Zhu, and Kelly Chen. Each of you contributed to a collaborative and enjoyable lab environment. Additionally, deepest thanks go to Dr. Vitor Cunha for his invaluable contribution to the enrichment project. His expertise and dedication in synthesizing the probes were instrumental in advancing this research. The quality and precision of his work provided the foundation for many key findings, and his support has been a cornerstone of progress. Thanks to the entire laboratory team for their support and contributions.

I would like to express my gratitude to Prof. Stephen Withers and Dr. Feng Liu at UBC for their invaluable assistance in screening the glycoside hydrolase (GH) enzymes, which was crucial for the success of the chemoenzymatic enrichment approach.

Prof. Jennifer Brodbelt and Dr. Edwin Escobar at the University of Texas at Austin, your collaboration on mass spectroscopic mapping of O-GlcNAc sites with UVPD greatly enriched my research. Thank you for your innovative approaches.

Gratitude is extended to Prof. Alberto Fernández-Tejada, Adrián Plata-Ruiz, Abhijit Saha, and Mirelys Sáenz at the CIC bioGUNE for their exceptional efforts in synthesizing the Nup peptides for the scFv project. Their collaboration and meticulous work significantly enhanced the quality and scope of the research, proving crucial for validating the functionality and specificity of the scFv.

Prof. Cornelia Bohne and Jessy Oake at the University of Victoria's VicPick provided invaluable assistance with folding kinetics experiments. Prof. Bohne's insightful discussions greatly enhanced my understanding of folding kinetics.

I am also grateful to Prof. Stephanie Olivier-Van Stichelen and Dr. Florian Malard at the Medical College of Wisconsin for their openness to our computational analysis of structurally relevant O-GlcNAc sites, significantly contributing to the success of this project.

Dr. Yanping Zhu has been a pillar of support with his countless pieces of advice and confidence-building encouragement. His willingness to entertain my ideas with respect and scientific integrity has been incredibly valuable. His constructive feedback and availability for discussions have been immensely helpful.

I deeply appreciate the friendship and support of Jil A. Busmann and Jun Bian in the lab. Jil's constant presence and readiness to listen made the long hours more bearable with her humor and support. Jun, though our friendship blossomed later in my PhD, quickly became one of my most cherished friends. Her vibrant energy and genuine kindness brought joy and warmth into my life when I needed it most.

I also thank the Canadian Light Source for training me in crystallography and beam line operations. Their support was essential to the success of my structural studies.

The staff at Simon Fraser University, particularly in the Department of Molecular Biology and Biochemistry and the Department of Chemistry, were always helpful. Their dedication to supporting students made a significant difference in my academic experience.

Finally, Erick, you have been my rock throughout this entire journey. Your constant support, patience, and love have been the main reasons I could persevere. To Dexter, who is forever missed, and Hugo, who continues to bring joy to my life, thank you for your boundless love and companionship. Dexter's memory and Hugo's playful spirit have been a source of immense comfort and happiness.

Table of Contents

Declaration of Committee	ii
Abstract	iii
Dedication	v
Acknowledgements	vi
Table of Contents	viii
List of Tables	xi
List of Figures	xii
List of Acronyms	xiv
Glossary	xvi
Chapter 1.	GENERAL INTRODUCTION
.....	1
1.1. O-GlcNAc modification: Importance	1
1.2. Chemical and Structural Properties of O-GlcNAc	2
1.2.1. Chemistry of O-GlcNAcylation	2
1.3. Biophysical Characterization of O-GlcNAcylated Proteins	4
1.3.1. Protein Stability and Folding	4
1.3.2. Aggregation and Phase Separation	4
1.3.3. Thermal and <i>Kinetic</i> Stability	6
1.3.4. Future Directions and Untapped Techniques	8
1.3.5. Untapped Techniques	9
1.4. Biochemical Characterization of O-GlcNAcylation	10
1.4.1. Hexosamine Biosynthetic Pathway	10
1.4.2. OGT and OGA Enzymes.....	12
1.4.3. Sequence specificity	15
1.5. Detection Methods for O-GlcNAc Modifications	16
1.5.1. Antibody-Based Detection Methods.....	16
1.5.2. Lectins and Glycan-Binding Proteins (GBPs)-Based Detection Methods	18
1.5.3. Labeling Techniques for O-GlcNAc Detection	20
1.6. Challenges and Gaps in O-GlcNAc Detection	22
1.6.1. Stoichiometry	22
1.6.2. Temporal and Spatial Resolution.....	24
1.6.3. Sensitivity and Specificity.....	25
1.6.4. Biophysical Properties	26
1.7. Conclusions	27
Chapter 2.	ENGINEERING A SCFV FOR O-GLCNAC DETECTION
.....	29
2.1. Background and Importance	29
2.2. Objective	34
2.3. Methods	35
2.3.1. Antibody Sequencing	35

2.3.2. Vector Design	36
2.3.3. Cloning	39
2.3.4. Protein Purification.....	42
2.3.5. Secondary Structure Analysis.....	44
2.3.6. PAGEs and Immunoblots.....	45
2.3.7. Peptides Synthesis and Analysis	46
2.3.8. BLI.....	47
2.3.9. Pulldown	47
2.3.10. Protein Crystallization	48
2.4. Results.....	48
2.4.1. Recombinant RL2 Production	48
2.4.2. scFv(RL2) Production	51
2.4.3. Binding of scFv(RL2) to O-GlcNAc	54
2.4.4. Pull-Down of O-GlcNAc-modified Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ using scFv(RL2).....	60
2.4.5. Crystallization of scFv(RL2)	62
2.5. Discussion.....	64
2.5.1. Recombinant full-length RL2 production	64
2.5.2. scFv(RL2) development.....	65
2.6. Acknowledgments	66
Chapter 3.CHEMOENZYMATIC METHOD FOR O-GLCNAC ENRICHMENT	68
3.1. Background and Importance	68
3.2. Objective.....	71
3.3. Methods	72
3.3.1. Synthesis and Characterization of Probes	72
3.3.2. GH Screening	74
3.3.3. GHA09 and GHA10 Expression.....	75
3.3.4. Kinetic Assays.....	75
3.4. Results.....	76
3.4.1. Metagenomic Screening	76
3.4.2. Enzyme Characterization and Kinetics	79
3.4.3. GH/OGA-Coupled Assay	82
3.4.4. Steady-state kinetics of LacNAc turnover	84
3.5. Discussion.....	87
3.6. Acknowledgments	89
Chapter 4.BIOPHYSICAL ANALYSIS OF O-GLCNACYLATION	91
4.1. Background and Importance	91
4.2. Objective.....	97
4.3. Methods	98
4.3.1. Cloning	98
4.3.2. Protein Production	99
4.3.3. Immunoblots.....	100

4.3.4.	Intact protein MS	100
4.3.5.	Site Mapping	101
4.3.6.	CD spectroscopy.....	101
4.3.7.	Intrinsic tryptophan fluorescence	102
4.3.8.	Lysate thermal denaturation	103
4.3.9.	Computational Analysis	104
4.3.10.	Molecular Dynamics.....	105
4.4.	Results.....	106
4.4.1.	Ph-P ¹²²⁷⁻¹⁵⁹⁸ thermal denaturation	106
4.4.2.	Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ lysate thermal denaturation.....	110
4.4.3.	Assessment of O-GlcNAcylation of Ph-P ¹³⁹⁷⁻¹⁵⁸⁹	112
4.4.4.	Isothermal Stability of Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ at Extreme Temp.....	114
4.4.5.	Biophysical Analysis	115
4.4.6.	Equilibrium Chemical Denaturation.....	117
4.4.7.	Folding Kinetics.....	118
4.4.8.	Computational Analysis of O-GlcNAc Modified Sites.....	119
4.4.9.	MD Simulations	122
4.5.	Discussion.....	125
4.6.	Acknowledgments	130
Chapter 5	CONCLUSION	
		132
5.1.	5.1 Summary of Key Findings	132
5.1.1.	Development of a Practical scFv for O-GlcNAc Detection.....	132
5.1.2.	Chemoenzymatic Enrichment Method for O-GlcNAc-Modified Proteins	132
5.1.3.	Biophysical Effects of O-GlcNAcylation on Protein Stability	132
5.2.	Implications of the Research.....	133
5.3.	Limitations of the Study.....	133
5.4.	5.4 Future Directions	133
5.4.1.	Enhancing the utility and practicality of the scFv for O-GlcNAc Detection	134
5.4.2.	Developing the chemoenzymatic enrichment method for O-GlcNAc-modified Proteins	135
5.4.3.	Expanding Biophysical Studies of O-GlcNAcylation on Protein Stability	136
5.5.	5.5 Concluding Remarks.....	137
REFERENCES		139
APPENDIX FULL GELS AND IMMUNOBLOTS		160

List of Tables

Table 1.1 Non-comprehensive list of O-GlcNAc antibodies.	16
Table 2.1 Recent developments of scFv targeting carbohydrates.	33
Table 2.2 Binding data of scFv and Ph-P with streptavidin beads.	62
Table 3.1 Metagenomic study derived GH enzymes. (Heins et al., 2014)	77
Table 3.2 Kinetic parameters of enzymes GHA09 and GHA10 for β -gal-4MU.....	81
Table 3.3 Kinetics of GHA09 and GHA10 using LacNAc-4MU as substrate.....	86

List of Figures

Figure 1.1 Chemical Structure of O-GlcNAc Modification.	2
Figure 1-1.2 Biosynthesis Pathway of UDP-GlcNAc.	11
Figure 1-1.3 Enzymatic Cycle of O-GlcNAcylation.	13
Figure 1.4 Structural representation of OGT (PDB: 7YEA)	14
Figure 2.1 Schematic comparison of IgM, IgG, and scFv structures....	31
Figure 2.2 RL2 light chain κ amino acid sequence.....	35
Figure 2.3 RL2 light chain κ amino acid sequence.....	36
Figure 2.4 Full amino acid sequence of the scFv(RL2)	37
Figure 2.5 Biotin ligase (birA) reaction.	38
Figure 2.6 Full amino acid sequence of the scFv(RL2)-V/A.	39
Figure 2.7 Vector designs for antibody expression.	41
Figure 2.8 Characterization of recombinant RL2 antibody.	50
Figure 2.9 scFv(RL2) characterization.	52
Figure 2.10 Predicted structure of scFv(RL2) highlighting CDR loops.	53
Figure 2.11 Native State Mass Spectrometry of Ph-P ¹³⁹⁷⁻¹⁵⁸⁹	55
Figure 2.12 Mechanism of α-N-gluconoylation of His-tagged Proteins in <i>E. coli</i>	56
Figure 2.13 Binding analysis of scFv to O-GlcNAc modified protein and peptides.	57
Figure 2.14 MALDI-ToF mass spectra of glycosylated Nup peptides.	59
Figure 2.15 Pull-down assay of O-GlcNAc modified Ph-P using scFv.	61
Figure 2.16 scFv(RL2) protein crystals.....	64
Figure 3.1 Enzymatic strategy for O-GlcNAc protein enrichment.	71
Figure 3.2 GH enzymes screening against biotinylated LacNAc.....	78
Figure 3.3 Standard curve for 4MU concentration.....	79
Figure 3.4 Characterization of GHs GHA09 and GHA10.	81
Figure 3.5 Substrate turnover by GHA09 and GHA10 in coupled assay.....	82
Figure 3.6 Reaction rates of GHA09 and GHA10 with LacNAc-4MU.	83
Figure 3.7 Kinetic analysis of GHs in Coupled assay for LacNAc turnover.....	85
Figure 4.1 Funnel of proteins' free energy landscape.	92
Figure 4.2 Energy profile during protein folding.....	93
Figure 4.3 O-GlcNAc prevents PhP aggregation as per Gambetta and Müller (2014)....	96
Figure 4.4 Schematic Ph-P domain structure and derived constructs.....	107
Figure 4.5 Effect of O-GlcNAc on Ph-P ¹²²⁷⁻¹⁵⁹⁸ thermal stability.	109
Figure 4.6 Analysis of recombinant Ph-P ¹²²⁷⁻¹⁵⁹⁸ purification.	110
Figure 4.7 Thermostability analysis of Ph-P ¹³⁹⁷⁻¹⁵⁹⁸	112
Figure 4.8 Assessment of O-GlcNAcylation of Ph-P ¹³⁹⁷⁻¹⁵⁸⁹	112
Figure 4.9 Isothermal assessment of Ph-P ¹³⁹⁷⁻¹⁵⁹⁸	114
Figure 4.10 Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ biophysical characterization.	115

Figure 4.11 Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ equilibrium chemical denaturaiton.....	117
Figure 4.12 Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ folding kinetics.....	119
Figure 4.13 Computational analysis of O-GlcNAc modified sites.....	120
Figure 4.14 MD model of CAT with and without O-GlcNAc modification.....	122
Figure 4.15 MD model of human LDHA enzyme with and without O-GlcNAc.....	123
Figure 4.16 MD model of human G6PD enzyme with and without O-GlcNAc.....	124
Figure 4.17 Ph-P ¹³⁹⁷⁻¹⁵⁸⁹ MD model.....	124
Figure 4.18 pPSE score ≥ 5 overlap with co-translational O-GlcNAc sites.....	127
Figure 4.19 Metabolic insight from computational analysis of O-GlcNAc sites.....	129

List of Acronyms

4MU	4-methylumbelliferon
Ab	Antibody
An	Antigen
BLI	Biolayer interferometry
BSA	Bovine serum albumin
CD	Circular dichroism
CHO-K1	Chinese hamster ovary-K1
CETSA	Cellular thermal shift assay
ELISA	Enzyme-linked immunosorbent assay
Gal	Galactose
B4GALT1	β -1,4-galactosyltransferase
GBP	Glycan binding protein
GH	Galactose hydrolase
Glc	Glucose
HDX	Hydrogen-deuterium exchange
His	Histidine
HPLC	High-performance liquid chromatography
LacNAc	<i>N</i> -acetyllactosamine
MS	Mass spectroscopy
IPTG	Isopropyl β -D-1-thiogalactopyranoside
KD	Dissociation constant
O-GlcNAc	O-linked β -N-acetylglucosamine
OGA	O-GlcNAcase
OGT	O-GlcNAc transferase
PEG	Polyethylene glycol
Ph-P	Polyhomeotic protein
Ph-P(G)	Polyhomeotic protein O-GlcNAc modified
RFU	Relative fluorescence units

SAM	Sterile alpha unit
SDS	Sodium dodecyl sulfate
S or Ser	Serine
PAGE	Polyacrylamide gel electrophoresis
scFv	Single-chain variable fragment (Fv)
T or Thr	Threonine
TLC	Thin-layer chromatography
T _M	Melting point
UVPD	Ultraviolet photodissociation

Glossary

4MU	4-methylumbelliflone, a fluorescent molecule often used as a substrate in enzymatic assays.
BLI	Biolayer interferometry, is an optical analytical technique used to measure biomolecular interactions in real-time without the need for labels. It works by detecting changes in the interference pattern of light reflected from a biosensor surface. When a biomolecule binds to the sensor surface, it causes a shift in the interference pattern, which can be measured and quantified. BLI is commonly used to determine kinetic parameters such as association and dissociation rates, and equilibrium binding constants (K_D) of interactions between proteins, antibodies, nucleic acids, and other biomolecules. This technique is valuable in various fields, including drug discovery, diagnostics, and molecular biology, for characterizing the strength and specificity of biomolecular interactions.
BSA	Bovine serum albumin, a protein derived from cows that is commonly used in biochemical assays.
CD	Circular dichroism, Circular Dichroism is a spectroscopic technique used to measure the difference in the absorption of left-handed and right-handed circularly polarized light by chiral molecules. It is particularly useful in studying the secondary structures of proteins, such as α -helices and β -sheets. By analyzing the CD spectra, researchers can gain insights into the folding, conformational changes, and stability of proteins under various conditions. This technique is essential for understanding protein structure-function relationships and for characterizing the effects of modifications, such as glycosylation, on protein conformation.
CETSA	Cellular thermal shift assay, a method to study protein stability and
DE3	A strain of <i>Escherichia coli</i> bacteria used for protein expression.
ELISA	Enzyme-linked immunosorbent assay, a plate-based assay technique for detecting and quantifying

	substances such as peptides, proteins, antibodies, and hormones.
B4GALT1	β -1,4-galactosyltransferase, an enzyme that transfers galactose residues to proteins or other molecules.
GH	Glycoside hydrolase, an enzyme that hydrolyzes glycosidic bonds in carbohydrates.
HDX	Hydrogen-deuterium exchange is a mass spectrometry technique used to study protein dynamics and interactions. By replacing hydrogen atoms in a protein with deuterium, researchers can track changes in protein conformation and stability. This method provides insights into protein folding, binding sites, and structural changes, making it valuable for understanding protein behavior under different conditions.
His	Histidine, an amino acid commonly used in protein purification tags (e.g., His-tag).
HPLC	High-performance liquid chromatography, a technique used to separate, identify, and quantify components in a mixture.
IPTG	Isopropyl β -D-1-thiogalactopyranoside, a chemical compound used to induce protein expression in bacteria.
OGA	O-GlcNAcase, an enzyme that removes O-GlcNAc modifications from proteins.
OGT	O-GlcNAc Transferase, an enzyme that adds O-GlcNAc modifications to proteins.
SAM	Sterile alpha motif, a protein domain involved in protein-protein interactions.
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis, a technique used to separate proteins based on their electrophoretic mobility.
scFv	Single-chain variable fragment, a type of antibody fragment.
UVPD	Ultraviolet photodissociation is a mass spectrometry technique that uses ultraviolet light to fragment molecules. This method provides detailed structural information about the molecules, including sequence and post-translational modifications, making it

particularly useful for proteomics and structural biology studies.

Chapter 1.

GENERAL INTRODUCTION

1.1. O-GlcNAc modification: Importance

The O-linked β -N-acetylglucosamine (O-GlcNAc) modification is a dynamic and essential modification found on serine and threonine residues of nuclear and cytoplasmic proteins.¹⁻⁵ O-GlcNAc appears to be principally installed post-translationally, though it has been found to occur co-translationally on some proteins.^{6,7} Unlike other glycosylation processes that occur in the endoplasmic reticulum (ER) and Golgi apparatus, O-GlcNAc modification takes place in the cytoplasm and nucleus,² where it plays a crucial role in regulating numerous cellular functions, including transcription,^{3,8-10} and stress response.¹¹⁻¹³ The transient nature of O-GlcNAc modification, which resembles serine and threonine phosphorylation in some respects, underscoring its importance in cellular signaling pathways.³

Detecting O-GlcNAc is critical for several reasons. Firstly, aberrant O-GlcNAcylation has been implicated in various diseases, including cancer,¹⁴ X-linked intellectual disability (XLID),¹⁵ type 2 diabetes mellitus,¹⁶ and neurodegenerative disorders: Alzheimer's,^{17,18} Parkinson's,^{19,20} amyotrophic lateral sclerosis (ALS),²¹ and Huntington's.²² By understanding the distribution and levels of O-GlcNAc modifications, researchers can gain insights into disease mechanisms and identify potential biomarkers for diagnosis and prognosis. Secondly, as noted above O-GlcNAc modification appears to be involved in the regulation of many cellular processes including those noted as well as many other such as cell cycle control,²³ autophagy,^{24,25} and circadian rhythm.²⁶ Monitoring these modifications can therefore provide valuable information on cellular function and response to environmental changes. Lastly, studying O-GlcNAc can reveal new therapeutic targets.

O-GlcNAcylation levels vary significantly across different tissues, reflecting its role in tissue-specific functions. Higher levels of O-GlcNAcylation are observed in metabolically active organs such as the brain,^{27,28} heart,²⁹ and pancreas,²⁸ where it plays a critical role

in regulating key cellular processes. In the brain, for instance, O-GlcNAcylation is involved in synaptic function³⁰ and neuronal survival,³¹ which is particularly relevant in the context of neurodegenerative diseases. The distribution pattern suggests that O-GlcNAcylation is closely linked to the metabolic demands and functional requirements of specific tissues.

However, the detection and characterization of O-GlcNAc modifications pose significant analytical challenges due to their substoichiometric and labile nature.³² Consequently, developing sensitive and specific detection tools for O-GlcNAc has become a focus in the field, driving advancements in the development of both analytical techniques and molecular probes.

1.2. Chemical and Structural Properties of O-GlcNAc

1.2.1. Chemistry of O-GlcNAcylation

O-GlcNAc is a specific type of post-translational modification in which an *N*-acetylglucosamine monosaccharide is covalently attached to the hydroxyl group of serine (ser) or threonine (thr) residues within proteins.¹ This modification is fundamentally based on the chemical structure of GlcNAc, an amino sugar derived from glucose. (Figure 1.1) In the GlcNAc molecule, the hydroxyl group at the second carbon (C2) of the glucose ring is replaced by an acetamido group (-NHCOCH₃). The overall chemical formula for GlcNAc is C₈H₁₅NO₆, and it typically exists in a six-membered pyranose ring form, which includes multiple hydroxyl groups (-OH), an acetal functionality making the ring and glycosidic bond, as well as the aforementioned acetamido group at C2.

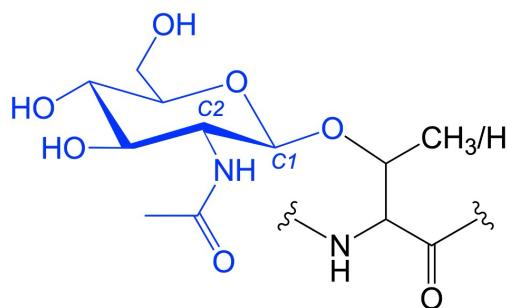


Figure 1.1 Chemical Structure of O-GlcNAc Modification.

Covalent GlcNAc residue to the hydroxyl group of a ser/thr residue in a protein. The GlcNAc is shown as a six-membered pyranose ring, with the C1 carbon forming a β -O-glycosidic bond with the ser/thr residue. The acetamido group is attached to the C2 position of the GlcNAc.

The key chemical feature of O-GlcNAcylation is the formation of a β -O-glycosidic bond between the anomeric carbon (C1) of the GlcNAc molecule and the hydroxyl group of a ser/thr residue in the target protein. This glycosidic bond is specifically a β -linkage, meaning that the GlcNAc is attached with the hydroxyl group at C1 in an equatorial position, which retains the β -configuration of the sugar. The formation of this bond is catalyzed by the enzyme O-GlcNAc transferase (OGT),³³ which uses UDP-GlcNAc (uridine diphosphate N-acetylglucosamine) as the donor substrate. During the reaction, OGT transfers the GlcNAc moiety from UDP-GlcNAc to the hydroxyl group of the serine or threonine residue, resulting in the inversion of stereochemistry of the monomeric center as the α -anomeric configurated GlcNAc unit of UDP-GlcNAc ends up being β -linked to the at the modified residue, releasing UDP (uridine diphosphate) as a byproduct. (Figure 1.3)

The glycosidic bond formed in O-GlcNAcylation is a covalent bond that, while strong, is relatively labile compared to other glycosidic linkages, allowing for the reversible nature of this modification. This reversibility is facilitated by the enzyme O-GlcNAcase (OGA),³⁴ which hydrolyzes the glycosidic bond, removing the GlcNAc moiety and restoring the unmodified serine or threonine residue. (Figure 1.3)

From a chemical perspective, the O-GlcNAc modification introduces several functional groups into the protein structure, including multiple hydroxyl groups and an acetamido group, which can engage in hydrogen bonding. These interactions can influence the protein's folding, stability, and interactions with other molecules. Additionally, the introduction of the polar GlcNAc moiety increases the local hydrophilicity of the modified protein, which can affect its solubility and how it interacts within aqueous environments.

1.3. Biophysical Characterization of O-GlcNAcylated Proteins

1.3.1. Protein Stability and Folding

Understanding the biophysical effects of O-GlcNAcylation is crucial for comprehending how these modifications influence protein stability, folding, and interactions. Insights into these effects can reveal disease mechanisms, identify therapeutic targets, and provide valuable information for biotechnological applications. O-GlcNAcylation can prevent protein aggregation, enhance thermal stability, and modulate protein-protein and protein-nucleic acid interactions.^{35,36}

1.3.2. Aggregation and Phase Separation

The protective role of O-GlcNAcylation in stabilizing proteins against aggregation is a well-established phenomenon, with its roots tracing back to studies from the mid-1990s. One of the earliest and most compelling examples is the modification of the small heat shock protein alpha B-crystallin.^{37,38} As early as 1996, it was demonstrated that alpha B-crystallin undergoes O-GlcNAc modification, particularly at T170, a modification that appears to be crucial for its function as a molecular chaperone.³⁸ This early work suggested that O-GlcNAcylation may play a key role in preventing the aggregation of alpha B-crystallin under stress conditions, a hypothesis that has since been supported by a growing body of evidence.³⁷⁻⁴⁰

Building on this foundational knowledge, subsequent studies have further demonstrated the broad protective effects of O-GlcNAc against protein aggregation. For instance, the transcription factor Sp1 was shown to resist thermal aggregation when modified by O-GlcNAc.⁴¹ This finding, while consistent with earlier observations in alpha B-crystallin, also highlighted a broader protective role for O-GlcNAc across different classes of proteins. Reductions in O-GlcNAc levels through chemical inhibition or genetic manipulation resulted in increased aggregation of Sp1, reinforcing the idea that O-GlcNAcylation serves as a critical stabilizing modification.

In the context of neurodegenerative diseases, the role of O-GlcNAcylation in preventing aggregation has become particularly significant.⁴² The microtubule-associated

protein tau, known for its aggregation into neurofibrillary tangles in Alzheimer's disease, has been a focal point of research. Tau, encoded by the MAPT gene, is derived from alternative splicing, which produces six major isoforms in the adult human brain. These isoforms differ by the number of microtubule-binding repeats (3R or 4R tau) and N-terminal inserts (0, 1, or 2), allowing for some diversity. As an intrinsically disordered protein (IDP), tau lacks a stable tertiary structure under physiological conditions, which contributes to its ability to bind microtubules and other cellular partners but also makes it susceptible to pathological aggregation.⁴³ O-GlcNAcylation of tau has been shown to inhibit its aggregation, independent of its phosphorylation status. This finding not only confirmed the long-standing hypothesis that O-GlcNAcylation stabilizes proteins but also suggested a specific protective mechanism that could be leveraged therapeutically.⁴⁴⁻⁴⁶ The reduction of tau aggregation through increased O-GlcNAc levels, observed both *in vitro* and *in vivo*, underscores the potential of O-GlcNAcylation as a key modulator in protein aggregation pathways.

The anti-aggregation effects of O-GlcNAcylation are not limited to tau. α -Synuclein, another protein central to neurodegenerative disease pathology, particularly Parkinson's disease, also exhibits reduced aggregation when O-GlcNAcylated.⁴⁷ This modification appears to prevent the peptide-dependent acceleration of α -synuclein aggregation, suggesting that O-GlcNAc may exert a broad anti-aggregation effect across various proteins susceptible to pathological aggregation.

The story of O-GlcNAcylation's role in preventing protein aggregation extends to the Polycomb group repressor polyhomeotic protein (Ph-P), where O-GlcNAcylation prevents nonproductive aggregation and maintains proper protein function.⁴⁸ The failure of Ph to aggregate in the absence of O-GlcNAcylation further highlights the critical role of this modification in ensuring the functional integrity of proteins under various cellular conditions.

A new development in the study of O-GlcNAcylation's protective role against protein aggregation is its emerging connection to liquid-liquid phase separation (LLPS). Traditionally, aggregation has been viewed primarily because of protein misfolding and subsequent insoluble aggregate formation. However, recent research has begun to reveal that LLPS—where proteins and other biomolecules condense into dynamic, membraneless compartments—can play a pivotal role in the early stages of aggregation.⁴⁹ Dysregulated LLPS may serve as a precursor to pathological protein aggregation,

particularly in neurodegenerative diseases. O-GlcNAcylation has been identified as a key regulator of this process, preventing the transition from the liquid-like state of phase-separated droplets to the solid-like state characteristic of protein aggregates. This modulation is especially evident in proteins like SynGAP, where O-GlcNAcylation has been shown to inhibit the interactions driving phase separation, thereby reducing the likelihood of aggregation.⁵⁰ This emerging relationship between O-GlcNAcylation and LLPS represents an exciting advancement in the aggregation field, offering a fresh perspective on how this modification can stabilize proteins by intervening at an earlier, more dynamic stage of the aggregation pathway. As more proteins are found to be regulated by this mechanism, it is becoming increasingly clear that O-GlcNAcylation may be critical in preventing aggregation by controlling phase behavior.

While the protective role of O-GlcNAcylation against protein aggregation has been recognized for decades, the recent discoveries linking it to the regulation of liquid-liquid phase separation represent a significant evolution in our understanding. This new perspective underscores how O-GlcNAcylation not only directly impacts the final stages of aggregation but also plays a crucial role in modulating the earlier, dynamic phase behaviors that precede aggregation. By stabilizing proteins and preventing their transition from phase-separated states to pathological aggregates, O-GlcNAcylation offers a powerful mechanism to maintain protein homeostasis. This dual role in both phase separation and aggregation highlight the modification's importance in cellular function and its potential as a therapeutic target in diseases characterized by protein aggregation. Continued exploration of this emerging connection will be vital in developing strategies to combat aggregation-related neurodegenerative diseases.

1.3.3. Thermal and Kinetic Stability

O-GlcNAcylation is increasingly recognized as involved protein thermal stability, a feature particularly vital for organisms exposed to fluctuating temperatures. Studies in ectotherms, such as *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Danio rerio*, have shown that the levels of O-GlcNAcylation tightly correlate with ambient temperature, suggesting a protective mechanism that ensures protein stability under thermal stress. Elevated O-GlcNAc levels were found to be necessary for proper development at higher temperatures, indicating that this modification plays a pivotal role in maintaining protein function when thermal fluctuations occur. The regulation of O-GlcNAc levels in response

to temperature shifts likely involves the modulation of the activities of OGT and OGA, the enzymes responsible for adding and removing the modification, respectively.³⁶ This dynamic regulation may enable ectotherms to rapidly adjust their proteomes to varying environmental conditions, thus preventing protein denaturation and aggregation under heat stress.

The application of cellular thermal shift assay (CETSA) has further elucidated the role of O-GlcNAcylation in enhancing protein thermal stability. CETSA, traditionally used to assess protein-ligand interactions,⁵¹ was adapted to measure the stabilizing effects of O-GlcNAcylation on proteins. For instance, O-GlcNAcylation of the innate immune receptor Nod2 was shown to significantly increase its thermal stability, as indicated by a higher melting temperature compared to its non-O-GlcNAcylated counterpart.⁵² This increase in thermal stability was observed not only in wild-type Nod2 but also in certain disease-associated variants, suggesting that O-GlcNAcylation could be a general mechanism for stabilizing proteins against thermal denaturation. The adaptation of CETSA for studying O-GlcNAc effects has provided a robust and quantitative method for probing the thermal stability of O-GlcNAc-modified proteins, thus advancing our understanding of this modification's protective role.

Thermal proteome profiling (TPP)⁵³ has provided a proteome-wide perspective on the impact of O-GlcNAcylation on protein thermal stability. Studies employing TPP have demonstrated that O-GlcNAcylation consistently increases the thermal stability of a broad range of proteins across different cellular contexts.^{54,55} By comparing the melting temperatures of proteins in O-GlcNAcylated and non-O-GlcNAcylated states, it was revealed that O-GlcNAcylation generally confers increased resistance to thermal denaturation, which may prevent protein misfolding and aggregation under heat stress. These findings underscore the systemic importance of O-GlcNAcylation in maintaining proteome stability, particularly under conditions that challenge protein homeostasis.

NMR studies by Yuzwa et al. (2014) on the tau protein fragment (residues 353–408) provides important insights into how O-GlcNAcylation might influence the structural dynamics of proteins, with potential implications for aggregation.⁴⁶ The data indicate that while O-GlcNAcylation at Ser400 does not induce significant global structural changes, it causes localized alterations in the dynamics near the modification site. These changes manifest as reduced conformational flexibility around Ser400, suggesting that O-GlcNAcylation may stabilize certain regions of the protein, making them less prone to the

conformational fluctuations that often precede aggregation. Although the study does not directly measure kinetic rates, the observed reduction in flexibility could imply a slower progression through the early stages of aggregation, such as nucleation, by maintaining the protein in a more soluble and less aggregation-prone state. This stabilization effect could indirectly influence the overall kinetics of tau aggregation, slowing the rate at which aggregation-prone conformers are formed.

The evidence strongly supports the role of O-GlcNAcylation as a factor in protein thermal stability and influencing protein dynamics. From ectotherms that rely on this modification to survive fluctuating environmental temperatures, to cellular models demonstrating the stabilizing effects of O-GlcNAcylation on specific proteins such as Nod2, O-GlcNAcylation serves as a protective mechanism against thermal stress. Advanced techniques like CETSA and TPP have further elucidated how O-GlcNAcylation increases the thermal stability of a wide range of proteins across the proteome, preventing denaturation and aggregation. Additionally, insights from NMR studies suggest that O-GlcNAcylation may also influence the conformational dynamics of proteins, potentially stabilizing them in less aggregation-prone states. While the exact kinetic effects of O-GlcNAcylation on protein folding and aggregation are still under investigation, the modification's ability to modulate both thermal stability and protein dynamics underscores its pivotal role in maintaining cellular homeostasis under stress conditions. As research continues to uncover the broader implications of O-GlcNAcylation, its significance in protecting against protein misfolding and aggregation becomes increasingly clear.

1.3.4. Future Directions and Untapped Techniques

Identifying relevant O-GlcNAc sites and detecting minor changes in their modification status remain significant challenges in the field of glycobiology. The complexity of the proteome and the stoichiometric nature of O-GlcNAcylation complicate these efforts, as many O-GlcNAc modifications occur at low abundance and are difficult to distinguish from other similar modifications. Overcoming these challenges requires the development of more sensitive and specific analytical techniques that can accurately measure and identify O-GlcNAc modifications in complex biological samples.

1.3.5. Untapped Techniques

Several advanced techniques hold promise for furthering our understanding of O-GlcNAcylation:

Enrichment Methods. Current enrichment techniques for O-GlcNAcylated proteins, such as lectin-based affinity⁵⁶ and metabolic labeling,⁵⁷ have greatly advanced the field, but their full potential remains untapped. Enhancing the specificity of these methods, particularly by developing more selective lectins or antibodies, could significantly improve the accuracy of O-GlcNAc protein isolation. Moreover, optimizing bioorthogonal chemistry techniques for capturing low-abundance modifications and integrating these methods with advanced proteomics could provide deeper insights into the O-GlcNAcylated proteome, revealing novel proteins and pathways influenced by this modification.

Hydrogen/Deuterium Exchange Mass Spectrometry (HDX-MS). This technique could provide insights into protein dynamics and conformational changes induced by O-GlcNAcylation. By measuring the exchange rates of hydrogen and deuterium in the protein backbone, HDX-MS could reveal subtle conformational shifts and identify regions of the protein affected by O-GlcNAc modifications.

Optical Tweezers. Optical tweezers could allow researchers to study the mechanical properties of single O-GlcNAcylated proteins and their interactions under force. This technique can measure the forces involved in protein folding and unfolding and could therefore provide a detailed understanding of how O-GlcNAcylation influences protein mechanics.

Computational Methods. Simulations and molecular dynamics studies offer detailed views of the structural and functional impacts of O-GlcNAcylation on proteins. These computational approaches could be used to model the dynamic behavior of O-GlcNAc-modified proteins, predicting how these modifications affect protein stability, interactions, and overall function.

Protein Hydration and Solvation Shell. O-GlcNAcylation is proposed to influence the hydration and solvation shell of proteins,⁵⁸ affecting their stability and interactions with other molecules. The addition of GlcNAc groups can alter the local water structure around the protein, impacting its solubility and interaction with other cellular components. Understanding these effects could provide deeper insights into the biophysical properties of O-GlcNAcylated proteins and their consequent roles in

cellular processes. This knowledge could help elucidate the mechanisms by which O-GlcNAcylation regulates protein function and contributes to cellular homeostasis.

1.4. Biochemical Characterization of O-GlcNAcylation

1.4.1. Hexosamine Biosynthetic Pathway

The hexosamine biosynthetic pathway (HBP) is a critical metabolic pathway responsible for the synthesis of UDP-GlcNAc, a key donor substrate for various glycosylation processes, including O-GlcNAcylation. HBP acts as an essential node linking nutrient availability to cellular signaling and metabolic regulation.

HBP begins with the enzyme glutamine:fructose-6-phosphate aminotransferase (GFAT), which catalyzes the conversion of fructose-6-phosphate, derived from glucose metabolism, and glutamine into glucosamine-6-phosphate.⁵⁹ This reaction is the rate-limiting step of the pathway and is tightly regulated by several factors, including substrate availability and feedback inhibition by the end product, UDP-GlcNAc. (Figure 1.2)

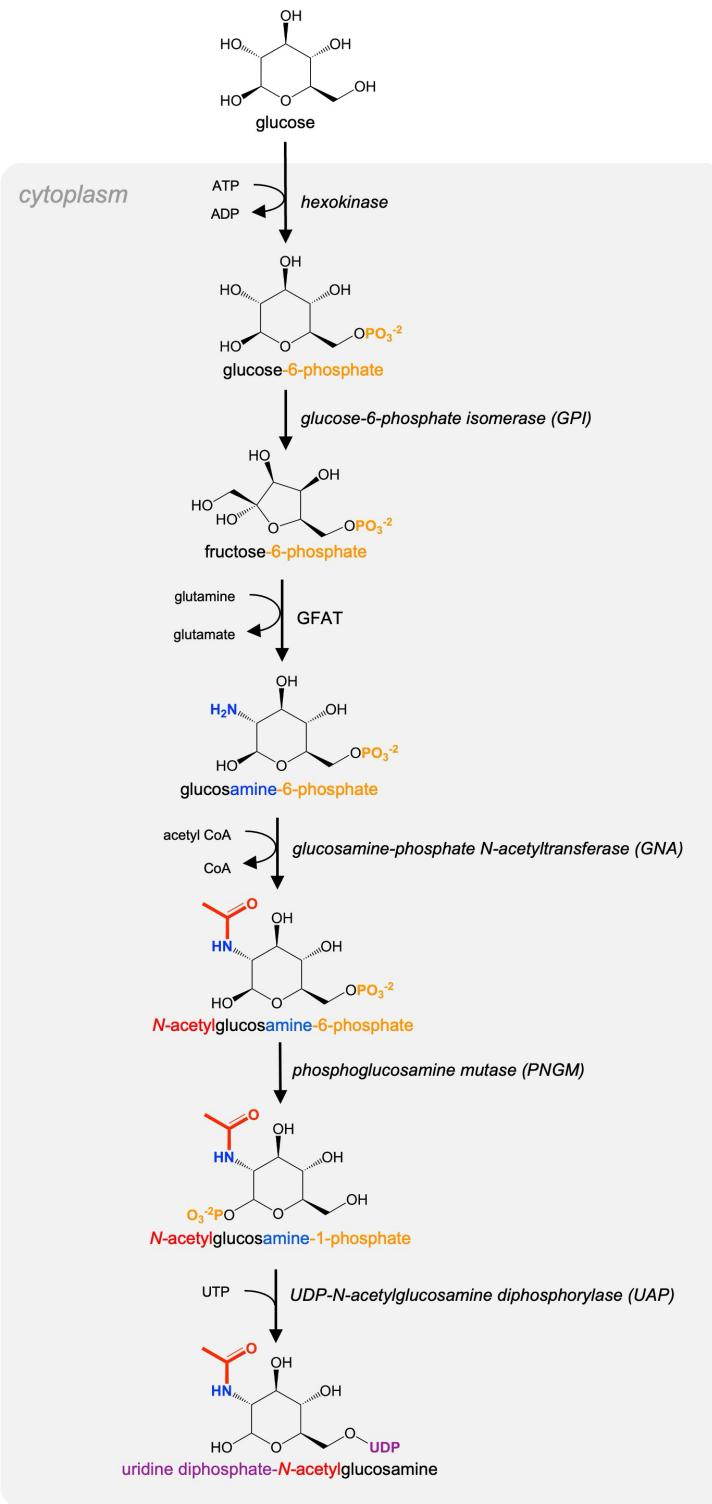


Figure 1-1.2 Biosynthesis Pathway of UDP-GlcNAc.

Multi-step enzymatic process leading to the formation of UDP-GlcNAc, starting from glucose. The key steps include the phosphorylation of glucose by hexokinase, the conversion of glucose-6-

phosphate to fructose-6-phosphate by glucose-6-phosphate isomerase (GPI), the amination to glucosamine-6-phosphate by glutamine: fructose-6-phosphate aminotransferase (GFAT), and subsequent acetylation, mutase activity, and diphosphorylation leading to the final product, UDP-GlcNAc. Each enzyme and intermediate is clearly indicated, highlighting the critical steps in the hexosamine biosynthetic pathway.

Following this, glucosamine-6-phosphate undergoes acetylation by glucosamine-phosphate N-acetyltransferase (GNA), resulting in the formation of *N*-acetylglucosamine-6-phosphate.⁶⁰ This acetylation step introduces an acetyl group from acetyl-CoA, further linking the HBP to the central metabolic pathways such as glycolysis and the citric acid cycle. (Figure 1.2) *N*-acetylglucosamine-6-phosphate is then isomerized to *N*-acetylglucosamine-1-phosphate by the enzyme phosphoglucosamine mutase (PNGM).⁶¹ This reaction involves the transfer of the phosphate group from the 6-position to the 1-position on the glucosamine moiety, which is a critical step in preparing the molecule for subsequent activation. (Figure 1.2)

The final step in the HBP is catalyzed by UDP-*N*-acetylglucosamine diphosphorylase (UAP), which catalyzes the uridylation of *N*-acetylglucosamine-1-phosphate using UTP to produce UDP-GlcNAc.⁶² This reaction not only generates the active form of GlcNAc but also serves as a key control point, as UTP availability can influence the overall flux through the pathway. (Figure 1.2)

Throughout these enzymatic steps, the HBP is highly sensitive to the availability of its substrates: glucose, glutamine, acetyl-CoA, and UTP. Each of these metabolites integrates signals from various metabolic states, thereby enabling the HBP to act as a sensor of the cell's nutritional and energy status. The level of UDP-GlcNAc produced by the HBP directly influences the extent of glycosylation modifications on proteins and other biomolecules, affecting a wide array of cellular processes. HBP is a metabolically integrative pathway that links nutrient sensing to the synthesis of UDP-GlcNAc, playing a vital role in cellular homeostasis through the regulation of glycosylation processes. The pathway's tight regulation and sensitivity to metabolic cues underscore its significance in maintaining cellular function under varying physiological conditions.

1.4.2. OGT and OGA Enzymes

O-GlcNAc modification, a dynamic and reversible post-translational modification, is governed by two critical enzymes: OGT and OGA. OGT catalyzes the addition of O-

GlcNAc to serine or threonine residues on target proteins,³³ while OGA removes the modification, thus maintaining a balance in O-GlcNAc cycling within the cell. (Figure 1.3)³⁴ The cycling of O-GlcNAc is tightly linked to nutrient availability and cellular stress responses, underscoring the regulatory importance of these enzymes.

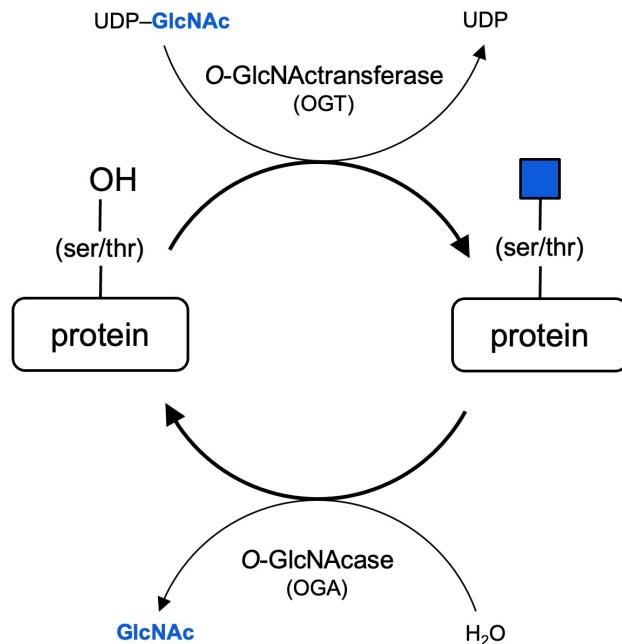


Figure 1-1.3 Enzymatic Cycle of O-GlcNAcylation.

The dynamic process of O-GlcNAcylation and its removal. O-GlcNAc transferase (OGT) catalyzes the addition of O-GlcNAc to serine/threonine residues on proteins using UDP-GlcNAc as a donor substrate, while O-GlcNAcase (OGA) removes the modification, restoring the unmodified protein. The cycle illustrates the reversible nature of this post-translational modification.

OGT, a glycosyltransferase from the CAZy family GT41 — a classification system that groups carbohydrate-active enzymes based on sequence similarity⁶³ — is located on the X chromosome and encodes a modular protein comprising several distinct domains: the N-terminal tetratricopeptide repeat (TPR) domain, the catalytic domain with two lobes (Cat1 and Cat2), an intervening domain (IntD), and a C-terminal phosphatidylinositol-binding region (PPO). The TPR domain of OGT is particularly important for substrate recognition, as it provides a scaffold for protein-protein interactions, ensuring that specific target proteins are modified. The catalytic mechanism of OGT involves the transfer of *N*-acetylglucosamine (GlcNAc) from the donor substrate UDP-GlcNAc to the hydroxyl group of serine or threonine residues, forming a β -O-glycosidic bond. This inverting transfer

mechanism is proposed to follow an ordered bi-bi mechanism, wherein UDP-GlcNAc must bind the active site of OGT before the acceptor protein or peptide.^{64,65}

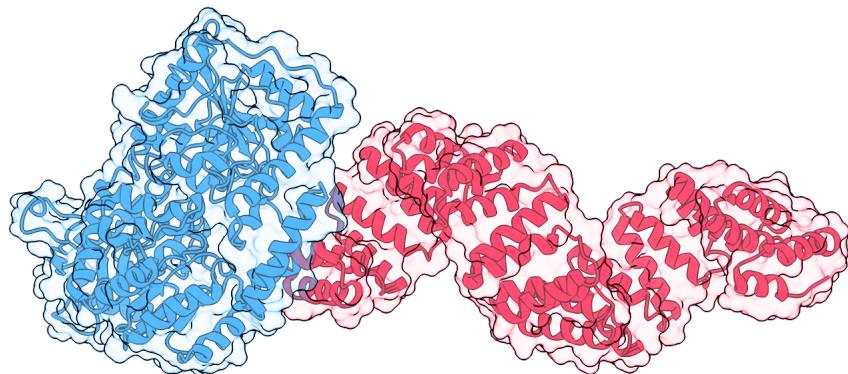


Figure 1.4 Structural representation of OGT (PDB: 7YEA)

Cryo-electron microscopy structure of O-GlcNAc transferase (OGT) from the PDB entry 7YEA. The catalytic domain, responsible for transferring the GlcNAc moiety from UDP-GlcNAc to the hydroxyl group of serine or threonine residues, is highlighted in blue. The tetratricopeptide repeat (TPR) domain, shown in red, mediates substrate recognition and binding. This modular organization exemplifies OGT's function as a glycosyltransferase within the CAZy family GT41, with distinct domains contributing to its enzymatic activity and specificity.⁶³

The structural basis of OGT's function has been elucidated through crystallographic studies, which have provided insights into its unique atomic architecture, particularly the role of the TPR domain in binding and orienting the acceptor peptide within the active site.⁶⁵ Despite these advances, certain mechanistic details, such as the identity of the general base responsible for deprotonating the catalytic serine, remain under investigation. Emerging evidence suggests that OGT's activity is influenced not only by its structural domains but also by its interactions with other proteins.^{66,67}

OGA, the enzyme responsible for removing O-GlcNAc, exists in two major isoforms: the long form (OGA-L) and the short form (OGA-S). These isoforms differ in their domain structures and subcellular localization, with OGA-L being nucleocytosolic and OGA-S being predominantly nuclear. The catalytic domain of OGA, belonging to the glycoside hydrolase CAZy family GH84,⁶³ catalyzes the hydrolysis of the glycosidic bond between GlcNAc and the serine or threonine residue, restoring the unmodified protein. OGA's catalytic mechanism involves a two-step process that proceeds through a transient oxazoline intermediate, facilitated by conserved aspartate residues acting as general acid/base catalysts.⁶⁴

Recent structural studies have provided a detailed view of OGA's active site, revealing its broad substrate specificity, which allows it to act on a wide range of O-GlcNAcylated proteins. This promiscuity is likely due to the enzyme's interaction with the amide backbone atoms within the bound glycopeptide, rather than specific side-chain interactions. Despite its broad specificity, OGA's activity can be modulated by the presence of adjacent post-translational modifications, such as phosphorylation, which can influence the efficiency of O-GlcNAc removal.⁶⁴ OGT and OGA are essential for the regulation of O-GlcNAcylation, a modification that plays a crucial role in cellular homeostasis. The detailed structural and mechanistic insights into these enzymes have advanced our understanding of their functions and opened new avenues for investigating their roles in health and disease.

1.4.3. Sequence specificity

OGT is known for its ability to glycosylate a wide range of substrates without adhering to a strict consensus sequence. Initial studies attempted to identify a clear consensus sequence for OGT, [TS][PT][VT]S/T[RLV][ASY], but these efforts have been met with limited success. For instance, a proposed sequence has been observed,⁶⁸ but many O-GlcNAc-modified proteins do not conform to this pattern, indicating the enzyme's broad specificity.

OGT recognizes its substrates through multiple factors, including the presence of serine or threonine residues within extended loops or intrinsically disordered regions (IDRs).⁶⁹ This substrate recognition is further complicated by the fact that OGT can interact with peptides much longer than the catalytic site can accommodate, as evidenced by the glycosylation of the RNA Polymerase II C-terminal domain (CTD) requiring interaction with an extended region of over 140 residues.⁷⁰ The TPR region of OGT, consisting of 13 repeats in its largest isoform, plays a crucial role in recognizing these extended substrates, contributing to the enzyme's broad substrate specificity.

Moreover, the amino acid composition beyond the immediate glycosylation site influences OGT's substrate specificity. OGT substrates tend to have a compositional bias with certain amino acids like alanine, proline, and threonine being more favorable in glycosylated regions, while others like glycine and asparagine are less frequent.⁷¹ This suggests that OGT's substrate recognition is influenced by both sequence-specific and

compositional elements, allowing it to glycosylate a wide array of protein targets. OGT can also glycosylate cysteine residues, a modification termed S-GlcNAcylation.⁷² This expands the understanding of OGT's substrate flexibility. Unlike O-GlcNAcase, which removes O-GlcNAc modifications from serine or threonine residues, the S-GlcNAc modification on cysteine is more stable and resistant to enzymatic removal,⁷³ further highlighting the unique and broad specificity of OGT.

Previous findings⁷¹ provide further insights into the flexibility of OGT's substrate recognition. This study employed both biochemical and structural analyses to demonstrate that OGT does not rely on a strict consensus sequence but rather utilizes multiple contact points along the substrate, particularly with the TPR domain, to guide substrate binding and catalysis. The study highlighted the enzyme's capacity to accommodate and modify diverse protein sequences, reinforcing the notion of its broad substrate specificity.

1.5. Detection Methods for O-GlcNAc Modifications

1.5.1. Antibody-Based Detection Methods

One of the primary methods for detecting O-GlcNAc modifications involves the use of specific antibodies. Pan-specific antibodies have been developed to recognize and bind to various sets of O-GlcNAc-modified proteins, facilitating their detection and characterization.⁷⁴ These antibodies can be used in various immune-logical techniques, such as immunoblotting, immunoprecipitation, and immunofluorescence microscopy. Monoclonal and polyclonal antibodies against O-GlcNAc have been designed to target the O-GlcNAc moiety. Monoclonal antibodies, which typically obtained from a clonal population of a hybridoma, provide consistency, making them well suited for detecting O-GlcNAc in complex biological samples.⁷⁵ Polyclonal antibodies, on the other hand, are typically purified using the antigen. These polyclonal antibodies can recognize O-GlcNAc on proteins in different contexts, offering greater versatility.^{75,76}

Table 1.1 Non-comprehensive list of O-GlcNAc antibodies.

Antibody	Antigen	Isotype	Reference
RL2	Pore complex-lamina fraction purified from rat liver nuclear envelopes.	IgG1	(Snow et al., 1987)

10D8	N-acetylglucosamine	IgM	(Yoshida et al., 1989)
HGAC39	Heat-killed, pepsin-treated group A. streptococci.	IgG	(Turner et al., 1990)
HGAC85			
CTD110.6	YSPTS(β -O-GlcNAc)PSK	IgM	(Comer et al., 2001)
14D09.D4			
18B10.C7			
01F05.D6	GSTPVS(β -O-GlcNAc)SANM	IgG	(Teo et al., 2010)
09D01.E4			

One notable monoclonal antibody in the field is the IgG RL2, widely used in research for detecting O-GlcNAc-modified proteins.⁷⁷ However, RL2 was developed to target nuclear envelope proteins associated with the nuclear pore complex (NPC), which are particularly rich in O-GlcNAc modifications.⁷⁸ This targeted development means that RL2's affinity for O-GlcNAc is largely focused on NPC-related proteins, thus overlooking a significant portion of O-GlcNAc-modified proteins outside of the NPC. Developed by immunizing mice with a pore complex-lamina fraction from rat liver nuclear envelopes, RL2 shows a clear preference for NPC proteins and tends to bind more effectively to higher molecular weight proteins. This suggests that RL2's epitope may be associated with specific protein elements characteristic of these NPC proteins. Consequently, RL2's effectiveness is limited when detecting O-GlcNAc modifications on non-NPC proteins and lower molecular weight. Furthermore, RL2's cross-reactivity can complicate experimental results, making it less ideal for studies requiring comprehensive detection of O-GlcNAc modifications across a diverse range of proteins. This limitation contrasts with newer antibodies that provide moderately broader specificity and sensitivity.

Another antibody widely used in the field is the IgM CTD110.6 –developed through the immunization of mice with synthetic glycopeptides designed to mimic O-GlcNAc modifications on serine and threonine residues.⁷⁹ This targeted approach aimed to produce an antibody that specifically recognizes O-GlcNAc without cross-reacting with other similar carbohydrate structures. As an IgM, CTD110.6 has a pentameric structure, leading to increased avidity. This means the antibody can strongly bind to proteins with multiple O-GlcNAc modifications, beneficial for detecting such proteins. But, this same feature reduces its effectiveness for proteins with only one or a few O-GlcNAc sites, especially if those sites are sparsely distributed across the protein's surface. In these cases, the antibody's binding affinity drastically is lower, leading to less reliable detection. These aspects underscore the importance of considering the antibody's design and its potential limitations when applying it in various experimental contexts.

Recent advancements have introduced a range of new antibodies designed to detect O-GlcNAc with varying degrees of specificity and binding affinity, as shown in the table above. However, despite the sophisticated methods used in their development, these newer antibodies have gained little traction in the research community. This limited adoption is largely due to persistent challenges such as variability in their binding effectiveness across different protein contexts and the potential for cross-reactivity. Furthermore, these antibodies often struggle with specificity when applied to complex biological samples, where O-GlcNAc modifications may not align perfectly with the designed epitopes. As a result, the research community continues to rely on older, more established antibodies like RL2 and CTD110.6, despite their known limitations, because the newer options have not yet demonstrated significant advantages that would justify a widespread shift in their usage.

While existing antibodies are indispensable for detecting O-GlcNAc modifications, their significant limitations—including high costs, variability, and cross-reactivity—underscore an urgent need for more accessible and reliable solutions. The current reliance on these tools often hampers research progress due to their inconsistent performance across different biological contexts. In contrast, alternatives like nanobodies or single chain variable fragments (scFvs), which can be produced in bacterial systems at a fraction of the cost, offer a promising and necessary shift. These lab-made options could not only reduce expenses but also provide research labs with more consistent, versatile tools, addressing the current gaps and potentially revolutionizing the study of O-GlcNAc modifications.

1.5.2. Lectins and Glycan-Binding Proteins (GBPs)-Based Detection Methods

In addition to antibodies, lectins and glycan-binding proteins (GBPs) represent another class of tools used to detect O-GlcNAc modifications. Lectins are naturally occurring proteins that bind specifically to carbohydrate moieties, making them valuable for detecting glycosylated proteins. GBPs, which include lectins, are proteins that recognize and bind to specific carbohydrate structures.

Unfortunately, lectins bind with low affinity to their glycan ligands. Nevertheless, lectins such as wheat germ agglutinin (WGA)⁸⁰ have been extensively used to detect O-

GlcNAc modifications. WGA binds to *N*-acetylglucosamine residues, making it a useful probe for O-GlcNAc detection.^{74,81} However, its ability to also bind to sialic acid reduces its specificity and can complicate the interpretation of results in complex biological samples.⁸² Succinylation of WGA, however, has been reported to eliminate binding to sialic acid.⁸³ By conjugating WGA to various reporters, such as fluorescent dyes or enzymes, researchers can visualize and quantify O-GlcNAc modifications in different biological samples using techniques like lectin blotting, and lectin histochemistry.^{81,84}

Despite the considerable use of WGA for detecting O-GlcNAc modifications, its limitations prompted the design of G5-lectibody immunoprecipitation strategy.⁵⁶ This method combines multiple O-GlcNAc-specific antibodies (CTD110.6, RL2, HGAC39, and HGAC85) with WGA to enrich O-GlcNAc-modified proteins. However, the design of such complex system introduces several critical pitfalls. The varying affinities of the antibodies and WGA can result in cross-reactivity and non-specific binding, leading to contamination and a skewed representation of the O-GlcNAc proteome. Additionally, the accessibility of the O-GlcNAc epitopes, potential loss of low-abundance proteins, and inherent variability in immunoprecipitation procedures further undermine the reliability and quantitative accuracy of this approach. More recently several new antibodies binding O-GlcNAc have been described,⁷⁴⁻⁷⁶ but these have not seen widespread use.

In the context of GBPs, Mariappa et al. (2015) engineered a glycan-binding protein from a catalytically inactive mutant of OGA (*Cp*OGA–D298N) derived from *Clostridium perfringens*.⁸⁵ OGA evolved to bind O-GlcNAc moieties on proteins, making it a promising candidate for engineering GBPs, as it inherently possesses high affinity and specificity for O-GlcNAc. Replacing the catalytic aspartate 289 responsible for proton transfer during glycosidic bond hydrolysis with asparagine, inactivates the catalytic activity while allowing *Cp*OGA–D298N to bind O-GlcNAc-modified proteins with nanomolar affinity without cleaving the bond. However, a significant limitation is that once *Cp*OGAD298N binds to O-GlcNAc, it remains bound and does not facilitate removal of the modification, which could restrict its utility in certain applications. These developments, while promising, still face challenges that may limit their widespread adoption. Compared to the established antibody-based methods, which remain the standard due to their accessibility, these alternatives present opportunities for future integration but require further refinement to address current gaps in detection specificity and accuracy.

1.5.3. Labeling Techniques for O-GlcNAc Detection

Another powerful approach for detecting O-GlcNAc modifications involves labeling techniques, such as isotopic labeling, metabolic labeling, and chemoenzymatic modification. These methods can provide high sensitivity and specificity, making them invaluable for detailed analysis of O-GlcNAcylation.

In the realm of *in vitro* labeling techniques for detecting O-GlcNAc modifications, two particularly effective methods stand out. One involves the use of GlcNAc-specific sulfotransferases to transfer a radioactive sulfate ³⁵S onto O-GlcNAc residues, which are then separated by SDS-PAGE and visualized through autoradiography.⁸⁶ Another method uses β-1,3-N-acetylgalactosaminyltransferase 2 (B3GALNT2) to attach an azide-labeled sugar, *N*-azidoacetylgalactosamine (GalNAz), to O-GlcNAc sites on proteins. The azide group on GalNAz is particularly useful because it acts as a chemical handle that can undergo a highly specific reaction.⁸⁷ This reaction allows for the azide group to be easily linked to a fluorescent tag or other detection molecules, enabling precise visualization of O-GlcNAcylation patterns in proteins. While primarily used *in vitro*, this method has also been adapted for *in vivo* applications in limited settings, expanding its potential for studying O-GlcNAc modifications within living organisms. This technique, while offering specificity in labeling and studying O-GlcNAc modifications, has limitations such as the need for specialized reagents and the potential for non-specific labeling in complex biological samples.

Chemoenzymatic labeling strategies have become essential *in vitro* tools for detecting and analyzing O-GlcNAc modifications. One such method focuses on a chemoenzymatic histology approach, where an engineered β-1,4-galactosyltransferase (B4GalT1) enzyme is used to tag O-GlcNAc residues in tissue specimens with a detectable marker. The engineered form of B4GalT1 carries a Y289L mutation, which alters the enzyme's substrate specificity, allowing it to efficiently use GalNAc-based monosaccharides as substrates. This modification enhances its utility in O-GlcNAc detection by enabling the incorporation of modified sugars that can be functionalized in a second chemical step to install detectable markers.⁸⁸ This technique allows researchers to map O-GlcNAc distribution patterns across various tissues, revealing differences between healthy and diseased states, such as in Alzheimer's disease and cancer.⁸⁸ The second method employs an engineered B4GalT1 enzyme to attach a galactose derivative

bearing an azide, alkyne, or ketone group to O-GlcNAc-modified proteins. These modifications enable, subsequent tagging with fluorescent or biotin labels through bioorthogonal chemistry. This enables not only the direct detection of these proteins in gels and mass spectrometry but also allows for their imaging within cells, making it possible to observe dynamic changes in O-GlcNAc levels.⁸⁹ A third strategy uses a mutant version of B4GALT1 to transfer a ketone group onto O-GlcNAc residues, which can then be linked to biotin for chemiluminescent detection, offering a rapid and sensitive alternative to traditional radioactive methods.³⁹ These chemoenzymatic approaches provide powerful and versatile tools for studying O-GlcNAc modifications in a controlled environment, enhancing our understanding of their roles in various biological processes.

In vivo labeling of O-GlcNAc-modified proteins involves the use of metabolic chemical reporters (MCRs), which are analogues of monosaccharides that incorporate bioorthogonal functionalities – chemical groups that can be added to molecules in living cells without interfering with normal biological processes – into cellular glycoproteins. One approach uses *N*-azidoacetylglucosamine (GlcNAz), which is metabolically incorporated into nuclear and cytoplasmic proteins. The azide groups in GlcNAz-labeled proteins can be selectively tagged with biochemical probes, facilitating the identification of O-GlcNAc-modified proteins.⁹⁰ Another approach involves 6-azido-6-deoxy-N-acetyl-glucosamine (6AzGlcNAc), a specific MCR for O-GlcNAcylated proteins that avoids incorporation into other glycosylation pathways, providing a more targeted labeling method.⁹¹ A third method uses 2-azido-2-deoxy-glucose (2AzGlc), which selectively labels intracellular O-GlcNAc modifications. However, 2AzGlc is not dynamically removed from protein substrates and can be toxic at higher concentrations, highlighting the need for careful optimization.⁹² Together, these approaches offer powerful tools for studying O-GlcNAc modifications in living cells, although they require careful consideration of specificity, toxicity, and metabolic pathways.

Two notable isotopic *in vivo* labeling techniques for detecting O-GlcNAc modifications involve the use of radiolabeled sugar nucleotides. One approach utilizes bovine milk galactosyltransferase (B4GALT1) and UDP-[³H] galactose to label O-GlcNAc residues on proteins in intact or permeabilized cells. B4GALT1 is a member of the β-1,4-galactosyltransferase family and is commonly referred to as bovine milk galactosyl-transferase when it forms part of the lactose synthase complex with α-lactalbumin, a protein found in milk. In this complex, α-lactalbumin modifies B4GALT1's substrate

specificity, enabling lactose synthesis.⁹³ This method is particularly effective for identifying O-GlcNAc-modified proteins in the cytoplasm. Another technique involves metabolic labeling with [³H] glucosamine, which is incorporated into O-GlcNAc-bearing proteins during their synthesis in cells.⁹⁴ This method, combined with subcellular fractionation, allows for the detection and analysis of O-GlcNAc on low-abundance proteins with high specificity and sensitivity. Both techniques provide powerful tools for studying O-GlcNAc modifications, though they require careful optimization and substantial amounts of protein for accurate analysis. Despite their advantages, labeling techniques can be complex to implement and require careful optimization to ensure specificity and efficiency. However, when combined with other detection methods, such as antibodies or lectins, labeling techniques offer a powerful complementary approach to studying O-GlcNAc modifications.

1.6. Challenges and Gaps in O-GlcNAc Detection

Despite significant advancements in the field, several challenges and gaps remain in the detection and characterization of O-GlcNAc modifications. These challenges are primarily related to measure stoichiometry of O-GlcNAc on specific proteins, temporal and spatial resolution of O-GlcNAc with cells and tissue samples, limited sensitivity and specificity. The *gap* in this context refers to the lack of robust tools and techniques that can simultaneously address these challenges. For example, there is a need for quantitative approaches that reliably determine the dynamic range of O-GlcNAc modifications across different cellular states and within subcellular compartments. Furthermore, the gap also includes the limited availability of specific antibodies or probes that can differentiate O-GlcNAc modifications from other glycosylation types or post-translational modifications. These limitations hinder a comprehensive understanding of O-GlcNAc's functional roles in complex biological systems.

1.6.1. Stoichiometry

One of the primary challenges in studying O-GlcNAc modifications is their substoichiometric nature. O-GlcNAc modifications often occur only on specific proteins at low abundance and only on specific subsets of proteins, making them difficult to detect amidst the vast array of other post-translational modifications. This low stoichiometry

complicates quantitative analysis, as traditional detection methods do not simultaneously detect glycosylated and unglycosylated sites with a protein of interest.

To address the challenge of substoichiometric O-GlcNAc modifications, various strategies have been attempted, each with its own limitations. The S-GlcNAc approach involves substituting the ser/thr O-GlcNAc site with cysteine, allowing for the addition of a thiol-reactive group that mimics O-GlcNAc but resistant to cleavage by OGA.^{72,73} While this method provides a stable and specific way to label glycosylation sites, it cannot be reversed, which limits the study of dynamic O-GlcNAc processes and restricts its application to static analyses.

In vitro labeling techniques using chemoenzymatic methods, such as those involving mutant β-1,4-B4GalT1, aim to enhance detection sensitivity by tagging O-GlcNAc residues with detectable markers.^{88,89} However, these methods are typically restricted to isolated proteins or cellular extracts, making them less effective for real-time or *in vivo* studies. While they offer some improvement in detection, they don't fully address the complexities of O-GlcNAc modifications within living systems.

In cell labeling approaches, like dual RNA aptamers, strive to modulate O-GlcNAc levels directly within living cells. The dual RNA aptamers are engineered to simultaneously bind to OGT and a specific target protein. By bringing OGT near the target protein, the aptamers enhance the likelihood that O-GlcNAc will be added to specific ser/thr residues on the protein. This allows for precise modulation of O-GlcNAcylation, either increasing the glycosylation of the target protein based on the aptamer design.⁹⁵ This method provides a more controlled way to study low-abundance proteins, but the complexity of designing aptamers that are both effective and specific can hinder their practical application. Additionally, the potential for off-target effects remains a significant concern.

Nanobody-based strategies attempt to selectively modulate O-GlcNAc levels on specific proteins using nanobodies fused to OGA or OGT. These nanobody constructs are designed by fusing a nanobody that specifically recognizes a target protein with OGA or OGT.⁹⁶ When the nanobody binds to its target protein, the attached OGA can remove existing O-GlcNAc sites, or the attached OGT can add new O-GlcNAc sites to specific proteins. This construct allows for targeted control over the O-GlcNAcylation state of the protein, providing a tool for studying the functional impact of O-GlcNAc modifications in living cells. While these fusions offer a targeted approach, they are not without their challenges, including specificity issues and the risk of unintended interactions. The

effectiveness of these strategies is often compromised by the inherent difficulties in achieving precise control over O-GlcNAc modifications.

While these methods represent efforts to tackle the issue of substoichiometric O-GlcNAc modifications, they all have significant limitations that prevent them from being fully reliable or comprehensive. The current tools, though innovative, fall short in various respects, underscoring the need for the development of new approaches that can more effectively address the challenges posed by O-GlcNAc modifications. Future research should focus on creating methods that offer greater specificity, sensitivity, and applicability across different biological contexts.

1.6.2. Temporal and Spatial Resolution

Achieving high temporal and spatial resolution in O-GlcNAc detection is crucial for understanding the dynamic nature of this modification. O-GlcNAc modifications can rapidly change in response to various cellular signals, necessitating techniques that can capture these modifications in real-time and at specific cellular locations. Current methods often lack the resolution required to map these changes accurately, hindering our ability to fully understand the temporal dynamics and spatial distribution of O-GlcNAcylation.

However, some progress has been made in addressing these challenges. For instance, Liu et al. (2016) developed a method that combines time-course metabolic labeling with ChIP-seq (chromatin immunoprecipitation followed by sequencing) to track the turnover of O-GlcNAc modifications on chromatin-associated proteins at specific genomic loci in *Drosophila* larvae, providing insights into how O-GlcNAcylation changes over time in response to cellular signals.¹⁰ Despite these advances, the resolution is limited by the inherent stability of the modification and the complexity of chromatin dynamics.

Similarly, Rexach et al. (2012) employed a chemoenzymatic method that uses an engineered B4GalT1 to transfer a radiolabeled or fluorescent galactose onto O-GlcNAc residues on the transcription factor CREB.⁹⁷ This allows for the precise spatial detection of O-GlcNAc modifications at specific protein sites, which can then be analyzed through methods such as immunoprecipitation and fluorescence microscopy, providing insights into how these modifications are regulated in response to neuronal activity. While this method offered precise detection at specific protein sites, it primarily focused on individual

proteins rather than capturing a broader proteomic landscape, thus limiting its spatial scope.

Another approach by Rexach et al. (2012) a chemoenzymatic strategy that utilizes a mutant β -1,4-galactosyltransferase to transfer a ketone group onto O-GlcNAc-modified proteins.⁹⁷ This ketone group is then reacted with a fluorescent or biotinylated probe, enabling the direct in-gel detection, proteomic analysis, and spatial visualization of O-GlcNAc modifications within live cells or tissue samples, helping to map where these modifications occur. Nonetheless, it still faces challenges in fully capturing the temporal dynamics of O-GlcNAc across different cellular environments. These examples illustrate the ongoing efforts to improve temporal and spatial resolution in O-GlcNAc detection, yet they also underscore the existing limitations. The field would benefit greatly from the development of new methodologies that can better capture the rapid and localized changes in O-GlcNAcylation, ultimately providing a more comprehensive understanding of this critical post-translational modification.

1.6.3. Sensitivity and Specificity

The detection sensitivity and specificity of O-GlcNAc modifications pose significant challenges. While antibodies and lectins can provide specific recognition of O-GlcNAc, they are not without limitations. Antibodies, lectins, and GBPs may cross-react with other glycans, leading to potential false positives. Additionally, the sensitivity of these detection methods must be sufficient to identify low-abundance O-GlcNAc modifications, which is often challenging in complex biological samples.

Recent advancements have aimed to improve the sensitivity and specificity of O-GlcNAc detection, addressing the challenges posed by low-abundance modifications and potential cross-reactivity. Zhu et al. (2020) developed a tandem bioorthogonal labeling strategy combined with SILAC (stable isotope labeling by amino acids in cell culture) to uncover endogenous co-translational O-GlcNAc modifications on nascent proteins.⁹⁸ This method involves two key steps: first, the incorporation of azide-labeled methionine analogs into nascent polypeptides during protein synthesis, followed by the tagging of O-GlcNAc-modified proteins using a bioorthogonal click chemistry reaction. SILAC is then used to quantitatively analyze these labeled proteins, providing high sensitivity and specificity by enriching only those proteins that are both newly synthesized and O-GlcNAc-modified.

This approach minimizes false positives through rigorous controls and precise timing of cotranslational modification, offering a substantial improvement over traditional antibody-based methods in complex biological samples.

While progress has been made in enhancing the sensitivity and specificity of O-GlcNAc detection, particularly through innovative approaches like tandem bioorthogonal labeling combined with SILAC, the field still faces critical challenges. The ability to detect low-abundance O-GlcNAc modifications with high precision remains essential for accurately mapping the functional roles of these modifications in complex biological systems. Despite these advancements, current methods still struggle with issues such as incomplete coverage, potential cross-reactivity, and the need for extensive controls to avoid false positives. Moving forward, the development of even more refined techniques—potentially integrating advanced proteomic technologies, higher-resolution imaging, and more robust quantitative methods—will be crucial. Such advancements will not only improve our understanding of O-GlcNAc's role in cellular processes but also open new avenues for exploring its implications in disease and therapeutic interventions.

1.6.4. Biophysical Properties

The biophysical properties of O-GlcNAc modifications introduce specific challenges that are deeply rooted in the fundamental principles of physics. O-GlcNAc is a small, polar modification that introduces subtle changes to the electrostatic landscape of proteins. The addition of the GlcNAc group, though charge-neutral, alters the dipole moment of the modified residue and its immediate surroundings, influencing the local electric field and potentially altering the protein's interaction with solvent molecules and other macromolecules. This change in polarity can affect the hydrogen bonding network within the protein, leading to shifts in protein conformation and dynamics that are critical to its function.

One advancement in understanding the biophysical effects of O-GlcNAc is the finding that O-GlcNAc modification of tau protein inhibits its aggregation without significantly altering the overall conformation of tau monomers.⁴⁶ The study demonstrated that O-GlcNAcylation enhances the solubility of tau by stabilizing its native state or preventing the formation of aggregation-prone intermediates, as observed through Förster resonance energy transfer (FRET) and nuclear magnetic resonance (NMR) spectroscopy.

This finding underscores the importance of O-GlcNAc in modulating protein stability and suggests that its influence on aggregation is a critical area for further research.

Moreover, the charge-neutral nature of O-GlcNAc, unlike the phosphorylation modification which adds negative charge, presents a unique challenge for detection and analysis. Traditional methods that rely on charge-based separations, such as electrophoresis, may struggle to differentiate O-GlcNAc-modified proteins from their unmodified counterparts, leading to difficulties in accurately quantifying and characterizing these modifications. The small size and polarity of O-GlcNAc also contribute to its lability, making it susceptible to loss during sample preparation and analysis, particularly under harsh conditions.⁹⁹ This lability necessitates the development of more refined analytical techniques that can preserve the integrity of O-GlcNAc modifications while providing high sensitivity and specificity.

The structural diversity introduced by O-GlcNAc further complicates biophysical studies. Proteins modified by O-GlcNAc may exhibit a wide range of conformations and dynamic behaviors, influenced by the specific site of modification and the surrounding protein context. Capturing this diversity without sacrificing sensitivity or specificity remains a significant gap in the field. Current analytical approaches often fall short in their ability to fully resolve the subtle biophysical changes induced by O-GlcNAc, leading to incomplete or inaccurate representations of its role in protein function.

To advance the field, there is a need for the development of biophysical techniques that can more precisely characterize the effects of O-GlcNAc on protein structure, dynamics, and interactions. Approaches that consider the polarity, charge, and conformational flexibility introduced by O-GlcNAc will be crucial in bridging these gaps. Integrating high-resolution structural methods, such as NMR spectroscopy and cryo-electron microscopy, with advanced computational modeling could provide deeper insights into the biophysical implications of O-GlcNAc modifications. By focusing on these fundamental properties, the field can move towards a more comprehensive understanding of how O-GlcNAc influences protein behavior at a molecular level.

1.7. Conclusions

O-GlcNAc is a critical post-translational modification that plays a vital role in various cellular processes, including transcription, stress response, and protein

homeostasis. Its significance is underscored by its involvement in numerous diseases, from neurodegenerative disorders to cancer, highlighting the necessity for robust detection and analysis methods. However, the study of O-GlcNAc modifications is fraught with challenges. The substoichiometric nature of O-GlcNAcylation, combined with its lability and structural diversity, complicates its detection and characterization, requiring innovative approaches that can accurately capture these modifications in real-time and within specific cellular contexts.

Advancements in detection techniques, such as the development of novel antibodies, metabolic labeling, and chemoenzymatic methods, have provided researchers with more powerful tools to study O-GlcNAc. Yet, these methods often fall short in terms of sensitivity, specificity, and the ability to fully capture the dynamic and complex nature of O-GlcNAcylation. While significant progress has been made, the current methodologies still face limitations that impede a comprehensive understanding of O-GlcNAc's role in cellular function and disease.

Moreover, the biophysical properties of O-GlcNAc present additional challenges. The modification's small, polar nature affects protein stability, folding, and interactions, but these changes are difficult to detect and quantify with existing tools. As research continues to uncover the intricate details of O-GlcNAc's impact on protein function, there is a pressing need for the development of more refined biophysical techniques that can provide high-resolution insights into these modifications.

While the field has made considerable strides in understanding and detecting O-GlcNAc modifications, significant gaps remain. Here, a *significant gap* refers to substantial unresolved questions or methodological limitations that, if addressed, could profoundly expand our understanding or enable transformative breakthroughs. Even though O-GlcNAc was first discovered over 40 years ago, each advance tends to reveal additional layers of complexity, ensuring that new lines of inquiry continue to emerge. Future research must focus on developing new methodologies that address the current limitations, particularly in achieving higher sensitivity, specificity, and resolution. By overcoming these challenges, we can advance our understanding of O-GlcNAc and its implications in health and disease, ultimately paving the way for new therapeutic strategies and diagnostic tools.

Chapter 2.

ENGINEERING A SCFV FOR O-GLCNAC DETECTION

The research presented in this chapter was supported by contributions from several individuals. Dr. Subramania Kolappan provided critical assistance with protein purification, particularly with the denaturing purification protocol for scFv(RL2). Adrian Plata-Ruiz synthesized the glycosylated peptides used in binding assays, enabling validation of scFv(RL2)'s specificity. The crystallization screening was conducted at the National Crystallization Center at the Hauptman-Woodward Medical Research Institute. Their contributions were instrumental to the development and characterization of the scFv described herein.

2.1. Background and Importance

The ability to detect O-GlcNAc modifications is crucial for advancing our understanding of the role of O-GlcNAc in cellular processes and disease mechanisms. Various techniques have been developed for O-GlcNAc, each with distinct advantages and limitations:

Lectin-based detection involves using lectins to bind O-GlcNAc residues,^{56,81} but the low specificity due to cross-reactivity and false positives is a significant limitation, often caused by lectins' low affinity for individual carbohydrate residues, despite high avidity from multivalent interactions.¹⁰⁰ Glycan binding proteins (GBP), like OGA catalytically dead mutant (*mutOGA*), can bind O-GlcNAc with high affinity,⁸⁵ but their complex production, time-intensive directed evolution, and difficulty distinguishing similar glycan structures limit their use.¹⁰¹ Chemoenzymatic labeling adds detectable tags to O-GlcNAc residues via engineered glycosyl-transferases,^{39,88-92} but the dependency on specialized, costly, and unstable enzymes makes this method less practical for routine use, further complicated by multi-step protocols.¹⁰² Mass spectrometry provides detailed molecular insights into O-GlcNAc modifications,^{32,103,104} yet challenges with oxonium ion formation during fragmentation and poor peptide sequence coverage hinder accurate site mapping, requiring additional complex steps to overcome these limitations.

Various polyclonal and monoclonal antibodies that can detect O-GlcNAc have been reported.^{75,77,79,105} Some of these have seen more common use where as others, such as 14D09.D4,⁷⁵ 10D8,¹⁰⁵ and HGAC39¹⁰⁶ have seen limited use in detecting O-GlcNAc-modified proteins. The most widely used antibodies are RL2⁷⁷ and CTD110.6⁷⁹. Newer antibodies targeting O-GlcNAc have demonstrated minimal improvements in specificity or reproducibility. This may rise in part from inertia but also from perceived problems with cross-reactivity and variable performance. To this point, these limitations have led researchers to combine antibodies with lectin-based techniques to enhance detection sensitivity.⁵⁶ This approach underscores the ongoing challenges with current antibodies, which have shown little improvement in specificity or performance. Despite attempts to develop new antibodies, the performance of these newer reagents remains unclear and have not been replicated by other researchers ultimately. Recently, Burt et al. (2021) developed a mixture of anti-O-GlcNAc monoclonal antibodies intended for the enrichment of native O-GlcNAc-modified peptides.⁷⁶ While these antibodies showed some ability to detect O-GlcNAc modifications in mouse brain tissue, their practical application is limited. The production process is complex, and the requirement for specialized mass spectrometry techniques hinders their accessibility for routine use. Moreover, their performance across different experimental settings remains uncertain, and there has been limited independent validation of their efficacy. These limitations highlight the ongoing need for more practical, versatile, and easily producible tools for O-GlcNAc detection.

RL2 has been widely used in O-GlcNAc research since the late 1980s,⁷⁷ and is similar in performance to CTD110.6,⁷⁹ another commonly used pan-specific anti-O-GlcNAc antibody. While both RL2 and CTD110.6 are effective in detecting O-GlcNAc modifications, they clearly recognize different sets of O-GlcNAc-modified proteins, plus there are important distinctions between them. RL2 is an IgG, whereas CTD110.6 is an IgM. (Figure 2.1) This distinction between antibody classes has important implications both for their use but also for exploiting ese on further antibody engineering. IgGs, such as RL2, are often preferred for detailed studies and further development because of their simpler and more consistent structure. IgGs are composed of two heavy and two light chains, with a well-defined hinge region¹⁰⁷ that allows for relatively straightforward identification of the variable regions—critical elements for antigen recognition.¹⁰⁸ The ability to easily map these regions makes IgGs particularly suitable for downstream antibody engineering. In contrast, IgMs like CTD110.6 are pentameric, consisting of five monomer units linked by a J chain.¹⁰⁷ This complex structure can complicate the process

of mapping the variable regions because interactions with antigens may extend across multiple monomers. Additionally, the larger quaternary structure of IgM antibodies can make them less stable and more prone to aggregation, posing challenges for further development.¹⁰⁸ Finally, IgMs typically show lower monomeric affinities to their target antigens.

RL2's development process laid the foundation for its unique characteristics. This involved enriching nuclear pore complex (NPC) from rat liver cells through high-salt extraction to obtain glycoproteins containing O-GlcNAc modifications.⁷⁷ This fraction of enriched nuclear pore glycoproteins was then used to immunize mice, prompting the generation of an immune response. Spleen cells from the immunized mice were fused with myeloma cells to produce hybridomas, which were then screened to identify those producing antibodies specific to O-GlcNAc-modified proteins. Some of the resulting hybridomas were then cultured as clonal populations and from among this one clone producing a particular promising anti-O-GlcNAc antibody was identified. RL2 produced by this clone, was then purified and characterized through immunoblotting and immunoabsorption, confirming its specificity for O-GlcNAc. This process established RL2 as a reliable tool for O-GlcNAc detection.

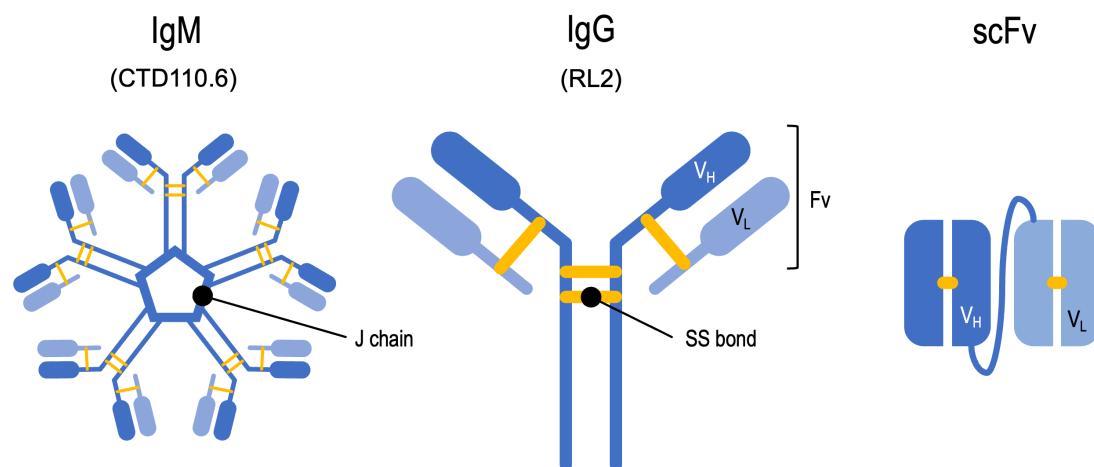


Figure 2.1 Schematic comparison of IgM, IgG, and scFv structures.

The IgM antibody is depicted as a pentamer, with five monomer units connected by a J chain, forming a complex structure. The IgG antibody consists of two heavy chains and two light chains, with variable regions (VH and VL) responsible for antigen binding. The scFv retains the variable regions of the IgG, connected by a flexible linker, resulting in a smaller and more versatile single-chain structure.

Nevertheless, given the advantages of RL2 as an IgG—particularly its simpler structure and high affinity for O-GlcNAc^{56,77}—it stands out as an ideal candidate for developing a more versatile and robust detection tool. This led us to the decision to develop a single-chain variable fragment (scFv) derived from RL2. A scFv is a smaller, engineered antibody fragment that retains the antigen-binding specificity of the original IgG. It consists of the variable regions of the heavy (V_H) and light (V_L) chains connected by a flexible linker.¹⁰⁹ (Figure 2.1) The scFv format offers several significant benefits over full-length antibodies:

- **Smaller size:** scFvs are typically only 25–30 kDa as compared to their parent IgG which are approximately 150 kDa. Their reduced size enhances their tissue penetration and reduces steric hindrance, making scFvs particularly useful in applications where staining of tissues is required or in cases where an epitope is in a sterically congested environment.¹⁰⁹
- **Increased stability:** Because they are generally composed of a single polypeptide, scFvs tend to be more stable than full-length antibodies, especially reducing conditions or in the presence of mild denaturants.¹⁰⁹
- **Ease of production:** scFvs can be easily produced in various expression systems, such as bacteria or yeast, making them much more cost-effective and accessible at large scale.¹⁰⁹
- **Versatility:** scFvs can be readily engineered to include functional tags, such as biotin or fluorescent markers, which expand their utility in different experimental contexts, facilitating detection, purification, and imaging.¹⁰⁹

The combination of RL2’s specificity and the inherent advantages of the scFv format presents a potentially powerful combination that could yield a robust and versatile tool for O-GlcNAc detection. The development of an RL2-derived scFv could represent an important practical advance in O-GlcNAc research, providing a tool that is not only specific and stable but also adaptable to a wide range of experimental needs.

While methods such as phage and yeast display have become dominant in recent years, particularly for carbohydrate-targeting scFvs (Table 2.1), the conceptual extraction and combination of V_H and V_L regions from existing monoclonal antibodies remains a robust method. The development of an RL2-derived scFv builds on this well-established method that has been used since the first scFv was developed by Huston et al. (1988),¹¹⁰

and it continues to be a reliable and frequently reported approach in scFv research through 2024. This method is particularly effective when the IgG antibody is already well characterized, as in the case of RL2's specificity for O-GlcNAc. Despite its long-standing use, there are no reports of scFvs, or any other antibody fragment, being used for against O-GlcNAc detection, making this application of RL2-derived scFv both timely and innovative. A summary of recent developments in scFv targeting carbohydrate antigens, demonstrating that while various methods are employed, the direct derivation of scFvs from existing antibodies remains a powerful tool. In this context, developing an RL2-derived could fill an important gap, setting the stage for the objective chief of this chapter.

Table 2.1 Recent developments of scFv targeting carbohydrates.

Target	Target composition	Development method	Reference
Alginate	ManA(β1-4)GulA(α1-4)] _n	V _H and V _L regions derived from existing antibody MFb	(Gao et al., 2020)
Disialoganglioside (GD2)	GalNAc-β1,4-(Neu5Ac)-β2,3-Gal-β1,4-Glc	V _H and V _L regions derived from existing antibody ch14.18	(Seitz et al., 2020)
F77	α1,2-fucose	V _H and V _L regions derived from existing antibody F77	(Grover et al., 2023)
Glucose polymer	β-1,3-glucans	V _H and V _L regions derived from existing murine antibody 2G8	(Di Mambro et al., 2022)
Glucuronoxylomannan (GXM)	αManp-(1-4)-[βXylp-(1-2)]-αManp-(1-4)-[βGlcA-(1-2)]-αManp	V _H and V _L regions derived from existing antibodies 2H1 and 18B7	(Sharma et al., 2020)
Heparan sulfate	GlcA-GlcNAc GlcA-GlcNS IdoA-GlcNS IdoA(2S)-GlcNS IdoA-GlcNS(6S) IdoA(2S)-GlcNS(6S)	Structure-guided engineering, phage display	(Damen et al., 2024)
Lewis Y antigen	Fucα1-2Galβ1-4[Fucα1-3]GlcNAcβ1	V _H and V _L regions derived from existing antibody	(J. Chen et al., 2021)
Lipoarabinomannan (LAM)	PI-Manp-(1-6)-Manp-(1-2)-Araf	Phage display using V _H and V _L regions derived from leporine antibodies	(Yan et al., 2021)
Lipopolysaccharide (LPS)	KDO-Heptose-Glc-Gal with variable O-antigen	Phage and yeast display using V _H and V _L regions derived antibodies library	(Fux et al., 2024)
N-glycolylneuraminic acid	Neu5Gc	Phage display using V _H and V _L regions derived from galline antibodies library	(H. Wang et al., 2021)
STn antigen	Neu5Ac-α2,6-GalNAc	V _H and V _L regions derived from existing antibody HsglG4	(Loureiro et al., 2020)
Thomsen-Friedenreich antigen	Galβ1-3GalNAc	V _H and V _L regions derived from existing antibody nemod-TF	(Dragon et al., 2023)
Tn antigen	GalNAc	Yeast display using V _H and V _L regions derived from existing antibody 237	(Sharma et al., 2020)

Tn antigen	GalNAc	Yeast display using V _H and V _L regions derived antibody library	(Guerrero-Ochoa et al., 2020)
------------	--------	--	-------------------------------

2.2. Objective

The primary objective of this chapter is to develop a scFv derived from the IgG RL2 monoclonal antibody that can be used to specifically recognize O-GlcNAc-modified proteins. RL2 was selected for scFv development primarily because it is an IgG1, a subclass known for its stability and high affinity for their ligands, which we envisioned would result in this antibody being easier to sequence compared to IgM antibodies like CTD110.6. Moreover, the consistent performance of RL2 in O-GlcNAc detection⁷⁷ further supports its suitability for engineering. By leveraging the benefits of scFvs, this work aims to create a new biochemical tool that can help address the current challenges associated with O-GlcNAc detection, such a tool would benefit from improved stability, ease of production, and the ability to incorporate various functional modifications. Functionalization, such as biotinylation, polymerization, fluorescent labeling, or nanoparticle conjugation, can be efficiently achieved through site-specific modification of engineered residues like cysteine or lysine.¹¹¹ In the scFv format, these modifications are straightforward due to the monomeric structure and low molecular weight (approximately 25 kDa), which enable readily expression in bacterial systems like *E. coli*. This allows for uniform attachment of functional groups without potential interference from glycosylation or the presence multiple disulfide bonds that are present in full-length IgG molecules like RL2.¹¹² For example, biotinylation of scFvs can be performed using chemical conjugation or enzyme-mediated methods such as BirA ligase,^{113,114} while fluorescent labeling can be achieved via simple maleimide-thiol chemistry, both of which offer high specificity and reproducibility.¹¹¹ These functional modifications, which are more challenging to introduce into full-length antibodies, allow scFvs to be more easily adapted for use in various detection assays and imaging applications. In summary, a scFv would be expected to serve as a valuable resource for further studies on the role of O-GlcNAc modifications in various biological and pathological processes.

2.3. Methods

2.3.1. Antibody Sequencing

Commercially available RL2 monoclonal antibody was subjected to detail mass spectrometry (MS) analysis to determine the complete amino acid sequences of both its heavy and light chains. This process involved enzymatic digestion involved multiple enzymatic digestions using trypsin, chymotrypsin, pepsin followed by MS and tandem MS (MS/MS) analysis of the resulting proteolytic peptides. The analysis revealed that the light chain is of the kappa (κ) type, (Figure 2.2) while the heavy chain is of the gamma (γ) 1 (IgG1) subclass (Figure 2.3). κ light chains are one of the two light chain types in antibodies, contributing to the antigen-binding site, while the γ 1 heavy chain indicates that RL2 is of the IgG1 subclass, known for its strong immune functions and versatility in research applications.¹⁰⁷

Figure 2.2 RL2 light chain κ amino acid sequence.

V _H domain	CDR 1	CDR 2	
QVQLQQSGTELVRPGTSVKVSCKTSGYAFTNYLIEWIKQRPGQGLEWIGVINPGSGITNY		CDR 2	60
NEFKDKATLTADKSSSTAYIQQLSSLTSDDSAVYFCARTPASFAYWGQGTIVTVSAAKTT	CDR 3		120
PPSVYPLAPGSAAQTNSMVTLGCLVKGYFPEPVTVWNNGSLSGVHTFPAVLQSDLYTL			180
SSSVTVPSSWPSETVTCNVAHPASSTKVDKKIVPRDCGCKPCICTVPEVSSVFIFPPKP			240
KDVLTTILTPKVTCVVVDISKDDPEVQFSWFVDDVEVHTAQTPREEQFNSTFRSVELP			300
IMHQDWLNGKEFKCRVNSAAFPIEKTISTKGRPKAPQVYTIPPPKEQMAKDVKVSLTC			360
MITDFFPEDITVEWQWNGQPAENYKNTQPIMDTDGSYFVYSKLNVQKSNNWEAGNTFTCSV			420
LHEGLHNHHTEKSLSHSPKG			440

The de novo sequencing was performed by Rapid Novor Inc., using their proprietary REpAb® technology. This technology employs high-resolution tandem mass spectrometry (MS/MS) to achieve comprehensive coverage and accurate sequence determination of antibody peptides. The peptides were analyzed using high-resolution MS with Thermo-Fisher Q Exactive™ and Orbitrap Fusion™ mass spectrometers. This analysis achieved full sequence coverage for both the heavy and light chains, with each amino acid residue being covered by at least five peptide scans. The identification of these

peptides was supported by significant fragment ions, adding to the confidence in the sequence accuracy.

Figure 2.3 RL2 light chain κ amino acid sequence.

V _L domain	CDR 1	CDR 2	
DIVMTQSPSSLTVTAGERVTMNCRSSQSLLHSENQKNYL	WYQQKPGQPPKLLIY	WTSTR	60
ESGVPDFRTGSGSGTDFTLTISSVQAEDLAVYYC	QNDYGYPLT	FGAGTKLELKRADAAPT	120
VSIFPPSSEQLTSGGASVVCFLNNFYPKDINVWKWIDGSERQNGVLNSWTQDSKDSTYS			180
MSSTLTLLTKEDEYERHNSYTCEATHKTSTSPIVKSFNRNEC			220

The variable regions, including the complementarity-determining regions (CDRs), were fully annotated using the Chothia scheme, a structural classification system that defines the CDR loops based on their three-dimensional conformations.¹¹⁵ These CDRs are key structural elements within the variable regions of both heavy and light chains, forming loops that directly interact with the antigen—a molecule specifically recognized by the antibody and capable of binding to it, though not all antigens necessarily elicit an immune response. All CDR loops interact with the antigen. CDR3 is the one that has the most interaction with the antigen but this is by virtue of its size compared to the other loops. I am not sure if this translates as the CDR loop that binds as they all do. These regions are the most diverse parts of the antibody, exhibiting a high degree of variability in their amino acid sequences. This variability allows the CDRs to create a wide range of binding sites with distinct shapes and chemical properties, enabling antibodies to bind specifically to a vast array of antigens with different structures and surfaces. This specificity is crucial for the immune system's ability to target various pathogens and foreign molecules. Challenging residues like leucine and isoleucine, which have identical masses, were confidently identified through a combination of W-ion detection, enzyme specificity, and statistical residue distribution. Additionally, modifications such as N-linked glycosylation at Asn290 on the heavy chain and pyroglutamate formation at the N-terminus were identified, offering a comprehensive characterization of the RL2 antibody.

2.3.2. Vector Design

The first step in the scFv(RL2) design was to identify and extract the V_H and V_L regions from the sequence analysis of the RL2 antibody, as these regions are crucial for

constructing the antigen-binding domain. (**Figure 2.4**) The abovementioned Chothia scheme was used to identify these regions within the antibody sequence. To further confirm the accuracy of the extracted V_H and V_L sequences, the NCBI Conserved Domain Database (CDD)¹¹⁶ was used to cross-verify the sequence of these domains. This analysis confirmed that the extracted sequences corresponded precisely to the expected variable regions necessary for constructing a functional scFv(RL2), thereby preserving the antibody's specificity for its target. Based on the domain analysis, the V_H and V_L regions were linked using the commonly used flexible linker $(G_4S)_3$. This linker is known to maintain the functional integrity of the scFv(RL2) while allowing the necessary flexibility between the variable domains. Although other linkers, such as $(GS)_4$ and $(GS)_5$, are sometimes employed depending on the desired stability and function, $(G_4S)_3$ is frequently preferred for its balance between flexibility and stability.^{117,118}

Figure 2.4 Full amino acid sequence of the scFv(RL2)

TEV site	V_H domain	
	ENLYFQGSSGQVQLQQSGTELVRPGTSVKVSCKTSGYAFTNYLIEWIKQRPGQGLEWIGV	60
	INPGSGITNYNEKFKDATALTADKSSTAYIQLSSLTSDDSAVYFCARTPASFAYWGQGT	120
	V_L domain	
	LTVVSASSGGGGGGGGGGGGSSSDIVMTQSPSSLVTAGERVTMNCRSSQSLLHSE	180
	NQKNYLTWYQQKPGQPPKLLIYTSTRESGVPDFTGSGSGTDFTLTISSVQAEDLAVYY	240
	AviTag	
	CQNDYGYPLTFGAGTKLELKGSGLNDIFEAKIEWHE	278

To facilitate downstream applications, an AviTag peptide sequence (GLNDIFEAKIEWHE) was added to the C-terminus of the scFv(RL2) to enable enzymatic biotinylation using the biotin ligase enzyme birA.¹¹⁴ AviTag, also known as the acceptor peptide (AP), was developed by Beckett et al. (1999) through phage display selection.¹¹⁹ This 14-mer peptide efficiently mimics the biotin acceptor function of the natural biotin carboxyl carrier protein (BCCP) domain. birA catalyzes the biotinylation of AviTag with a specificity constant (k_{cat}/K_m) comparable to that of the full BCCP, providing a more compact and manageable substrate for site-specific biotinylation without compromising the efficiency of the reaction. (**Figure 2.5**) birA specifically recognizes the AviTag and catalyzes the attachment of biotin to a specific lysine residue within the tag, resulting in the covalent linking of free biotin to the lysine within the AviTag peptide sequence. This biotinylation process effectively functionalizes the scFv(RL2), meaning it imparts a specific biochemical

capability to the scFv, allowing it to interact with streptavidin-based systems and enabling its use in targeted assays and purification techniques.

Biotinylation serves as a versatile tool for functionalizing scFv(RL2) in various applications. Namely, in immunoblotting, the biotinylated scFv(RL2) could be detected using commercially available streptavidin conjugated to dyes, simplifying the detection process without the need for direct chemical conjugation. For pull-down assays, the biotinylated scFv could be immobilized on streptavidin or avidin-coated beads, allowing the selective isolation of O-GlcNAc-modified proteins from complex samples. The strong biotin-streptavidin interaction ensures stable attachment of the scFv(RL2), and the multimerization capability of streptavidin enhances avidity by presenting multiple scFv(RL2) molecules, increasing capture efficiency.¹²⁰⁻¹²²

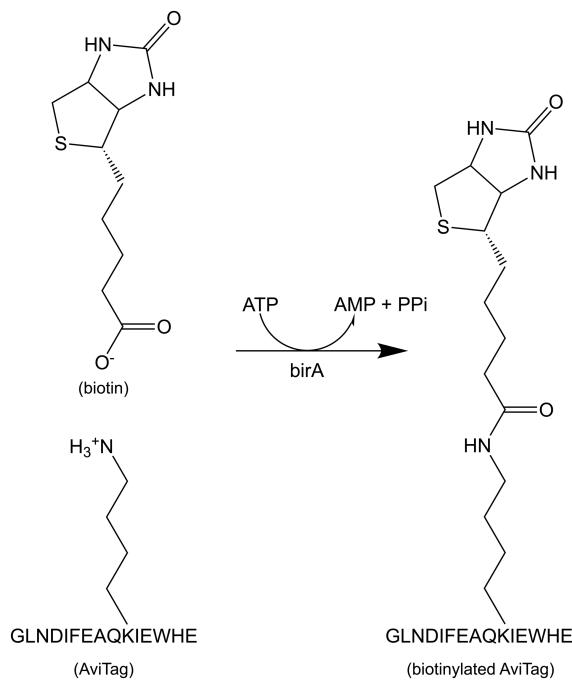


Figure 2.5 Biotin ligase (birA) reaction.
Covalent linking of free biotin to the lysine in AviTag peptide sequence.

Further, biotinylated scFv(RL2) can be used in various other functionalizations, such as anchoring it to sensor surfaces in biosensors for real-time detection,^{123,124} or conjugating it to nanoparticles or other functional molecules.¹²⁵⁻¹²⁷ These functionalizations broaden the utility of the scFv(RL2), enabling its application in diverse fields like diagnostics, targeted delivery, and molecular imaging, and providing a flexible toolkit for a range of experimental needs.

To minimize potential structural impact on the scFv(RL2), a short GSS linker was included before the AviTag.¹¹⁸ Additionally, an *N*-terminal His tag was introduced via the expression vector to aid in purification, with a tobacco etch virus (TEV) protease cleavage site (ENLYFQG)¹²⁸ inserted between the His-tag and the scFv(RL2), allowing for possible removal of the His-tag in future applications.

To pre-empt potential solubility issues often encountered in scFv expression,^{129,130} specific cysteine residues at positions 32, 106, 170, and 241 of scFv(RL2) were mutated to valine and alanine (C32V, C106A, C170V, and C241A), resulting in the construct scFv(RL2)-V/A. (**Figure 2.6**)

Figure 2.6 Full amino acid sequence of the scFv(RL2)-V/A.

TEV site	V _H domain	
ENLYFQGSSG	QVQLQQSGTELVRPGTSVKVS V KTSGYAFTNYLIEWI K QRP G GLEWIGV	60
	▲	
INPGSGITNYNEKF K D K ATLTADKSS S TAYIQL L SSLTS D DSAVY F AARTPASFAYWGQGT		120
	▲	
LVTVSA S SSGGGGSGGGGS G GGGSSSDIVMTQSP S SLTVAGERVTMN V RSSQSLLHSE		180
	▲	
NQKNYL T WYQQ K P G QPPK L LI Y WTSTRESGV P DRFTG G SG S GTDFTLT I SSV Q AED L AVYY		240
AviTag		
AQNDYGYPLTFGAGTKLELKGSSGLNDIFEAQKIEWHE		278
	▲	

This mutation strategy is commonly employed to reduce the risk of improper disulfide bond formation and aggregation,¹³⁰⁻¹³⁴ which can occur in the reducing environment of bacterial cytoplasm. By substituting these cysteines with non-polar residues, the goal was to enhance the solubility and stability of the scFv(RL2) in aqueous buffers, thereby facilitating its downstream applications. The goal was to replace these cysteines with non-polar residues to promote hydrophobic interactions within the protein core, thereby potentially improving its solubility and stability. The scFv(RL2)- V/A was thus designed as an attempt to mitigate these solubility issues and support its use in downstream applications.

2.3.3. Cloning

Before designing the scFv(RL2), I expressed the full RL2 antibody as a proof of principle and a control for binding studies. This approach also enabled a direct comparison between the recombinant RL2 and the commercially available RL2, ensuring similar

performance. To achieve this, both the light and heavy chain sequences were cloned into a single mammalian expression vector, pGenLenti, provided by GenScript Biotech Corp. (**Figure 2.7a**) The sequences were inserted between BamHI and Nhel sites, placing them under the control of a CMV promoter. I achieved expression of both RL2 chains from a single vector by incorporating a ribosome-skipping, self-cleaving T2A sequence (EGRGSLLTCGDVEENPGP) between them, which allows the translation of both chains from a single transcript.¹³⁵ The T2A sequence belongs to a group of 2A peptides that induce the ribosome to skip forming a peptide bond between glycine and proline near the C-terminus. This leads to the upstream protein gaining additional amino acids at its C-terminus, while the downstream protein starts with an extra proline at its N-terminus.¹³⁶ While the exact mechanism behind 2A-peptide-induced cleavage remains unclear, it is thought to involve the ribosome skipping over the glycy-prolyl bond formation rather than undergoing true proteolytic cleavage.^{137,138} To improve secretion and increase antibody titers, I incorporated an H7 secretion signal (MEFGLSWVFLVALFRGVQC) for the heavy chain and an L1 secretion signal (MDMRVPAQLLLLWLSGARC) for the light chain, positioning each at the N-terminus of their respective genes. These specific peptides were selected based on their demonstrated ability to enhance antibody production, with the H7/L1 combination leading to a more than 2-fold increase in the production of antibodies such as Rituxan, Avastin, and Humira compared to other heavy and light chain signal peptides.¹³⁹

The designed scFv(RL2) construct was cloned into the pETDuet-1 expression vector^{140,141} by GenScript Biotech Corp., with the sequence optimized for expression in *E. coli*. (**Figure 2.7b**) To enhance expression efficiency in *E. coli*, the scFv(RL2) gene sequence was optimized to match the preferred codon usage of *E. coli* while maintaining the original amino acid sequence. Rare codons that could slow translation or lead to premature termination were replaced with synonymous codons that are more frequently used in *E. coli*. Additionally, the mRNA secondary structure near the ribosome-binding site was adjusted to minimize potential folding that could hinder efficient translation initiation. The sequence was further screened to remove cryptic regulatory elements such as premature transcription terminators, ribosomal pause sites, and internal Shine-Dalgarno-like sequences that could interfere with proper expression. These optimizations aimed to enhance translation efficiency, increase protein yield, and minimize potential issues related to expression in the bacterial system. The scFv(RL2) gene was inserted between the BamHI and NotI restriction sites, resulting in the addition of a His-tag to the N-terminus

from the vector. Additionally, birA (EC 6.3.4.15) was cloned between the NdeI and Xhol sites, positioning an S-tag—an oligopeptide derived from pancreatic ribonuclease A—¹⁴² at the C-terminus of birA. Both genes are under the control of two independent T7 promoters, allowing for efficient and regulated expression. The pETDuet-1 vector also includes ampicillin resistance for selection and an inducible lacI repressor for tight regulation of gene expression. This vector was specifically designed to co-express scFv(RL2) and birA, with the aim of facilitating *in vivo* biotinylation of the AviTag on the scFv(RL2). The presence of birA facilitates the addition of biotin to the AviTag, streamlining subsequent purification and use in various assays. This co-expression approach avoids the metabolic burden associated with the use of multiple vectors in bacterial expression systems.¹⁴³

As an additional strategy to address potential solubility issues, the scFv(RL2) gene was cloned into the pMAL-c5E vector¹⁴⁴ (GenScript Biotech Corp.; **Figure 2.7c**) between the NdeI and EcoRI restriction sites. This construct fused the scFv(RL2) to the maltose-binding protein (MBP), creating an scFv(RL2)-MBP fusion protein with an enterokinase (EK) cleavage site (DDDDK) positioned between the MBP and the scFv-(RL2).

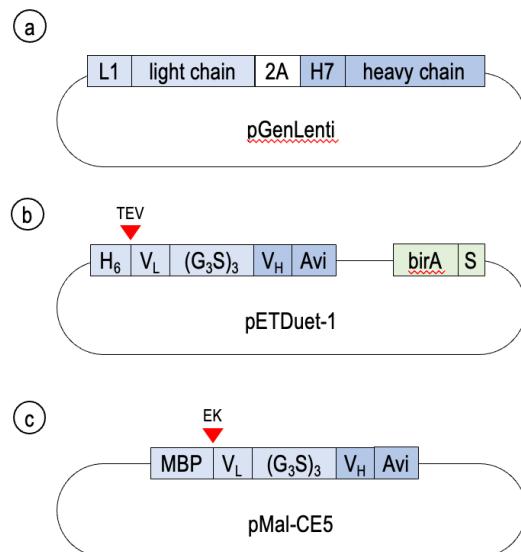


Figure 2.7 Vector designs for antibody expression.

Design of a vector containing the gene for the full-length RL2. (a) The light chain gene is preceded by an L1 sequence peptide, and the heavy chain gene is preceded by an H7 sequence peptide. Both chains are cloned in a single vector (pGenLenti) using the 2A ribosome sequence, which flanks the light and heavy chain genes, allowing them to be expressed as a single cluster. (b) Vector design for the scFv(RL2) gene. The V_L and V_H genes are flanked by a flexible linker and are cloned in the bacterial pETDuet-1 vector. The gene includes a His-tag at the N-terminus and an AviTag at

the C-terminus. Additionally, the gene comes with a birA biotin ligase to promote biotinylation of the AviTag *in celulo*. (c) Vector design for scFv(RL2) expression using the pMal-CE5 vector. In this design, the scFv gene is fused to an *N*-terminal MBP tag, which facilitates purification using an amylose or dextrin resin.

The pMAL-c5E vector also contains ampicillin resistance and the lacI gene, with expression driven by the tac promoter. To further optimize the construct, the scFv(RL2) was also cloned into a modified pMAL-c5E vector where the EK cleavage site was altered to DDDDR, a modification reported to enhance cleavage efficiency.¹⁴⁵ Additionally, another construct was generated using a pMAL-c5E vector with a TEV protease cleavage site (ENLYFQG), providing multiple options to improve the expression and solubility of the scFv(RL2).

2.3.4. Protein Purification

The full-length RL2 pGenLenti vector was transfected into CHO-K1 cells (CCL-61™ ATCC) using Lipofectamine™ 3000 (ThermoFisher Scientific), following the manufacturer's instructions. The cells were cultured in Dulbecco's Modified Eagle Medium F12 (DMEM F12; Gibco) supplemented with 10% fetal bovine serum (FBS; HyClone™), 50 IU/mL penicillin, and 50 µg/mL streptomycin, and incubated at 37°C with 5% CO₂. Media was collected 24, 48, and 72 h post-transfection. Although pGenLenti is designed for lentiviral-mediated stable expression, this experiment utilized transient transfection to quickly assess RL2 production. A stable cell line using lentiviral transduction is planned for future studies to ensure long-term expression. A HiTrap protein-G HP 1 mL column (Cytiva) was pre-equilibrated in wash buffer (citric acid•sodium phosphate, pH 7.0, containing 1 mg/mL lysozyme and 1 EDTA-free protease inhibitor tablet from Roche). The medium was then passed through the column, which was subsequently washed with 10 column volumes of wash buffer. Elution was performed with wash buffer adjusted to pH 3.0, and the eluted antibody was dialyzed overnight in 4 L of 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl at 4°C.

The scFv(RL2) expression vectors were transformed into One Shot™ BL21(DE3) chemically competent *E. coli* cells (Invitrogen C600003). *E. coli* BL21(DE3) was selected as the expression host due to its high-level recombinant protein production capabilities and well-established use in antibody fragment expression. This strain carries a chromosomal T7 RNA polymerase under lacUV5 control, allowing for IPTG-inducible

expression of scFv(RL2) from the T7 promoter present in the pETDuet-1 vector. Additionally, BL21(DE3) lacks the Lon protease and has reduced OmpT protease activity, minimizing potential degradation of recombinant proteins. Alternative strains optimized for disulfide bond formation, such as SHuffle, were considered, but initial tests did not yield significant improvements in solubility or expression levels for scFv(RL2). Given its reliable expression levels, ease of use, and broad compatibility with bacterial protein production workflows, BL21(DE3) was chosen as the most suitable host for scFv(RL2) expression. The cells were grown at 37 °C until an OD₆₀₀ of 0.7 was reached, at which point protein expression was induced by adding 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cultures were then incubated for 16 h at 16 °C. Harvested cells (20 g) were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mg/mL lysozyme, 1 EDTA-free protease inhibitor tablet from Roche), lysed by sonication, and centrifuged at 45,000 g for 1 h. The supernatant was passed through either a 1 mL HiTrap HP nickel column or a 1 mL MBPTrap HP dextrin column (Cytiva), depending on the construct, using the ÄKTA Purifier fast protein liquid chromatography (FPLC) system, both pre-equilibrated with the lysis buffer. The columns were washed with 10 column volumes of wash buffer specific to each construct type: for pETDuet constructs, the wash buffer contained 50 mM Tris (pH 7.5), 200 mM NaCl, and 10 mM imidazole; for pMAL constructs, it contained 50 mM Tris (pH 7.5) and 200 mM NaCl. Proteins were eluted using a gradient of either imidazole (0.1 to 0.5 M) or maltose (0.01 to 0.2 M) over 50 minutes, respectively. For pETDuet constructs, the elution buffer consisted of 50 mM Tris (pH 7.5), 200 mM NaCl, and 1 M imidazole, while for pMAL constructs, it included 50 mM Tris (pH 7.5), 200 mM NaCl, and 200 mM maltose. Fractions containing purified proteins were pooled, dialyzed overnight into the final buffer (20 mM Tris, pH 7.5, 200 mM NaCl), and concentrated to approximately 1 mg/mL.

For constructs exhibiting low solubility, purification under denaturing conditions was performed. The protocol followed was like the above procedure up to the resuspension of the cells. The harvested cells were resuspended in ice-cold 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2 % (v/v) Tween 20, and 30 mM dithiothreitol (DTT). As before, the cells were lysed by sonication and centrifuged at 45,000 g for 1 hour. The supernatant was then resolubilized in 6 M guanidinium hydrochloride (GndHCl), 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), and 20 mM imidazole. After resuspension, the solution was passed through a 1 mL HiTrap HP nickel column pre-equilibrated with the same buffer. The column was washed with 10 column volumes of the

resuspension buffer, and the proteins were eluted with 6 M GndHCl, 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM TCEP, and 300 mM imidazole. The purification under denaturing conditions of the scFv(RL2) were greatly assisted by Dr. Subramania Kolappan, whose expertise in protein purification was invaluable for this study.

The eluted protein was then subjected to a stepwise dialysis process to gradually remove the denaturant as previously described.¹⁴⁶ Initially, it was dialyzed for 8-16 h against 4 L of 6 M urea, 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 1 mM DTT. This was followed by dialysis against 4 L of the same buffer containing 4 M urea, and subsequently against 2 M urea. Finally, the protein was dialyzed against 4 L of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.5 mM oxidized glutathione for 8 h to facilitate the formation of disulfide bonds in scFv(RL2). The final dialysis step was conducted in 4 L of 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. This gradual refolding process was crucial to ensure that the scFv(RL2) transitioned from a linearized state to a properly folded conformation.¹⁴⁷

2.3.5. Secondary Structure Analysis

The scFv(RL2) purified under denaturing conditions and subsequently refolded was analyzed by circular dichroism (CD) spectroscopy to assess the integrity of its secondary structure. Prior to CD analysis, the sample was dialyzed into 20 mM Tris-HCl (pH 7.5) to remove any residual salts, as salts can interfere with CD measurements by affecting light scattering and the absorbance baseline. Far-ultraviolet CD experiments were conducted using a Chirascan Plus circular dichroism spectrometer (Applied Photophysics) with quartz cuvettes of a 2 mm path length, operated at 25 °C. CD spectral data were collected over a wavelength range of 195 to 300 nm with a 1.0 nm interval, using 500 µL of scFv(RL2) prepared as described. A CD spectrum was recorded to evaluate the secondary structure content of the protein. In parallel, the scFv(RL2) was also sent for native-state MS using direct injection to verify whether the observed mass corresponded to that of a correctly biotinylated scFv(RL2). Direct injection enables rapid introduction of the sample into the mass spectrometer without chromatographic separation, making it well-suited for purified protein samples.

2.3.6. PAGEs and Immunoblots

The eluted fractions from the protein G column were analyzed by native-polyacrylamide gel electrophoresis (PAGE) gel using a tricine-based, sodium dodecyl sulfate (SDS)-free running buffer to assess the molecular weight and purity of the RL2 antibody samples. For protein quantification, the concentration of recombinant RL2 was initially measured using a Nanodrop spectrophotometer immediately before flash-freezing the purified samples. This allowed for a quick assessment of protein yield. Before experimental use, the samples were thawed and quantified again using a Bradford assay to ensure accurate protein concentration measurements, as freezing and thawing can sometimes affect protein solubility and recovery. This two-step quantification approach ensured that the protein concentrations used in the gel analyses accurately reflected the available protein in solution at the time of the experiments. Native-PAGE was performed to detect the heterodimeric form of the antibody at its expected molecular weight of approximately 180 kDa. To further evaluate the integrity of the antibody, a denaturing SDS-PAGE was conducted under similar conditions. In the SDS-PAGE, the antibody was expected to separate into its individual heavy (~50 kDa) and light (~25 kDa) chains, allowing for the confirmation of the correct assembly and purity of the RL2 antibody.

To further confirm the biotinylation, 1 µg of the sample was loaded onto a 10% SDS-PAGE gel, followed by electrophoresis at 70 V for 1.5 h. Proteins were then transferred onto a 0.45 µm nitrocellulose membrane using the TransBlot Turbo (Bio-Rad) semidry transfer system, following the manufacturer's instructions. The membrane was blocked with PBS containing 4% BSA for 1 h before being incubated with IRDye® 800CW streptavidin (LI-COR) overnight. The blot was then imaged using the LI-COR gel imaging system to detect the biotinylated scFv(RL2), as per the manufacturer's instructions.

To test whether the scFv(RL2) binds to O-GlcNAc, an immunoblotting approach was employed using the O-GlcNAc-modified Polyhomeotic (Ph-P) sterile alpha motif (SAM) domain construct (Ph-P^{1397–1589}) as the target antigen. Ph-P is a well-characterized O-GlcNAc-modified protein,^{8,9,35} and the glycosylation sites on the Ph-P^{1397–1589} construct were mapped as part of a collaboration.¹⁴⁸ Following the expression and purification of Ph-P^{1397–1589} (more details on Section 4.3.1), its O-GlcNAc modification was confirmed by treating the construct with the bacterial OGA *Bacteroides thetaiotaomicron* (*Bt*) GH84. Both the wild-type (WT) enzyme, capable of cleaving O-GlcNAc residues, and the catalytic mutant (D242A), which lacks enzymatic activity, were used. Ph-P^{1397–1589} construct was

incubated with 1 µg of each enzyme variant in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 1 mM DTT for 4 h at 25 °C. Successful removal of O-GlcNAc by the WT enzyme was expected to result in a reduced signal, while the D242A mutant served as a negative control. Following the incubation, samples were analyzed by SDS-PAGE and subsequent transfer onto nitrocellulose membranes. The presence of O-GlcNAc modification and the His-tag was then verified using immunoblotting with mouse anti-O-GlcNAc (Biolegend; CTD110.6) and rabbit monoclonal anti-His antibodies (Invitrogen; 21HCLC).

After confirming that Ph-P1397–1589 was O-GlcNAc-modified, an immunoblotting assay was performed using scFv(RL2) to assess its binding specificity. In this assay, 1 µg of Ph-P1397–1589 was transferred to a nitrocellulose membrane as described previously. The membrane was then incubated with scFv(RL2) overnight at 4 °C to allow binding. Following this, the membrane was incubated with IRDye®-conjugated streptavidin for 1 h, as previously outlined, to detect the biotinylated scFv(RL2). Imaging was performed using the LI-COR fluorescent gel imaging system. The results confirmed the specific binding of scFv(RL2) to the O-GlcNAc-modified Ph-P^{1397–1589}, demonstrating the potential of scFv(RL2) for identifying O-GlcNAc modifications in target proteins.

2.3.7. Peptides Synthesis and Analysis

The glycopeptides used in this study were synthesized by Adrian Plata-Ruiz, using chemoenzymatically prepared serine-O-GlcNAc intermediates that I provided. Adrian then incorporated these intermediates into the final glycopeptide products for the nuclear pore proteins Nup62^{471–480}, Nup98^{241–254}, Nup153^{961–973}, allowing us to bypass the need for OGT in the synthesis. Following synthesis, I performed HPLC analysis to verify the quality and modification status of the glycosylated peptides.

For the HPLC analysis, an Agilent 1200 series HPLC equipped with an Agilent XDB-C18 Eclipse reversed-phase column was used to purify the glycopeptides. The glycopeptides were eluted with a gradient of 10–50% acetonitrile in water containing 0.1% trifluoroacetic acid, and fractions were collected and lyophilized. The resulting products, with purity exceeding 95%, were further analyzed by matrix-assisted laser desorption/ionization (MALDI)-time of flight (ToF) MS to confirm their molecular weight and glycosylation status.

While I have confirmed the quality of the glycosylated peptides, I have not yet tested the unglycosylated versions in the binding assays. Future experiments will incorporate these unglycosylated peptides to provide a comparative analysis and further validate the specificity of scFv(RL2) for O-GlcNAc-modified peptides. I am also awaiting additional protocol details from Adrian to ensure the reproducibility of the glycopeptide synthesis for subsequent assays.

2.3.8. BLI

The binding affinity of the scFv(RL2) to the O-GlcNAc-modified Ph-P^{1397–1589} was quantitatively measured using biolayer interferometry (BLI) on the Octet RED96e eight-channel system (Sartorius). High-precision streptavidin biosensors (catalog no. 18-5019) were used for the analysis. Prior to the experiment, the sensors were pre-equilibrated in a read buffer composed of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.1% BSA for 30 min to ensure stability and reduce nonspecific binding. The scFv(RL2) was then loaded onto the biosensors by incubating them in a solution of 200 nM biotinylated scFv(RL2) for 5 min. After loading, the sensors were rinsed in the read buffer, followed by baseline collection for 60 seconds. The loaded sensors were subsequently dipped into wells containing 0.1 to 2.5 mM O-GlcNAc-modified Ph-P^{1397–1589}, to monitor the association phase for 800 s. After association, the sensors were moved to wells containing only the read buffer to monitor the dissociation phase for an additional 300 s. Controls included a unmodified Ph-P^{1397–1589} sample to account for baseline drift and a sample lacking the scFv(RL2) to monitor nonspecific binding. For data analysis, the response at equilibrium (R_{eq}) was calculated, and the data were fit to a binding model using the Octet software (v.12). This analysis provided real-time kinetic parameters offering a detailed quantitative measurement of the scFv(RL2)'s affinity for the O-GlcNAc-modified Ph-P^{1397–1589}. To further validate the specificity of the scFv(RL2), O-GlcNAc modified peptides from nuclear pore proteins (Nup) Nup62, Nup98, and Nup158 were tested. These peptides were loaded onto the streptavidin sensors in a similar manner, and their interaction with the scFv(RL2) was assessed using the same BLI protocol.

2.3.9. Pulldown

To evaluate the binding specificity and capacity of the scFv(RL2) for O-GlcNAc-modified proteins, a pull-down assay was performed using scFv(RL2) immobilized on

NanoLINK® streptavidin-coated magnetic beads (Vector Labs). Streptavidin-coated magnetic beads (100 µL) were equilibrated by washing three times with 500 µL of PBS (pH 7.4). Biotinylated scFv(RL2) (35 µg) was added to the washed beads and incubated at room temperature for 30 min on a rotator to facilitate immobilization. Following incubation, the beads were separated using a magnetic separator, and the supernatant was collected to assess any unbound scFv(RL2). The beads were then washed three times with PBS to remove any remaining unbound scFv.

After immobilization, the beads were incubated with O-GlcNAc-modified or non-modified Ph-P¹³⁹⁷⁻¹⁵⁸⁹ (70 µg) in 500 µL of PBS. The reaction was incubated for 1-3 hours at room temperature on a rotator to allow for binding. Following incubation, the beads were separated using a magnetic separator, and the second supernatant was collected to assess the amount of unbound Ph-P¹³⁹⁷⁻¹⁵⁸⁹. The beads were then washed with PBS to remove any non-specifically bound protein. Bound proteins were eluted by boiling the beads in SDS loading buffer and analyzed by SDS-PAGE. The efficiency of the pull-down and the specificity of scFv(RL2) binding to O-GlcNAc-modified Ph-P were assessed by comparing the amount of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ bound in the presence and absence of O-GlcNAc modification.

2.3.10. Protein Crystallization

scFv(RL2) was purified under denaturing conditions as specified previously, and refolded into a buffer containing 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. The refolded scFv was concentrated to 20 mg/mL and prepared for crystallization. Approximately 500 µL of the 20 mg/mL scFv(RL2) sample was submitted to the National Crystallization Center at the Hauptman-Woodward Medical Research Institute in Buffalo, New York. Crystallization screening was performed using high-throughput methods, testing a variety of conditions to identify those favorable for crystal formation.

2.4. Results

2.4.1. Recombinant RL2 Production

I confirmed the expression of the recombinant RL2 in CHO-K1 cells through the analysis of native- and SDS-PAGE. To achieve this, the full-length RL2 pGenLenti vector

was transiently transfected into CHO-K1 cells using Lipofectamine™ 3000. Cells were cultured in DMEM F12 medium supplemented with 10% FBS, penicillin, and streptomycin, and media was collected at 24, 48, and 72 hours post-transfection. The antibody was purified from the media using a HiTrap protein-G HP column, equilibrated with wash buffer containing citric acid-sodium phosphate, followed by elution at pH 3.0. The eluted protein was dialyzed into 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl overnight to remove residual salts and adjust the buffer conditions.

Gels obtained native- and SDS-PAGE demonstrated successful expression of my RL2 construct, showing bands corresponding to the full RL2 antibody in the native gel and the heavy and light chains of similar intensity in the denaturing gel. (**Figure 2.8a**) For native-PAGE, tricine-based running buffer (SDS-free) was used to detect the heterodimeric antibody (~180 kDa), while SDS-PAGE under denaturing conditions separated the heavy (~50 kDa) and light (~25 kDa) chains.

However, the bands for the recombinantly expressed antibody were fainter compared to the commercially available antibody, despite both being at the same concentration. This discrepancy could potentially be attributed to factors such as differences in PTM, such as glycosylation, when expressed in CHO-K1 cells. For native-PAGE, tricine-based running buffer (SDS-free) was used to detect the heterodimeric antibody (~180 kDa), while SDS-PAGE under denaturing conditions separated the heavy (~50 kDa) and light (~25 kDa) chains. It is well-established that glycosylation can hinder Coomassie Brilliant Blue dye binding by reducing interactions with protein residues, leading to weaker staining.¹⁴⁹ However, this is a hypothesis, and further analysis, such as enzymatic deglycosylation followed by re-staining, would be required to confirm whether glycosylation is indeed responsible for the reduced band intensity. Additionally, variations in the formulation of the commercial antibody could also contribute to these differences. The commercial RL2 was confirmed through sequencing and analysis to consist of a kappa light chain and gamma 1 (IgG1) heavy chain, both integral to its antigen-binding and functional properties, ensuring accurate comparisons with the recombinant construct. An alternative approach to further investigate this discrepancy would be to use a more sensitive staining method, such as silver staining, which has a lower detection limit than Coomassie. This could help determine whether the faint bands are due to differences in staining efficiency rather than actual protein abundance.

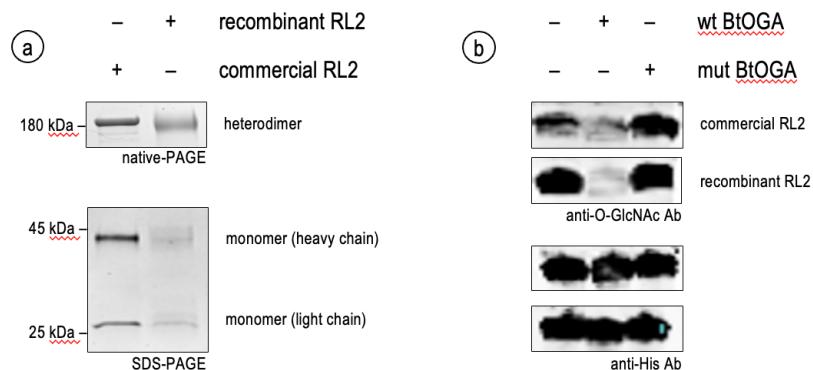


Figure 2.8 Characterization of recombinant RL2 antibody.

SDS-PAGE and Native PAGE of Commercial and Recombinant RL2. (a) A cropped section of a gel obtained by SDS-PAGE. The top image is gel obtained from a native-PAGE comparing the commercial RL2 with the one recombinantly expressed. The lower image is from a gel obtained by reducing SDS-PAGE and two bands corresponding to the heavy and light chains, respectively, for both the commercial and recombinant RL2 antibodies. (b) Western Blot Using Recombinant RL2: This panel presents a blot using the recombinant RL2 to detect O-GlcNAc modified and unmodified proteins. The top section shows the results for commercial RL2 compared to recombinant RL2. Below are the bands corresponding to the His tag on the Ph-P¹³⁹⁷⁻¹⁵⁸⁹ construct protein. wt. BtOGA was preincubated with Ph-P to remove O-GlcNAc, and the catalytically dead mutant BtOGA served as a negative control.

Immunoblotting was performed to evaluate the functionality of recombinant RL2 compared to commercial RL2 antibody. Both antibodies were tested at the same concentration, using a 1:3000 dilution of a 0.5 mg/mL stock. The commercial RL2 followed the manufacturer's recommended dilution, and the recombinant RL2 was prepared to match this for a direct comparison. Their ability to detect O-GlcNAc-modified and unmodified Ph-P¹³⁹⁷⁻¹⁵⁸⁹ construct was analyzed. The results demonstrated that recombinant RL2 showed comparable specificity to the commercial RL2, effectively distinguishing between the O-GlcNAc-modified and unmodified Ph-P construct. The commercial antibody exhibited strong preference for binding to the glycosylated form, while the recombinant RL2 performed similarly, validating its potential as a functional alternative to the commercial antibody (Figure 2.8b). These results suggest that the recombinant RL2 retains specificity for O-GlcNAc modifications, which could be attributed to the controlled expression system in CHO-K1 cells, potentially offering a consistent alternative to commercially stored antibodies.

2.4.2. scFv(RL2) Production

Several scFv(RL2) constructs were generated in an attempt to improve solubility and expression. Initially, the scFv(RL2) expressed in the pETDuet-1 vector proved to be insoluble. Specifically, after transformation of the pETDuet-1 construct into One Shot™ BL21(DE3) chemically competent E. coli cells, the culture was grown to an OD₆₀₀ of 0.7 at 37°C and induced with 1 mM IPTG for 16 h at 16°C. Cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 1 mg/mL lysozyme, followed by sonication and centrifugation at 45,000 g for 1 h to separate the soluble and insoluble fractions. The scFv(RL2) was not detected in the soluble fraction, but instead was found in the insoluble fraction. This was confirmed by SDS-PAGE and streptavidin blot analysis of both the soluble and insoluble fractions. (data not shown) Attempts to improve solubility by introducing mutations (C32V, C106A, C170V, and C241A) did not lead to a significant change, as the scFv(RL2)-V/A still partitioned into the insoluble fraction. (data not shown)

To address this issue, I explored the use of an MBP-tag, which successfully increased solubility, allowing the scFv(RL2) to remain in the soluble fraction. The scFv(RL2) gene was cloned into the pMAL-c5E vector, which fused the scFv(RL2) to an MBP tag. Expression was induced as described above, and purification was performed using MBPTrap HP dextrin columns pre-equilibrated with lysis buffer. However, when attempting to cleave the MBP tag using enterokinase, the enzyme was unable to effectively cleave the tag, possibly due to structural hindrances or poor enzyme efficiency in this context. (data not shown) Because viral proteolytic enzymes are generally more efficient, I replaced the EK cleavage site with a TEV protease site, which allowed for successful cleavage of the MBP tag. Unfortunately, after TEV cleavage, the scFv(RL2) once again became insoluble, precipitating out of solution. (data not shown)

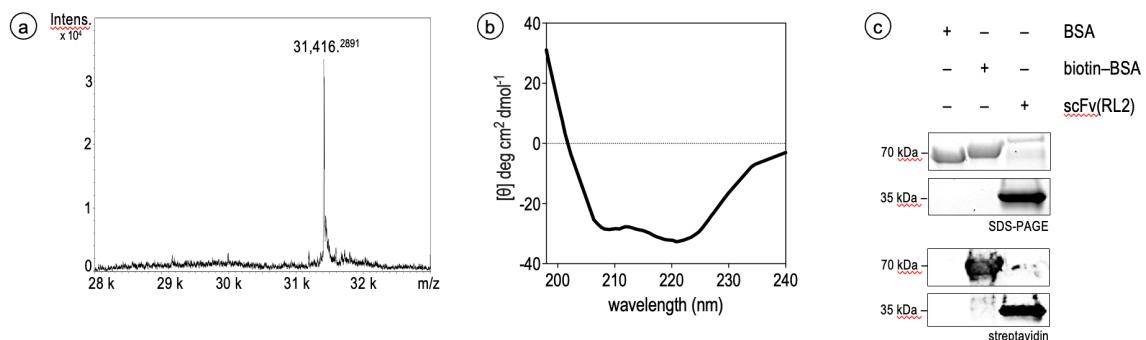


Figure 2.9 scFv(RL2) characterization.

MS analysis of scFv(RL2) (a) The MS spectrum of the refolded scFv(RL2) reveals a molecular weight of 31,416 Da, closely aligning with the theoretical value for biotinylated scFv(RL2). The mass loss of 18 Da during biotinylation is attributed to the removal of 1 Da from lysine (loss of a hydrogen atom) and 17 Da from biotin (loss of a hydroxyl group), confirming the successful biotinylation of the scFv(RL2). (b) CD spectra of refolded scFv(RL2). The CD spectra show typical α -helical and β -sheet features, indicative of the correct folding and stable secondary structure of the refolded protein. (c) SDS-PAGE and streptavidin blot of scFv(RL2). The SDS-PAGE gel and streptavidin blot confirm the expression and biotinylation of scFv(RL2). The scFv(RL2) migrated as expected and produced a strong signal, comparable to the positive control (biotinylated-BSA), validating its successful biotinylation and appropriate folding for further functional assays.

Faced with these solubility challenges, I opted to move forward with the scFv(RL2) construct from the pETDuet-1 system, incorporating a denaturing purification protocol with the help of Dr. Subramania Kolappan. Cells were lysed in buffer containing 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2% Tween 20, and 30 mM DTT, followed by resolubilization of the pellet in 6 M guanidinium hydrochloride, 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM TCEP, and 20 mM imidazole. The solubilized fraction was purified using a HiTrap HP nickel column, and the protein was gradually refolded through stepwise dialysis against decreasing concentrations of denaturant, as described in the methods section. This approach enabled the successful refolding and purification of the scFv(RL2) from the insoluble fraction. MS analysis of the refolded scFv(RL2) yielded a molecular weight of 31,416 Da, closely matching the theoretical value for biotinylated scFv(RL2). During the biotinylation process, 1 Da is lost from lysine due to the removal of a hydrogen atom, and 17 Da is lost from biotin due to the removal of a hydroxyl group, resulting in a total mass loss of 18 Da. This gives the final calculated mass of 31,416 Da (31,434 Da - 18 Da = 31,416 Da). (**Figure 2.9a**) This MS data supports the successful biotinylation of the scFv(RL2).

Additionally, CD spectra of the refolded scFv(RL2) confirmed the correct folding of the protein, displaying characteristic α -helical and β -sheet structures, indicating the presence of a stable secondary structure. (**Figure 2.9b**) CD measurements were performed using a Chirascan Plus spectrometer with a 2 mm path-length cuvette, and the protein was dialyzed into 20 mM Tris-HCl (pH 7.5) prior to analysis to minimize light scattering artifacts. In the streptavidin blot, the scFv(RL2) migrated as expected, with a clear signal at the anticipated molecular weight, and was comparable to the positive control (biotin-BSA), further confirming its biotinylation. (**Figure 2.9c**) The bands were observed at the correct molecular weight, consistent with the predicted size of the scFv(RL2), suggesting that the protein was properly refolded and biotinylated.

Accordingly, the scFv(RL2) was judged as being suitable for subsequent functional assays.

To further evaluate structural features and CDR positioning of the scFv(RL2), I submitted the amino acid sequence to RoseTTAFold for predictive modeling. (**Figure 2.10**) This method integrates sequence, distance, and structural data through a three-track network to accurately predict protein structures.¹⁵⁰ The submitted sequence incorporated the flexible (G₄S)₃ linker and the C-terminal AviTag sequence, with cysteine mutations included to match the final construct. The resulting models demonstrated appropriate folding and positioning of the CDR loops, with both loops oriented on the surface in a manner consistent with anticipated binding configurations. Inspection of the predicted structures also revealed that the lysine residue within the AviTag sequence remains solvent-exposed, suggesting minimal potential for interaction with the antigen-binding surface. This observation is promising, as it supports the intended functionality of the scFv(RL2) construct by preserving antigen accessibility while minimizing steric hindrance.

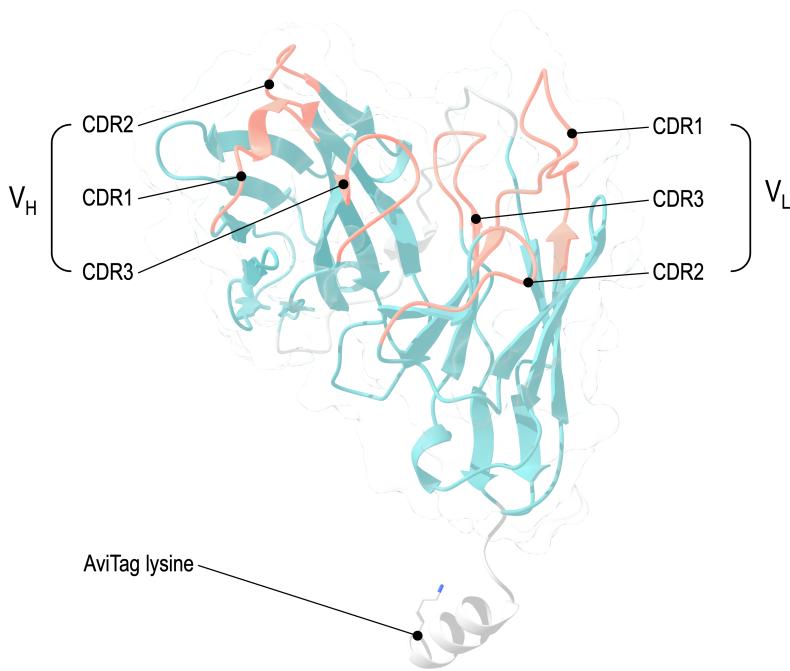


Figure 2.10 Predicted structure of scFv(RL2) highlighting CDR loops.

Model of scFv(RL2) generated using RoseTTAFold,¹⁵⁰ with CDR loops highlighted in pink to indicate their surface orientation and accessibility. The solvent-exposed lysine residue from the AviTag sequence is also shown, confirming minimal interaction with the antigen-binding surface.

2.4.3. Binding of scFv(RL2) to O-GlcNAc

To assess the binding capabilities of scFv(RL2) against O-GlcNAc-modified proteins, Ph-P^{1397–1589} was chosen as a candidate substrate for initial characterization. The Ph-P1397–1589 construct was expressed in *E. coli* using a pETDuet-1 vector and purified via Ni-NTA affinity chromatography. Prior to binding assays, the glycosylation status of the construct was confirmed through bacterial OGA (*Bacteroides thetaiotaomicron* GH84) digestion and immunoblotting using anti-O-GlcNAc antibodies to validate the presence of glycosylation. The Ph-P^{1397–1589} had been previously described, but to ensure the sample was suitable for binding assays, native-state MS was employed to confirm the molecular weights of the unglycosylated and glycosylated forms. MS analysis was performed via direct injection into a high-resolution mass spectrometer to allow for accurate detection of molecular weight differences without chromatographic separation. The multiple peaks in the native mass spectra for O-GlcNAcylated Ph-P^{1397–1589} likely arise from adduct formation and deconvolution artifacts. Noncovalent adducts from salts, buffer components, or solvent molecules can remain bound to the protein during ionization, causing small mass shifts and contributing to peak broadening. Additionally, the deconvolution algorithm may not fully resolve closely spaced glycoform states, leading to a distribution of peaks rather than distinct single peaks for each glycosylation state. These factors collectively account for the observed family of peaks in the spectra.

The unglycosylated Ph-P^{1397–1589} showed the expected molecular weight at 22,042.08 Da, with an additional peak corresponding to a mass shift of +178 Da. This mass shift was attributed to α-N-gluconoylation of the His-tag, a common modification in *E. coli* expression systems caused by the reaction of glucose-6-phosphate with the His-tag's N-terminal amine group, as detailed in **Figure 2.12**. The glycosylated form of Ph-P^{1397–1589} displayed five distinct peaks corresponding to O-GlcNAc modifications, with each glycosylation adding ~203 Da, consistent with mapping experiments confirming up to five glycosylation sites.

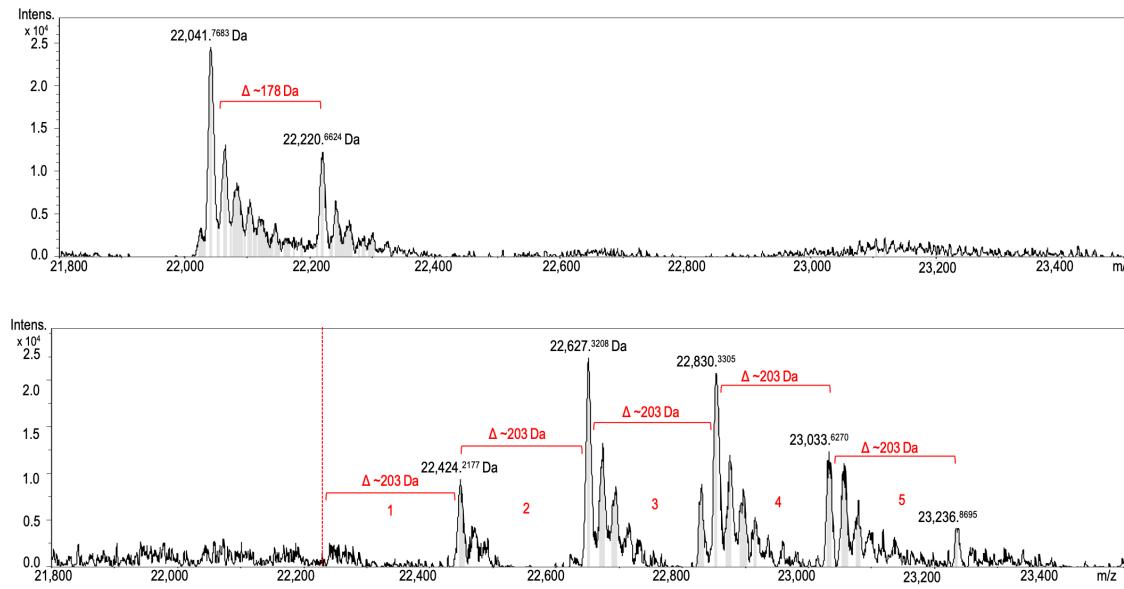


Figure 2.11 Native State Mass Spectrometry of Ph-P^{1397–1589}

The spectra depict the molecular weights of unglycosylated and glycosylated forms of the Ph-P^{1397–1589} construct. (Top) The unglycosylated Ph-P^{1397–1589} shows the expected molecular weight at 22,042.08 Da with a secondary peak corresponding to a mass shift of +178 Da. (Bottom) The glycosylated Ph-P^{1397–1589} spectrum displays five primary species, each representing a different glycosylation stoichiometry. The peaks correspond to the addition of 1, 2, 3, 4, and 5 O-GlcNAc units, with each glycosylation event contributing approximately +203 Da, indicating increasing glycosylation states, with five modifications being the most prominent.

This shift can be attributed to the spontaneous α -N-gluconoylation of the His-tag, a known spontaneous modification in *E. coli* expression systems, as outlined in **Figure 2.12**.^{151,152} MS analysis of the glycosylated Ph-P^{1397–1589} revealed multiple peaks corresponding to different glycosylation states, with the highest number of modifications being 5 O-GlcNAc units. This aligns with mapping experiments performed in collaboration.¹⁴⁸ These results provided a well-characterized analyte for scFv(RL2) binding studies, confirming that the glycosylation states were well-represented.

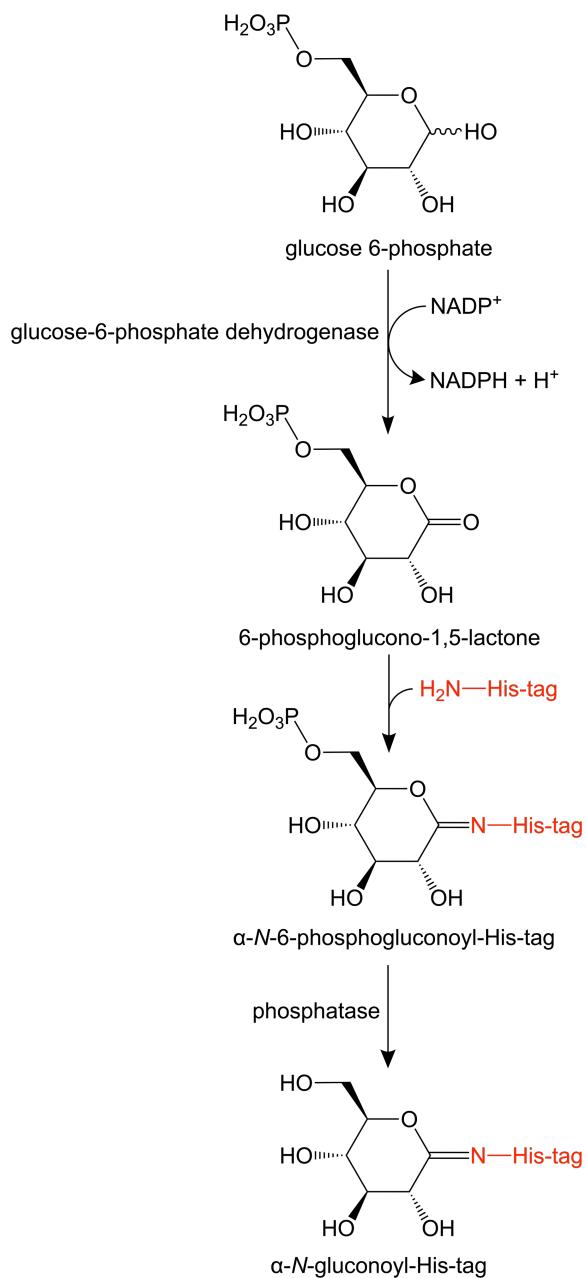


Figure 2.12 Mechanism of α -N-gluconylation of His-tagged Proteins in *E. coli*

The reaction illustrates how His-tagged proteins expressed in *E. coli* can undergo spontaneous α -N-gluconylation.¹⁵¹ Starting with glucose-6-phosphate, the enzyme glucose-6-phosphate dehydrogenase (G6PD) converts it to 6-phosphoglucono-1,5-lactone. This lactone spontaneously reacts with the N-terminal amine group of the His-tag, forming an intermediate. Bacterial cellular phosphatases subsequently reduce the phosphate group, yielding the final α -N-gluconoyl modification on the His-tag, which explains the observed +178 Da modification.

The binding of scFv(RL2) to O-GlcNAc-modified $\text{Ph-P}^{1397-1589}$ was tested using biolayer interferometry (BLI), a label-free optical technique that monitors real-time molecular interactions by detecting shifts in light interference patterns.¹⁵³ For these

experiments, scFv(RL2) was biotinylated in vitro through co-expression with the biotin ligase birA in the pETDuet-1 system. Following purification via Ni-NTA chromatography, the biotinylated scFv(RL2) was dialyzed into 20 mM Tris-HCl (pH 7.5), 150 mM NaCl to remove residual imidazole and immobilized onto streptavidin-coated biosensors. Assays were conducted using the Octet RED96e platform, with a read buffer containing 0.1% BSA to minimize nonspecific binding. In this assay, biotinylated scFv(RL2) was immobilized onto streptavidin-coated biosensors. As the glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ analyte is introduced and binds to the immobilized scFv(RL2), BLI measures the resulting changes in the optical thickness of the biosensor surface. These changes are recorded as shifts in the interference pattern, providing data on both the association and dissociation phases of the binding event. By analyzing these shifts, we can extract kinetic parameters such as association (k_{on}), dissociation (k_{off}), and equilibrium dissociation constants (K_D), offering quantitative insights into the strength and dynamics of the scFv's interaction with O-GlcNAc-modified Ph-P¹³⁹⁷⁻¹⁵⁸⁹.

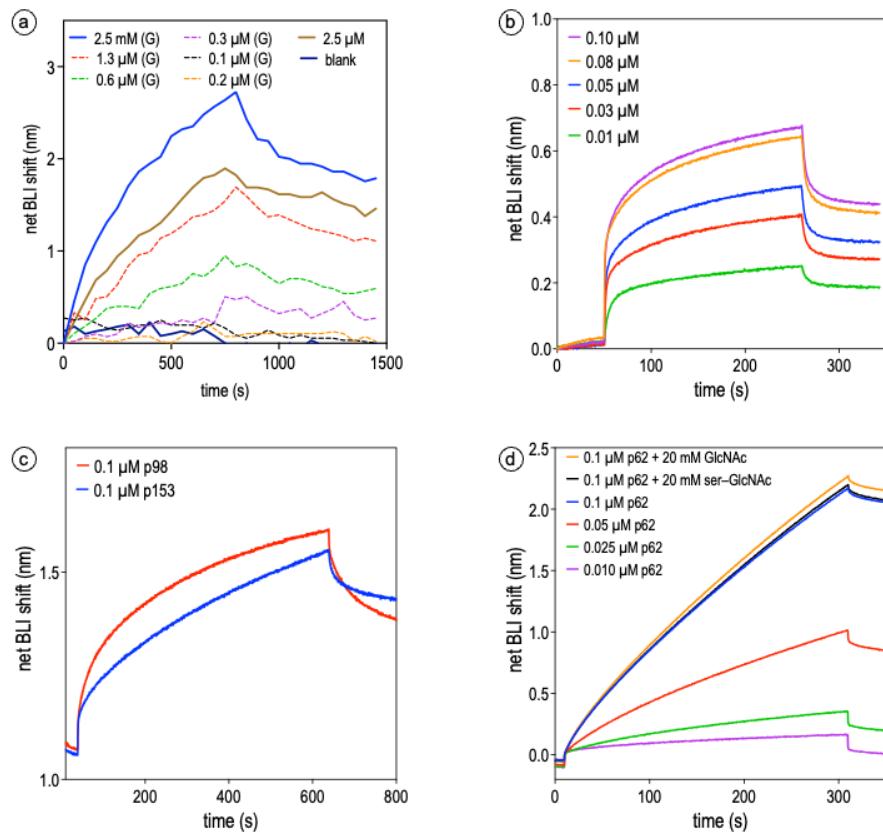


Figure 2.13 Binding analysis of scFv to O-GlcNAc modified protein and peptides.
Graph for scFv Binding to Ph-P¹³⁹⁷⁻¹⁵⁸⁹ (a) Relative binding versus time for immobilized scFv(RL2) on biotinylated surface. The ligands tested include glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ –denoted (G) at

different concentrations, unglycosylated Ph-P^{1397–1589}, and a buffer as a blank, demonstrating scFv(RL2)'s binding specificity to O-GlcNAc modified Ph-P^{1397–1589}. (b) BLI Graph for Ph-P^{1397–1589} concentration gradient. Graph showing the relative binding versus time for Ph-P^{1397–1589} at concentrations from 0.1 μM to 0.01 μM. The scFv(RL2) concentration was 200 nM, and the results are shown as duplicate runs, further verifying scFv(RL2)'s binding to O-GlcNAc modified Ph-P^{1397–1589}. (c-d) BLI Graph for glycosylated nuclear pore protein peptides. (c) BLI data for glycosylated nuclear pore protein peptides (Nup98^{241–254}, Nup153^{961–973}, and Nup62^{471–480}), indicating scFv(RL2) binding to these O-GlcNAc modified peptides. (d) Includes co-incubation of Nup62^{471–480} with 20 mM GlcNAc or 20 mM serine-GlcNAc to compete in the binding assay.

The binding of scFv(RL2) to both glycosylated and unglycosylated Ph-P^{1397–1589} showed an expected trend, where the glycosylated form exhibited a higher affinity, supporting the specificity of scFv(RL2) for O-GlcNAc. (**Figure 2.13a**) Controls included blank runs with buffer alone, as well as unglycosylated Ph-P^{1397–1589} to account for nonspecific interactions. However, anomalous binding was observed at low concentrations of glycosylated analyte, which may suggest cooperative binding effects or sensor saturation. High background signals were also noted, likely due to calibration or sensor surface issues. Repeating these assays with fresh sensors and recalibrating the instrument may help reduce variability in future experiments. However, at lower concentrations of the glycosylated Ph-P, the unglycosylated form displayed stronger binding than anticipated, suggesting possible non-specific interactions. This anomalous binding behavior, particularly noticeable at low glycosylated Ph-P concentrations, may indicate complex cooperative effects or saturation of the sensor surface at higher analyte concentrations.

Additionally, high background signals were observed across all binding curves in **Figure 2.13a**. These elevated baseline readings likely resulted from an issue with the Octet instruments calibration or sensor surface conditions during the experiments. Future controls, such as using irrelevant proteins (e.g., BSA), could help confirm the specificity of scFv(RL2) binding. The background signal persisted regardless of the analyte concentration, complicating the interpretation of the binding data. Given this background issue, additional controls such as blank runs using only buffer or an irrelevant protein like BSA would be necessary to assess and subtract non-specific interactions.¹⁵³ Furthermore, recalibration of the instrument or repeating the assay with fresh sensors may help mitigate these background effects and clarify the binding behavior of scFv(RL2).

In addition to Ph-P^{1397–1589}, binding tests were conducted using glycosylated nuclear pore protein peptides, such as Nup62^{471–480}, Nup98^{241–254}, and Nup153^{961–973}, which further demonstrated the scFv(RL2)'s ability to recognize O-GlcNAc modifications.

(Figure 2.13c-d) These peptides were synthesized chemoenzymatically, with glycosylated serine intermediates prepared and incorporated during synthesis. HPLC purification ensured >95% purity, while MALDI-ToF mass spectrometry validated molecular weights consistent with the glycosylated forms. Immunoblotting using anti-O-GlcNAc antibodies further confirmed glycosylation status prior to binding experiments. Nup62⁴⁷¹⁻⁴⁸⁰, Nup98²⁴¹⁻²⁵⁴, and Nup153⁹⁶¹⁻⁹⁷³, were analyzed to ensure their purity and modification prior to binding assays. HPLC traces (not shown) confirmed that all peptides were of high purity, exceeding 95%, and MALDI mass spectrometry validated their appropriate molecular weights for glycosylated forms (**Figure 2.14**). In addition to Ph-P¹³⁹⁷⁻¹⁵⁸⁹, these glycosylated peptides were used in binding tests, demonstrating scFv(RL2)'s ability to specifically recognize O-GlcNAc modifications.

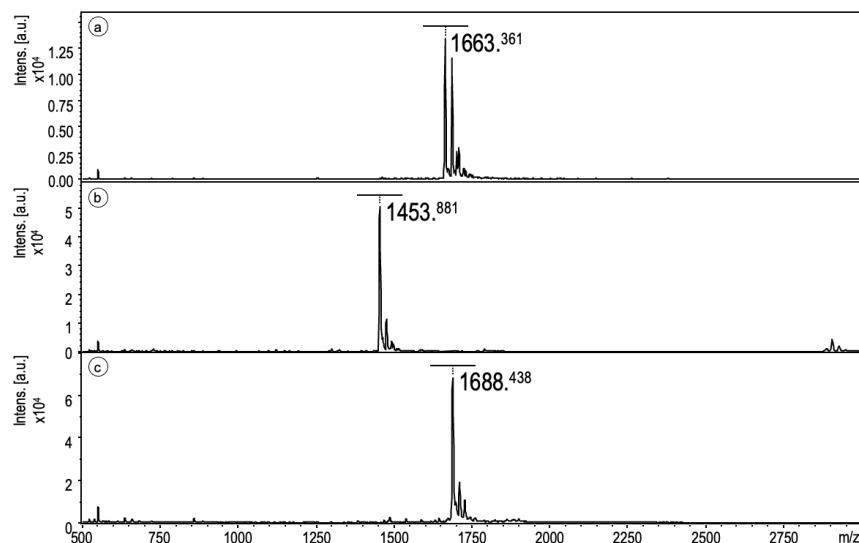


Figure 2.14 MALDI-ToF mass spectra of glycosylated Nup peptides.

MS spectra of glycosylated (a) Nup153⁹⁶¹⁻⁹⁷³, (b) Nup62⁴⁷¹⁻⁴⁸⁰, and (c) Nup98²⁴¹⁻²⁵⁴, consistent with the theoretical mass of the glycosylated form.

To further investigate the specificity of scFv(RL2), competition experiments were performed by preincubating the scFv(RL2) with 20 mM of free GlcNAc or serine-O-GlcNAc before introducing the glycosylated Nup62⁴⁷¹⁻⁴⁸⁰ peptide. (**Figure 2.13d**) Free sugars were incubated with scFv(RL2) for 30 minutes to test their ability to compete for binding. Despite this, no inhibition was observed, suggesting that scFv(RL2) may recognize a more complex epitope involving additional structural or glycan features. The goal of these assays was to determine whether the free sugars could compete with the glycosylated peptide for scFv(RL2) binding and inhibit the interaction. However, neither GlcNAc nor

serine-O-GlcNAc effectively inhibited the binding of the scFv to the glycosylated peptide. This lack of inhibition suggests that factors other than simple sugar competition may be influencing the binding, possibly indicating that scFv(RL2) recognizes a more complex epitope than just the O-GlcNAc moiety. It is important to note that these results are based on a single experiment, making it difficult to draw definitive conclusions, and further replicates and optimized conditions will be needed to better understand these findings.

Although the results support the specificity of scFv(RL2) for O-GlcNAc, more experiments are required to fully dissect the binding dynamics. Testing with full-length RL2 antibody alongside irrelevant glycoproteins could further validate the findings. Additionally, exploring saturation kinetics at higher glycosylated analyte concentrations may clarify cooperative effects observed at lower concentrations. It is unlikely that additional factors, such as the complex nature of glycosylation or other PTM like α -N-gluconoylation, significantly interfere with the analysis, but proper controls—including testing with full-length RL2 antibody—are necessary to confirm these findings.

2.4.4. Pull-Down of O-GlcNAc-modified Ph-P¹³⁹⁷⁻¹⁵⁸⁹ using scFv(RL2)

To assess the binding specificity of scFv(RL2) against O-GlcNAc-modified proteins, a pull-down assay was performed using scFv(RL2) immobilized on streptavidin-coated magnetic beads. The scFv(RL2) was biotinylated in vitro through co-expression with the biotin ligase birA in the pETDuet-1 system. The purified scFv was dialyzed into 20 mM Tris-HCl (pH 7.5), 150 mM NaCl to remove residual imidazole and prepare it for biotinylation. Streptavidin-coated beads were equilibrated in phosphate-buffered saline (PBS, pH 7.4) by washing them three times with 500 μ L of PBS to remove storage buffer and ensure compatibility with the assay conditions. The biotinylated scFv(RL2) (35 μ g) was added to the pre-equilibrated beads and incubated for 30 minutes at room temperature on a rotator to promote binding. The binding efficiency of the scFv(RL2) was first confirmed by analyzing the flow-through and eluates via SDS-PAGE. As shown in **Figure 2.15**, the first flow-through contained a faint band, indicating that a portion of the scFv remained unbound after the initial immobilization step. This suggests partial saturation of the streptavidin-coated beads, as approximately 19.5 μ g of the loaded scFv was successfully bound to the beads.

Following immobilization, unbound scFv was washed off the beads using three washes with PBS, each performed by separating the beads using a magnetic separator and discarding the supernatant. This washing step ensured that non-specifically bound material was removed before the Ph-P binding assays.

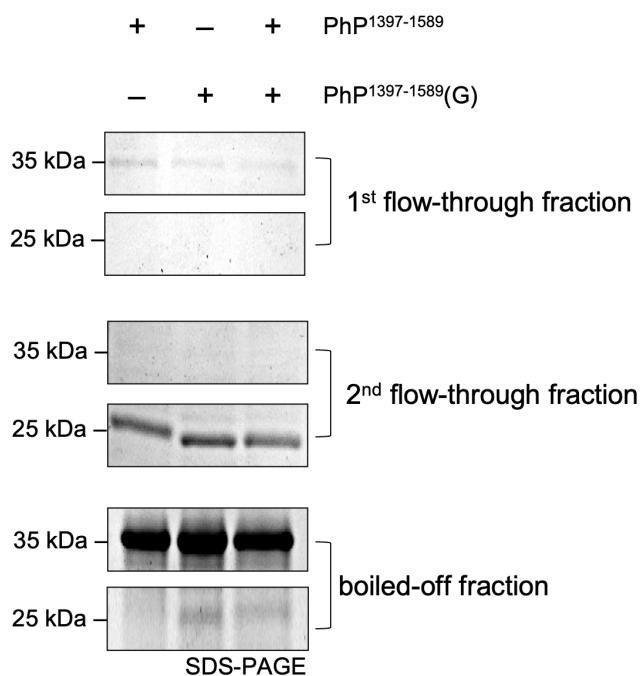


Figure 2.15 Pull-down assay of O-GlcNAc modified Ph-P using scFv.

This figure shows SDS-PAGE crops from a pull-down assay using scFv immobilized on streptavidin-coated magnetic beads. The top panel displays the first flow-through, indicating the amount of scFv bound after 30 minutes of incubation. The middle panel shows the second flow-through after a 1-3 hour incubation with modified and non-modified Ph-P. The bottom panel reveals the remaining Ph-P on the beads after washing and boiling.

Following the immobilization, the beads were incubated with either O-GlcNAc-modified Ph-P1397-1589 or non-modified Ph-P. Ph-P constructs were prepared by expressing the proteins in *E. coli* and purifying them via Ni-NTA affinity chromatography. Glycosylation status was validated using bacterial OGA treatment and immunoblotting with anti-O-GlcNAc antibodies. Unglycosylated Ph-P was similarly purified and confirmed to lack O-GlcNAc modifications. For the binding assay, 70 µg of each construct was diluted in PBS and added to the scFv(RL2)-streptavidin bead complex. The mixture was incubated for 1-3 hours at room temperature on a rotator to allow specific binding. After 1-3 hours of incubation, the flow-through collected from this step revealed faint bands,

indicating that a portion of Ph-P remained unbound. Quantification of the bound Ph-P indicated that 15.4 µg of the loaded O-GlcNAc-modified Ph-P was captured by the scFv-streptavidin bead complex, demonstrating specific interaction between scFv(RL2) and the O-GlcNAc modification.

Table 2.2 Binding data of scFv and Ph-P with streptavidin beads.

scFv loaded to streptavidin-coated magnetic beads	35.0 µg
scFv bound to streptavidin-coated magnetic beads	19.5 µg
An (Ph-P, or Ph-P(G)) loaded to scFv-streptavidin beads complex	70.0 µg
An (Ph-P, or Ph-P(G)) bound to scFv-streptavidin beads complex	15.4 µg

In contrast, the non-modified Ph-P showed minimal binding to the scFv(RL2)-streptavidin beads, confirming the specificity of the scFv for O-GlcNAc modifications. To remove non-specifically bound material, the beads were washed three times with PBS after incubation. Bound proteins were eluted by boiling the beads in SDS-PAGE loading buffer, which dissociated the streptavidin-biotin interaction and released the bound proteins for analysis. The subsequent washing and elution steps further confirmed the retention of O-GlcNAc-modified Ph-P on the beads, as evidenced by the presence of clear bands in the eluted fraction following SDS-PAGE analysis.

Despite some variability in band intensity between the modified and non-modified Ph-P, the results consistently demonstrated the scFv(RL2)'s ability to selectively capture O-GlcNAc-modified proteins. These findings validate the functionality of the scFv in binding assays and provide a strong basis for its use in detecting O-GlcNAc modifications in more complex protein mixtures. Future optimization could involve increasing scFv concentrations to reduce unbound material in the flow-through or optimizing incubation times to improve overall binding efficiency. Additional controls using irrelevant glycoproteins could also help validate specificity further.

2.4.5. Crystallization of scFv(RL2)

To gain structural insights into the scFv(RL2), crystallization trials were performed to identify conditions conducive to crystal formation. The screening was conducted at the National Crystallization Center (Hauptman-Woodward Medical Research Institute) using the scFv(RL2) at a concentration of 20 mg/mL. The scFv(RL2) protein was prepared by

refolding it from inclusion bodies following denaturing purification. Refolding was achieved through stepwise dialysis, gradually removing guanidinium hydrochloride, as previously described. After refolding, the scFv(RL2) was concentrated to 20 mg/mL using Amicon Ultra-15 centrifugal filters. The final buffer consisted of 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl. Purity was confirmed by SDS-PAGE, which showed a single band corresponding to the expected molecular weight of scFv(RL2). Before crystallization trials, the protein solution was filtered through a 0.22 μ m membrane to remove any aggregates or particulates.

Crystals were likely successfully obtained under the following conditions: 0.05 M HEPES, pH 6.8, 12.5% PEG 3,350, and a mixture of carbohydrates from the Silver Bullet C10 condition, including β -cyclodextrin, D-(+)-cellobiose, D-(+)-maltotriose, D-(+)-melezitose hydrate, D-(+)-raffinose pentahydrate, and stachyose hydrate. The crystallization screening employed the sitting-drop vapor diffusion method. Each drop consisted of 1 μ L of the scFv(RL2) solution mixed with 1 μ L of reservoir solution, equilibrated against 500 μ L of reservoir solution at 18°C. Drops were visually inspected daily using a light microscope for crystal formation, and initial hits were observed after 7–10 days of incubation.

The crystals formed were rectangular or tablet-like, with well-defined edges, indicative of ordered crystal lattice formation. The presence of these oligosaccharides in the crystallization condition is particularly interesting given the scFv(RL2)'s specificity for O-GlcNAc moiety, as these carbohydrates may play a role in stabilizing the crystal lattice or mimic interactions between the scFv(RL2) and glycosylated epitopes.

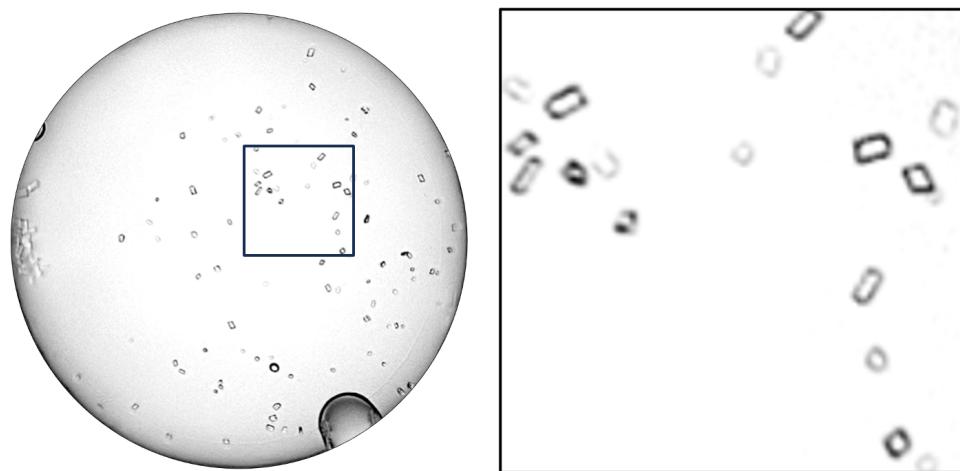


Figure 2.16 scFv(RL2) protein crystals.

Light microscope image at 4X magnification, displaying scFv(RL2) crystals that were seeded for obtaining an X-ray structure. A zoomed-in section is shown to the right, highlighting the rectangular or tablet-like shape of the crystals for clearer visibility.

Crystallizing scFv(RL2) is an important step toward solving its structure, which would provide detailed insights into the molecular architecture of the variable regions that confer its specificity for O-GlcNAc. Structural data on scFv(RL2) would help to visualize how these regions are arranged to recognize and bind to O-GlcNAc, shedding light on the mechanisms behind its affinity and selectivity. Additionally, the structural knowledge could serve as a basis for engineering efforts to evolve scFv(RL2) into a more specific or higher-affinity binder. These efforts might involve refining interactions with glycosylated targets or introducing mutations to improve its stability and performance in different experimental settings. Understanding the structural details would also be valuable for exploring potential co-crystallization experiments with O-GlcNAc-modified peptides, which could offer further insights into the recognition and binding modes of scFv(RL2) and guide its evolution as a more precise tool for detecting O-GlcNAc modifications.

2.5. Discussion

This chapter outlines the development and characterization of scFv(RL2) targeting O-GlcNAc-modified proteins. Key results include the successful production and functional validation of recombinant full-length RL2 and scFv(RL2), with assays alluding to specificity for O-GlcNAc-modified substrates.

2.5.1. Recombinant full-length RL2 production

Recombinant RL2 expressed in CHO-K1 cells showed comparable specificity to commercial RL2 antibodies, despite slightly fainter SDS-PAGE bands, likely due to post-translational modifications.¹⁴⁹ This result affirms the recombinant antibody's utility in detecting O-GlcNAc. What sets this approach apart is its streamlined nature: by using a single vector containing both the heavy and light chain sequences linked by a T2A ribosome-skipping sequence, this method eliminates the need for complex co-transfection ratios often required in traditional systems that rely on separate plasmids.^{135,154} This simplifies transient expression processes, allowing labs to easily produce recombinant antibodies without the hassle of balancing vector ratios. Furthermore, the incorporation of

optimized signal peptides (H7 for the heavy chain and L1 for the light chain) ensures enhanced secretion efficiency,¹³⁹ potentially increasing antibody titers more reliably than conventional systems, such as bicistronic approaches,¹⁵⁵ where an internal ribosome entry site (IRES) is used to express two genes from a single transcript. However, this often results in lower efficiency for the downstream gene.^{155,156} This construct, therefore, stands as a valuable tool for the field, offering an efficient and user-friendly method for producing RL2 in mammalian expression systems.

2.5.2. scFv(RL2) development

Efforts to produce scFv(RL2) in *E. coli* faced solubility issues, with scFv(RL2) primarily aggregating in the insoluble fraction, a common challenge in bacterial hosts where the cytoplasm does not favor proper folding of mammalian proteins.¹⁵⁷ Mutagenesis aimed at improving solubility was unsuccessful. However, attaching an MBP-tag improved solubility, though scFv(RL2) became insoluble after the MBP was cleaved. A denaturing purification protocol followed by gradual refolding recovered functional scFv(RL2), validated through multiple assays. This reflects the broader issue in recombinant protein production, where even soluble proteins may be misfolded or aggregated, necessitating careful screening for both solubility and functional competence to ensure recombinant protein quality.¹⁵⁸ Ensuring the solubility and proper folding of recombinant proteins like scFv(RL2) is critical for their reliability in basic scientific research, as these tools lay the groundwork for applications such as therapeutics. Without functional and properly folded proteins in research, any downstream applications, including potential therapeutic use, would be compromised, highlighting the necessity of robust expression and purification strategies in the early stages of development.

BLI experiments showed that scFv(RL2) binds more strongly to O-GlcNAc-modified Ph-P^{1397–1589} than the unglycosylated form, confirming specificity for O-GlcNAc. The binding affinity showed a concentration-dependent response, suggesting a potentially complex interaction mechanism due to multiple O-GlcNAc sites on Ph-P^{1397–1589}. However, high background signals in these assays suggest that further optimization is needed, including the calibration of the instrument and the use of blanks to reduce background interference. Glycosylated nuclear pore peptides (Nup62^{471–480}, Nup98^{241–254}, and Nup153^{961–973}) were also tested. Competition assays with GlcNAc and serine-O-GlcNAc showed no inhibition, possibly, yet unlikely, suggesting that the structural context of the

glycosylation site might impact scFv(RL2) binding. These results are preliminary, based on a single experiment, and further investigation is needed to validate these findings through replicates and proper controls.

Crystallization of scFv(RL2) was achieved under conditions containing a variety of oligosaccharides. The presence of these carbohydrates in the crystallization condition is intriguing, given scFv(RL2)'s specificity for O-GlcNAc, as they might stabilize the protein or mimic interactions with glycosylated targets. Structural data from these crystals would provide valuable insights into the variable regions responsible for O-GlcNAc recognition and could inform the engineering of scFv variants with enhanced affinity and specificity.^{110,159,160} Additionally, solving the structure of scFv(RL2) in complex with O-GlcNAc-modified peptides could offer critical information about the molecular interactions driving its specificity and suggest strategies for improving its performance as a detection tool. Overall, while scFv(RL2) shows potential for detecting O-GlcNAc, further experiments are required to optimize its binding, understand its interaction mechanism in greater detail, and explore its broader applicability in glycosylation research.

2.6. Acknowledgments

This journey would not have been possible without the support, guidance, and encouragement of many individuals, and I am deeply grateful to each of them.

I would like to express my heartfelt gratitude to my supervisor, Prof. David J. Vocablo. His invaluable mentorship, insightful advice, and unwavering support have been instrumental in shaping this research. Thank you for challenging me and always being available to provide guidance and feedback.

A special thank you goes to Dr. Subramania Kolappan for his critical assistance with the denaturing purification protocol for scFv(RL2). His expertise allowed me to overcome key experimental challenges, and this work would not have been possible without his contributions.

I am immensely grateful to Adrian Plata-Ruiz for synthesizing the glycosylated peptides used in this project. His skillful work was essential for validating the scFv, and I look forward to working with the unglycosylated peptides he provided for future experiments. I would also like to thank Dr. Jacob Brockerman for his valuable discussions

on antibody design, and Dr. Yanping Zhu and Dr. Dustin King, whose insights greatly contributed to the success of this project.

Acknowledgment is also due to the National Crystallization Center at the Hauptman-Woodward Medical Research Institute for their support in crystallization screening, which played a pivotal role in advancing this research.

To my fellow lab members, thank you for your camaraderie and technical assistance. It has been a privilege to work alongside such a brilliant group of scientists. I would also like to express my gratitude to Jun Bian for her continuous coaching in chemical characterization techniques, and to the technical staff at Simon Fraser University for their support throughout my research.

This chapter is a product of many contributions, and I am fortunate to have had such a dedicated and supportive team along the way. Thank you all.

Chapter 3.

CHEMOENZYMATIC METHOD FOR O-GLCNAC ENRICHMENT

The development of this chemoenzymatic enrichment strategy was made possible through collaborations with several researchers. Dr. Vitor Cunha contributed to the synthesis of key biotinylated probes used in this study. Dr. Stephen Withers and Dr. Feng Liu provided expertise and assistance in screening glycoside hydrolases, which were integral to the enzymatic approach. Additionally, the galactosyltransferases (B4GALT1s) used in this method were generously provided by Prof. Xi Chen and his team at the University of California, Davis. These contributions played a vital role in shaping the methodology presented in this chapter.

3.1. Background and Importance

O-GlcNAcylation typically is a low-abundance, dynamic PTM and therefore presents significant challenges in terms of its detection and enrichment. Its dynamic nature refers to the cycling of O-GlcNAc on and off proteins, governed by the reciprocal actions of OGT^{161,162} and OGA^{163,164} enzymes, which regulate cellular responses to internal and external stimuli.^{5,36,64} This cycling makes it difficult to capture O-GlcNAc-modified proteins at any given time. Further, it being generally substoichiometric,¹⁶⁵ where only a small percentage of potential sites are modified, complicates the detection of this modification, requiring highly sensitive techniques to enable accurate analysis. These challenges underscore the need for advanced detection methods to study O-GlcNAcylation's crucial roles in cellular processes like transcription,^{97,166} translation,⁵ and cell cycle regulation.^{85,167}

Enrichment methods for O-GlcNAcylated proteins generally fall into two categories: noncovalent and covalent. Noncovalent techniques, like lectin weak affinity chromatography (LWAC), use the natural affinity between WGA and GlcNAc, making them easier to perform while preserving the native O-GlcNAc modification.¹⁶⁸ O-GlcNAc-specific antibodies such as RL2⁷⁷ and CTD110.6⁷⁹, as well as combination of several O-GlcNAc

binding proteins,^{169,170} have also been used alongside WGA lectin in immunoprecipitation strategies.⁵⁶ These methods facilitate selective enrichment of proteins with O-GlcNAc modifications by capturing them through these established interactions. An alternative noncovalent method involves using a catalytically dead mutant of *Clostridium perfringens* OGA (*CpOGA*^{D298N}) that binds O-GlcNAc sites with two-digit micromolar affinity for threonine-modified and single-digit micromolar affinity for serine-modified peptides.⁸⁵ While this method takes advantage of the natural binding properties of the enzyme, it doesn't offer any substantial improvements over existing techniques. The process is straightforward but still relies on the limitations of affinity-based capture methods, meaning it doesn't necessarily increase specificity or efficiency compared to other noncovalent approaches.

In covalent methods, the goal is to create a stable bond between the O-GlcNAc modification and functional materials to aid in isolating and enriching these modified proteins. One approach uses periodate to oxidize the O-GlcNAc structure, generating aldehyde groups that can then be captured on hydrazide resin.¹⁷¹ While this reduces nonspecific binding, the method is not particularly efficient due to the O-GlcNAc structure being difficult to oxidize. O-GlcNAc lacks the cis-diol groups, which are more readily oxidized. As a result, attempts to oxidize and enrich O-GlcNAc-modified proteins using standard periodate-based methods face limitations, leading to reduced efficiency in isolating these modified proteins. Moreover, this method also enriches many other glycosylated proteins. Recent attempts to improve enrichment methods include the development of anti-O-GlcNAc antibodies.⁷⁶ While these antibodies have shown some promise in enriching O-GlcNAc-modified peptides, their reliance on specialized mass spectrometry techniques and limited validation across different biological contexts restrict their practical utility. This underscores the ongoing need for more versatile and efficient strategies for enriching O-GlcNAc-modified proteins.

This limitation has prompted methods like bioorthogonal labeling, which bypasses oxidation. For instance, feeding cells azide-functionalized GlcNAc analogs allows proteins to be tagged via Staudinger ligation.⁹⁰ While this approach adds more steps, it offers a selective alternative for tagging O-GlcNAc-modified proteins. Similarly, other methods have used *Bos taurus* β-1,4-galactosyltransferase 1 (B4GalT1; accession number P08037) to attach an azido-labeled galactose to O-GlcNAc, followed by click chemistry to tag the residue with biotin.⁴⁰ Although effective, the multi-step nature introduces

complexity without greatly improving specificity. An alternative enzyme, β -1,3-N-acetylgalactosaminyltransferase 2 (B3GalNT2), has been used in a similar manner.⁸⁷ A variation of these methods also uses B4GalT1 but includes a photocleavable biotin tag.¹⁷² Here, a galactose analog is enzymatically attached to O-GlcNAc, and the modification is then labeled with biotin. The main difference is the added photocleavage step for releasing enriched peptides, which increases complexity but again without great improvement in specificity or efficiency compared to azido-based methods. Xu et al. (2020) modified the B4GalT1 method by oxidizing the attached galactose to form an aldehyde for enrichment.¹⁷³ However, the reactivity of aldehydes in complex samples poses issues, as the electrophilic carbonyl group can react with biomolecules like proteins and nucleic acids, leading to irreversible modifications and cross-linking. Similarly, Chen et al. (2021) employed sodium periodate for galactose oxidation, but the approach still faces challenges from unwanted interactions driven by the aldehyde groups, complicating the analysis of O-GlcNAcylation in biological samples.¹⁷⁴ Considering the persistent limitations surrounding chemical reactivity and the recovery of native O-GlcNAc structures, I sought to address these challenges by exploring a new direction based on prior findings regarding the catalytic promiscuity of various GalT enzymes.¹⁷⁵ These observations led me to develop a strategy that avoids the complications of requiring post-modification structural changes that destroy the native O-GlcNAc residue. Such a method could be useful to facilitate purification of O-GlcNAc modified proteins to enable their biophysical and biochemical study.

Earlier work showed B4GalT1 can tolerate the transfer of galactose residues that were modified at the 6-hydroxyl group with bulky chemical groups, like biotin.¹⁷⁵ This work precede the approaches taken by Xu et al. (2020) and Chen et al. (2021) and suggests a streamlined method where pre-modified galactose can be attached directly to O-GlcNAc sites.^{173,174} This strategy could avoid the chemical reactivity challenges previously highlighted, offering a more efficient path for protein enrichment. Crystallographic data further support this possibility, revealing that the active site of β 4B4GalT11 is structured with a relatively exposed binding surface, particularly at the C6 position of UDP-Gal, which projects outward and allows significant solvent accessibility.^{176,177} This structural feature accounts for the enzyme's tolerance toward modifications at the 6-hydroxyl group, reinforcing the notion that larger, pre-modified galactose substrates can be transferred.

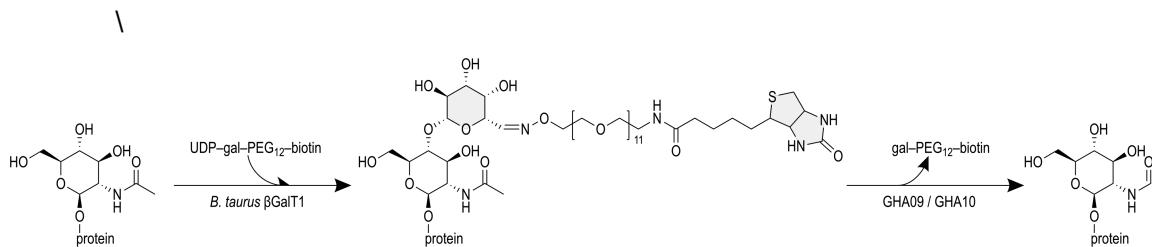


Figure 3.1 Enzymatic strategy for O-GlcNAc protein enrichment.

β4B4GalT11-mediated attachment of galactose functionalized with a PEG₁₂-biotin linker to O-GlcNAc residues. Following this step, a galactosidase specifically hydrolyzes the galactose, restoring the native O-GlcNAc modification while retaining biotin for downstream enrichment. This two-step reaction provides a reversible method for O-GlcNAc protein enrichment with minimal steric hindrance, enhancing the efficiency of subsequent analysis.

Drawing from previous insights into β4B4GalT11's tolerance for transferring galactose derivatives pre-functionalized with bulky moieties, I developed an approach where a galactose unit, conjugated with a PEG₁₂ linker and biotin, is enzymatically attached to O-GlcNAc residues. This PEG₁₂ spacer is predicted to provide adequate length and flexibility, minimizing steric hindrance during subsequent steps. The biotinylated proteins can then be enriched using affinity methods, such as with streptavidin-coated beads, ensuring specificity and efficiency. However, the chief novelty lies in the reversibility of this modification. This reversibility ensures that unlike conventional enrichment strategies, which often modify O-GlcNAc irreversibly, our approach enables the recovery of O-GlcNAc-modified proteins in their native state for later biochemical or biophysical analysis. Through high-throughput metagenomic screening of a library of bacterial glycoside hydrolases (GH), we identified enzymes two capable of cleaving the biotinylated galactose, leaving the O-GlcNAc site unaltered and intact. This strategy circumvents the irreversible alterations commonly seen in other methods, preserving the native glycan state for further analysis. The enzymatic screening process will be elaborated on in later sections. (Figure 3.2)

3.2. Objective

The objective of this chapter is to address the limitations associated with existing O-GlcNAc enrichment techniques by developing a chemoenzymatic approach that improves both specificity and efficiency while preserving the native state of O-GlcNAc. Traditional methods, such as periodate oxidation and noncovalent interactions with lectins or antibodies, face limitations in either having limited specificity or causing irreversible

modification of O-GlcNAc sites. These limitations complicate the ability to purify and study O-GlcNAc modified proteins from a biochemical and biophysical perspective. To overcome these challenges, this chapter builds on past advancements in existing knowledge of the enzymatic promiscuity of different GaIT enzyme, specifically leveraging the ability of B4GalT1 to transfer galactose derivatives that are modified at the 6-hydroxyl group. The method focuses on the enzymatic attachment of galactose functionalized with a PEG₁₂-biotin linker to O-GlcNAc residues. The PEG₁₂ spacer minimizes steric hindrance, enabling the use of solid phase affinity enrichment methods, such as streptavidin-coated beads. The biotinylation strategy not only ensures efficient protein enrichment but also enables reversibility through a subsequent enzymatic deglycosylation step, where GHs identified through high-throughput metagenomic screening selectively cleave the galactose moiety, restoring the native O-GlcNAc structure.

This approach directly addresses the need for methods that avoid irreversible modification of O-GlcNAc, complications arising from undesired reactivity of reactive functional groups, and the current inability to efficiently recover native glycan structures. Through this strategy, the enrichment and analysis of O-GlcNAc-modified proteins should become more precise, laying the groundwork for more accurate studies on the roles of O-GlcNAcylation. This method not only overcomes the limitations of prior techniques but also provides a new platform for studying the effects of O-GlcNAc modification on protein function.

3.3. Methods

3.3.1. Synthesis and Characterization of Probes

Before detailing the synthesis of the probes used in this study, I would like to express my gratitude to Drs. Roger Ashmus and Vitor Cunha, whose expertise in synthesizing the biotinylated substrates was crucial for advancing the screening and characterization of GHs.

All chemicals used in these procedures were obtained from Sigma-Aldrich and used without further purification unless otherwise noted. Nuclear magnetic resonance (NMR) spectra were recorded using Bruker AVII 600 and AVIII 400 spectrometers. HPLC with MS purification was performed on an Agilent 1100 series instrument using HPLC-

grade solvents. High-resolution MS (HRMS) analysis was conducted on a Bruker maXis ToF LC/MS/MS instrument using either positive or negative electrospray ionization (ESI).

Synthesis of 4MU- β -LacNAc. The probe 4MU- β -LacNAc was synthesized by first dissolving 4-methylumbelliferyl 2-acetamido-2-deoxy- β -d-glucopyranoside (10.0 mg, 0.0264 mmol) in a HEPES buffer solution (2.5 mL, 100 mM, pH 7.0) containing 30 mM MgCl₂ and DMSO (70 μ L). To this solution, UDP-galactose disodium salt (29 mg, 0.048 mmol), BSA (5.0 mg), alkaline phosphatase (50 U), and bovine β -(1,4)-galactosyltransferase (6000 mU) were added. The reaction was allowed to proceed at 37 °C for 36 h. Once the reaction was complete, the mixture was ultrafiltered using an Amicon® Ultra filter with a 10 kDa cutoff to remove proteins and enzymes. The filtrate was purified via preparative HPLC using a Zorbex 5 Eclipse Plus C18 column (21.1 \times 250 mm, 5 μ m) at a flow rate of 25 mL/min. The mobile phase was composed of water (A) and MeCN (B) with 0.1% formic acid. A gradient was applied: 5% B for 2 minutes, followed by 10–50% B over 10 minutes, and 90% B for 4 minutes. This yielded 4-MU- β -LacNAc (11.0 mg, 75%) as a white solid.

Synthesis of UDP-6-biotinyl-oxyamine-Gal. The synthesis of UDP-6-biotinyl-oxyamine-Gal was adapted from the method by Butler et al. (2001).¹⁷⁵ UDP-Gal (4.5 mg, 7.5 μ mol) was dissolved in Na₂HPO₄/NaH₂PO₄ buffer (1.0 mL, 25 mM, pH 6.0) containing 0.5 mM CuSO₄. Catalase from bovine liver (5400 U) and galactose oxidase from *Dactylium dendroides* (9 U) were added, and the mixture was stirred vigorously at room temperature for 72 hours. Following the reaction, the mixture was ultrafiltered using an Amicon® Ultra filter with a 10 kDa cutoff. Biotin-dPEG₁₁-oxyamine•HCl (10 mg, 12.1 μ mol), dissolved in 100 μ L of DMSO, was then added to the filtrate. The reaction mixture was stirred for 5 h at room temperature. Subsequently, the mixture was cooled to 0 °C, and NaCNBH₃ (20 mg, 0.32 mmol), dissolved in 200 μ L of ice-cold buffer, was added. The reaction was incubated at –20 °C for 72 h. The crude product was purified by preparative HPLC using an XDB C18 column (21.2 \times 150 mm, 5 μ m) at a flow rate of 15 mL/min. The mobile phase consisted of water (A) and MeCN (B) with 0.1% formic acid. The gradient used was 10% B for 2 minutes, 10–30% B over 21 minutes, and 95% B for 2 minutes. This resulted in UDP-6-biotinyl-oxyamine-Gal (0.78 mg, 8%) as a colorless wax.

3.3.2. GH Screening

I would like to extend my sincere thanks to Dr. Stephen Withers and Dr. Feng Liu for their invaluable assistance in the screening of glycoside hydrolases.

Cell culturing. In 96-well deep format plates, cultures were initiated by inoculating 0.8 mL of ZYP-5052 media, supplemented with 50 mg/L carbenicillin, with 2 μ L of glycerol stock. Cultures were incubated at 37 °C with shaking at 225 rpm for s. Cells were subsequently harvested by centrifugation at 3000 rpm for 30 minutes. The supernatant was decanted, and the resulting cell pellets were resuspended in 0.2 mL of lysis buffer (50 mM Na-HEPES, 1% Triton X-100, 0.25 mg/mL lysozyme, benzonase, and cOmplete™ Protease Inhibitor EDTA-free, pH 7.0). Cell lysis was facilitated by agitation with glass beads for 2 hours at room temperature.

Enzymatic screening assays. For the enzymatic assays, 20 μ L of each lysate was transferred to 96-well half-area plates (Corning 3694) and mixed with 30 μ L of 50 mM Na-HEPES (pH 7.0) containing 167 μ M of substrate, resulting in a final substrate concentration of 100 μ M. When biotin-PEG₄-LacNAc-4MU was used as the substrate, 1.3 nM of *Streptomyces plicatus* β -hexosaminidase (*SpHex*)¹⁷⁸ was added to the reaction mixture. Fluorescence (excitation: 360 nm, emission: 465 nm) was recorded at multiple time points using a Beckman Coulter DTX 880 Multimode Detector.

TLC analysis of GHA09 and GHA10 activities on biotin-PEG₄-LacNAc-4MU.

To assess the enzymatic activities of GHA09 and GHA10 on the substrate biotin-PEG₄-LacNAc-4MU, incubations were conducted at 37 °C with a final substrate concentration of 5 mM. The reaction mixtures were monitored using TLC to identify the conversion of substrate to product. The mobile phase utilized was a solvent mixture of water:iso-propanol:ethyl acetate (1:2:2, v/v/v). TLC analysis revealed the formation of the product, GlcNAc-4MU, indicated by an R_f value of 0.11 after 3 hours for both GHA09 and GHA10 reactions. Furthermore, the initial substrate, biotin-PEG₄-LacNAc-4MU, exhibited an R_f value of 0.63 and was completely consumed after 16 hours of incubation with GHA10, confirming the enzymatic activity of both variants and their ability to hydrolyze the glycosidic bond in the substrate, and also suggesting GHA10 may have greater catalytic efficiency toward this substrate.

3.3.3. GHA09 and GHA10 Expression

The expression vectors for GHA09 and GHA10 were kindly provided by Dr. Stephen Withers and Dr. Feng Liu.

The vectors were transformed into One Shot™ BL21(DE3) chemically competent *E. coli* cells (Invitrogen C600003). Following transformation, the cells were cultured at 37 °C until an optical density at 600 nm (OD_{600}) of 0.7 was reached. Protein expression was then induced by the addition of 1 mM IPTG, and the cultures were incubated for an additional 16 h at 16 °C. Harvested cells (20 g) were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mg/mL lysozyme, and one EDTA-free protease inhibitor tablet from Roche). The suspension was lysed by sonication and subsequently centrifuged at 45,000 g for 1 hour to separate cellular debris from the soluble fraction. The supernatant was then passed through a 1 mL HiTrap HP nickel column (Cytiva), which was pre-equilibrated with the lysis buffer. The column was washed with 10 column volumes of wash buffer (50 mM Tris, pH 7.5, 200 mM NaCl, and 10 mM imidazole) to remove unbound proteins. GHA09 and GHA10 proteins, which contain a His tag, were eluted using a gradient of imidazole (0.1 to 0.5 M) over 50 minutes, with the elution buffer comprising 50 mM Tris (pH 7.5), 200 mM NaCl, and 1 M imidazole. Fractions containing purified GHA09 and GHA10 proteins were pooled, dialyzed overnight into the final buffer (20 mM Tris, pH 7.5, 200 mM NaCl), and concentrated to approximately 1 mg/mL for subsequent assays.

3.3.4. Kinetic Assays

To characterize the kinetic parameters of GHA09 and GHA10, β-galactosidase-4-methylumbelliflone (β-gal-4MU) was used as the substrate in a fluorescence-based assay. Enzymatic reactions were set up by preincubating the 0.1 μM and 0.01 μM of GHA09 and GHA10, respectively, purified enzymes in a reaction buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl. Although the optimal pH for GHA09 and GHA10 activity is around 6 (**Figure 3.4**), the assays were performed at pH 7.5 to accommodate the fluorescence properties of the 4-methylumbelliflone (4MU) leaving group. The pKa of 4MU is approximately 7.79, and only its deprotonated anionic form is fluorescent. At pH 6, only about 1.56% of 4MU would be ionized, resulting in insufficient fluorescence signal for accurate measurements. By conducting the reactions at pH 7.5, we strike a balance between maintaining reasonable enzyme activity and ensuring adequate fluorescence

detection for the assay. The preincubation was carried out at 30 °C for 15 minutes in a sealed plate to allow thermal equilibration. Substrate solutions of β -gal-4MU were prepared in the same buffer, with final concentrations ranging from 0.03 mM to 2 mM, providing a range for determining Michaelis-Menten kinetics.

In a 384-well microplate, 10 μ L of 2 μ M enzyme solution was mixed with 10 μ L of substrate solution for each reaction, with each substrate concentration tested in triplicate. The reactions were initiated immediately after mixing, and enzyme activity was monitored by measuring the release of fluorescent 4MU at 465 nm (excitation at 360 nm) using a Synergy Neo2 hybrid multi-mode plate reader. Fluorescence data were collected continuously over a 10 min period to capture the initial reaction rates (V_0), calculated from the linear portion of the progress curves.

The initial rates at varying substrate concentrations were analyzed using the Michaelis-Menten equation to derive key kinetic parameters, including V_{max} and K_m , using Graph-Pad Prism. These parameters provided insights into the catalytic efficiency of GHA09 and GHA10. To further assess the substrate specificity of both enzymes, similar assays were conducted using Gal-4MU and GlcNAc-4MU. This comparative analysis allowed for a thorough evaluation of the enzymes' substrate preferences.

3.4. Results

3.4.1. Metagenomic Screening

The GH enzymes screened were identified through a systematic metagenomic screening,¹⁷⁹ characterizing a diverse range of glycoside hydrolases. This screening involved cloning metagenomic DNA fragments into expression vectors, transforming them into *E. coli* hosts, and culturing the transformants in ZYP-5052 media supplemented with carbenicillin. After 16 hours of incubation, cells were harvested, lysed using a Na-HEPES-based buffer with lysozyme, Triton X-100, benzonase, and protease inhibitors, and subjected to bead-based agitation to ensure complete lysis. The lysates were then prepared for enzymatic assays by centrifugation to remove insoluble debris.

These enzymes were evaluated using a biotin-PEG₄-LacNAc-4MU probe, designed to mimic O-GlcNAc-modified substrates relevant to enrichment protocols. The probe was synthesized by coupling biotin-PEG₄ to LacNAc-4MU, followed by purification

using HPLC with a gradient of acetonitrile in water containing 0.1% formic acid. The final purity of the probe was confirmed by HPLC-MS, ensuring its suitability for screening assays.

Their original metagenomic designations were retained (**Table 3.1**). Screening results with the biotinylated LacNAc fluorescent probe revealed that out of 194 GH enzymes, BA09 (GHA09) and BA10 (GHA10) exhibited the highest activity (Figure 3.2), as determined by relative fluorescence emitted following the liberation of the 4MU fluorophore. Fluorescence-based activity assays were conducted using a final substrate concentration of 100 µM in Na-HEPES buffer (pH 7.0) at 30°C. Reactions were initiated by mixing 20 µL of lysate with 30 µL of substrate solution in 96-well plates. Fluorescence was monitored at excitation and emission wavelengths of 360 nm and 465 nm, respectively, using a multimode plate reader. Initial rates of 4MU release were calculated from the linear portion of the fluorescence progress curves.

Table 3.1 Metagenomic study derived GH enzymes. (Heins et al., 2014)

	1	2	3	4	5	6	7	8	9	10	11	12
AA	SDJGI_31	SDJGI_139	SDJGI_35	SDJGI_137	SDJGI_42	SDJGI_34	SDJGI_33	SDJGI_83	SDJGI_136	SDJGI_79	SDJGI_86	SDJGI_80
AB	SDJGI_182	SDJGI_134	SDJGI_141	SDJGI_181	SDJGI_135	SDJGI_85	SDJGI_185	SDJGI_138	SDJGI_82	SDJGI_78	SDJGI_87	SDJGI_183
AC	SDJGI_30	SDJGI_40	SDJGI_84	SDJGI_140	SDJGI_81	SDJGI_53	SDJGI_170	SDJGI_162	SDJGI_108	SDJGI_52	SDJGI_187	SDJGI_197
AD	SDJGI_73	SDJGI_128	SDJGI_180	SDJGI_151	SDJGI_177	SDJGI_117	SDJGI_55	SDJGI_156	SDJGI_123	SDJGI_112	SDJGI_163	SDJGI_67
AE	SDJGI_90	SDJGI_44	SDJGI_101	SDJGI_95	SDJGI_106	SDJGI_121	SDJGI_178	SDJGI_100	SDJGI_150	SDJGI_60	SDJGI_109	SDJGI_179
AF	SDJGI_77	SDJGI_43	SDJGI_92	SDJGI_131	SDJGI_74	SDJGI_126	SDJGI_158	SDJGI_120	SDJGI_66	SDJGI_171	SDJGI_62	SDJGI_75
AG	SDJGI_64	SDJGI_152	SDJGI_116	SDJGI_125	SDJGI_132	SDJGI_153	SDJGI_157	SDJGI_159	SDJGI_22	SDJGI_24	SDJGI_47	SDJGI_54
AH	SDJGI_63	SDJGI_69	SDJGI_76	SDJGI_45	SDJGI_36	SDJGI_175	SDJGI_176	SDJGI_193	SDJGI_10	SDJGI_11	SDJGI_13	SDJGI_14
BA	SDJGI_15	SDJGI_16	SDJGI_17	SDJGI_19	SDJGI_29	SDJGI_4	SDJGI_6	SDJGI_8	SDJGI_206	SDJGI_205	SDJGI_214	SDJGI_215
BB	SDJGI_208	SDJGI_210	SDJGI_211	SDJGI_37	SDJGI_103	SDJGI_110	SDJGI_113	SDJGI_115	SDJGI_82	SDJGI_122	SDJGI_87	SDJGI_130
BC	SDJGI_144	SDJGI_145	SDJGI_146	SDJGI_147	SDJGI_81	SDJGI_53	SDJGI_170	SDJGI_160	SDJGI_108	SDJGI_52	SDJGI_187	SDJGI_197
BD	SDJGI_190	SDJGI_191	SDJGI_192	SDJGI_196	SDJGI_177	SDJGI_117	SDJGI_56	SDJGI_156	SDJGI_123	SDJGI_68	SDJGI_163	SDJGI_88
BE	SDJGI_89	SDJGI_93	SDJGI_96	SDJGI_99	SDJGI_142	SDJGI_143	SDJGI_39	SDJGI_41	SDJGI_32	SDJGI_38	SDJGI_26	SDJGI_201
BF	SDJGI_1	SDJGI_202	SDJGI_204	SDJGI_21	SDJGI_114	SDJGI_133	SDJGI_119	SDJGI_186	SDJGI_173	SDJGI_194	SDJGI_127	SDJGI_94
BG	SDJGI_97	SDJGI_91	SDJGI_48	SDJGI_125	SDJGI_161	SDJGI_59	SDJGI_157	SDJGI_169	SDJGI_23	SDJGI_118	SDJGI_124	SDJGI_148
BH	SDJGI_154	SDJGI_155	SDJGI_164	SDJGI_45	SDJGI_36	SDJGI_175	SDJGI_176	SDJGI_193	SDJGI_10	SDJGI_11	SDJGI_13	SDJGI_14

Of the 194 GHs tested, approximately 68% exhibited minimal or no activity, with fluorescence signals close to background levels. The remaining enzymes displayed

variable activity levels, but GHA09 and GHA10 consistently produced the strongest signals, with relative fluorescence values significantly above other candidates. This strong activity indicated efficient hydrolysis of the glycosidic bond in the biotin-PEG₄-LacNAc-4MU substrate.

Further specificity analyses revealed that GHA09 and GHA10 preferentially hydrolyzed biotin-PEG₄-LacNAc-4MU over other fluorogenic glycosides. TLC assays were performed to validate these findings by monitoring the conversion of biotin-PEG₄-LacNAc-4MU to its hydrolysis product, GlcNAc-4MU. Incubations were carried out at 37°C, and reaction progress was assessed using a water:isopropanol:ethyl acetate (1:2:2, v/v/v) mobile phase. Distinct R_f values for the substrate (0.63) and product (0.11) confirmed substrate conversion by both GHA09 and GHA10, with complete substrate consumption observed after 16 hours of incubation with GHA10, suggesting higher catalytic efficiency for this enzyme.

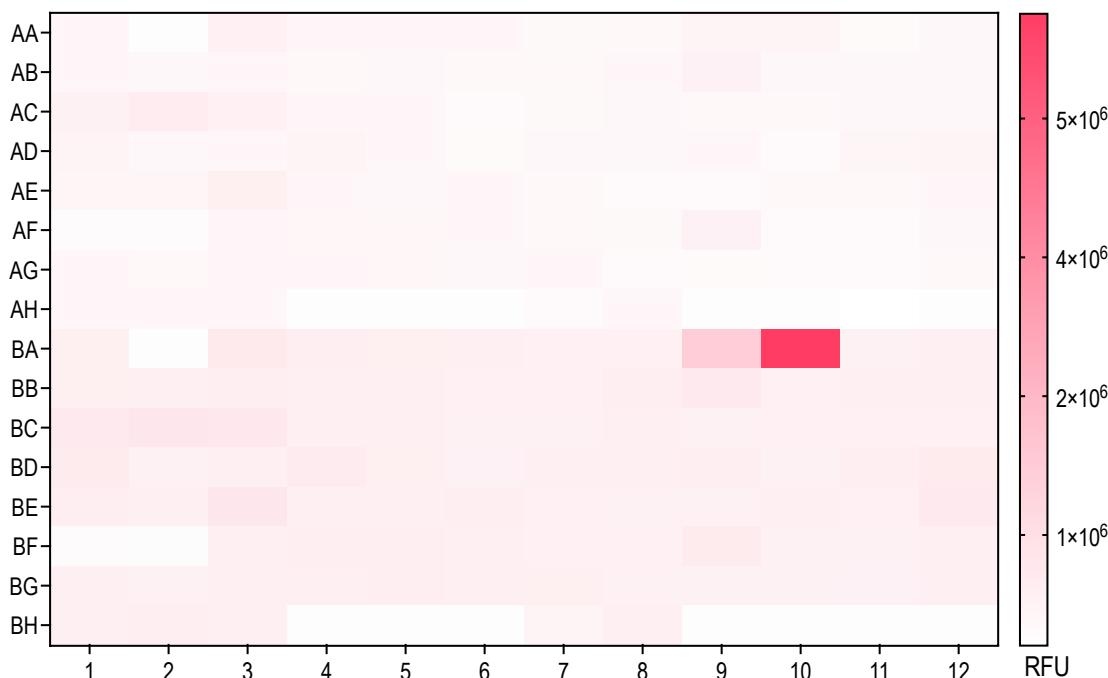


Figure 3.2 GH enzymes screening against biotinylated LacNAc.

Metagenomically derived GH enzymes screening against the biotin-PEG₄-LacNAc-4MU. Enzyme activity was assessed to evaluate the potential for processing O-GlcNAc-modified substrates, with higher intensities indicating greater enzymatic activity.

3.4.2. Enzyme Characterization and Kinetics

To facilitate the measurement of enzymatic activity, I first established a standard curve for 4MU concentration. (**Figure 3.3**) This was necessary because, in the assays, 4MU is released when the carbohydrate portions of the substrate are hydrolyzed by GHA09 and GHA10. By measuring the fluorescence of known concentrations of 4MU, I created a standard curve for quantifying the amount of 4MU produced during the enzymatic reactions. The standard curve was generated by preparing serial dilutions of 4MU in 50 mM Tris-HCl (pH 7.5) containing 200 mM NaCl. Fluorescence was measured at excitation and emission wavelengths of 360 nm and 465 nm, respectively, using a Synergy Neo2 hybrid multi-mode plate reader. Each data point was averaged across triplicate readings, and linear regression analysis yielded a correlation coefficient (R) of 0.993, confirming the accuracy of the curve for quantifying fluorescence signals.

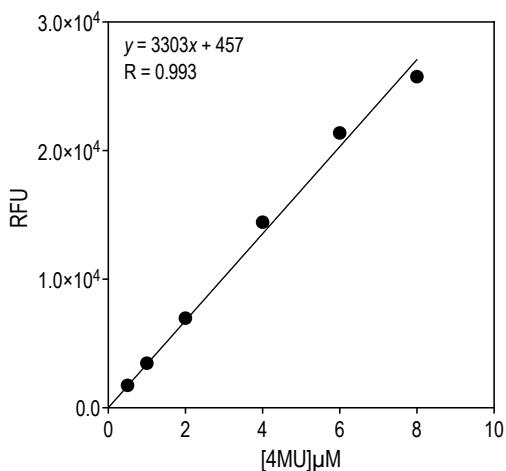


Figure 3.3 Standard curve for 4MU concentration

This graph shows the relationship between relative fluorescence units (RFU) and 4MU concentration, demonstrating good linearity with an $R = 0.993$. The measurements were taken in a buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl at a temp of 30°C.

GHA09 and GHA10 were produced in *E. coli* and purified using affinity chromatography. The yields obtained were approximately 35 mg/L for GHA09 and 48 mg/L for GHA10. Following expression, the proteins were purified using Ni-NTA chromatography with an imidazole gradient ranging from 0.1 to 0.5 M. The eluted proteins were dialyzed overnight into a final buffer containing 20 mM Tris-HCl (pH 7.5) and 200 mM NaCl to remove residual imidazole. Purity was confirmed by SDS-PAGE, with single

bands observed at the expected molecular weights for both GHA09 and GHA10, indicating >95% purity.

SDS-PAGE analysis confirmed the purity of both enzymes, with both GHA09 and GHA10 reaching a purity of >95% (**Figure 3.4a**). This level of purity supported advancing to subsequent activity assays. To assess substrate specificity of the enzymes, GHA09 and GHA10 were incubated with a variety of labeled carbohydrates, including α - and β -gal, α -glucose, and α - and β -GlcNAc. (**Figure 3.4b**) Substrates were prepared in Na-HEPES buffer (pH 7.0) and tested at final concentrations of 100 μ M in 96-well plates. Enzyme reactions were initiated by adding 10 μ L of purified enzyme solution to 40 μ L of substrate solution, and fluorescence was monitored at 360 nm (excitation) and 465 nm (emission) using a plate reader. All reactions were performed in triplicate to ensure data reproducibility.

Hydrolyzing only β -gal-4MU, as seen in the results, (**Figure 3.4b**) indicates that GHA09 and GHA10 are selective for β -galactosidic bonds. These data suggest that GHA09 and GHA10 should cleave off only the galactose residue, leaving the GlcNAc residue in place. This specificity is essential for application in the proposed strategy, where preserving the O-GlcNAc moiety is essential for enrichment. The pH optimum for both GHA09 and GHA10 was determined to be 6.0-6.5 (**Figure 3.4c**). To determine the pH profile, reactions were set up in a series of buffers covering a pH range from 4.5 to 8.0, each containing 50 mM of the corresponding buffering system (e.g., citrate, MES, HEPES). Enzymatic activity was assessed using β -gal-4MU as the substrate, and fluorescence measurements were recorded under standard conditions. Maximal activity was observed in MES buffer at pH 6.0-6.5, confirming this as the optimal pH range for hydrolysis. This narrow pH range supports the proper protonation of key catalytic residues in the active site, such as aspartic or glutamic acid, which are essential for the hydrolysis reaction catalyzed by these enzymes.¹⁸⁰ Maintaining this pH in future assays was important to achieve optimal enzyme activity, as small deviations from this range could lead to reduced activity or specificity.

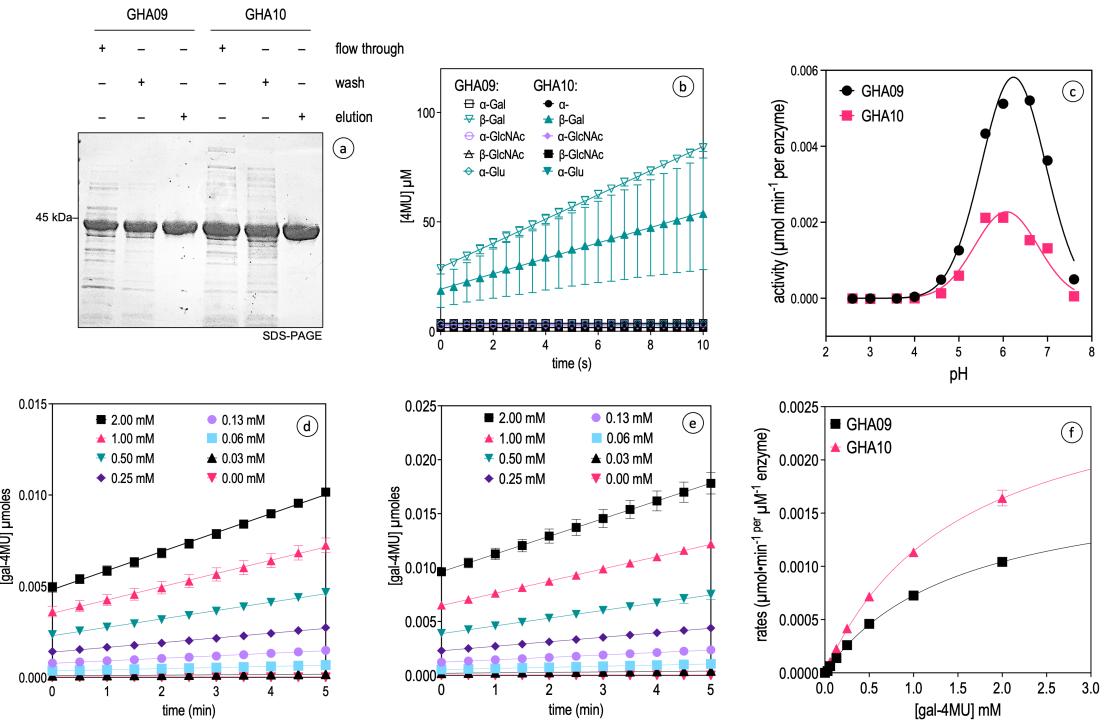


Figure 3.4 Characterization of GHs GHA09 and GHA10.

SDS-PAGE gel (a) analysis showing ~95% purity for GHA09 and GHA10. (b) Substrate specificity analysis demonstrating β-gal-4MU turnover by both enzymes. (c) pH profile for 0.1 μM GHA09 and 0.01 μM GHA10, with optimal activity at pH 6.0-6.5. (d and e) Initial rates of β-gal-4MU hydrolysis by GHA09 (d) and GHA10 (e) across a range of substrate concentrations at pH 7.5. (f) Michaelis-Menten kinetics derived from the data in panels (d) and (e). The measurements were taken in a buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl at a temp of 30°C.

Table 3.2 Kinetic parameters of enzymes GHA09 and GHA10 for β-gal-4MU.

Parameters	GHA09	GHA10	Units
V_{max}	0.0018 ± 0.0001	0.0029 ± 0.0001	$\mu\text{mol min}^{-1} \mu\text{M}^{-1} \text{enzyme}$
K_m	1.5 ± 0.1	1.5 ± 0.1	mM
k_{cat}	0.018 ± 0.001	0.029 ± 0.001	min^{-1}
k_{cat}/K_m	0.012 ± 0.001	0.019 ± 0.002	$\text{min}^{-1} \text{mM}^{-1}$

The initial rates of β-gal-4MU hydrolysis by GHA09 and GHA10 were measured across a range of substrate concentrations (**Figure 3.2d** and **3.2e**). Using these rates, Michaelis-Menten kinetics were applied to derive key enzymatic parameters, including V_{max} , K_m , and k_{cat} (**Table 3.1**). Reactions for kinetic analysis were set up in a 384-well plate, with each reaction containing 0.1 μM GHA09 or 0.01 μM GHA10 and substrate concentrations ranging from 0.03 mM to 2 mM. Fluorescence was monitored continuously

over a 10-minute period to capture the initial rates. The data were fitted to the Michaelis-Menten equation using GraphPad Prism to calculate the kinetic parameters. Based on these data GHA10 shows superior catalytic efficiency.

3.4.3. GH/OGA-Coupled Assay

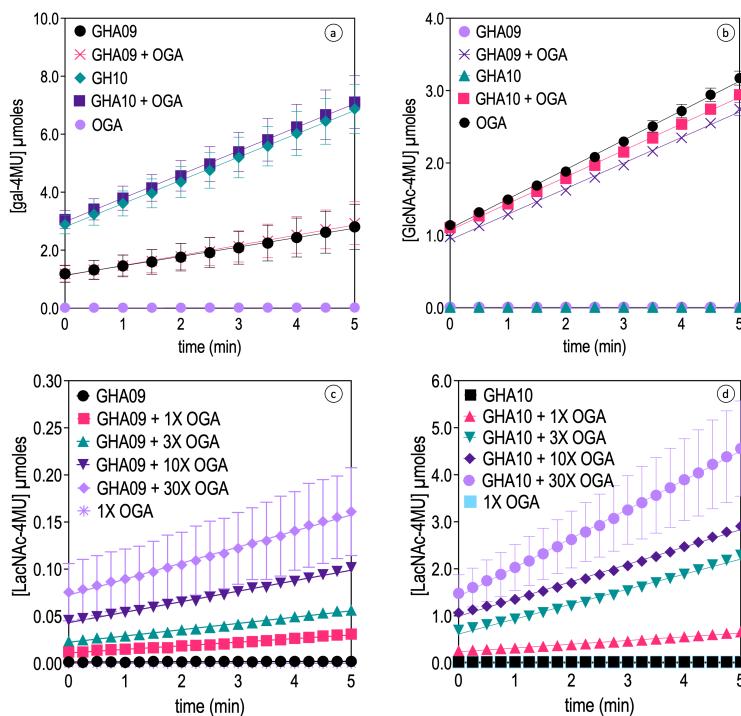


Figure 3.5 Substrate turnover by GHA09 and GHA10 in coupled assay.

Initial rates of GHA09 and GHA10 with (a) 5 mM Gal-4MU and (b) 5 mM GlcNAc-4MU (enzyme concentration of 0.1 μM). (c and d) 5 mM LacNAc-4MU turnover by GHA09 (0.1 μM) (c) and GHA10 (0.1 μM) (d) occurred only when both one of GHA09 or GHA10, and OGA (1x equals 0.1 μM) are present. Experiments were done in triplicate. The measurements were taken in a buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl at a temp of 30°C.

To evaluate the activity of GHA09 and GHA10 on a more relevant substrate for the proposed experiments using a coupled assay, I first tested β-gal-4MU in **Figure 3.5a**. For these assays, 5 mM solutions of β-gal-4MU were prepared in 50 mM Tris-HCl (pH 7.5) with 200 mM NaCl. Reactions were initiated by adding 0.1 μM of GHA09 or GHA10, and fluorescence was measured over time at excitation and emission wavelengths of 360 nm and 465 nm, respectively, using a Synergy Neo2 hybrid plate reader. Initial rates were calculated from the linear portion of the fluorescence progress curves. In this panel, substrate turnover was observed in both samples containing GH—whether GHA09 or

GHA10—regardless of the presence of OGA, confirming that both enzymes can efficiently hydrolyze β -gal-4MU on their own.

In Figure 3.5b, GlcNAc-4MU was tested as a substrate. Similar reaction conditions were used, except that 5 mM GlcNAc-4MU replaced β -gal-4MU as the substrate. Turnover was assessed in reactions containing only GHA09 or GHA10, as well as in reactions supplemented with 0.1 μ M OGA. Turnover was only observed in samples containing OGA, indicating that GHA09 and GHA10 are unable to hydrolyze GlcNAc-4MU.

The results for LacNAc-4MU turnover are shown in Figures 3.5c and 3.5d, representing GHA09 and GHA10, respectively. In both panels, turnover of LacNAc-4MU only occurred when both GH and OGA were present, confirming that the coupled action of both enzymes is necessary for liberation of the fluorescent 4MU moiety. Reactions were prepared with 5 mM LacNAc-4MU as the substrate, 0.1 μ M of GHA09 or GHA10, and varying concentrations of OGA. Incubations were carried out in a buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl at 30°C. Fluorescence was recorded continuously, and the initial rates of LacNAc-4MU turnover were determined as described above.

The reaction rate for the turnover of LacNAc-4MU turnover by GHA09 and GHA10 in the presence of varying concentrations of OGA are shown in Figure 3.6.

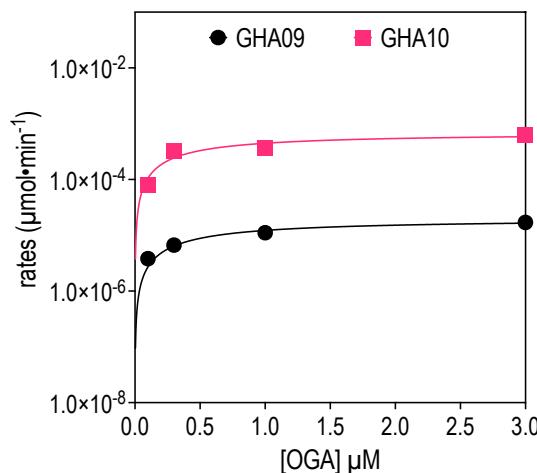


Figure 3.6 Reaction rates of GHA09 and GHA10 with LacNAc-4MU.

Reaction rates of 0.1 μ M GHA09 and GHA10 hydrolyzing LacNAc-4MU plotted against varying OGA concentrations. The data indicate that as OGA concentration increases, the reaction rate rises until a saturation point is reached at approximately three times the GH concentration. Beyond this point, the reaction becomes independent of OGA, with further turnover driven solely by GH activity. This illustrates that OGA is initially rate-limiting, but after sufficient OGA is present, the reaction rate is governed entirely by GH catalytic efficiency.

As the concentration of OGA was increased, a clear trend emerged: the reaction rate increased until the OGA concentration reached approximately three times that of the GH. This suggests that OGA initially limits the reaction due to its role in cleaving the GlcNAc residue, freeing the β -galactosidic bond for hydrolysis by the GH. Beyond this point, OGA concentration no longer impacts the reaction rate, as the GH becomes the rate-limiting enzyme in the coupled assay. To determine the optimal OGA concentration, reactions were conducted with OGA concentrations ranging from 0.05 μ M to 1.5 μ M, and the rate of LacNAc-4MU hydrolysis was monitored. At this point, the rate of the reaction plateaued, indicating that the reaction became independent of OGA concentration. Beyond this threshold, the reaction rate was solely defined by GH activity. Based on these data, I concluded that a concentration of 1 μ M OGA in the coupled enzyme assay is sufficient to ensure that the assay accurately reports on the GH enzyme activity at the concentrations used.

3.4.4. Steady-state kinetics of LacNAc turnover

Once OGA was confirmed not to be rate-limiting, I proceeded with a further kinetic analysis to quantify the efficiency and catalytic potential of GHA09 and GHA10 in the coupled assay. This analysis was essential for understanding their role in O-GlcNAc-modified protein enrichment.

For the kinetic experiments, LacNAc-4MU solutions were prepared at concentrations ranging from 0.05 mM to 5 mM in a buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl. Reactions were initiated by adding 0.1 μ M of GHA09 or GHA10 in the presence of 0.3 μ M OGA to ensure optimal turnover. Fluorescence was recorded in triplicate at excitation and emission wavelengths of 360 nm and 465 nm, respectively, using a Synergy Neo2 hybrid multi-mode plate reader. Initial reaction rates were calculated from the linear portion of the fluorescence progress curves.

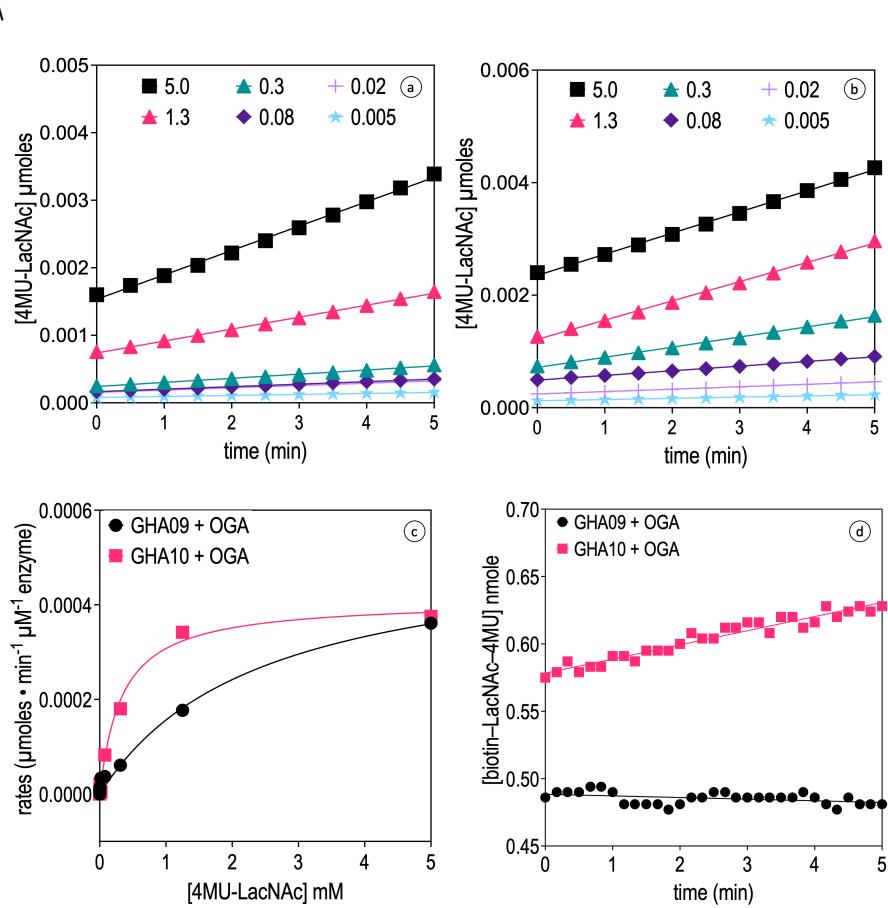


Figure 3.7 Kinetic analysis of GHs in Coupled assay for LacNAc turnover

Initial velocity measurements of LacNAc turnover in the showing GH-OGA coupled assay, demonstrating substrate processing by 0.1 μM GHA09 (a) and 0.1 μM GH10 (b). (c) Michaelis-Menten kinetics derived from the initial rates in panels (a) and (b), providing key kinetic parameters for both enzymes. (d) Initial velocity data for LacNAc-PEG₁₂-biotin-4MU cleavage by GHA09 and GH10 at varying substrate concentrations, ranging from micromolar to nanomolar, to assess the enzymes' catalytic efficiency. The measurements were taken in a buffer containing 50 mM Tris-HCl (pH 7.5) and 200 mM NaCl at a temp of 30°C.

Figures 3.4a and 3.4b present the initial rates of LacNAc turnover, showing that both GHA09 and GHA10 efficiently cleave LacNAc when combined with OGA, underscoring the critical role of this enzyme combination for effective substrate processing. Data points were fitted to the Michaelis-Menten equation using GraphPad Prism, providing estimates of V_{\max} , K_m , and k_{cat} . This approach allowed for a comparative analysis of the catalytic efficiency of both enzymes under identical conditions. To further quantify their catalytic efficiency, Michaelis-Menten kinetics were derived from these initial rates, as shown in Figure 3.4c. This analysis reveals key differences in the kinetic

parameters of GHA09 and GHA10, offering deeper insights into their relative efficiency. (Table 3.2)

Table 3.3 Kinetics of GHA09 and GHA10 using LacNAc-4MU as substrate.

Parameters	GHA09	GHA10	Units
V_{max}	0.533 ± 0.07	0.41 ± 0.02	nmol min ⁻¹ nM ⁻¹ enzyme
K_m	2.4 ± 0.8	0.33 ± 0.06	mM
k_{cat}	5.3 ± 0.7	4.1 ± 0.2	min ⁻¹
k_{cat}/K_m	2.2 ± 0.8	12.4 ± 2.4	min ⁻¹ mM ⁻¹

Reactions for LacNAc-PEG₁₂-biotin-4MU cleavage were prepared with substrate concentrations ranging from 0.05 μM to 5 μM. Due to limited substrate availability (0.8 μM), enzyme activity was assessed under suboptimal conditions. Reaction mixtures containing 0.1 μM GHA09 or GHA10 and 0.3 μM OGA were incubated at 30°C, and fluorescence was measured over time. Initial velocities were calculated but were constrained by the low substrate concentration, impacting the ability to derive comprehensive kinetic parameters for GHA09 and GHA10 under these conditions.

For GHA09, the V_{max} was determined to be 0.533 ± 0.07 nmol per min⁻¹ per nM⁻¹, with a K_m of 2.4 ± 0.8 mM, indicating that this enzyme requires a relatively high substrate concentration to achieve half-maximal velocity. In contrast, GHA10 exhibited a lower V_{max} of 0.41 ± 0.02 nmol per min⁻¹ per nM⁻¹ but with a significantly lower K_m of 0.33 ± 0.06 mM, suggesting that GHA10 can operate efficiently even at lower substrate concentrations. The turnover rate (k_{cat}) further highlights the catalytic disparity between the two enzymes. GHA09 exhibited a k_{cat} of 5.3 ± 0.7 min⁻¹, whereas GHA10's turnover rate was 4.1 ± 0.2 min⁻¹. When evaluating the catalytic efficiency via the k_{cat}/K_m ratio, GHA09 showed a modest value of 2.2 ± 0.8 min⁻¹ mM⁻¹, while GHA10 displayed a much higher efficiency of 12.4 ± 2.4 min⁻¹ mM⁻¹, making it the more efficient enzyme for LacNAc hydrolysis. These results highlight that while GHA09 is capable of LacNAc turnover, GHA10 is more efficient at lower substrate concentrations and exhibits overall superior catalytic performance. This distinction is particularly relevant for their application in the enrichment of O-GlcNAc-modified proteins, where GHA10's ability to function effectively with lower substrate levels can be leveraged for more sensitive and efficient substrate processing.

Compared to β -galactosidases reported in the literature, both GHA09 and GHA10 exhibit lower catalytic turnover rates. Wild-type β -galactosidases typically have k_{cat} values ranging from 50 to 100 min⁻¹,¹⁸¹⁻¹⁸³ and engineered variants with His-tags have been reported to show lower activity but still retain k_{cat} values significantly higher than those observed here. Similarly, β -galactosidases from various microbial sources display catalytic efficiencies (k_{cat}/K_m) in the range of 100–500 min⁻¹ mM⁻¹,¹⁸³⁻¹⁸⁵ whereas GHA10, despite being more efficient than GHA09, exhibits a much lower efficiency of 12.4 min⁻¹ mM⁻¹. This suggests that while GHA10 operates effectively at low substrate concentrations, it does not match the catalytic performance of industrial β -galactosidases. However, its ability to function under low-substrate conditions may still provide advantages in specialized applications, such as O-GlcNAc enrichment.

Preliminary analysis of both GHA09 and GH10's ability to cleave LacNAc-PEG₁₂-biotin-4MU, as shown in **Figure 3.4d**, was limited by substrate availability. Due to the low substrate concentration (approximately 0.8 μ M), I was only able to measure a single initial velocity curve for each enzyme. Under these conditions, GH10 exhibited detectable turnover, while GHA09 showed no measurable activity. The substrate concentration appeared to be insufficient for both enzymes, with only GH10 showing any turnover. These preliminary findings emphasize the need for further experiments with higher substrate concentrations to fully assess the kinetic parameters of both GHA09 and GH10. Future work will focus on optimizing substrate levels and repeating these measurements to gain a comprehensive understanding of their catalytic efficiency.

3.5. Discussion

The chemoenzymatic method described in this chapter addresses key challenges in the detection and enrichment of stoichiometric,³² dynamically regulated O-GlcNAcylated proteins.¹⁶⁵ By using β 4B4GALT11 to transfer bulky galactose derivatives, such as PEG₁₂-biotin,¹⁷⁵ to O-GlcNAc residues, this method offers an efficient, reversible enrichment strategy that preserves the native O-GlcNAc structure. Unlike traditional approaches—such as periodate oxidation and azido-based labeling, which irreversibly modify the O-GlcNAc site—this technique could allow for the removal of the galactose moiety via glycosidase-mediated cleavage, restoring the native O-GlcNAc modification. The use of PEG₁₂-biotin could minimize steric hindrance during affinity enrichment, and

high-throughput glycoside hydrolase screening ensures the identification of optimal enzymes for precise, reversible enrichment and analysis of O-GlcNAc-modified proteins.

Coupled enzyme assay and specificity. The results from the GH/OGA-coupled assays establish that OGA enables the selective hydrolysis of LacNAc-4MU, with both GHA09 and GHA10 demonstrating efficient turnover only when OGA is present. This coupled assay ensures that OGA is not rate-limiting, allowing for direct monitoring of GH activity. The enzyme-specific cleavage of β -gal-4MU and LacNAc-PEG₁₂-biotin-4MU by GHA09 and GHA10 highlights the substrate specificity of these enzymes and their potential utility in O-GlcNAc-modified protein enrichment workflows. The kinetic analyses of these enzymes, particularly the Michaelis-Menten parameters, further validate their catalytic efficiency, with GHA10 showing superior performance, as evidenced by its higher turnover rates and k_{cat}/K_m values.

Compared to other glycoside hydrolase families, the rate constants (k_{cat} and k_{cat}/K_m) observed for GHA09 and GHA10 are markedly lower than those of industrially relevant GH enzymes.¹⁸¹⁻¹⁸⁵ For instance, members of GH2, GH35, and GH42 families—commonly associated with β -galactosidase activity—often exhibit k_{cat} values exceeding 100 min⁻¹ and catalytic efficiencies in the range of 100–500 min⁻¹ mM⁻¹.¹⁸³ In contrast, GHA09 and GHA10 demonstrate significantly lower turnover rates, with GHA10 achieving a k_{cat} of only 4.1 min⁻¹ and a k_{cat}/K_m of 12.4 min⁻¹ mM⁻¹. This suggests that while these enzymes are capable of LacNAc hydrolysis, they do not match the catalytic efficiency of classical β -galactosidases. However, their ability to function effectively under low-substrate conditions may still provide advantages in specialized applications, such as O-GlcNAc enrichment, where enzyme selectivity and controlled cleavage may be more critical than high turnover rates.

Future work and optimization. While the results presented here demonstrate the potential of this chemoenzymatic method, certain limitations—such as low substrate concentrations in the LacNAc-PEG₁₂-biotin-4MU assays—hindered a comprehensive analysis of GHA09 and GHA10 activity. Fortunately, Dr. Cuhna has provided enough of the probe to allow for rerunning these experiments with optimal substrate concentrations, ensuring a thorough analysis of both enzymes' activity. Future work will focus on optimizing substrate availability, repeating kinetic measurements, and expanding the analysis to include a broader range of glycan substrates. A critical next step will involve testing the ability of this method to specifically pull-down O-GlcNAc-modified proteins. To

do this, I will mix purified O-GlcNAc-modified protein(s) with non-glycosylated, purified counterparts and assess whether the method can selectively isolate only the O-GlcNAc-modified proteins. This could be achieved using magnetic streptavidin-coated beads to pull down the biotinylated O-GlcNAc proteins, verifying the specificity of the method. The success of this approach will demonstrate the utility of the method for isolating O-GlcNAc-modified proteins from complex mixtures, while leaving non-glycosylated proteins unbound.

This chemoenzymatic method holds significant promise for improving O-GlcNAc enrichment. The successful identification of GHs that preserve the O-GlcNAc moiety opens new avenues for studying this post-translational modification. Additionally, future comparisons with GH enzymes beyond β -galactosidases, particularly those with known O-GlcNAc hydrolase activity, may help contextualize the efficiency of GHA09 and GHA10 in broader glycan-modifying applications. Understanding how these enzymes compare in selectivity and efficiency to those in GH84, GH89, and GH20 families could further refine their applicability in O-GlcNAc research. Further optimization, particularly in pull-down experiments and kinetic characterization, will be critical for refining the method and enhancing its applicability in both simplified and complex biological systems.

3.6. Acknowledgments

I owe my deepest gratitude to Prof. David J. Vocadlo, whose sharp intellect and steady guidance steered this research from the abstract to the tangible. His wisdom, patience, and occasional tough love were indispensable in making this work what it is.

I am deeply appreciative to Dr. Stephen Withers and Dr. Feng Liu at the University of British Columbia for their invaluable support in screening glycoside hydrolases. Their expertise and contributions were pivotal in identifying the enzymes that laid the foundation for the success of this project.

And to Dr. Vitor Cunha, for expertly synthesizing the probes that became the backbone of this project—thank you. Without your skillful hands and generous collaboration, much of this work would have remained an idea waiting for form.

A special nod goes to Prof. Xi Chen and the University of California, Davis, for providing the galactosyltransferases (B4GALT1s) that enabled me to push the boundaries of this study.

My heartfelt thanks also go to Jun Bian, who could always be counted on for deep dives into chemical characterization and for providing support when things got technically tricky. To Dr. Sha Zhu, your expertise and input helped smooth out experimental bumps and refine the details that matter.

Last but by no means least, to my lab mates: your camaraderie, humor, and shared curiosity made even the toughest days in the lab a joy. You all played a crucial role in moving this work forward, and for that, I'm profoundly grateful.

Chapter 4.

BIOPHYSICAL ANALYSIS OF O-GLCNACYLATION

The research presented in this chapter was made possible through the contributions of several collaborators. Dr. Florian Malard performed the molecular dynamics (MD) simulations, providing key computational insights into the structural effects of O-GlcNAcylation. Prof. Stephanie Olivier-Van Stechelen contributed to the computational analysis of O-GlcNAcylation sites and their solvent exposure, offering critical expertise in interpreting the structural and thermodynamic implications of glycosylation. The stopped-flow fluorescence experiments were conducted in collaboration with Prof. Cornelia Bohne's team at the University of Victoria, whose technical expertise was invaluable for analyzing protein folding kinetics. Dr. Edwin Escobar and Prof. Jennifer Brodbelt performed mass spectrometry analyses for O-GlcNAc site mapping, ensuring precise identification of glycosylation modifications. Their collective contributions significantly enriched the biophysical characterization of O-GlcNAcylation presented in this chapter.

4.1. Background and Importance

As established earlier, O-GlcNAcylation is a dynamic PTM involved in numerous cellular processes¹⁸⁶, and alterations in this modification can contribute to various diseases^{5,187}. To fully understand this connection, it is important to examine how O-GlcNAcylation affects proteins at the molecular level, particularly regarding their stability. The stability of proteins is a key factor in their function, but the term can have different meanings depending on the context. In fields like cell biology and biochemistry, stability often refers to the degradation rate of a protein within the cell—that is, its susceptibility to proteolytic pathways. In contrast, in biophysics, stability pertains to thermodynamic stability—the ability of a protein to maintain its native, folded conformation under varying conditions.

The stability of proteins at the molecular level can be described through the free energy landscape model. This model illustrates the energetic states a protein may occupy

during folding or unfolding.¹⁸⁸⁻¹⁹⁰ The landscape is depicted as a funnel-shaped surface, where each point represents a state of the protein, ranging from unfolded and misfolded conformations to the native, functional state at the lowest energy point. (**Figure 4.1**) As proteins navigate this landscape, they encounter energy barriers that must be overcome to reach the most stable configuration. In thermodynamic equilibrium, proteins follow a path that minimizes free energy, systematically moving toward the native state. This process is represented on the left side of the funnel in **Figure 4.1**, where the distribution indicates that lower energy correlates with a higher prevalence of protein states. Conversely, under kinetic control, proteins may become ensnared in local minima, trapped by kinetic barriers that hinder their progression toward the global minimum. This scenario is depicted on the right side of the funnel in (**Figure 4.1**), showing proteins in a kinetic trap. Kinetic trapping often leads to proteins becoming ensnared in less favorable states such as aggregates, yet it can also serve a protective role by stabilizing bioactive states against aggregation.¹⁹¹

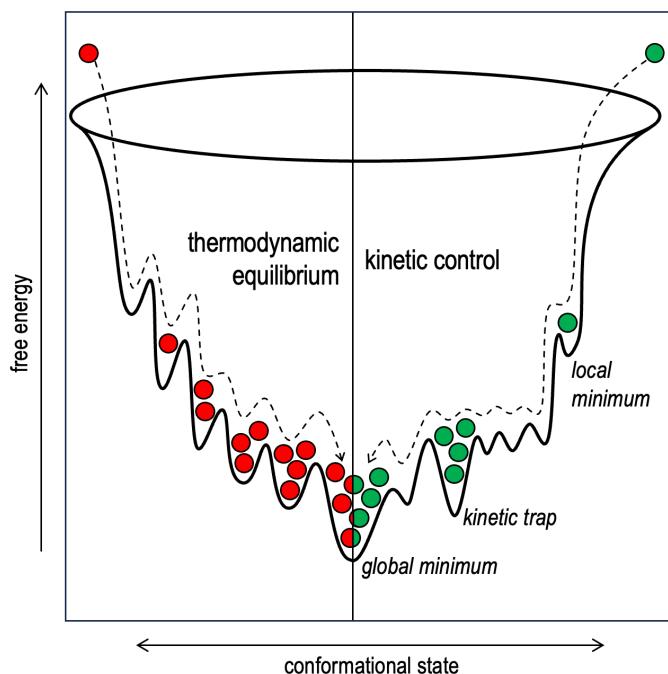


Figure 4.1 Funnel of proteins' free energy landscape.

This figure illustrates a 3D funnel-shaped representation of the protein free energy landscape. The left side of the funnel shows a set of proteins (red dots) progressing through thermodynamic equilibrium, depicted by a distribution indicating that lower energy states contain more proteins. The right side of the funnel (in green) displays proteins under kinetic control, including a group trapped in a local minimum indicative of a kinetic barrier. Both sides converge at the bottom at a global minimum, representing the native state of the proteins.

The energy profile of protein folding, delineating the transition between misfolded, transition, and native states, highlights the crucial role of activation energy (E_A) in navigating these stages. (Figure 4.2) Activation energy facilitates the overcoming of kinetic barriers, enabling the protein to transition from less stable states to its most stable, native conformation. This concept is depicted in Figure 4.2 through a graph that visualizes how kinetic barriers can trap proteins in local minima, illustrating the interplay between kinetic control and thermodynamic stability in the protein folding process. Further, the graph demonstrates changes in free energy (ΔE), illustrating how shifts in thermodynamic control dictate the movement of the protein from higher-energy misfolded states to the energetically favorable native state. This view provided by the free energy landscape not only illustrates the complex interplay between kinetic barriers and thermodynamic pathways but also underscores the essential mechanisms by which proteins achieve their stable, native conformations through intricate folding pathways.

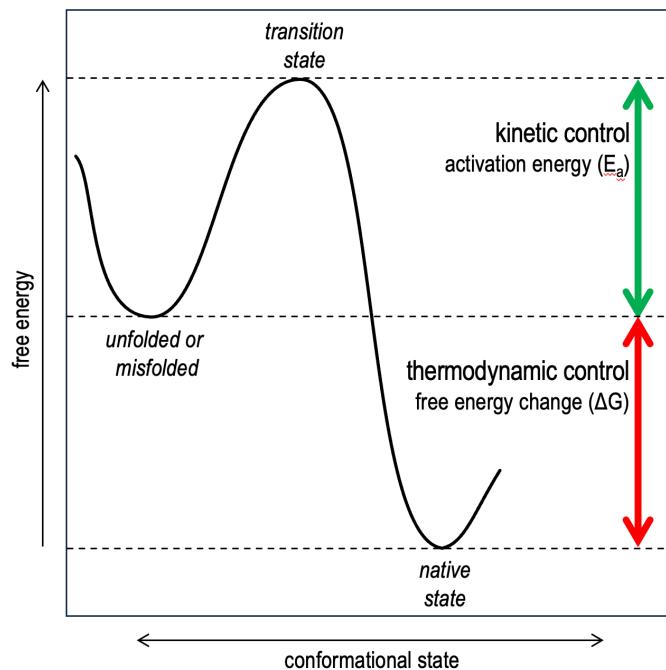


Figure 4.2 Energy profile during protein folding.

This graph presents the typical energy profile associated with protein folding. It identifies key states including misfolded, transition, and native conformations. The plot illustrates the activation energy required to transition between these states, with arrows indicating the direction of protein folding towards the energetically favorable native state.

Perturbations such as PTMs can reshape a protein's energy landscape, altering the height of energy barriers or creating new intermediate states. For example, glycosylation—a PTM where carbohydrate moieties are added to specific amino acid

residues—has been shown to stabilize proteins both thermodynamically and kinetically.¹⁹² Studies on the SH3 domain demonstrated that *N*-linked glycosylation increases the folding temperature and enhances protein stability, with the effect proportional to the number of attached glycans. This stabilization arises primarily from an increase in the enthalpy of the unfolded state due to glycosylation, which raises the energy level of the unfolded state and effectively deepens the energy well of the native state. Furthermore, the specific site of glycosylation plays a crucial role in modulating protein energetics. Introducing an enhanced aromatic sequon—a structural motif involving an aromatic amino acid like phenylalanine near the glycosylation site—can further stabilize the native state.¹⁹³ This is achieved through specific interactions between the aromatic side chain, the attached glycan, and nearby residues, forming a compact and stable structure. Such site-specific modifications highlight how PTMs can fine-tune protein stability by influencing both the thermodynamic depth of the native state's energy well and the kinetic barriers to unfolding.^{192,193}

In contrast, ubiquitination—a PTM where ubiquitin is attached to lysine residues—can destabilize proteins by altering their energy landscapes.¹⁹⁴ Site-specific ubiquitination has been found to decrease thermodynamic stability by destabilizing the folded state, increasing the population of partially unfolded states. This destabilization facilitates recognition and degradation by the proteasome, demonstrating how PTMs can also promote protein turnover by modulating stability. These examples illustrate that PTMs can influence both thermodynamic stability, which is closely tied to a protein's ability to maintain its structure under physiological and stress conditions, and kinetic stability, which refers to how long a protein remains folded before transitioning to less favorable states like unfolding or aggregation. By modifying specific residues, PTMs like glycosylation and ubiquitination can significantly impact how a protein moves through its energy landscape.

By analogy to these other systems, one can reason that O-GlcNAcylation may enhance the thermodynamic and kinetic stability of proteins, influencing their folding pathways and resistance to denaturation. This modification could smooth the energy landscape or lower energy barriers, making it easier for the protein to remain in its native folded state or resist unfolding. By potentially increasing the energy of unfolded states or stabilizing specific structural motifs, O-GlcNAcylation might help proteins avoid misfolding or aggregation. Although few examples have been documented, one case is the O-GlcNAcylation of the intrinsically disordered microtubule-associated protein tau (MAPT)—

which tends to aggregate in Alzheimer's disease—has been observed to reduce its propensity to aggregate.^{46,195,196} This suggests that O-GlcNAcylation may help maintain tau in a stable soluble form, preventing the adaptation of a conformation that leads to oligomerization and downstream formation of toxic aggregates.⁴⁶

Similarly, in Parkinson's disease (PD), decreased levels of O-GlcNAcylation on the intrinsically disordered protein α -synuclein have been linked to increased aggregation of this protein, which forms the characteristic pathology seen in PD known as Lewy bodies.^{47,197,198} Enhancing O-GlcNAcylation of α -synuclein has been shown to reduce its aggregation propensity, indicating a potential role in enhancing its solubility. In Huntington's disease, the mutant huntingtin protein forms harmful aggregates, and studies indicate that O-GlcNAcylation of huntingtin can reduce its aggregation and toxicity.¹⁹⁹⁻²⁰¹ This effect might be due to O-GlcNAcylation preventing misfolding of huntingtin. These examples illustrate how O-GlcNAcylation of specific proteins can influence their stability and aggregation behavior, potentially affecting disease progression in the examples noted above. Consistent with the idea that O-GlcNAcylation can affect protein aggregation and stabilize proteins, studies have shown that this modification slows the thermal aggregation of proteins such as TGF-beta-activated kinase 1 (TAB1),⁴⁵ specificity protein 1 (Sp1),⁴¹ tumor protein p53, and²⁰² polyhomeotic-proximal chromatin protein (Ph-P).³⁵

Conceptual support for this idea arises from studies of ectothermic organisms, whose body temperatures vary with the environment, that have shown elevated levels of O-GlcNAcylation contribute to their ability to develop and survive at higher temperatures.³⁶ Thus, one might reason that increased O-GlcNAc levels in *Caenorhabditis elegans* could enhance protein stability under heat stress, aiding proper development and survival. A logical corollary is that O-GlcNAcylation may protect proteins against thermal denaturation by stabilizing their folded structures. Studies using cellular thermal shift assays (CETSA) revealed that O-GlcNAcylated Nod2 has an increased melting temperature compared to its non-modified form, though the effect was modest and not entirely compelling.⁵² Nevertheless, this increase in melting temperature indicates that modified Nod2 remains folded and functional at higher temperatures than the unmodified form and is consistent with the overall concept of O-GlcNAc affecting protein stability.

Collectively, these examples indicate that O-GlcNAcylation is associated with enhanced protein stability, possibly by influencing folding pathways and reducing the likelihood of misfolding and aggregation. However, while these pioneering studies suggest

a stabilizing role for O-GlcNAcylation, detailed folding studies are lacking and the specific molecular mechanisms by which the addition of the GlcNAc moiety influences protein folding and stability are not fully understood. The precise interactions, structural changes, and effects on the energy landscape of protein folding induced by O-GlcNAcylation remain to be elucidated.

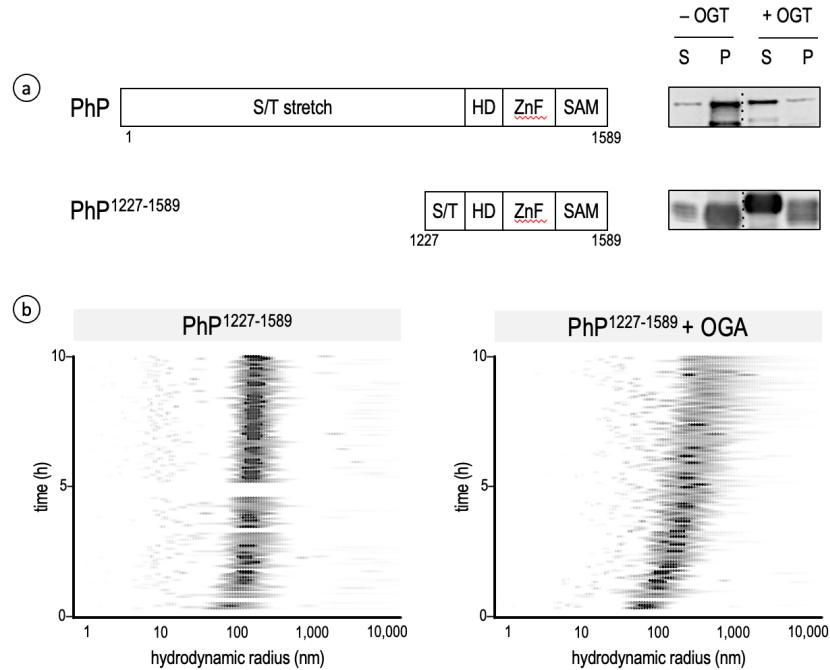


Figure 4.3 O-GlcNAc prevents PhP aggregation as per Gambetta and Müller (2014). Schematic of the full-length PhP protein showing key domains (a): serine/threonine-rich (S/T) stretch, homeodomain (HD), zinc finger (ZnF), and SAM domain. To the right, immunoblot analysis of soluble (S) and insoluble (P) fractions from *Drosophila* S2 cells expressing PhP, either co-expressed with OGT or in cells where endogenous OGT was conditionally deleted (OGT-deficient, OGT^{mat/zyg-}). The absence of O-GlcNAcylation results in PhP aggregation into the insoluble fraction. Below, a similar analysis for PhP¹²²⁷⁻¹⁵⁸⁹. (b) DLS analysis of purified PhP¹²²⁷⁻¹⁵⁸⁹ shows a dispersed state when O-GlcNAcylated and aggregation after treatment with *C. perfringens* OGA, as indicated by increased particle size. Figure adapted from Gambetta and Müller (2014).⁴⁸

To address this gap, I focused on the Polyhomeotic-proximal chromatin protein (UniProt accession number P39769; Ph-P; **Figure 4.3a**), a crucial component of the Polycomb Group (PcG) proteins involved in chromatin remodeling and gene repression.²⁰³ Ph-P plays an essential role in maintaining the repressed state of developmental genes, thereby regulating cellular identity and differentiation.^{85,204} Importantly, Ph-P is known to be O-GlcNAcylated,^{8,10,48,205} but the impact of this modification on its thermodynamic

stability and folding dynamics has not been thoroughly investigated. However, O-GlcNAcylation of a site near the N-terminus of the SAM (sterile alpha motif) domain of Ph-P has been shown to reduce its aggregation propensity. By studying Ph-P, I aimed to elucidate how O-GlcNAcylation may affect protein stability at the molecular level, providing insights that could extend to other proteins and contribute to a broader understanding of O-GlcNAc-mediated regulation.

4.2. Objective

In addition to investigating the effects of O-GlcNAcylation on Ph-P, I spearheaded the design of computational studies aimed at identifying protein targets where O-GlcNAc modifications might induce measurable biophysical effects, particularly within structured protein regions. Using a computational approach, performed by Dr. Florian Malard, we focused on determining whether O-GlcNAc modifications could influence protein structure, in a way that it would influence function. From this analysis, we identified three likely cases: catalase (CAT), glucose-6-phosphate dehydrogenase (G6PD), and lactate dehydrogenase A (LDHA). These observations suggest: the possibility that O-GlcNAcylation could affect the thermodynamic stability of enzymes such as CAT, G6PD, and LDHA, potentially influencing their metabolic roles. That is, O-GlcNAc might initiate a feedback mechanism where, under certain conditions, the organism adjusts the O-GlcNAcylation of these enzymes to regulate metabolic pathways. This scenario could lead to metabolic feedback that impacts the production of O-GlcNAc itself, potentially aiding the organism in adapting to environmental challenges. Though speculative this idea serves as a starting point for a more detailed discussion later in the text, aiming to explore how metabolic adaptations might be linked to protein modifications.

The primary objective of chapter 4 is to investigate the biophysical effects of O-GlcNAcylation on the Ph-P protein, aiming to elucidate the molecular mechanisms by which O-GlcNAcylation influences protein thermodynamic stability and folding dynamics. By focusing on Ph-P, we seek to understand how the addition of the O-GlcNAc moiety affects protein conformation, stability, and aggregation propensity at the molecular level. To achieve this, I first assess the thermodynamic stability of Ph-P with and without O-GlcNAcylation using biophysical techniques such as circular dichroism (CD) spectroscopy and fluorescence thermal denaturation. These methods allowed me to measure changes in the protein's secondary structure and monitor thermal unfolding processes, providing

insights into the stabilizing effects conferred by O-GlcNAcylation. Next, I examined the folding and unfolding kinetics of Ph-P through stopped-flow fluorescence and equilibrium denaturation experiments. These studies enable us to observe protein folding pathways and determine how O-GlcNAcylation influences the kinetics of folding and unfolding. Understanding these dynamics is crucial for elucidating the mechanisms by which O-GlcNAcylation affects protein stability.

Additionally, I evaluated the impact of O-GlcNAcylation on protein aggregation by investigating the aggregation propensity of Ph-P under thermal and chaotropic denaturation conditions. By comparing the aggregation behavior of glycosylated and non-glycosylated Ph-P, I aim to determine whether O-GlcNAcylation reduces misfolding and aggregation, which are critical factors in protein stability and function. To complement our experimental studies, we conduct computational modeling, including molecular dynamics (MD) simulations, to visualize structural changes induced by O-GlcNAcylation on Ph-P. These simulations provide atomic-level insights into how the O-GlcNAc moiety interacts with specific amino acid residues, potentially stabilizing certain conformations and influencing the protein energy landscape. We explored the generality of the effect O-GlcNAcylation's effects on protein stability by performing computational analyses to identify and investigate other candidate proteins that may be similarly affected by this modification. This broader approach aims to determine whether the stabilizing effects observed in Ph-P are applicable to other proteins, thereby enhancing our understanding of the general role of O-GlcNAcylation in protein stability.

By accomplishing these objectives, this study seeks to fill the gap in knowledge regarding the molecular basis of O-GlcNAc-mediated protein stabilization. The insights gained from this research could contribute to a broader understanding of protein homeostasis and inform potential therapeutic strategies for diseases associated with protein misfolding and instability.

4.3. Methods

4.3.1. Cloning

A *Drosophila melanogaster* Ph-P^{1227–1589} gene segment (NCBI accession number CAA45211; **Figure 4.1b**) featuring a truncated N-terminal stretch to reduce molecular

size, was initially cloned and expressed, following the protocols outlined previously.⁴⁸ This segment, which becomes insoluble without co-expression with OGT, demonstrated O-GlcNAc-dependent protection against aggregation. For optimal expression, this segment was codon-optimized, synthesized, and subcloned into the pET28a(+) vector using NdeI and BamHI restriction sites by Genscript. In parallel, the full-length *Homo sapiens* OGT^{1–1046} gene (NCBI accession O15294.3) was integrated into the pMal-c2X vector's (New England Biolabs) multiple cloning site as described previously.²⁰⁶ These steps were aimed at enhancing the expression efficiency and purity of the protein for in-depth studies on the role of O-GlcNAcylation in protein stability. Due to persistent issues with low yield and purity in the Ph-P^{1227–1589} construct, an alternative construct, Ph-P^{1397–1589} (**Figure 4.1c**), was subsequently designed.

4.3.2. Protein Production

The Ph-P-pET28a constructs and the OGT-pMal-c2X plasmid were co-transformed into One Shot™ BL21(DE3) chemically competent *E. coli* cells (Invitrogen C600003) via heat shock transformation.²⁰⁶ Approximately 200 ng of each plasmid DNA were used for the co-transformation. To ensure the maintenance of both plasmids, transformed cells were selected on LB agar plates containing kanamycin (50 µg/mL) for the pET28a vector and ampicillin (100 µg/mL) for the pMal-c2X vector. This co-expression system allowed simultaneous expression of Ph-P and OGT for subsequent experiments. The cells were grown at 37°C until an OD₆₀₀ of 0.7 was reached, at which point protein expression was induced by adding 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). The cultures were then incubated for 16 h at 16°C. Harvested cells (20 g) were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mg/mL lysozyme, 1 EDTA-free protease inhibitor tablet from Roche), lysed by sonication, and centrifuged at 45,000 g for 1 h. The supernatant was passed through a 1 mL HiTrap HP nickel column (Cytiva) pre-equilibrated with lysis buffer. The column was washed with 10 column volumes of wash buffer containing 50 mM Tris, pH 7.5, 200 mM NaCl, and 10 mM imidazole. Proteins were eluted using a gradient of imidazole (0.1 to 0.5 M) over 50 minutes at a flow rate of 1 mL per min. The elution buffer consisted of 50 mM Tris, pH 7.5, 200 mM NaCl, and 1 M imidazole. Fractions of 2 mL containing purified proteins were pooled, dialyzed overnight into the final buffer (20 mM Tris, pH 7.5, 200 mM NaCl), and concentrated to approximately 1 mg/mL.

4.3.3. Immunoblots

To confirm that the Ph-P–OGT co-expression system was glycosylating the Ph-P protein constructs, samples were treated with either the wild-type or the D242A catalytically dead mutant form of *Bacteroides thetaiotaomicron* OGA (*Bt*OGA) for 4 hours at 25 °C. Subsequently, I loaded 1 µg of each sample onto a 10% SDS-PAGE gel and performed electrophoresis at 70 V for 1.5 hours. Following electrophoresis, the proteins were transferred to 0.45 µm nitrocellulose membranes using the TransBlot Turbo (Bio-Rad) semi-dry transfer system, following the manufacturer's instructions. To detect specific glycosylation, I blocked the blots in a PBS buffer containing 4% BSA for one hour, applied primary antibodies including a mouse anti-O-GlcNAc antibody (BioLegend, CTD110.6) and a rabbit monoclonal Anti-His antibody (Invitrogen; 21HCLC), and incubated the blots overnight at 4°C. The blots were then washed with PBST (phosphate buffered saline containing 0.1% Tween 20) and incubated with secondary IRDye® 680RD Goat anti-Mouse IgG secondary antibody (LICOR) for 1 h at room tempe, followed by another wash with PBST. Visualization was achieved using LI-COR fluorescent secondary antibodies, according to the gel imaging system's protocols.

4.3.4. Intact protein MS

Glycosylated and non-glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ constructs were adjusted to 0.2 mg/ml in 0.1% formic acid. A 1 µl aliquot of each protein solution was injected into a Waters Acquity UPLC BEH C8 column (130 Å, 1.7 µm, 2.1 × 50 mm) at 35°C. Proteins were separated using a gradient from 2% to 80% solvent B (acetonitrile with 0.1% formic acid) over 15 minutes, followed by 2 minutes at 80% B, at a flow rate of 200 µl/min. MS analysis was performed using an Bruker Maxis Impact ultra-high resolution quadrupole time-of-flight (QTOF) LC/MS/MS system (Bruker Daltonics). Ionization conditions included a spray voltage of 3.6 kV, capillary temperature of 325°C, and sheath gas flow rate of 25. Mass spectra were collected over an m/z range of 500–5,000 with a resolution of 40,000 m/z. Data were processed using Bruker Daltonics deconvolution software using standard parameters.

4.3.5. Site Mapping

To prepare the protein samples for O-GlcNAc site mapping analysis, I first separated the O-GlcNAc-modified Ph-P^{1397–1589} using SDS-PAGE (10% gel). Following electrophoresis, I stained the gels with Coomassie Brilliant Blue 250 (Sigma Aldrich) to visualize Ph-P^{1397–1589} protein bands. I carefully excised the bands corresponding to the Ph-P proteins from the gel, and then proceeded to destain the gel pieces to remove excess stain. After destaining, the proteins within these gel pieces were reduced with 10 mM dithiothreitol and alkylated with 55 mM iodoacetamide. Following destaining, I performed in-gel trypsin digestion directly within the gel pieces to break down the proteins into peptides. Once digested, I sent these gel pieces containing the peptides to our collaborators for detailed LC-MS analysis.

Our collaborators extracted the peptides from the gel using a solution of 0.1% formic acid in 70% acetonitrile, ensuring efficient recovery of the peptides for subsequent mass spectrometric analysis. Then, they conducted the LC-MS using a Dionex RSLC 3000 nano-LC system and a Thermo Scientific Lumos Orbitrap mass spectrometer, employing techniques such as HCD, UVPD, and EThcD for fragmentation. They used Byonic software for the identification and site localization of O-GlcNAcylated peptides, with the Delta Mod Score and PEP 2D score providing measures of confidence in peptide identification. Each glycopeptide was manually validated to ensure accuracy. This process was crucial for mapping the glycosylation of Ph-P^{1397–1589} and understanding the O-GlcNAcylation patterns within the co-expression system. The detailed findings and methodologies from this collaborative effort have been reported,¹⁴⁸ highlighting the specialized analytical work that facilitated our research.

4.3.6. CD spectroscopy

Far-ultraviolet (UV) CD experiments were performed using a Chirascan Plus spectrometer (Applied Photophysics), operated with Chirascan Pro-Data SX software (v2.5.0). Measurements were taken using quartz cuvettes with a 2.0 mm path length. Ph-P^{1397–1589} was prepared at a concentration of 5 mg/mL in 10 mM Tris buffer at pH 7.5. Spectral data were collected at 25°C with a wavelength interval of 1.0 nm, scanning from 195 to 300 nm for 500 µL samples. The resulting millidegree output was plotted as a function of wavelength using GraPh-Pad Prism (v10). For the thermal stability

assessment, Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was diluted to 0.1 mg/mL in 10 mM Tris buffer at pH 7.5. CD measurements were performed at 222 nm – ideal for monitoring protein α -helical content– using the same 2 mm quartz cuvettes as in the far-UV CD experiments. A temperature ramp was conducted from 20°C to 95°C at a rate of 1°C/min, with data collected every 0.2 – 0.5 °C. The temperature was controlled using a Peltier element (Quantum Northwest), and the temperature of the sample was monitored via a temperature probe. The unfolding profile of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was evaluated based on the collected data.

4.3.7. Intrinsic tryptophan fluorescence

Fluorescence spectroscopy experiments were performed using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was prepared at a concentration of 0.1 mg/mL in 10 mM Tris buffer at pH 7.5. Samples were placed in quartz cuvettes with a 2 mm path length, and intrinsic tryptophan fluorescence was monitored by exciting the sample at 280 nm and recording emission spectra from 300 to 400 nm. For the thermal stability assessment, Ph-P¹³⁹⁷⁻¹⁵⁸⁹ samples were prepared at 0.1 mg/mL in 10 mM Tris buffer (pH 7.5). It is important to note that the pKa of Tris buffer is temp-sensitive, which leads to fluctuations in pH during the thermal stability experiments. These pH changes may convolute the analysis of the protein's stability. While for this study we assume that the temp effects on protein stability are more significant than any potential pH shifts, addressing this issue in greater depth would require further investigation to fully understand the impact of these fluctuations.

Fluorescence emission was monitored at a fixed excitation wavelength of 280 nm, and the emission at 330 nm was recorded to detect changes in the protein tertiary structure. The temp ramp was conducted from 20°C to 95°C at a rate of 1°C/min, with data collected every 0.2°C. The temperature was controlled using the built-in Peltier element of the Cary Eclipse, and the unfolding profile of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was plotted as a function of temperature to evaluate the thermal denaturation process. Fluorescence was monitored in a bandwidth corresponding to tryptophan's intrinsic fluorescence (emission at 330 nm), which reflects changes in protein tertiary structure and can indicate whether the protein remains in solution or aggregates during unfolding. A two-state folding model was applied to the thermal denaturation data to determine the transition midpoint (T_m) from the resulting sigmoidal curve.

For the urea denaturation experiment, Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was prepared at a concentration of 0.1 mg/mL in 10 mM Tris buffer at pH 7.5. Increasing concentrations of urea (up to 8 M) were added to the protein solution. Samples were aliquoted into Corning™ 384-Well Solid Black Polystyrene Microplates (Fisher Sci), at a volume of 50 µL per well. The intrinsic tryptophan fluorescence was monitored using a BioTek fluorescence plate reader (BioAgilytix), with excitation at 280 nm and emission at 330 nm. Each sample was allowed to equilibrate for 1 h at room temperature before measurements were taken. Experiments were conducted in triplicates to ensure reproducibility and accuracy. The unfolding of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ in the presence of urea was assessed by observing changes in fluorescence intensity, indicative of alterations in the protein's tertiary structure as it interacts with the denaturant. Fluorescence data were analyzed using a Boltzmann equation to extract the denaturation midpoint (C_m), providing insight into the protein's stability and solubility under chemical denaturation.

To further explore the dynamic folding and unfolding mechanisms of Ph-P¹³⁹⁷⁻¹⁵⁸⁹, stopped-flow fluorescence experiments were conducted at the University of Victoria by the team of Prof. Cornelia Bohne. Using a Cary Eclipse fluorescence spectrophotometer, these experiments were designed to capture rapid conformational changes in the protein in response to varying urea concentrations. The protein was excited at 295 nm, and emissions were recorded using a WG360 filter, specifically chosen to track changes in tryptophan fluorescence that might indicate changes in folding. For these studies, Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was diluted in 10 mM Tris buffer at pH 7.5, and rapidly mixed with urea solutions ranging from 0 to 4 M to yield a final concentration of 0.1 mg/mL in the mixing conditions. This setup enabled the observation of both folding and unfolding processes over time scales of 2, 20, and 50 seconds. The rapid mixing essential for capturing these transient states was achieved using a stopped-flow apparatus (Applied Photophysics PMT), ensuring precise synchronization of sample mixing and data acquisition.

4.3.8. Lysate thermal denaturation

As previously described in section 4.3.2, the expression of Ph-P¹²²⁷⁻¹⁵⁸⁹ and Ph-P¹³⁹⁷⁻¹⁵⁸⁹ constructs in *E. coli* was produced under optimized conditions. Following protein production, cells were harvested and resuspended in phosphate-buffered saline (PBS) containing protease inhibitors (Roche), then centrifuged at 5000 g for 10 min to collect the cell pellet. This pellet was resuspended in 500 µL PBS with protease inhibitor cocktail

tablet (Roche). The resuspended were then subjected to a temperature gradient from 40°C to 68.5°C in a thermocycler (BioRad), with each temperature held for 5 min. After incubation, the samples were immediately placed on ice for 5 min. Cells were then lysed by sonication and centrifuged at 20,000 g for 20 minutes. Approximately 20 µg of protein from the resulting supernatants where then loaded in separate wells onto an SDS-PAGE gel, prepared with 5x Laemmli loading dye (100 mM Tris-HCl, pH 6.8, 4% SDS, 12% glycerol, 0.008% bromophenol blue, 2% β-mercaptoethanol), and boiled for 5 minutes. Immunoblotting was subsequently performed as described in section 4.3.3, to assess the expression levels and thermostability of constructs Ph-P¹²²⁷⁻¹⁵⁸⁹ and Ph-P¹³⁹⁷⁻¹⁵⁸⁹ under these conditions.

4.3.9. Computational Analysis

A computational analysis was undertaken to identify potential O-GlcNAc sites on protein structures that might influence their thermodynamic stability. In collaboration with Prof. Stephanie Olivier-Van Stechelen and Dr. Florian Malard at the Medical College of Wisconsin, we utilized AlphaFold2²⁰⁷ as the primary database for our analysis. AlphaFold2 was selected not only for its extensive collection of known 3D structures but also for its predictions on proteins where no structures have yet been resolved. This approach allowed us to expand our cohort significantly, incorporating predicted structures alongside experimentally determined ones. To enrich our analysis, we integrated UniProt²⁰⁸ annotations for these proteins, which encompass the entire human proteome. This integration was crucial for identifying structured regions within the proteins, enabling a targeted searching for O-GlcNAc sites within these structured regions. The third component of our analysis involved the O-GlcNAc database,⁶⁸ which catalogs O-GlcNAc sites that have been mapped to specific residues. This database allowed us to examine O-GlcNAc on a large set of proteins and allowing us to select those that may harbor O-GlcNAc sites on structured regions.

To quantify the extent to which O-GlcNAc is solvent exposed at these sites, we utilized the prediction-aware part-sphere exposure (pPSE) metric developed by Bludau et al. (2022).²⁰⁹ (**Figure 4.11c**) pPSE calculates how “buried” an amino acid is within a protein, considering not just the solvent accessibility but also the orientation of side chains and the prediction error associated with the relevant region of each AlphaFold model. Specifically, pPSE was calculated using a 12 Å radius centered on the α carbon of the O-

GlcNAc-modified amino acid, with a 70° angle oriented towards the β carbon or the side chain of the amino acid, reflecting the average size of amino acids and their side chain flexibility. This pPSE metric, alongside the solvent accessible surface area (SASA), provided a dual approach to evaluate to extent to which the environment around each O-GlcNAc sites was structured. The higher the pPSE (or lower the SASA score), the more other amino acids are near the modified amino acid, indicating a more structured environment. These computational tools provided a robust method to sift through nearly 10,000 known O-GlcNAcylated proteins. This method was particularly adept at highlighting sites on structured regions that we reasoned might exert a measurable thermodynamic stabilizing (or destabilizing) effect on proteins, making it a good approach for identifying promising candidates for further biophysical study.

4.3.10. Molecular Dynamics

Following the computational analysis to identify potential O-GlcNAc sites on structured regions of proteins, we concentrated our molecular dynamics (MD) studies on the top three hits. These candidates were selected based on the highest confidence in O-GlcNAc mapping, as quantified by a metric developed by Wulff-Fuentes et al. (2021),⁶⁸ the O-GlcNAc score (S). The selected proteins were also around 30 kDa in size, making them suitable candidates for further exploratory studies in *E. coli* dual expression systems due to their manageable size. Additionally, these proteins were annotated as enzymes, providing an opportunity to measure the functional impact of O-GlcNAcylation more distinctly. The O-GlcNAc score was designed to quantify the exhaustiveness of each entry in the O-GlcNAc database based on the available literature. This score, S(x), is calculated as the sum of several normalized factors, each reflecting different aspects of the literature on a given protein entry:

$$S(x) = R(x)^{\text{norm}} + C(x)^{\text{norm}} + T(x)^{\text{norm}} + fA(x)^{\text{norm}} + IA(x)^{\text{norm}} + B(x)^{\text{norm}}$$

Equation 4.1 Protein O-GlcNAc site mapping confidence score equation.

Given x as the list of all protein entries and x as a single entry, and considering the list of references N_x and P as the number of protein entries documented in an index i of N_x :

- R represents the length of the list of references N_x .

- C is the sum of per-year citations for each index i of N_x .
- T is the time span between the first and last reference publication.
- fA and lA are the number of distinct first and last authors, respectively, within N_x .
- B is a bonus term computed for each index i of N_x and averaged over R. Higher P_i negatively impacts B whereas higher C positively impacts B_i .

Each factor contributes within a [0,1] interval, with the total score $S(x)$ theoretically ranging up to six, scaled up to 100 for user-friendly interpretation in our Explore panel.

These scoring criteria allowed us to prioritize proteins not only based on structural data but also considering the depth and breadth of literature evidence supporting their modification by O-GlcNAc. By selecting enzymes as candidates, I aimed to assess how O-GlcNAcylation may influence protein functionality.

Molecular dynamics (MD) simulations were conducted by Dr. Malard using the GROMACS software package, applying the CHARMM force field to both O-GlcNAc-modified and unmodified protein models. The proteins were solvated in cubic boxes with TIP3P water, ensuring a minimum buffer of 10 Å around the protein to the box edges, and the system was neutralized by adding counterions. Energy minimization was performed using the steepest descent method to remove any steric clashes, followed by 100 picoseconds (ps) of NVT (number, volume, temp) equilibration using the V-rescale thermostat and 100 ps of NPT equilibration with the Parrinello-Rahman barostat to stabilize pressure and temperature. The production phase of the MD simulations extended for 100 ps, recording snapshots every 10 ps to capture the dynamic behavior and potential conformational changes, providing detailed insights into how O-GlcNAcylation influences protein stability and folding dynamics.

4.4. Results

4.4.1. Ph-P¹²²⁷⁻¹⁵⁹⁸ thermal denaturation

I selected the Ph-P¹²²⁷⁻¹⁵⁹⁸ construct protein due to its relatively small size (approximately 20 kDa), which facilitates easier expression and purification in *E. coli*. Additionally, this construct demonstrated a similar O-GlcNAc dependence as the full-length protein, as shown previously.⁴⁸ The light scattering studies (**Figure 4.3b**) provided a foundation for further analysis of Ph-P¹²²⁷⁻¹⁵⁹⁸ O-GlcNAc dependent aggregation behavior and structural properties.

To generate Ph-P¹²²⁷⁻¹⁵⁹⁸, the construct was codon-optimized for *E. coli* expression and subcloned into the pET28a(+) vector. The expression system included an N-terminal His tag to facilitate purification using nickel-affinity chromatography. The Ph-P¹²²⁷⁻¹⁵⁹⁸ construct was co-expressed with the OGT-pMal-c2X vector in BL21(DE3) chemically competent cells, enabling simultaneous expression of Ph-P and OGT for glycosylation. Following transformation, cells were grown in LB medium supplemented with kanamycin (50 µg/mL) and ampicillin (100 µg/mL) to maintain plasmid selection. Expression was induced with 1 mM IPTG when cultures reached an OD₆₀₀ of 0.7, followed by incubation at 16°C for 16 hours to promote protein folding and glycosylation. Cells were harvested by centrifugation and lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mg/mL lysozyme, and an EDTA-free protease inhibitor cocktail. The lysate was clarified by centrifugation at 45,000 g for 1 hour, and the supernatant was applied to a 1 mL HiTrap HP nickel-affinity column pre-equilibrated with lysis buffer. After washing with buffer containing 10 mM imidazole, proteins were eluted using a gradient of 10–500 mM imidazole. Fractions containing Ph-P¹²²⁷⁻¹⁵⁹⁸ were pooled, dialyzed into 20 mM Tris-HCl (pH 7.5) with 200 mM NaCl, and concentrated to ~1 mg/mL.

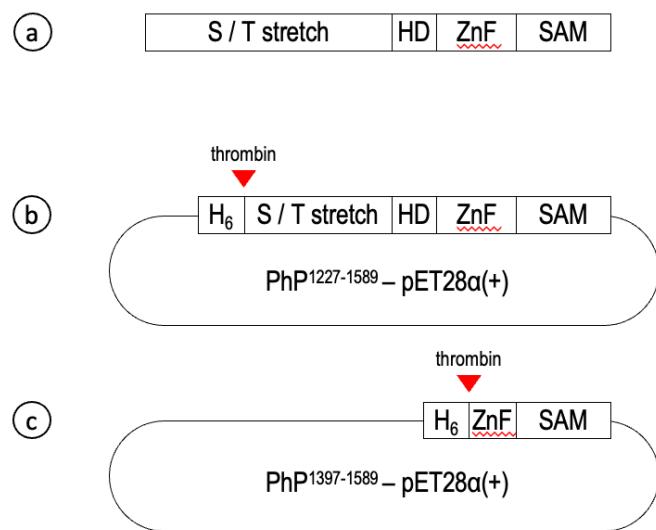


Figure 4.4 Schematic Ph-P domain structure and derived constructs.

Full-length Ph-P showing the domain organization (a), including a serine/threonine (S/T) stretch rich in serine and threonine residues, a homeodomain (HD), a zinc finger (ZnF), and a sterile alpha motif (SAM). (b) Construct Ph-P¹²²⁷⁻¹⁵⁸⁹ derived from Gambetta and Müller (2014),⁴⁸ featuring a truncated N-terminal S/T stretch to reduce molecular size, subcloned into the pET28a(+) vector. (c) Construct Ph-P¹³⁹⁷⁻¹⁵⁸⁹, designed to include only a portion of the ZnF and the complete SAM, subcloned into the pET28a(+) vector. This construct emphasizes a streamlined domain composition focusing on essential structural elements for functional studies.

Given this reasoning, I assessed the effectiveness of a dual OGT-substrate expression system to generate glycosylate recombinant Ph-P¹²²⁷⁻¹⁵⁹⁸ within *E. coli*. (**Figure 4.4**) To confirm glycosylation, samples were treated with wild-type or catalytically dead mutant (D242A) BtOGA enzymes in the presence or absence of the inhibitor thiamet-G. Immunoblots were performed with anti-O-GlcNAc (CTD110.6) and anti-His antibodies to detect glycosylated and total protein, respectively. This step ensured that the Ph-P¹²²⁷⁻¹⁵⁹⁸ was successfully glycosylated. Analysis of the protein products by immunoblot show a clear band detected by anti-O-GlcNAc antibody (CTD110.6) indicating O-GlcNAc-modification of Ph-P¹²²⁷⁻¹⁵⁹⁸. (**Figure 4.5a**) Treatment with bacterial *BtOGA*²¹⁰ significantly diminished the intensity of these bands, except in the presence of *BtOGA* inhibitor thiamet-G, confirming that Ph-P¹²²⁷⁻¹⁵⁹⁸ is glycosylated. This evidence supports the functionality of the expression system in producing O-GlcNAcylated Ph-P¹²²⁷⁻¹⁵⁹⁸, crucial for maintaining protein stability.

To evaluate the thermostability of Ph-P¹²²⁷⁻¹⁵⁹⁸, I performed a thermal gradient denaturation experiment using recombinant expression of Ph-P¹²²⁷⁻¹⁵⁹⁸ using both dual expression system to generate glycosylated Ph-P¹²²⁷⁻¹⁵⁹⁸ and non-glycosylated Ph-P¹²²⁷⁻¹⁵⁹⁸. (**Figure 4.5b**) Cell lysates were subjected to a temperature gradient from 37°C to 70°C in a thermocycler. After incubation at each temperature for 5 minutes, the samples were immediately cooled on ice and centrifuged at 20,000 g for 20 minutes. Supernatants containing soluble protein were analyzed by SDS-PAGE and immunoblotting. The bacterial cell samples were subjected to boiling in the presence of protease inhibitors for five minutes at specified temperatures, then cooled to 4°C. The supernatant was then collected and analyzed by immunoblot. (**Figure 4.5b**) This setup aimed to provide a more natural environment for the protein. The results demonstrate that the O-GlcNAc-modified Ph-P¹²²⁷⁻¹⁵⁹⁸ exhibits increased resistance to heat-induced precipitation, maintaining stability up to 68.5°C. In contrast, the non-O-GlcNAc modified Ph-P¹²²⁷⁻¹⁵⁹⁸ begins to precipitate even at the initial incubation temperature. These observations highlight the crucial role of O-GlcNAc modifications in enhancing the thermal stability and aggregation resistance of the Ph-P¹²²⁷⁻¹⁵⁹⁸. Although cells do not typically encounter extreme temperatures, thermal stability assays provide a practical way to assess how modifications like O-GlcNAc influence protein folding and stability. Increased resistance to thermal denaturation often correlates with greater stability under physiologically relevant stresses, such as molecular crowding, oxidative stress, or proteasomal degradation. This relationship is particularly evident in ectothermic organisms, where O-GlcNAcylation has

been shown to help proteins withstand temperature fluctuations. For example, studies in *Caenorhabditis elegans* demonstrate that elevated O-GlcNAc levels enhance protein stability, aiding survival under heat stress.³⁶ These findings suggest that, even in homeothermic systems, O-GlcNAcylation may serve as a general mechanism for maintaining protein integrity under diverse cellular stresses. By analyzing heat-induced unfolding, we can infer how O-GlcNAcylation contributes to protein resilience and function in different environmental and physiological contexts.

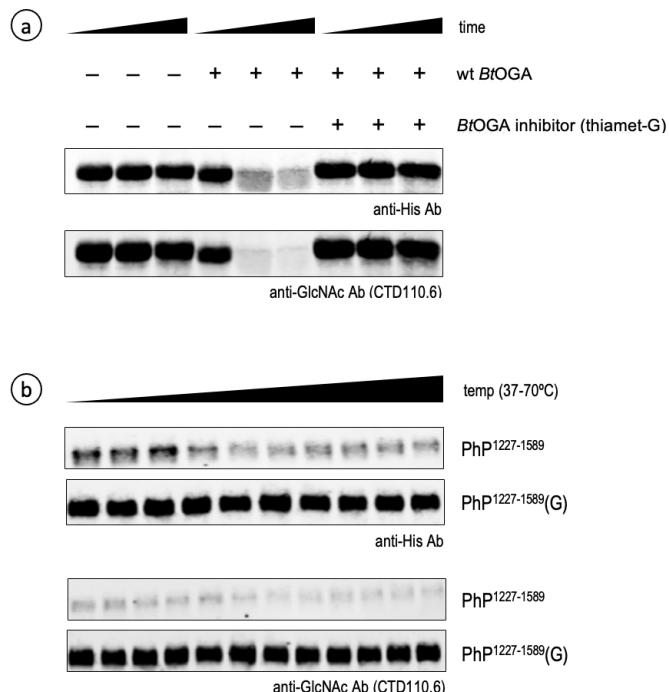


Figure 4.5 Effect of O-GlcNAc on Ph-P¹²²⁷⁻¹⁵⁹⁸ thermal stability.

(a) Verification of O-GlcNAc modification of Ph-P¹²²⁷⁻¹⁵⁹⁸ produced in *E. coli*. Treatment of glycosylated Ph-P¹²²⁷⁻¹⁵⁹⁸ within *E. coli* with wt *BtOGA* and thiamet-G in various conditions confirm this protein is glycosylated. Top panel shows total protein detection (anti-His blot), and the bottom panel assesses the levels of O-GlcNAc (anti-O-GlcNAc blot). (b) Thermal stability of Ph-P¹²²⁷⁻¹⁵⁹⁸ is evaluated using a temp gradient of 37 – 70°C. Anti-His blots (top panel) detect total protein, while anti-O-GlcNAc blots (bottom panel) monitor glycosylation. Immunoblots were done in technical duplicates and biological triplicates.

Despite these promising results, the initial construct Ph-P¹²²⁷⁻¹⁵⁹⁸ based on Gambetta's research⁴⁸ proved problematic to purify due to its low yield. During purification using a His-tag nickel affinity column. SDS-PAGE analysis revealed only a very faint band corresponding to the Ph-P¹²²⁷⁻¹⁵⁹⁸, overshadowed by numerous other bands indicative of contaminating proteins. (**Figure 4.6**) To identify these contaminants, protein bands were excised from the SDS-PAGE gel, destained, and subjected to in-gel trypsin digestion. The

resulting peptides were analyzed by LC-MS, which confirmed the identity of the contaminants as unrelated proteins, including molecular chaperones, metabolic enzymes, and transcription factors. The results confirmed that these were not fragments of Ph-P¹²²⁷⁻¹⁵⁹⁸ but rather unrelated proteins, including various chaperones (DnaK, Ch60, SlyD), transcription-related proteins (RpoB, RL15, LacI), metabolic enzymes (GlmS, NadK, ArnA, Ndh, DapB, RlmG), and others such as MalE, Maltoporin, PK-2, and Crp.

For these reasons of low yield and purity further use of Ph-P¹²²⁷⁻¹⁵⁹⁸ was abandoned.

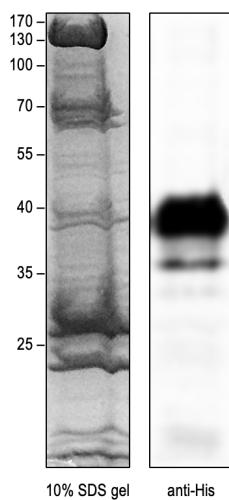


Figure 4.6 Analysis of recombinant Ph-P¹²²⁷⁻¹⁵⁹⁸ purification.

(Left) SDS-PAGE gel showing protein bands that were sent for MS analysis, revealing multiple contaminating proteins alongside a faint band corresponding to Ph-P¹²²⁷⁻¹⁵⁹⁸. (Right) Anti-His blot for Ph-P¹²²⁷⁻¹⁵⁹⁸ after recombinant expression and purification. The band for Ph-P¹²²⁷⁻¹⁵⁹⁸ is notably smaller than the contaminating proteins, highlighting the challenge of achieving a high yield of purified Ph-P¹²²⁷⁻¹⁵⁹⁸.

4.4.2. Ph-P¹³⁹⁷⁻¹⁵⁸⁹ lysate thermal denaturation

With the aim of addressing the poor expression of Ph-P¹²²⁷⁻¹⁵⁹⁸ construct, I designed Ph-P¹³⁹⁷⁻¹⁵⁸⁹, a smaller, single-domain polypeptide derived from the C-terminal SAM domain known to be structured.²¹¹ The thermal stability of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was examined. (Figure 4.4) To produce this construct, the Ph-P¹³⁹⁷⁻¹⁵⁸⁹ sequence was codon-optimized for E. coli and subcloned into the pET28a(+) vector using NdeI and BamHI restriction sites. The vector included an N-terminal His tag to facilitate purification. The Ph-P¹³⁹⁷⁻¹⁵⁸⁹ construct was co-expressed with OGT in BL21(DE3) cells under the same conditions as Ph-P¹²²⁷⁻¹⁵⁹⁸: cultures were grown in LB medium with kanamycin (50

$\mu\text{g/mL}$) and ampicillin (100 $\mu\text{g/mL}$), and expression was induced with 1 mM IPTG at an OD₆₀₀ of 0.7. Cells were incubated at 16°C for 16 hours to promote proper folding and glycosylation. Following cell lysis and centrifugation, the supernatant was passed through a 1 mL HiTrap HP nickel-affinity column pre-equilibrated with lysis buffer. Proteins were eluted using a gradient of 10–500 mM imidazole. Elution fractions were pooled, dialyzed into 20 mM Tris-HCl (pH 7.5) with 200 mM NaCl, and concentrated to ~1 mg/mL.

In contrast to the Ph-P¹²²⁷⁻¹⁵⁹⁸ construct, this new construct exhibited both a high yield and a purity of approximately 95%, as assessed by Coomassie staining of SDS-PAGE gels, with protein concentrations quantified using a NanoDrop spectrophotometer. However, the results for both the glycosylated and unglycosylated forms neither aggregated, indicating that this construct resists aggregation regardless of its glycosylation status. To evaluate thermostability, Ph-P1397-1589 samples were subjected to a temperature gradient ranging from 37°C to 70°C in a thermocycler. Supernatants collected after centrifugation were analyzed by immunoblotting. Anti-His blots confirmed equal loading of total protein across all temperature points, while anti-O-GlcNAc blots monitored glycosylation status.

While anti-O-GlcNAc blots confirmed the presence of glycosylation, the lack of differential stability prompted further examination to confirm the glycosylation status of these constructs. The glycosylated and non-glycosylated forms of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ were subjected to intact protein mass spectrometry to confirm the addition of O-GlcNAc moieties. Samples were prepared at a concentration of 0.2 mg/mL in 0.1% formic acid and separated on a UPLC C8 column using a gradient of 2–80% acetonitrile with 0.1% formic acid. MS spectra were deconvoluted to determine mass shifts corresponding to glycosylation. These results verified that glycosylation occurred on the expected residues.

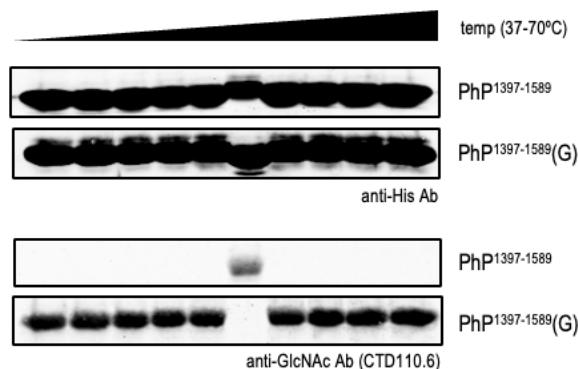


Figure 4.7 Thermostability analysis of Ph-P¹³⁹⁷⁻¹⁵⁸⁹.

Experiments conducted on the Ph-P¹³⁹⁷⁻¹⁵⁸⁹ construct to evaluate its thermal stability. Anti-His blots (top panel) confirm equal amounts of total protein across the temp range, while anti-O-GlcNAc blots (bottom panel) monitor glycosylation status. The data show that Ph-P¹³⁹⁷⁻¹⁵⁸⁹ is stable irrespective of glycosylation. (NOTE: Lane 6, samples were inadvertently swapped; O-GlcNAcylated was switched with non-O-GlcNAcylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹.)

4.4.3. Assessment of O-GlcNAcylation of Ph-P¹³⁹⁷⁻¹⁵⁸⁹

Intact mass spectrometry analysis of purified Ph-P¹³⁹⁷⁻¹⁵⁸⁹ protein revealed a primary peak at 22,042 Da, consistent with the predicted molecular weight of the unglycosylated protein (21904.94 Da). Additionally, a secondary peak at 22,221 had a mass increase of approximately 178 Da (179 Da) consistent with spontaneous α -N-gluconylation of the His-tag, a modification commonly observed in proteins expressed with in *E. coli* that have a polyhistidine tag.^{151,152} For the O-GlcNAc-modified Ph-P¹³⁹⁷⁻¹⁵⁸⁹, the spectrum showed five distinct peaks, indicating the presence of between 1 and 5 O-GlcNAc units, with each glycosylation event contributing a mass increase of +203 Da. (Figure 4.5)

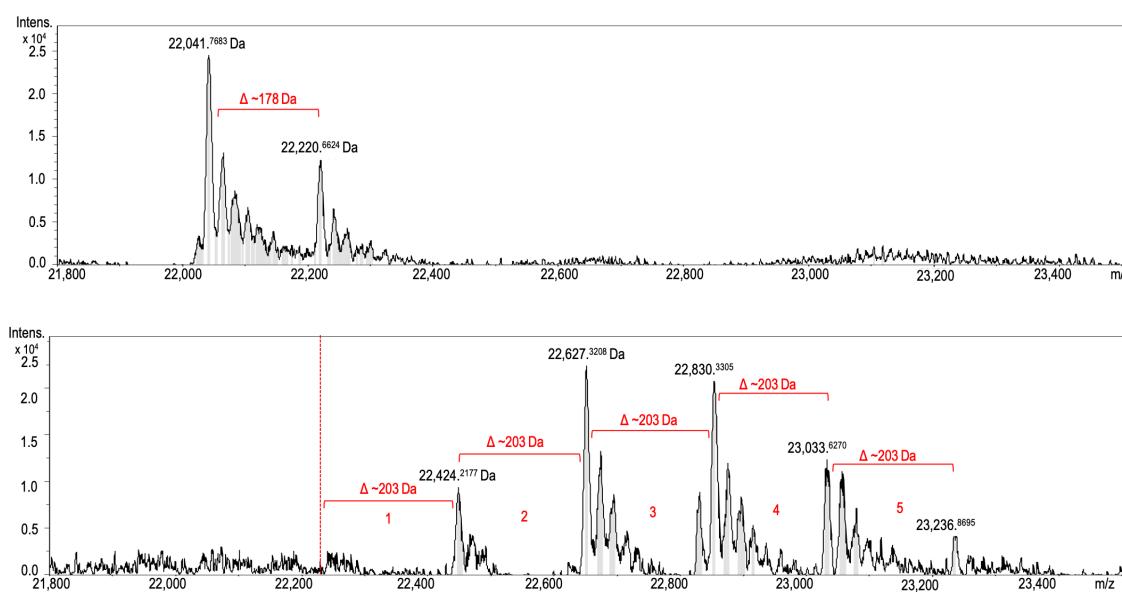


Figure 4.8 Assessment of O-GlcNAcylation of Ph-P¹³⁹⁷⁻¹⁵⁸⁹.

The spectra depict the molecular weights of unglycosylated and glycosylated forms of the Ph-P¹³⁹⁷⁻¹⁵⁸⁹ construct. (Top) The unglycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ shows the expected molecular weight at 22,042.08 Da with a secondary peak corresponding to a mass shift of +178 Da. (Bottom) The glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ spectrum displays five primary species, each representing a different glycosylation stoichiometry. The peaks correspond to the addition of 1, 2, 3, 4, and 5 O-GlcNAc units, with each glycosylation event contributing approximately +203 Da, indicating increasing glycosylation states, with five modifications being the most prominent.

For mass spectrometry analysis, purified protein samples were prepared at a concentration of 0.2 mg/mL in 0.1% formic acid. A 1 μ L aliquot of each sample was injected into a Waters Acquity UPLC BEH C8 column, using a gradient of 2–80% solvent B (acetonitrile with 0.1% formic acid) over 15 minutes, followed by a 2-minute hold at 80% solvent B. Separation was conducted at 35°C with a flow rate of 200 μ L/min. Proteins were analyzed using a Bruker Maxis Impact ultra-high resolution quadrupole time-of-flight (QTOF) LC/MS/MS system, with ionization conditions optimized for high-sensitivity detection. Specifically, the spray voltage was set to 3.6 kV, the capillary temperature to 325°C, and the sheath gas flow rate to 25. Mass spectra were collected over an m/z range of 500–5,000 with a resolution of 40,000 m/z. Spectral data were deconvoluted using Bruker Daltonics software to determine the molecular weights of the intact proteins.

Detailed ultraviolet photodissociation (UVPD)-MS analysis confirmed the five O-GlcNAc sites on Ph-P^{1397–1589}, precisely mapping the glycosylation and validating the initial findings. For site mapping, glycosylated Ph-P^{1397–1589} samples were digested with trypsin after reduction and alkylation of cysteine residues. The resulting peptides were analyzed by LC-MS using a Thermo Scientific Lumos Orbitrap mass spectrometer equipped with UVPD capabilities. A Dionex RSLC 3000 nano-LC system was used to separate peptides, employing a C18 column with a flow rate of 300 nL/min. The gradient consisted of 5–35% acetonitrile with 0.1% formic acid over 45 minutes. UVPD fragmentation was performed with 213 nm laser pulses to induce peptide dissociation, enabling the identification of glycosylation sites. Glycopeptides were identified using Byonic software, with manual validation performed to ensure accurate site mapping. Confidence in glycosylation assignments was supported by Delta Mod Scores and PEP 2D scores, as well as secondary confirmation through electron-transfer/higher-energy collision dissociation (EThcD).

	G1	G2					
1	MGSSHHHHHH	SSGLVPRGSH	MGVGSGETNG	LGTGGIVGVD	AMALVDRLDE	50	
51	AMAEEKMQTE	ATPKLSESFP	ILGASTEVPP	MSLPVQAAIS	APSPLAMPLG	100	
101	SPLSVALPTL	APLSVVTSGA	APKSSEVNGT	DRPPISSWSV	DDVSNFIREL	150	
151	PGCQDYVDDF	IQQEIDGQAL	LLLKEKHLVN	AMGMKLGPAL	KIVAKVESIK	200	
201	EVPPPGEAKD	PGAQ				204	148

4.4.4. Isothermal Stability of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ at Extreme Temp

After confirming through MS analysis that Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was indeed glycosylated, I performed a more detailed and stringent analysis of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ by using more harsh conditions. I conducted an isothermal heating experiment at 100°C for a time course ranging from 0 to 60 minutes to evaluate whether prolonged exposure to high temperatures would reveal any stabilizing arising effect from O-GlcNAcylation. (**Figure 4.6**) and used SDS-PAGE used to quantify the remaining soluble protein over time. No significant difference was observed between the glycosylated and non-glycosylated proteins.

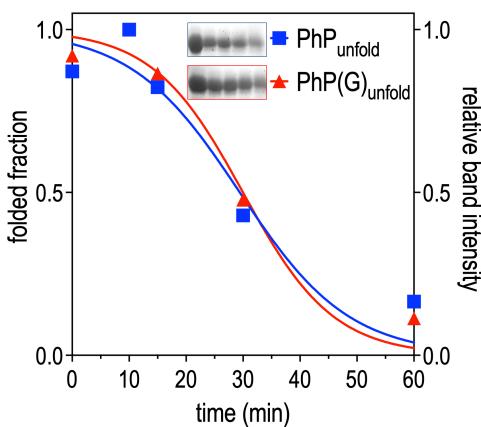


Figure 4.9 Isothermal assessment of Ph-P¹³⁹⁷⁻¹⁵⁸⁹.

Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was treated at 100°C on a time course to assess stability and the effect of O-GlcNAcylation under extreme conditions. Minimal differences in stability, e.g. aggregation, were observed between glycosylated and unglycosylated forms. Protein levels were quantified using SDS-PAGE gel and band intensity analysis.

For the isothermal heating experiment, samples of glycosylated and non-glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ were prepared at a concentration of 1 mg/mL in 20 mM Tris-HCl (pH 7.5) with 200 mM NaCl. Aliquots of 50 µL were placed in PCR tubes, and a thermocycler (Bio-Rad C1000 Touch) was used to maintain a constant temperature of 100°C. Time points were collected at 0, 5, 15, 30, 45, and 60 minutes by rapidly transferring tubes to an ice bath to stop further denaturation. Samples were centrifuged at 20,000 g for 10 minutes to separate soluble protein from aggregated material. The supernatant was collected and mixed with 5X Laemmli sample buffer, then boiled for 5 minutes to prepare for SDS-PAGE analysis.

SDS-PAGE gels were run using 12% acrylamide resolving gels with a 4% stacking gel at 70 V for 1.5 hours. Gels were stained with Coomassie Brilliant Blue R-250 and destained in a methanol-acetic acid solution. Band intensities were quantified using ImageJ software to determine the percentage of soluble protein remaining at each time point. Relative intensity values were normalized to the 0-minute time point for both glycosylated and non-glycosylated forms to allow for direct comparison.

4.4.5. Biophysical Analysis

After thermal denaturation experiments failed to reveal any significant effect of O-GlcNAcylation on Ph-P¹³⁹⁷⁻¹⁵⁸⁹, I turned to alternative biophysical methods to discern any more subtle effects that might have been overlooked. To do so, I employed CD spectroscopy and intrinsic tryptophan fluorescence.

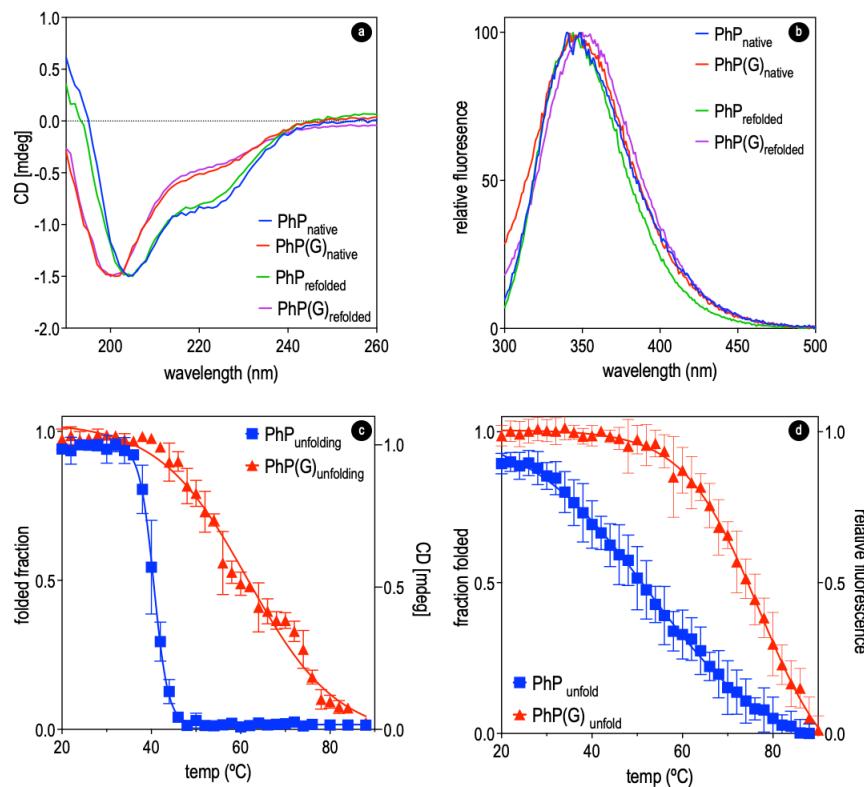


Figure 4.10 Ph-P¹³⁹⁷⁻¹⁵⁸⁹ biophysical characterization.

CD Spectra. (a) CD spectra of Ph-P¹³⁹⁷⁻¹⁵⁸⁹, both glycosylated and non-glycosylated, measured from 190-260 nm. After boiling for 10 minutes and refolding at 4°C, the spectra for the refolded and native proteins align for both forms. (b) Intrinsic tryptophan fluorescence. Fluorescence scan spectra for both modified and non-modified Ph-P¹³⁹⁷⁻¹⁵⁸⁹ under native and refolded conditions show identical Gaussian distributions, with no observable differences. (c) CD melt profile. CD monitoring of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ melting shows a T_m shift of 25°C for the O-GlcNAcylated form, indicating enhanced

thermal stability. (d) Fluorescence melt profile. Fluorescence monitoring of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ melting shows similar stabilization effects for the O-GlcNAcylated form, with similar results to the CD melt profile.

CD spectra were measured using a Chirascan Plus CD spectrometer (Applied Photophysics) with quartz cuvettes of 2 mm path length. Protein samples (0.5 mg/mL) were prepared in 20 mM Tris-HCl (pH 7.5) with 200 mM NaCl. Far-UV scans were recorded from 190–260 nm with a 1 nm step size at 25°C. For thermal denaturation, the samples were heated from 20°C to 95°C at 1°C/min, with changes in α-helical content monitored at 222 nm. Refolding experiments were conducted by cooling samples to 4°C for 30 minutes, and spectra were re-recorded to evaluate structural recovery. Thermal melt data were fitted to sigmoidal curves using GraphPad Prism to calculate T_m values.

CD spectra of Ph-P¹³⁹⁷⁻¹⁵⁸⁹, both glycosylated and non-glycosylated, were recorded to assess secondary structure (**Figure 4.10a**). This was done both in their native states and after subjecting the proteins to 100°C for 10 minutes, followed by refolding at 4°C. The refolded spectra for both forms aligned closely with their native spectra, suggesting refolding of these proteins to their native state. Notably, the CD spectrum of the O-GlcNAcylated differed to the non-glycosylated version, pointing to a potential alteration in its secondary structure. The shift of the peak around 210 nm to 200 nm suggests changes in α-helical content, while the less pronounced signal at 222 nm indicates a possible reduction in β-sheet content. These changes imply that O-GlcNAcylation may subtly influence the structural balance between α-helices and β-sheets in Ph-P¹³⁹⁷⁻¹⁵⁸⁹. When conducting a thermal ramp to monitor changes in α-helical content, measured at 222 nm, an increase in the melting temp (T_m) of these helices approximately 25°C for the O-GlcNAcylated form, indicating that O-GlcNAc stabilized the protein from thermal unfolding (**Figure 4.10c**).

Fluorescence measurements were performed using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies). Protein samples (0.1 mg/mL) were excited at 280 nm, and emission spectra were recorded from 300–500 nm with a 5 nm slit width. Thermal denaturation experiments were conducted by heating samples from 20°C to 95°C at 1°C/min, and fluorescence at 330 nm was monitored to track unfolding. Data points were collected every 0.2°C, and the resulting curves were analyzed using a two-state folding model to determine T_m values.

For the fluorescence analysis, both the glycosylated and non-glycosylated forms of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ were examined using intrinsic tryptophan fluorescence, with spectral scans covering the typical emission range for tryptophan residues (300–500 nm) (**Figure 4.7b**). Unlike the CD results, the fluorescence spectra for the glycosylated and non-glycosylated forms were identical in both their native and refolded states, indicating that both forms remain soluble or collapse into a compact globular form. While the fluorescence spectra *did not show differences in these folded states*, during the temp ramp experiment, *changes were observed as the samples were heated*. These changes monitored the unfolding process, with the glycosylated protein exhibited a T_m shift of approximately 25°C, indicating increased thermal stability conferred by O-GlcNAcylation (**Figure 4.7d**). The changes in the fluorescence spectra during heating were plotted into a fitted sigmoidal curve to analyze the thermal stability.

4.4.6. Equilibrium Chemical Denaturation

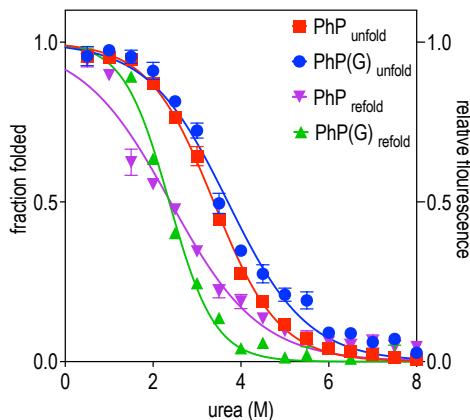


Figure 4.11 Ph-P¹³⁹⁷⁻¹⁵⁸⁹ equilibrium chemical denaturaiton.

Equilibrium urea denaturation curves of glycosylated and non-glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹. Both forms of the protein were subjected to increasing concentrations of urea to assess their chemical stability.

After assessing thermal stability, I moved on to evaluate the chemical stability of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ using equilibrium urea denaturation experiments. The goal was to determine if O-GlcNAcylation impacted the susceptibility of the protein to chemical denaturation. Both glycosylated and non-glycosylated forms were mixed with increasing concentrations of urea. Samples were incubated with urea concentrations ranging from 0 to 8 M for 1 hour at room temperature to allow for equilibrium. Fluorescence was monitored using a BioTek

fluorescence plate reader with excitation at 280 nm and emission at 330 nm to track changes in the protein's tertiary structure. While the overall concept of testing native and refolded proteins remained the same, the results (**Figure 4.9**), as determined by a paired t-test and one-way ANOVA, both yielding p-values higher than 0.5.

4.4.7. Folding Kinetics

We next performed stopped-flow experiments to analyze the folding and unfolding of Ph-P¹³⁹⁷⁻¹⁵⁸⁹. This method was selected for its ability to capture rapid conformational changes in proteins, particularly in folding pathways that involve intermediate states. Stop flow technique is particularly useful in monitoring fast reactions over very short time scales (milliseconds to seconds), allowing precise measurement of folding kinetics. The protein (0.1 mg/mL) was rapidly mixed with urea solutions ranging from 0 to 4 M in a stopped-flow apparatus (Applied Photophysics PMT). The intrinsic tryptophan fluorescence was monitored at an excitation wavelength of 295 nm, with emission recorded using a WG360 filter. Fluorescence data were captured every 0.1 seconds for 50 seconds to track the folding and unfolding dynamics in real-time.

In this experiment, the intrinsic tryptophan fluorescence of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ was used as a reporter for possible structural changes during folding and unfolding, with data collected from 0 to 50 seconds after mixing with varying concentrations of urea. The resulting fluorescence traces were fitted with single-exponential decay models to extract rate constants (k_{obs}), which were plotted against urea concentration to evaluate the folding kinetics under denaturant conditions. This approach enables the detection of transient intermediates that may otherwise be undetectable by traditional methods.

The exponential decay observed in the stopped-flow signal, suggests that the folding and refolding of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ do not follow a simple, single-step mechanism. As shown in **Figure 4.12** (left), the fluorescence traces for both glycosylated and non-glycosylated Ph-P¹³⁹⁷⁻¹⁵⁸⁹ clearly deviate from a linear progression at certain urea concentrations, particularly at higher concentrations such as 4 M urea. These traces exhibit a distinct exponential decay pattern during the 0–50 second interval, further supporting the idea that the folding process involves intermediate states and is not a straightforward single-step mechanism. This characteristic decay at higher urea

concentrations suggests a more complex folding pathway, likely involving multiple transitions between intermediate conformations.

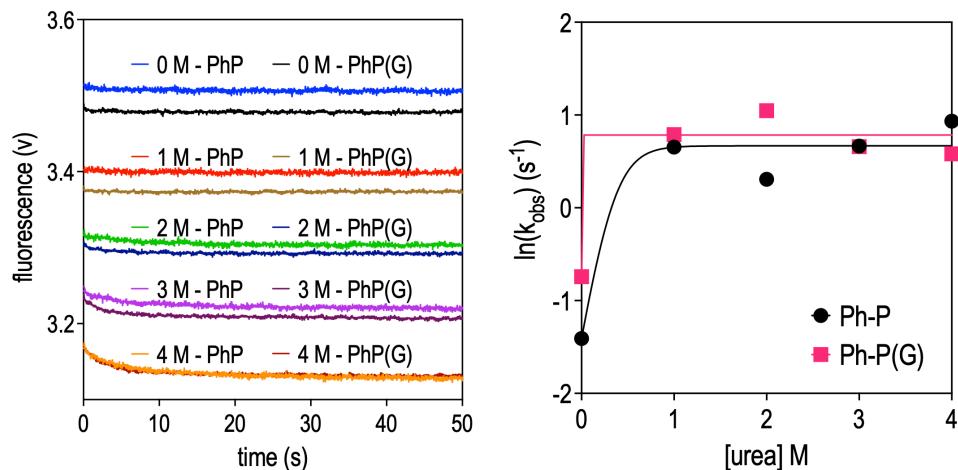


Figure 4.12 Ph-P¹³⁹⁷⁻¹⁵⁸⁹ folding kinetics.

(Left) Fluorescence traces recorded over 0 to 50 seconds for both glycosylated and non-glycosylated forms of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ during urea-induced denaturation, using stop-flow measurements. The fluorescence intensity changes reflect the folding process of each protein form. (Right) The observed rate constants (k_{obs}) were derived from fitting a single-exponential decay model to the fluorescence traces for each condition. The k_{obs} values were plotted as a function of urea concentration, providing insight into the folding kinetics and how glycosylation may affect the folding mechanism under increasing denaturant concentrations.

The unfolding process of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ appears to occur more rapidly between 0 and 1 M urea, with the glycosylated form showing a slightly faster unfolding rate compared to the non-glycosylated form. Fitting and analysis of fluorescence data were performed using GraphPad Prism software, and the resulting k_{obs} values were derived by applying single-exponential decay models. These k_{obs} values were subsequently plotted against the corresponding urea concentrations to visualize trends in folding and unfolding behavior. In **Figure 4.12** (right), the k_{obs} values suggest that glycosylation may influence the speed of the unfolding process, though this observation requires further validation or repetition of the experiment to confirm the consistency of the results. While these initial data suggest a potential effect of O-GlcNAcylation on accelerating the early stages of unfolding, additional experiments will be necessary to accurately interpret these findings.

4.4.8. Computational Analysis of O-GlcNAc Modified Sites

Putting aside the promising results from Ph-P¹³⁹⁷⁻¹⁵⁸⁹, I shifted my focus to identifying potential O-GlcNAc modification sites that could have a significant effect on

thermodynamic stability. To achieve this, I conducted a computational analysis in collaboration with Prof. Stephanie Olivier-Van Stechelen and Dr. Florian Malard. The primary aim of this analysis was to explore whether O-GlcNAcylation at specific sites on proteins might influence their structural stability, potentially impacting their biological function. This analysis first examined whether O-GlcNAc modifications exhibited any preferential targeting of serine or threonine residues. Using structural predictions generated by AlphaFold2, alongside annotations from UniProt and the O-GlcNAc database, we evaluated the solvent accessibility and structural environment of O-GlcNAc-modified residues across a range of proteins.

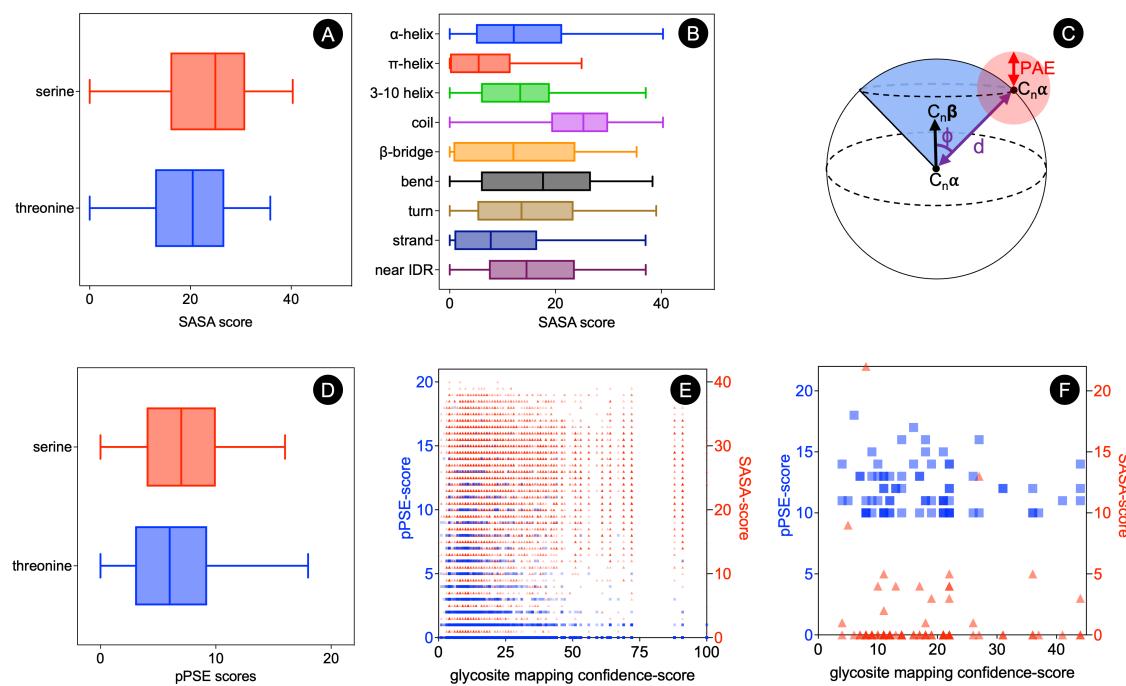


Figure 4.13 Computational analysis of O-GlcNAc modified sites.

Box and whisker plot comparing SASA scores (a) for serine and threonine glycosylated sites, showing no significant difference in solvent accessibility. (b) Analysis of SASA scores across different protein regions, indicating no clear pattern of solvent exposure for structurally relevant O-GlcNAc sites. (c) Diagram explaining the concept of pPSE score, which assesses the spatial landscape of solvent exposure and includes the PAE (Predicted Aligned Error) from AlphaFold.²⁰⁹ (d) Box and whisker plot screening pPSE scores for serine and threonine glycosylated sites, with results like SASA scores. (e) Scatter plot of pPSE scores versus glycosite mapping confidence score for 10,000 sites, identifying potentially buried and relevant O-GlcNAc sites with high modification confidence. (f) Scatter plot focusing on sites with pPSE scores of 10 or higher, highlighting significantly buried sites.

Two key metrics were employed: SASA to measure solvent exposure, and the prediction-aware part-sphere exposure (pPSE) score (**Figure 4.9c**) to assess how buried or exposed these glycosylation sites were within the protein structure. Initial results showed no significant difference in solvent accessibility between glycosylated serine and threonine residues, suggesting that O-GlcNAcylation is not dependent on whether the site is in an exposed or buried environment. (**Figure 4.9a** and **Figure 4.9d**) The SASA analysis further indicated that O-GlcNAc modifications can occur across a variety of protein environments, including α -helices, β -sheets, and coil regions, with no clear preference for specific secondary structures. (**Figure 4.9b**)

The pPSE analysis offered deeper insights into the structural context of glycosylation sites. Several O-GlcNAc sites, while buried within the protein structure, were still confidently identified as modified, suggesting that O-GlcNAcylation may not be strictly dependent on solvent exposure. These findings raised the possibility that dynamic changes in protein conformation or transient structural states could enable glycosylation in regions not traditionally accessible to solvent. The results shown in **Figure 4.9 (e)** and **(f)** further elucidate the complexity of O-GlcNAc modifications in relation to solvent exposure. The scatter plot in **Figure 4.9 (e)** compares pPSE scores with glycosite mapping confidence scores for approximately 10,000 glycosylation sites. This analysis identified several O-GlcNAc-modified sites that, despite being structurally buried, exhibited high modification confidence. Interestingly, no clear correlation was observed between pPSE scores and SASA values, indicating that solvent exposure is not a defining factor for glycosylation confidence.

This lack of correlation was anticipated, as pPSE and SASA measure different aspects of protein structure. These results suggest that O-GlcNAc modifications can occur in structurally complex environments, further supporting the idea that factors beyond solvent accessibility, such as protein dynamics, play a critical role. In **Figure 4.9 (f)**, a closer examination of glycosylation sites with pPSE scores of 10 or higher highlights several deeply buried but confidently modified sites. These findings underscore that O-GlcNAc modifications are not limited to solvent-exposed regions and can occur even in areas deeply embedded within the protein structure. This observation opens exciting possibilities for further investigation into the role of such buried glycosylation sites.

Overall, the computational analysis revealed that O-GlcNAc modifications are not limited by solvent exposure or structural motifs, indicating that these modifications could

play a more nuanced role in protein stability than previously thought. These results lay the groundwork for future biophysical studies to experimentally validate the potential influence of O-GlcNAcylation on protein stability.

4.4.9. MD Simulations

Our computational analysis examined the effects of O-GlcNAc modifications on several human enzymes involved in cellular metabolism. These analyses provided insights into how glycosylation might influence enzyme structure and function, with varying degrees of impact across different enzymes. **Figure 4.10** presents the MD model of human catalase, a key enzyme in hydrogen peroxide detoxification. The unmodified catalase (**Figure 4.10a**) shows Serine 114 near the heme group as a potential O-GlcNAc modification site. Upon modification, significant conformational changes were observed (**Figure 4.10c**), leading to a strained configuration of the heme group, which protrudes from the protein structure (**Figure 4.10d**). These changes suggest a destabilizing effect of the modification, which could impair the enzyme's ability to efficiently reduce hydrogen peroxide.

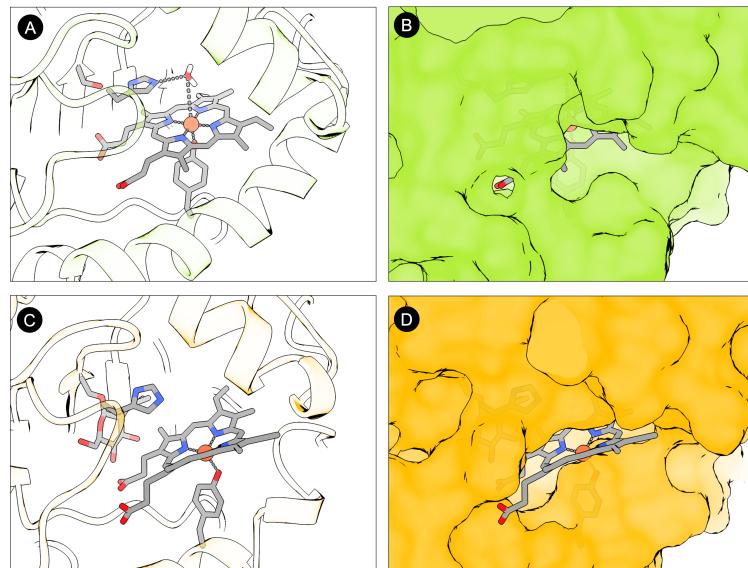


Figure 4.14 MD model of CAT with and without O-GlcNAc modification.

(a) Crystal structure of the unmodified human catalase enzyme (PDB ID: 1DGF), with the heme prosthetic group in focus. Serine 114, which is subject to modification, can be seen in the background. A water molecule is coordinated above the heme, potentially indicating the position typically occupied by the HOOH molecule. (b) Space-filling model of the unmodified catalase, showing the heme fitting into a constricted active site pocket with a minimal opening, suggesting limited accessibility for the HOOH molecule. (c) MD model of the catalase enzyme with O-GlcNAc

modification at Serine 114. The modification induces a significant conformational change, displacing the heme group into a strained configuration. (d) Space-filling model of the modified catalase, illustrating the heme protruding from the protein structure, indicative of the destabilizing effect of the O-GlcNAc modification.²¹²

In contrast, O-GlcNAc modification on LDHA at Serine 137, shown in **Figure 4.11**, caused only minor structural displacements of NAD⁺ and pyruvate (**Figure 4.11c**). The overall structure remained largely unchanged (**Figure 4.11d**), suggesting that O-GlcNAcylation may have minimal immediate effects on LDHA function, although slight shifts in the active site could still impact enzyme kinetics under specific conditions.

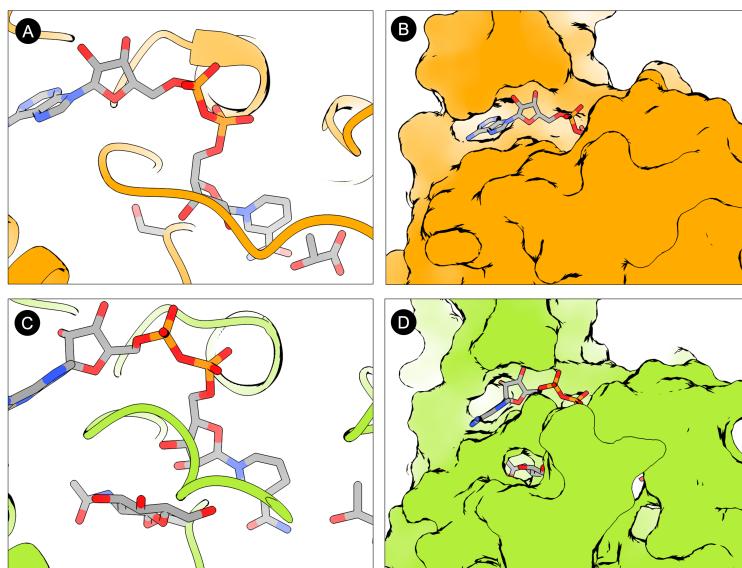


Figure 4.15 MD model of human LDHA enzyme with and without O-GlcNAc.
 (a) Crystal structure of the unmodified human LDHA enzyme (PDB ID: 4JNK), with NAD⁺ and pyruvate in focus. S137, which is subject to modification, is highlighted in the background. (b) Space-filling model of the unmodified LDHA structure, showing the tight interaction between NAD⁺ and pyruvate within the active site. (c) MD model of the LDHA enzyme with O-GlcNAc modification at S137. The modification causes slight displacements of NAD⁺ and pyruvate, resulting in minor structural changes. (d) Space-filling model of the modified LDHA, illustrating the subtle shifts in NAD⁺ and pyruvate positions due to the O-GlcNAc modification, with minimal overall structural perturbation.²¹³

Finally, the modification of G6PD at Serine 84 (**Figure 4.12**) resulted in a subtle displacement of NADP⁺ (**Figure 4.12b**). While this modification did not cause major structural changes, its proximity to the cofactor binding site suggests that even minor perturbations could influence the enzyme's activity.

These results demonstrate that O-GlcNAc modifications can have varying effects on enzyme structure, ranging from significant conformational changes to minimal shifts, depending on the enzyme and the location of the modification site.

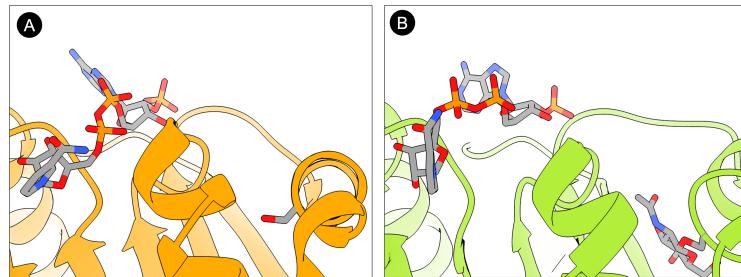


Figure 4.16 MD model of human G6PD enzyme with and without O-GlcNAc.

Crystal structure of the unmodified human G6PD enzyme, (a) highlighting the position of NADP+. S84, which is near NADP+ and subject to modification, is shown in the background. (b) MD model of the G6PD enzyme with O-GlcNAc modification at S84. The modification causes a very slight displacement of NADP+, indicating a potential relevance due to the proximity of the modification site.²¹⁴

Coming back to Ph-P¹³⁹⁷⁻¹⁵⁸⁹, I analyzed the results from the MD simulations I had requested to explore the impact of O-GlcNAc modifications at the three glycosylation sites (S61, S69, and S114).

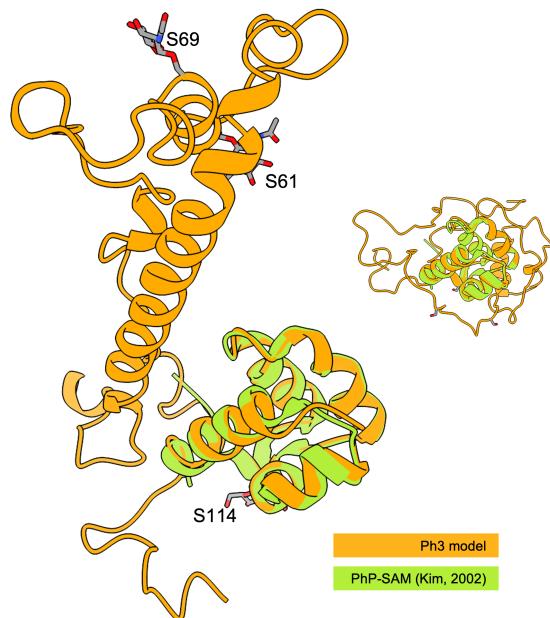


Figure 4.17 Ph-P¹³⁹⁷⁻¹⁵⁸⁹ MD model.

MD model of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ modified with all three glycosites O-GlcNAc modified (S61, S69, and S114) shown in orange, indicating the formation of a coil structure post-modification. The smaller model to the left shows the unmodified Ph-P¹³⁹⁷⁻¹⁵⁸⁹, which remains disordered. The Ph-P-SAM domain X-ray solved structure is depicted in lime.²¹¹

As shown in Figure 4.13, the MD model reveals that O-GlcNAcylation leads to the formation of a coil structure (highlighted in orange), in contrast to the unmodified Ph-P1397-1589 model, which remains disordered. This structural change indicates that O-GlcNAcylation likely promotes a more ordered conformation, suggesting a stabilizing effect on the protein. These results highlight that glycosylation has a direct influence on the secondary structure of Ph-P1397-1589, shifting it from a disordered state to a more stable, coil-like structure, which could have implications for its overall stability and function.

4.5. Discussion

This chapter delved into the biophysical effects of O-GlcNAcylation, focusing on Ph-P as a key model system. My results, consistent across several analytical methods, reveal that O-GlcNAcylation confers notable stability to Ph-P¹³⁹⁷⁻¹⁵⁸⁹, particularly in thermal conditions. The observed increase in the T_m of the glycosylated form, evident from circular CD and fluorescence thermal denaturation experiments, underlines the protective effect of this PTM. This aligns with previous studies that have suggested O-GlcNAcylation as a mechanism for enhancing protein resilience against denaturation and aggregation under stress.^{36,48,52,215} One possible mechanism for O-GlcNAcylation-mediated stabilization lies in the direct effects of the first GlcNAc residue on protein folding energetics. Studies on N-glycosylation, such as those by Culyba et al. (2011),¹⁹³ have shown that the initial GlcNAc attachment alone can enhance stability by shifting the folding equilibrium toward the native state. This effect is partly due to favorable glycan-protein interactions that reduce the entropy of the unfolded state, thereby stabilizing the folded conformation. While N-glycosylation occurs at asparagine residues within a sequon, the fundamental principle may extend to O-GlcNAcylation, where the single GlcNAc attached to serine or threonine could serve a similar function. The first GlcNAc may act as a steric constraint that limits excessive backbone flexibility, reduces transient unfolding, and decreases solvent exposure of aggregation-prone regions. Additionally, it could introduce local structural order, particularly in intrinsically disordered domains like those observed in Ph-P, stabilizing them against misfolding and aggregation. This aligns with our experimental findings, where O-GlcNAcylation increases Ph-P's thermal stability and decreases its aggregation propensity. Although O-GlcNAc lacks the extended glycan chain present in N-glycans, its localized effects on protein folding landscapes may still contribute significantly to structural stabilization. However, the stabilizing effects of O-GlcNAcylation

were not observed in lysate thermal denaturation-like studies. This may be due to the rapid folding and unfolding events in this protein, which might occur too quickly for these methods to detect significant differences. It is possible that the transient nature of these processes masks the impact of O-GlcNAcylation, suggesting that more sensitive or time-resolved techniques may be needed to fully capture the stabilizing effects under these conditions.

Further, the stabilizing effect of O-GlcNAcylation was more pronounced during thermal denaturation compared to chemical denaturation with urea. This discrepancy may arise from the inherent differences in protein folding and unfolding pathways under thermal versus chemical stress. Thermal denaturation often involves distinct intermediate states that may be more susceptible to stabilization by O-GlcNAcylation, while urea-induced denaturation follows a different pathway, potentially bypassing the exposure of these intermediates. Supporting this, previous studies have shown that chemically denatured proteins are generally more expanded than their thermally denatured counterparts, due to different backbone-solvent interactions.^{207,216} This variation in unfolding intermediates could explain why O-GlcNAcylation enhances thermal stability more effectively than chemical stability. These observations highlight the complexity of protein stability mechanisms, suggesting that O-GlcNAcylation interacts with specific dynamic processes that are more relevant to thermal stress.

However, it's important to exercise caution when interpreting the stabilizing effects of O-GlcNAcylation on Ph-P,¹³⁹⁷⁻¹⁵⁸⁹ given that its SAM domain can polymerize.²¹¹ This polymerization adds another layer of complexity to the stability mechanisms, suggesting that the dynamics may be far more intricate than they first appear. More research is needed to fully understand how these factors interact, and any conclusions drawn should be approached carefully.

Computational analysis. The computational analysis extended our exploration to other proteins, providing broader insights into how O-GlcNAcylation might affect thermodynamic stability across diverse proteins, including human enzymes involved in metabolic processes. One striking aspect of the analysis was the identification of O-GlcNAcylation sites in buried regions of proteins. These results challenge the traditional view that glycosylation occurs predominantly on solvent-exposed regions. The finding that O-GlcNAc modifications are present even at buried sites opens new questions about the

mechanisms that allow these modifications to occur. It suggests that protein dynamics or transient exposure during folding may facilitate glycosylation at these locations.

One possible explanation for how these O-GlcNAc sites in structurally buried regions could be modified by OGT is that O-GlcNAcylation may occur co-translationally. This process, which has been observed previously^{98,217} allows OGT to access certain sites before the protein fully folds, providing a pathway for modification even in regions that later become buried. However, the overlap between high pPSE sites and known co-translational datasets was limited, as shown in the Venn diagram (**Figure 4.14**), indicating that additional, yet-undiscovered mechanisms may also contribute to this phenomenon. More research is needed to fully explore these potential pathways.

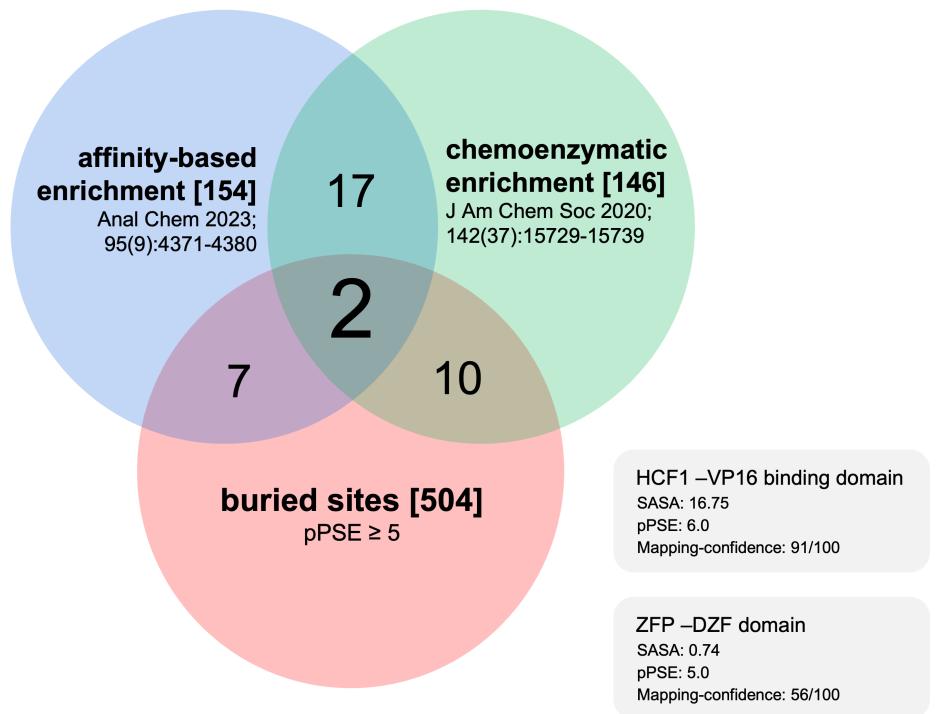


Figure 4.18 pPSE score \geq 5 overlap with co-translational O-GlcNAc sites.
Venn diagram showing the overlap of 504 sites with a pPSE score equal to or greater than 5, compared with datasets from two studies on co-translationally modified O-GlcNAc sites.^{98,217} Only two sites, corresponding to HCF1-VP16 binding protein and ZFP-DZF domain, are shared between the datasets.

Metabolic Insight from the Computational Analysis. In addition to the structural insights, our computational analysis revealed several metabolic enzymes—such as CAT, LDHA, and G6PD—that showed potential O-GlcNAc modification sites with implications for their stability and function. The MD simulations for these enzymes suggest

that glycosylation could influence their catalytic efficiency by altering the structural dynamics of their active sites. For instance, the modification of catalase at Serine 114 resulted in significant conformational changes, likely affecting its capacity to detoxify hydrogen peroxide, while the modifications in LDHA and G6PD showed more subtle, but still potentially impactful, shifts in cofactor binding or active site conformation.

These findings suggest a potential feedback loop where O-GlcNAcylation modulates key metabolic pathways by stabilizing enzymes under certain conditions. This regulatory mechanism may be particularly relevant in cellular stress responses or the fine-tuning of metabolic flux, where O-GlcNAcylation adjusts enzyme activity in response to environmental or metabolic cues. **Figure 4.15** illustrates this concept, showing key metabolic pathways such as glycolysis, the tricarboxylic acid (TCA) cycle, and redox reactions where enzymes like G6PD, LDHA, and CAT are highlighted. These enzymes, identified through computational analysis, are proposed to undergo thermodynamically stabilizing O-GlcNAc modifications, underscoring the interconnected role of O-GlcNAcylation in maintaining metabolic balance and cellular resilience to oxidative stress.

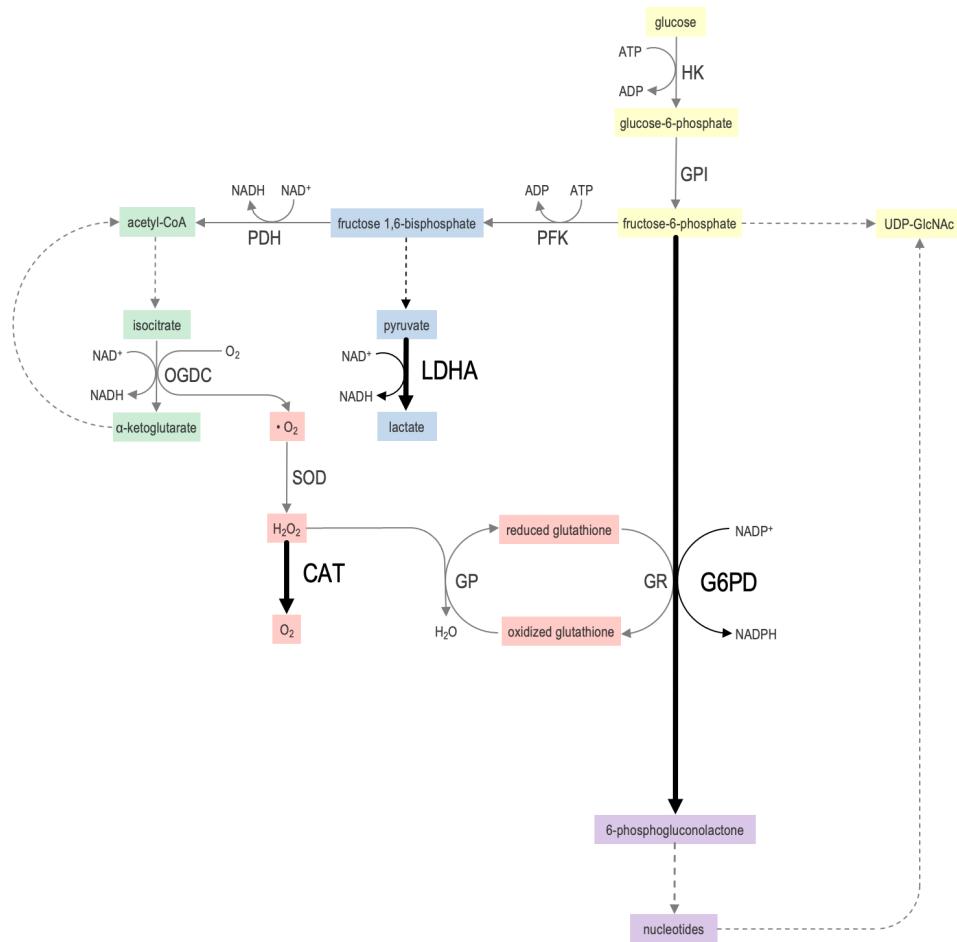


Figure 4.19 Metabolic insight from computational analysis of O-GlcNAc sites.

Metabolic pathways with computational analysis hits potentially relevant to O-GlcNAc metabolism. Highlighted enzymes, including G6PD, LDHA, and CAT, are identified as sites with potentially thermodynamically stabilizing O-GlcNAc modifications. The pathways are color-coded to indicate their functions—blue for glycolysis, green for the tricarboxylic acid (TCA) cycle, red for redox reactions including glutathione metabolism, and purple for the pentose phosphate pathway. This schematic underscores the potential feedback loops and the role of O-GlcNAcylation in modulating metabolic flux and enhancing cellular responses to oxidative stress, emphasizing the interconnected nature of these pathways.

Overall, this chapter provides strong evidence that O-GlcNAcylation plays a stabilizing role in protein structure, with effects observed across both experimental and computational approaches. The results demonstrate that O-GlcNAcylation significantly enhances the thermal stability of Ph-P¹³⁹⁷⁻¹⁵⁸⁹, particularly under physiological stress conditions, supporting its role as a protective post-translational modification. Notably, O-GlcNAc modifications were identified not only in solvent-exposed regions but also in structurally buried sites, suggesting that this modification may occur during transient folding states or even co-translationally. This raises intriguing questions about how OGT

accesses these buried regions, highlighting a potential complexity in the regulatory mechanisms underlying O-GlcNAcylation. Furthermore, computational analysis expanded these insights to include key metabolic enzymes, such as G6PD, LDHA, and CAT, where O-GlcNAcylation appears to influence their stability and catalytic activity, as illustrated in **Figure 4.15**. These findings suggest a broader role for O-GlcNAcylation in modulating enzyme function within metabolic pathways, potentially serving as a feedback mechanism under cellular stress conditions. Altogether, this chapter underscores the intricate dynamics of O-GlcNAcylation and its far-reaching implications for protein stability and cellular regulation, setting the stage for future research to further unravel these complex mechanisms.

4.6. Acknowledgments

I owe the deepest thanks to Prof. David J. Vocadlo, my supervisor, for his unwavering support, guidance, and for continually reminding me to trust the process when things felt most challenging. This project, one of the hardest yet most rewarding I've ever undertaken, couldn't have been possible without his patience and insightful feedback.

Special thanks to Prof. Dustin T. King, whose pioneering work on the Ph-P constructs provided a foundation for much of what I've accomplished here. Your collaborative spirit and enthusiasm for the research have been a true source of inspiration.

To Prof. Cornelia Bohne and the VicPick labs team at the University of Victoria—thank you for your invaluable expertise and assistance with the stopped-flow experiments, which were crucial to the progress of this work.

I'd also like to give a huge shout-out to Dr. Florian Malard and Prof. Stephanie Olivier-Van Stechelen for their brilliant computational analysis contributions. You both brought a unique perspective that elevated the work, and your guidance along the way was indispensable.

Another huge thank you goes to Dr. Edwin Escobar and Prof. Jennifer Brodbelt for their incredible expertise and collaboration in mapping O-GlcNAc sites. Their mastery of mass spectrometry brought a whole new level of clarity to this project, and without their contributions, the detailed glycosylation insights would have been impossible to achieve.

The precision and care they brought to this work made a massive impact, and I'm incredibly grateful for their part in helping bring this research to life.

And, of course, I can't forget to thank all the people who made this journey a little more fun—whether that was sharing in the ups and downs, offering a word of encouragement, or just helping me get through the days. I'm forever grateful for the support I've received from my friends, family, and colleagues.

Chapter 5.

CONCLUSION

5.1. 5.1 Summary of Key Findings

5.1.1. Development of a Practical scFv for O-GlcNAc Detection

The research successfully engineered a single-chain variable fragment (scFv) derived from the RL2 antibody to detect O-GlcNAc modifications. This scFv demonstrated specificity and binding affinity comparable to the commercial RL2 antibody. Functional assays confirmed its ability to distinguish between O-GlcNAc-modified and unmodified proteins, validating its potential as a reliable and cost-effective tool for detecting O-GlcNAc modifications.

5.1.2. Chemoenzymatic Enrichment Method for O-GlcNAc-Modified Proteins

A reversible chemoenzymatic method was developed to enrich O-GlcNAc-modified proteins while preserving their native state. By using β -1,4-galactosyl-transferase (β 4GALT1) to attach a biotinylated galactose derivative to O-GlcNAc residues and identifying glycoside hydrolases (GHA09 and GHA10) capable of selectively removing the attached galactose, the method overcomes limitations of previous techniques. This approach provides an efficient and specific strategy for the enrichment and analysis of O-GlcNAc-modified proteins.

5.1.3. Biophysical Effects of O-GlcNAcylation on Protein Stability

Using the Polyhomeotic-proximal chromatin protein (Ph-P) as a model, the study demonstrated that O-GlcNAcylation enhances thermal stability and reduces aggregation propensity. Circular dichroism spectroscopy and intrinsic tryptophan fluorescence assays showed that the glycosylated form of Ph-P exhibited increased resistance to thermal denaturation compared to its unmodified counterpart. Computational analyses supported

these findings, suggesting that O-GlcNAcylation influences the protein's energy landscape and folding kinetics.

5.2. Implications of the Research

The findings have significant implications for the field of O-GlcNAc research. The engineered scFv offers a practical alternative for detecting O-GlcNAc modifications, potentially simplifying and improving detection methods in various biological contexts. The chemoenzymatic enrichment method enhances the ability to study O-GlcNAc-modified proteins without altering their native state, facilitating downstream analyses such as mass spectrometry. The biophysical characterization of Ph-P contributes to understanding how O-GlcNAcylation modulates protein stability, which is crucial for elucidating its regulatory roles in cellular processes.

5.3. Limitations of the Study

While the research presents significant advancements, certain limitations exist. The scFv faced challenges with solubility and proper folding in bacterial expression systems, requiring denaturing purification and refolding that may limit its practicality. The glycoside hydrolases used in the chemoenzymatic method showed varying catalytic efficiencies, and substrate availability was sometimes limited, necessitating further optimization. The biophysical studies were conducted under in vitro conditions, and the translation of these findings to the complex in vivo environment requires additional investigation.

5.4. Future Directions

While this research provides significant advancements in understanding O-GlcNAcylation's biochemical properties and its effects on protein stability, several critical areas remain unexplored or incompletely understood. The following directions outline potential pathways for extending this work into fundamental biology, translational research, and clinical applications.

5.4.1. Enhancing the utility and practicality of the scFv for O-GlcNAc Detection

- **Structural characterization to understand the binding mechanism:** Determining how the scFv interacts with O-GlcNAc at a molecular level will provide insights for rational engineering. Techniques such as X-ray crystallography or cryo-electron microscopy (cryo-EM) can be used to obtain high-resolution structural data. Additionally, molecular dynamics simulations and computational docking studies can help model interactions and identify key residues involved in binding. These approaches will guide targeted mutations to improve binding affinity and specificity.
- **Improving affinity and specificity:** While the scFv demonstrates effective O-GlcNAc recognition, enhancing its binding properties can make it more robust for experimental applications. This can be achieved through directed evolution or phage display screening to select variants with stronger or more selective binding to O-GlcNAc-modified proteins. Site-directed mutagenesis based on structural insights can also be employed to fine-tune interactions and eliminate potential cross-reactivity with non-O-GlcNAc structures.
- **Addressing solubility and expression challenges:** The scFv exhibited solubility issues in bacterial expression systems, requiring denaturing purification and refolding. To overcome this, alternative expression systems, such as mammalian or insect cell expression, should be explored to improve proper folding. Additionally, engineering the scFv for enhanced solubility, such as by introducing hydrophilic mutations or fusing it to solubility-enhancing tags (e.g., GST or SUMO), may reduce aggregation and improve yield.
- **Enhancing stability and scalability:** For widespread application, the scFv must maintain its functional integrity over extended periods and under various conditions. Stability can be improved by introducing disulfide bonds or other stabilizing mutations. Lyophilization and formulation studies can help develop long-term storage conditions without loss of activity. Scaling up production for routine laboratory and potential diagnostic use will also require optimization of purification workflows and cost-effective expression strategies.

By addressing these aspects, the scFv can be developed into a more practical and reliable tool for detecting O-GlcNAc modifications, expanding its use in biochemical research and potential clinical applications.

5.4.2. Developing the chemoenzymatic enrichment method for O-GlcNAc-modified Proteins

- **Improving the catalytic efficiency and substrate specificity of glycoside hydrolases (GHA09 and GHA10):** The current hydrolases exhibit varying degrees of efficiency in removing the biotinylated galactose derivative, which can impact the overall effectiveness of the enrichment method. Engineering more efficient variants through site-directed mutagenesis or directed evolution could enhance their activity and selectivity. Structural characterization of these enzymes through X-ray crystallography or computational modeling can provide insights into active site dynamics, guiding rational design for improved function.
- **Expanding enzyme specificity to accommodate diverse O-GlcNAcylated proteins:** Different proteins may present O-GlcNAc modifications in structurally unique environments, potentially affecting enzymatic accessibility. Engineering or screening for hydrolases with broader specificity while retaining high selectivity for modified residues can increase the method's applicability across different protein types. Additionally, optimizing reaction conditions such as buffer composition, temperature, and pH can maximize enzymatic efficiency.
- **Scaling up the method for high-throughput applications:** To facilitate large-scale studies of O-GlcNAc-modified proteins, the chemoenzymatic method must be adapted for high-throughput workflows. This can involve miniaturizing the protocol for automated liquid-handling systems, developing 96-well plate-based enrichment assays, or integrating it with mass spectrometry-based proteomics pipelines. Establishing streamlined protocols for large-scale enrichment will allow for more comprehensive proteomic analyses.
- **Ensuring full reversibility and minimal sample loss:** While the method is designed to be reversible, further validation is needed to confirm that glycoside hydrolase treatment fully restores proteins to their native state without unintended modifications or degradation. Testing different enzyme concentrations and incubation times will help optimize conditions for maximal recovery. Additionally,

refining purification steps to minimize sample loss—such as using optimized affinity resins or alternative elution strategies—can enhance overall yield and reproducibility.

By refining these aspects, the chemoenzymatic enrichment method can become a more efficient and widely applicable tool for studying O-GlcNAcylation, ultimately improving our ability to analyze and characterize O-GlcNAc-modified proteins in complex biological systems.

5.4.3. Expanding Biophysical Studies of O-GlcNAcylation on Protein Stability

- **Conducting *in vivo* studies to validate the physiological relevance of O-GlcNAcylation's stabilizing effects:** While the biophysical experiments demonstrated enhanced thermal stability of Ph-P upon O-GlcNAcylation *in vitro*, it is critical to determine whether similar effects occur in a cellular context. Future studies could involve expressing wild-type and O-GlcNAc-deficient Ph-P mutants in model organisms such as *Drosophila melanogaster* or mammalian cell lines, followed by assessments of protein stability, aggregation propensity, and functional integrity under physiological conditions. Additionally, using cellular thermal shift assays (CETSA) or proteomic stability profiling (such as pulse-chase experiments) could provide quantitative data on the half-life and degradation kinetics of O-GlcNAcylated versus unmodified proteins *in vivo*.
- **Investigating the effects of O-GlcNAcylation on a broader range of proteins:** Given the observed stabilization of Ph-P, it is essential to determine whether this is a generalizable effect across different O-GlcNAc-modified proteins or specific to Ph-P. A systematic approach could involve selecting structurally diverse proteins that are known to undergo O-GlcNAcylation and performing similar biophysical analyses. These proteins should include intrinsically disordered proteins (IDPs), structured globular proteins, and proteins prone to aggregation, such as those implicated in neurodegenerative diseases (e.g., tau, α -synuclein, huntingtin). Comparing their thermal stability, folding kinetics, and aggregation behavior with and without O-GlcNAcylation will help establish the broader impact of this modification on protein stability.

- **Elucidating the molecular mechanisms by which O-GlcNAcylation stabilizes proteins:** Although O-GlcNAcylation appears to enhance protein stability, the exact mechanisms remain unclear. Future studies should focus on determining whether the stabilization arises from direct hydrogen bonding between the GlcNAc moiety and the protein backbone, increased hydrophilicity reducing aggregation, or alterations in the protein's energy landscape. Nuclear magnetic resonance (NMR) spectroscopy, hydrogen-deuterium exchange mass spectrometry (HDX-MS), and molecular dynamics simulations could provide detailed insights into the structural and energetic changes induced by O-GlcNAcylation.
- **Exploring how O-GlcNAcylation influences protein-protein interactions:** Since many proteins function as part of larger complexes, it is important to investigate whether O-GlcNAcylation modulates protein-protein interactions. Future experiments could use techniques such as surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), and co-immunoprecipitation (Co-IP) to assess how O-GlcNAcylation affects binding affinity and complex stability. These studies could help determine whether O-GlcNAcylation serves as a regulatory switch for protein-protein interactions, potentially impacting cellular signaling networks and epigenetic regulation.

By expanding these biophysical investigations, future research can provide a more comprehensive understanding of the stabilizing effects of O-GlcNAcylation, its molecular underpinnings, and its implications for protein function in both normal physiology and disease contexts.

5.5. 5.5 Concluding Remarks

This thesis presents innovative approaches to overcome existing challenges in studying O-GlcNAcylation. The development of a practical scFv enhances the toolkit for detecting O-GlcNAc modifications, potentially improving accuracy and efficiency. The chemoenzymatic enrichment method provides a reversible strategy that preserves protein native states, offering advantages over irreversible methods. The biophysical characterization of Ph-P reveals that O-GlcNAcylation can enhance protein stability, contributing to understanding the molecular mechanisms of this modification.

Collectively, these contributions advance the understanding of O-GlcNAcylation and open new avenues for research. By providing novel tools and methodologies, the work lays the foundation for future studies aimed at elucidating the precise roles of O-GlcNAcylation in protein function, stability, and disease processes. Continued exploration of O-GlcNAcylation's roles will enhance fundamental understanding of cellular regulation and may lead to targeted interventions for diseases where O-GlcNAc plays a critical role.

REFERENCES

- 1 Torres, C. R. & Hart, G. W. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem* **259**, 3308-3317 (1984).
- 2 Hart, G. W. Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu Rev Biochem* **66**, 315-335 (1997).
<https://doi.org/10.1146/annurev.biochem.66.1.315>
- 3 Hardivillé, S. & Hart, Gerald W. Nutrient Regulation of Signaling, Transcription, and Cell Physiology by O-GlcNAcylation. *Cell Metabolism* **20** (2014).
<https://doi.org/10.1016/j.cmet.2014.07.014>
- 4 Bond, M. R. & Hanover, J. A. A little sugar goes a long way: The cell biology of O-GlcNAc. *Journal of Cell Biology* **208** (2015).
<https://doi.org/10.1083/jcb.201501101>
- 5 Yang, X., Qian, K., Yang, X. & Qian, K. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nature Reviews Molecular Cell Biology* **2017 18:7 18** (2017). <https://doi.org/10.1038/nrm.2017.22>
- 6 Zhu, Y. et al. O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. *Nature Chemical Biology* **2015 11:5 11** (2015).
<https://doi.org/10.1038/nchembio.1774>
- 7 Xu, S., Yin, K. & Wu, R. Combining Selective Enrichment and a Boosting Approach to Globally and Site-Specifically Characterize Protein Co-translational O-GlcNAcylation. *Anal Chem* **95**, 4371-4380 (2023).
<https://doi.org/10.1021/acs.analchem.2c04779>
- 8 Sinclair, D. A. R. et al. Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). *Proceedings of the National Academy of Sciences* **106** (2009).
<https://doi.org/10.1073/pnas.0904638106>
- 9 Gambetta, M. C., Oktaba, K. & Müller, J. Essential Role of the Glycosyltransferase Sxc/Ogt in Polycomb Repression. *Science* **325** (2009).
<https://doi.org/10.1126/science.1169727>

- 10 Liu, T.-W. *et al.* Genome-wide chemical mapping of O-GlcNAcylated proteins in *Drosophila melanogaster*. *Nature Chemical Biology* 2016 13:2 **13** (2016). <https://doi.org/10.1038/nchembio.2247>
- 11 Groves, J. A., Lee, A., Yildirir, G. & Zachara, N. E. Dynamic O-GlcNAcylation and its roles in the cellular stress response and homeostasis. *Cell Stress Chaperones* **18**, 535-558 (2013). <https://doi.org/10.1007/s12192-013-0426-y>
- 12 Kazemi, Z., Chang, H., Haserodt, S., McKen, C. & Zachara, N. E. O-Linked β-N-acetylglucosamine (O-GlcNAc) Regulates Stress-induced Heat Shock Protein Expression in a GSK-3β-dependent Manner. *Journal of Biological Chemistry* **285** (2010). <https://doi.org/10.1074/jbc.M110.131102>
- 13 Martinez, M. R., Dias, T. B., Natov, P. S. & Zachara, N. E. Stress-induced O-GlcNAcylation: an adaptive process of injured cells. *Biochemical Society Transactions* **45** (2017). <https://doi.org/10.1042/BST20160153>
- 14 Chou, T. Y., Hart, G. W. & Dang, C. V. c-Myc Is Glycosylated at Threonine 58, a Known Phosphorylation Site and a Mutational Hot Spot in Lymphomas. *Journal of Biological Chemistry* **270** (1995). <https://doi.org/10.1074/jbc.270.32.18961>
- 15 Pravata, V. M. *et al.* Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability. *Proceedings of the National Academy of Sciences* **116** (2019). <https://doi.org/10.1073/pnas.1900065116>
- 16 Yki-Järvinen, H. *et al.* UDP-N-acetylglucosamine transferase and glutamine: fructose 6-phosphate amidotransferase activities in insulin-sensitive tissues. *Diabetologia* 1997 40:1 **40** (1997). <https://doi.org/10.1007/s001250050645>
- 17 Griffith, L. S., Mathes, M. & Schmitz, B. β-Amyloid precursor protein is modified with O-linked N-acetylglucosamine. *Journal of Neuroscience Research* **41** (1995). <https://doi.org/10.1002/jnr.490410214>
- 18 Cheng, Z. *et al.* Discovery of 4-(Arylethynyl)piperidine Derivatives as Potent Nonsaccharide O-GlcNAcase Inhibitors for the Treatment of Alzheimer's Disease. *Journal of Medicinal Chemistry* (2024). <https://doi.org/10.1021/acs.jmedchem.4c01132>
- 19 Gamage, K. *et al.* O-GlcNAc Modification of α-Synuclein Can Alter Monomer Dynamics to Control Aggregation Kinetics. *ACS Chemical Neuroscience* (2024). <https://doi.org/10.1021/acschemneuro.4c00301>

- 20 Bork, K. *et al.* N-Propanoylmannosamine interferes with O-GlcNAc modification of the tyrosine 3-monooxygenase and stimulates dopamine secretion. *Journal of Neuroscience Research* **86** (2008). <https://doi.org/10.1002/jnr.21526>
- 21 Lüdemann, N. *et al.* O-Glycosylation of the Tail Domain of Neurofilament Protein M in Human Neurons and in Spinal Cord Tissue of a Rat Model of Amyotrophic Lateral Sclerosis (ALS). *Journal of Biological Chemistry* **280** (2005). <https://doi.org/10.1074/jbc.M504395200>
- 22 Grima, J. *et al.* Mutant Huntingtin Disrupts the Nuclear Pore Complex. *Neuron* **94** (2017). <https://doi.org/10.1016/j.neuron.2017.03.023>
- 23 Haltiwanger, R. & Philipsberg, G. Mitotic Arrest with Nocodazole Induces Selective Changes in the Level of O-Linked N-Acetylglucosamine and Accumulation of Incompletely Processed N-Glycans on Proteins from HT29 Cells. *Journal of Biological Chemistry* **272** (1997). <https://doi.org/10.1074/jbc.272.13.8752>
- 24 Wang, P. & Hanover, J. A. Nutrient-driven O-GlcNAc cycling influences autophagic flux and neurodegenerative proteotoxicity. *Autophagy* **9** (2013). <https://doi.org/10.4161/auto.23459>
- 25 Zhu, Y. *et al.* Pharmacological Inhibition of O-GlcNAcase Enhances Autophagy in Brain through an mTOR-Independent Pathway. *ACS Chemical Neuroscience* **9** (2018). <https://doi.org/10.1021/acschemneuro.8b00015>
- 26 Li, M.-D. *et al.* O-GlcNAc Signaling Entrains the Circadian Clock by Inhibiting BMAL1/CLOCK Ubiquitination. *Cell Metabolism* **17** (2013). <https://doi.org/10.1016/j.cmet.2012.12.015>
- 27 Lagerlöf, O. & Lagerlöf, O. O-GlcNAc cycling in the developing, adult and geriatric brain. *Journal of Bioenergetics and Biomembranes* **2018 50:3** **50** (2018). <https://doi.org/10.1007/s10863-018-9760-1>
- 28 Baumann, D. *et al.* Role of nutrient-driven O-GlcNAc-post-translational modification in pancreatic exocrine and endocrine islet development. *Development* **147** (2020). <https://doi.org/10.1242/dev.186643>
- 29 Jones, S. P. A bittersweet modification: O-GlcNAc and cardiac dysfunction - PubMed. *Circulation research* **96** (2005). <https://doi.org/10.1161/01.RES.0000168039.61228.67>

- 30 Fan, J. *et al.* O-GlcNAc transferase in astrocytes modulates depression-related stress susceptibility through glutamatergic synaptic transmission. *The Journal of Clinical Investigation* **133** (2023). <https://doi.org/10.1172/JCI160016>
- 31 Su, C. & Schwarz, T. L. O-GlcNAc Transferase Is Essential for Sensory Neuron Survival and Maintenance. *Journal of Neuroscience* **37** (2017). <https://doi.org/10.1523/JNEUROSCI.3384-16.2017>
- 32 Burt, R. A. *et al.* Mapping the O-GlcNAc Modified Proteome: Applications for Health and Disease. *Front Mol Biosci* **9**, 920727 (2022). <https://doi.org/10.3389/fmolb.2022.920727>
- 33 Hart, G. W., Haltiwanger, R. S., Holt, G. D. & Kelly, W. G. Nucleoplasmic and Cytoplasmic Glycoproteins. *Ciba Foundation Symposium 145 - Carbohydrate Recognition in Cellular Function* (2007). <https://doi.org/10.1002/9780470513828.ch7>
- 34 Lv, P. *et al.* O-GlcNAcylation modulates liquid-liquid phase separation of SynGAP/PSD-95. *Nat Chem* **14**, 831-840 (2022). <https://doi.org/10.1038/s41557-022-00946-9>
- 35 Gambetta, M. C. & Muller, J. O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. *Dev Cell* **31**, 629-639 (2014). <https://doi.org/10.1016/j.devcel.2014.10.020>
- 36 Radermacher, P. T. *et al.* O-GlcNAc reports ambient temperature and confers heat resistance on ectotherm development. *Proceedings of the National Academy of Sciences* **111** (2014). <https://doi.org/10.1073/pnas.1322396111>
- 37 Krishnamoorthy, V., Donofrio, A. J. & Martin, J. L. O-GlcNAcylation of αB-crystallin regulates its stress-induced translocation and cytoprotection. *Mol Cell Biochem* **379**, 59-68 (2013). <https://doi.org/10.1007/s11010-013-1627-5>
- 38 Roquemore, E. P., Chevrier, M. R., Cotter, R. J. & Hart, G. W. Dynamic O-GlcNAcylation of the small heat shock protein alpha B-crystallin. *Biochemistry* **35**, 3578-3586 (1996). <https://doi.org/10.1021/bi951918j>
- 39 Khidekel, N. *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J Am Chem Soc* **125**, 16162-16163 (2003). <https://doi.org/10.1021/ja038545r>
- 40 Wang, Z. *et al.* Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and

- electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **9**, 153-160 (2010). <https://doi.org/10.1074/mcp.M900268-MCP200>
- 41 Lim, K. H. & Chang, H. I. O-linked N-acetylglucosamine suppresses thermal aggregation of Sp1. *FEBS Lett* **580**, 4645-4652 (2006).
<https://doi.org/10.1016/j.febslet.2006.07.040>
- 42 Hanover, J. A. & Wang, P. O-GlcNAc cycling shows neuroprotective potential in *C. elegans* models of neurodegenerative disease. *Worm* **2**, e27043 (2013).
<https://doi.org/10.4161/worm.27043>
- 43 Yuzwa, S. A. & Vocadlo, D. J. O-GlcNAc modification and the tauopathies: insights from chemical biology. *Curr Alzheimer Res* **6**, 451-454 (2009).
<https://doi.org/10.2174/156720509789207967>
- 44 Smet-Nocca, C. *et al.* Identification of O-GlcNAc sites within peptides of the Tau protein and their impact on phosphorylation. *Mol Biosyst* **7**, 1420-1429 (2011).
<https://doi.org/10.1039/c0mb00337a>
- 45 Yuzwa, S. A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol* **8**, 393-399 (2012).
<https://doi.org/10.1038/nchembio.797>
- 46 Yuzwa, S. A., Cheung, A. H., Okon, M., McIntosh, L. P. & Vocadlo, D. J. O-GlcNAc modification of tau directly inhibits its aggregation without perturbing the conformational properties of tau monomers. *J Mol Biol* **426**, 1736-1752 (2014).
<https://doi.org/10.1016/j.jmb.2014.01.004>
- 47 Marotta, N. P., Cherwien, C. A., Abeywardana, T. & Pratt, M. R. O-GlcNAc modification prevents peptide-dependent acceleration of α -synuclein aggregation. *Chembiochem* **13**, 2665-2670 (2012).
<https://doi.org/10.1002/cbic.201200478>
- 48 Gambetta, M. C. & Müller, J. O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. *Dev Cell* **31**, 629-639 (2014).
<https://doi.org/10.1016/j.devcel.2014.10.020>
- 49 Wang, B. *et al.* Liquid–liquid phase separation in human health and diseases. *Signal Transduction and Targeted Therapy* **2021 6:1** **6** (2021).
<https://doi.org/10.1038/s41392-021-00678-1>
- 50 Li, X., Lv, P., Du, Y., Chen, X. & Liu, C. Emerging roles of O-glycosylation in regulating protein aggregation, phase separation, and functions. *Current Opinion in Chemical Biology* **75** (2023). <https://doi.org/10.1016/j.cbpa.2023.102314>

- 51 Jafari, R. *et al.* The cellular thermal shift assay for evaluating drug target interactions in cells. *Nature Protocols* 2014 9:9 **9** (2014).
<https://doi.org/10.1038/nprot.2014.138>
- 52 Drake, W. R., Hou, C. W., Zachara, N. E. & Grimes, C. L. New use for CETSA: monitoring innate immune receptor stability via post-translational modification by OGT. *J Bioenerg Biomembr* **50**, 231-240 (2018). <https://doi.org/10.1007/s10863-018-9754-z>
- 53 Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nature Protocols* 2015 10:10 **10** (2015). <https://doi.org/10.1038/nprot.2015.101>
- 54 King, D. T. *et al.* Thermal Proteome Profiling Reveals the O-GlcNAc-Dependent Meltome. *J Am Chem Soc* **144**, 3833-3842 (2022).
<https://doi.org/10.1021/jacs.1c10621>
- 55 Wang, G. *et al.* Multi-comparative Thermal Proteome Profiling Uncovers New O-GlcNAc Proteins in a System-wide Method. *Analytical Chemistry* (2023).
<https://doi.org/10.1021/acs.analchem.2c03371>
- 56 Lee, A. *et al.* Combined Antibody/Lectin Enrichment Identifies Extensive Changes in the O-GlcNAc Sub-proteome upon Oxidative Stress. *Journal of Proteome Research* **15** (2016). <https://doi.org/10.1021/acs.jproteome.6b00369>
- 57 Maynard, J. C. & Chalkley, R. J. Methods for Enrichment and Assignment of N-Acetylglucosamine Modification Sites. *Mol Cell Proteomics* **20**, 100031 (2021).
<https://doi.org/10.1074/mcp.R120.002206>
- 58 Saha, A. *et al.* Advances in chemical probing of protein O-GlcNAc glycosylation: structural role and molecular mechanisms. *Chemical Society Reviews* **50** (2021).
<https://doi.org/10.1039/D0CS01275K>
- 59 Malathy K, K. P. Metabolism of glycosaminoglycans in alloxan diabetic rats-- glutamine fructose-6-phosphate aminotransferase & UDPG dehydrogenase activities of liver. *Indian J Biochem* **8**, 147-149 (1971).
- 60 Davidson, E., Blumenthal, H. & Roseman, S. Glucosamine metabolism: II. Studies on glucosamine 6-phosphate N-acetylase. *Journal of Biological Chemistry* **226** (1957). [https://doi.org/10.1016/S0021-9258\(18\)64811-2](https://doi.org/10.1016/S0021-9258(18)64811-2)
- 61 Mengin-Lecreux, D. & Heijenoort, J. v. Characterization of the Essential Gene glmM Encoding Phosphoglucosamine Mutase in Escherichia coli. *Journal of Biological Chemistry* **271** (1996). <https://doi.org/10.1074/jbc.271.1.32>

- 62 Herscovics, A. & Orlean, P. Glycoprotein biosynthesis in yeast. *The FASEB Journal* **7** (1993). <https://doi.org/10.1096/fasebj.7.6.8472892>
- 63 Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **50**, D571-D577 (2022).
<https://doi.org/10.1093/nar/gkab1045>
- 64 King, D. T., Males, A., Davies, G. J. & Vocadlo, D. J. Molecular mechanisms regulating O-linked N-acetylglucosamine (O-GlcNAc)-processing enzymes. *Current Opinion in Chemical Biology* **53** (2019).
<https://doi.org/10.1016/j.cbpa.2019.09.001>
- 65 Joiner, C. M., Li, H., Jiang, J. & Walker, S. Structural characterization of the O-GlcNAc cycling enzymes: insights into substrate recognition and catalytic mechanisms. *Current Opinion in Structural Biology* **56** (2019).
<https://doi.org/10.1016/j.sbi.2018.12.003>
- 66 Levine, Z. G. *et al.* O-GlcNAc Transferase Recognizes Protein Substrates Using an Asparagine Ladder in the Tetra-tripeptide Repeat (TPR) Superhelix. *Journal of the American Chemical Society* **140** (2018).
<https://doi.org/10.1021/jacs.7b13546>
- 67 Hu, C.-W. *et al.* Electrophilic probes for deciphering substrate recognition by O-GlcNAc transferase. *Nature Chemical Biology* **2017 13:12** **13** (2017).
<https://doi.org/10.1038/nchembio.2494>
- 68 Wulff-Fuentes, E. *et al.* The human O-GlcNAcome database and meta-analysis. *Sci Data* **8**, 25 (2021). <https://doi.org/10.1038/s41597-021-00810-4>
- 69 Trinidad, J. C. *et al.* Global Identification and Characterization of Both O-GlcNAcylation and Phosphorylation at the Murine Synapse *. *Molecular & Cellular Proteomics* **11** (2012). <https://doi.org/10.1074/mcp.O112.018366>
- 70 Lu, L. *et al.* Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. *Biochemistry* **55** (2016).
<https://doi.org/10.1021/acs.biochem.5b01280>
- 71 Chong, P. A., Nosella, M. L., Vanama, M., Ruiz-Arduengo, R. & Forman-Kay, J. D. Exploration of O-GlcNAc transferase glycosylation sites reveals a target sequence compositional bias. *The Journal of Biological Chemistry* **299** (2023).
<https://doi.org/10.1016/j.jbc.2023.104629>
- 72 Maynard, J. C., Burlingame, A. L. & Medzihradszky, K. F. Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), A New Post-translational Modification in

- Mammals. *Molecular & Cellular Proteomics* **15** (2016).
<https://doi.org/10.1074/mcp.M116.061549>
- 73 Alteen, M. G., Tan, H. Y. & Vocadlo, D. J. Monitoring and modulating O-GlcNAcylation: assays and inhibitors of O-GlcNAc processing enzymes. *Current Opinion in Structural Biology* **68** (2021). <https://doi.org/10.1016/j.sbi.2020.12.008>
- 74 Reeves, R. A., Lee, A., Henry, R. & Zachara, N. E. Characterization of the specificity of O-GlcNAc reactive antibodies under conditions of starvation and stress. *Analytical Biochemistry* **457** (2014).
<https://doi.org/10.1016/j.ab.2014.04.008>
- 75 Teo, C. F. et al. Glycopeptide specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nature chemical biology* **6** (2010).
<https://doi.org/10.1038/nchembio.338>
- 76 Burt, R. A. et al. Novel Antibodies for the Simple and Efficient Enrichment of Native O-GlcNAc Modified Peptides. *Mol Cell Proteomics* **20**, 100167 (2021).
<https://doi.org/10.1016/j.mcpro.2021.100167>
- 77 Snow, C. M., Senior, A. & Gerace, L. Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. *J Cell Biol* **104**, 1143-1156 (1987).
<https://doi.org/10.1083/jcb.104.5.1143>
- 78 Ruba, A., Yang, W., Ruba, A. & Yang, W. O-GlcNAc-ylation in the Nuclear Pore Complex. *Cellular and Molecular Bioengineering* **9** (2016).
<https://doi.org/10.1007/s12195-016-0440-0>
- 79 Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. *Anal Biochem* **293**, 169-177 (2001).
<https://doi.org/10.1006/abio.2001.5132>
- 80 Liske, R. & D., F. Specificity of the Agglutinin in Extracts of Wheat Germ. *Nature* **217** (1968). <https://doi.org/10.1038/217860a0>
- 81 Kelly, W. G. & Hart, G. W. Glycosylation of chromosomal proteins: localization of O-linked N-acetylglucosamine in Drosophila chromatin. *Cell* **57**, 243-251 (1989).
[https://doi.org/10.1016/0092-8674\(89\)90962-8](https://doi.org/10.1016/0092-8674(89)90962-8)
- 82 Monsigny, M., Roche, A. C., Sene, C., Maget-Dana, R. & Delmotte, F. Sugar-lectin interactions: how does wheat-germ agglutinin bind sialoglycoconjugates? *Eur J Biochem* **104**, 147-153 (1980). <https://doi.org/10.1111/j.1432-1033.1980.tb04410.x>

- 83 Monsigny, M. *et al.* Properties of succinylated wheat-germ agglutinin. *Eur J Biochem* **98**, 39-45 (1979). <https://doi.org/10.1111/j.1432-1033.1979.tb13157.x>
- 84 Whiteheart, S. W. *et al.* Glycosyltransferase probes. *Methods Enzymol* **179**, 82-95 (1989). [https://doi.org/10.1016/0076-6879\(89\)79116-3](https://doi.org/10.1016/0076-6879(89)79116-3)
- 85 Mariappa, D. *et al.* A mutant O-GlcNAcase as a probe to reveal global dynamics of protein O-GlcNAcylation during *Drosophila* embryonic development. *Biochem J* **470**, 255-262 (2015). <https://doi.org/10.1042/BJ20150610>
- 86 Wu, Z. L. *et al.* Detecting O-GlcNAc using in vitro sulfation. *Glycobiology* **24**, 740-747 (2014). <https://doi.org/10.1093/glycob/cwu037>
- 87 Wu, Z. L., Tatge, T. J., Grill, A. E. & Zou, Y. Detecting and Imaging O-GlcNAc Sites Using Glycosyltransferases: A Systematic Approach to Study O-GlcNAc. *Cell Chem Biol* **25**, 1428-1435 e1423 (2018).
<https://doi.org/10.1016/j.chembiol.2018.07.007>
- 88 Aguilar, A. L., Hou, X., Wen, L., Wang, P. G. & Wu, P. A Chemoenzymatic Histology Method for O-GlcNAc Detection. *Chembiochem* **18**, 2416-2421 (2017). <https://doi.org/10.1002/cbic.201700515>
- 89 Clark, P. M. *et al.* Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J Am Chem Soc* **130**, 11576-11577 (2008).
<https://doi.org/10.1021/ja8030467>
- 90 Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci U S A* **100**, 9116-9121 (2003).
<https://doi.org/10.1073/pnas.1632821100>
- 91 Chuh, K. N., Zaro, B. W., Piller, F., Piller, V. & Pratt, M. R. Changes in metabolic chemical reporter structure yield a selective probe of O-GlcNAc modification. *J Am Chem Soc* **136**, 12283-12295 (2014). <https://doi.org/10.1021/ja504063c>
- 92 Zaro, B. W., Batt, A. R., Chuh, K. N., Navarro, M. X. & Pratt, M. R. The Small Molecule 2-Azido-2-deoxy-glucose Is a Metabolic Chemical Reporter of O-GlcNAc Modifications in Mammalian Cells, Revealing an Unexpected Promiscuity of O-GlcNAc Transferase. *ACS Chem Biol* **12**, 787-794 (2017).
<https://doi.org/10.1021/acscchembio.6b00877>
- 93 Holt, G. D., Haltiwanger, R. S., Torres, C. R. & Hart, G. W. Erythrocytes contain cytoplasmic glycoproteins. O-linked GlcNAc on Band 4.1. *J Biol Chem* **262**, 14847-14850 (1987).

- 94 Roquemore, E. P., Chou, T. Y. & Hart, G. W. Detection of O-linked N-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins. *Methods Enzymol* **230**, 443-460 (1994). [https://doi.org/10.1016/0076-6879\(94\)30028-3](https://doi.org/10.1016/0076-6879(94)30028-3)
- 95 Zhu, Y. & Hart, G. W. Dual-specificity RNA aptamers enable manipulation of target-specific O-GlcNAcylation and unveil functions of O-GlcNAc on β-catenin. *Cell* **186** (2023). <https://doi.org/10.1016/j.cell.2022.12.016>
- 96 Ge, Y. et al. Target protein deglycosylation in living cells by a nanobody-fused split O-GlcNAcase. *Nature Chemical Biology* **2021** *17*:5 **17** (2021). <https://doi.org/10.1038/s41589-021-00757-y>
- 97 Rexach, J. E. et al. Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nature Chemical Biology* **8** (2012). <https://doi.org/10.1038/nchembio.770>
- 98 Zhu, Y. et al. Tandem Bioorthogonal Labeling Uncovers Endogenous Cotranslationally O-GlcNAc Modified Nascent Proteins. *Journal of the American Chemical Society* **142** (2020). <https://doi.org/10.1021/jacs.0c04121>
- 99 Ma, J. & Hart, G. W. Analysis of Protein O-GlcNAcylation by Mass Spectrometry. *Current protocols in protein science* **87** (2017). <https://doi.org/10.1002/cpp.24>
- 100 Kitov, P. I. & Bundle, D. R. On the Nature of the Multivalency Effect: A Thermodynamic Model. (2003). <https://doi.org/10.1021/ja038223>
- 101 Ward, E. M., Kizer, M. E. & Imperiali, B. Strategies and Tactics for the Development of Selective Glycan-Binding Proteins. *ACS Chemical Biology* **16** (2021). <https://doi.org/10.1021/acschembio.0c00880>
- 102 Tian, Y., Ma, S. & Wen, L. Towards chemoenzymatic labeling strategies for profiling protein glycosylation. *Current Opinion in Chemical Biology* **80** (2024). <https://doi.org/10.1016/j.cbpa.2024.102460>
- 103 Ma, J. & Hart, G. W. O-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics* **11**, 8 (2014). <https://doi.org/10.1186/1559-0275-11-8>
- 104 Thompson, J. W., Griffin, M. E. & Hsieh-Wilson, L. C. Methods for the Detection, Study, and Dynamic Profiling of O-GlcNAc Glycosylation. *Methods Enzymol* **598**, 101-135 (2018). <https://doi.org/10.1016/bs.mie.2017.06.009>
- 105 Yoshida, N., Mortara, R. A., Araguth, M. F., Gonzalez, J. C. & Russo, M. Metacyclic neutralizing effect of monoclonal antibody 10D8 directed to the 35- and 50-kilodalton surface glycoconjugates of Trypanosoma cruzi. *Infection and Immunity* **57** (1989). <https://doi.org/10.1128/iai.57.6.1663-1667.1989>

- 106 Turner, J. R. *et al.* Cytologic assessment of nuclear and cytoplasmic O-linked N-acetylglucosamine distribution by using anti-streptococcal monoclonal antibodies. *Proceedings of the National Academy of Sciences* **87** (1990).
<https://doi.org/10.1073/pnas.87.15.5608>
- 107 Stanfield, R. L. & Wilson, I. A. Antibody Structure. *Microbiology Spectrum* **2** (2014). <https://doi.org/10.1128/microbiolspec.aid-0012-2013>
- 108 Rouet, R., Jackson, K. J. L., Langley, D. B. & Christ, D. Next-Generation Sequencing of Antibody Display Repertoires. *Frontiers in Immunology* **9** (2018).
<https://doi.org/10.3389/fimmu.2018.00118>
- 109 Bemani, P., Mohammadi, M. & Hakakian, A. ScFv Improvement Approaches. *Protein & Peptide Letters* **25** (2018).
<https://doi.org/10.2174/0929866525666171129225436>
- 110 Huston, J. S. *et al.* Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **85** (1988). <https://doi.org/10.1073/pnas.85.16.5879>
- 111 Zielke, F. M., Rutjes, F. P. J. T., Zielke, F. M. & Rutjes, F. P. J. T. Recent Advances in Bioorthogonal Ligation and Bioconjugation. *Topics in Current Chemistry* **2023** *381*:6 **381** (2023). <https://doi.org/10.1007/s41061-023-00445-6>
- 112 Kitten, O. & Martineau, P. Antibody alternative formats: antibody fragments and new frameworks. *Med Sci (Paris)* **35**, 1092-1097 (2019).
<https://doi.org/10.1051/medsci/2019217>
- 113 Wang, H., Zhao, F., Han, X. & Yang, Z. Production and characterization of a biotinylated single-chain variable fragment antibody for detection of parathion-methyl. *Protein Expr Purif* **126**, 1-8 (2016).
<https://doi.org/10.1016/j.pep.2016.05.005>
- 114 Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods in molecular biology (Clifton, N.J.)* **1266** (2015).
https://doi.org/10.1007/978-1-4939-2272-7_12
- 115 Chothia, C. & Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology* **196** (1987).
[https://doi.org/10.1016/0022-2836\(87\)90412-8](https://doi.org/10.1016/0022-2836(87)90412-8)
- 116 Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Research* **51** (2023). <https://doi.org/10.1093/nar/gkac1096>

- 117 Chen, X., Zaro, J. & Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Advanced drug delivery reviews* **65** (2013).
<https://doi.org/10.1016/j.addr.2012.09.039>
- 118 Chichili, V. P. R., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein–protein interactions. *Protein Science : A Publication of the Protein Society* **22** (2013). <https://doi.org/10.1002/pro.2206>
- 119 Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science* **8** (1999).
<https://doi.org/10.1110/ps.8.4.921>
- 120 Dübel, S. et al. Bifunctional and multimeric complexes of streptavidin fused to single chain antibodies (scFv). *Journal of Immunological Methods* **178** (1995).
[https://doi.org/10.1016/0022-1759\(94\)00257-W](https://doi.org/10.1016/0022-1759(94)00257-W)
- 121 Lian, H. et al. A novel multimeric sCD19-streptavidin fusion protein for functional detection and selective expansion of CD19-targeted CAR-T cells. *Cancer Medicine* **11** (2022). <https://doi.org/10.1002/cam4.4657>
- 122 Ikonomova, S. P., Le, M. T., Kalla, N. & Karlsson, A. J. Effect of linkers on immobilization of scFvs with biotin–streptavidin interaction. *Biotechnology and Applied Biochemistry* **65** (2018). <https://doi.org/10.1002/bab.1645>
- 123 Wartchow, C. A. et al. Biosensor-based small molecule fragment screening with biolayer interferometry. *Journal of Computer-Aided Molecular Design* **2011** *25*:7 **25** (2011). <https://doi.org/10.1007/s10822-011-9439-8>
- 124 Kapcan, E., Lake, B., Yang, Z. & Rullo, A. F. Methods to Validate Binding and Kinetics of “Proximity-Inducing” Covalent Immune-Recruiting Molecules. *Current Protocols in Chemical Biology* **12** (2020). <https://doi.org/10.1002/cpch.88>
- 125 Dundas, C. M. et al. Streptavidin–biotin technology: improvements and innovations in chemical and biological applications. *Applied Microbiology and Biotechnology* **2013** *97*:21 **97** (2013). <https://doi.org/10.1007/s00253-013-5232-z>
- 126 Liu, W. et al. Theoretical basis, state and challenges of living cell-based drug delivery systems. *Theranostics* **14** (2024). <https://doi.org/10.7150/thno.99257>
- 127 Lim, K. H., Huang, H., Pralle, A. & Park, S. Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnology and Bioengineering* **110** (2013). <https://doi.org/10.1002/bit.24605>
- 128 Carrington, J. C. & Dougherty, W. G. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing.

- Proceedings of the National Academy of Sciences of the United States of America* **85** (1988). <https://doi.org/10.1073/pnas.85.10.3391>
- 129 Kang, T. H. & Seong, B. L. Solubility, Stability, and Avidity of Recombinant Antibody Fragments Expressed in Microorganisms. *Frontiers in Microbiology* **11** (2020). <https://doi.org/10.3389/fmicb.2020.01927>
- 130 Lee, C. C., Perchiacca, J. M. & Tessier, P. M. Toward aggregation-resistant antibodies by design. *Trends in Biotechnology* **31** (2013). <https://doi.org/10.1016/j.tibtech.2013.07.002>
- 131 Kiguchi, Y. et al. The VH framework region 1 as a target of efficient mutagenesis for generating a variety of affinity-matured scFv mutants. *Scientific Reports* **2021** *11*:1 **11** (2021). <https://doi.org/10.1038/s41598-021-87501-7>
- 132 Sun, W. et al. A combined strategy improves the solubility of aggregation-prone single-chain variable fragment antibodies. *Protein Expression and Purification* **83** (2012). <https://doi.org/10.1016/j.pep.2012.02.006>
- 133 Wörn, A. & Plückthun, A. An intrinsically stable antibody scFv fragment can tolerate the loss of both disulfide bonds and fold correctly. *FEBS Letters* **427** (1998). [https://doi.org/10.1016/S0014-5793\(98\)00463-3](https://doi.org/10.1016/S0014-5793(98)00463-3)
- 134 Proba, K., Wörn, A., Honegger, A. & Plückthun, A. Antibody scFv fragments without disulfide bonds, made by molecular evolution. *Journal of Molecular Biology* **275** (1998). <https://doi.org/10.1006/jmbi.1997.1457>
- 135 Chng, J. et al. Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *mAbs* **7** (2015). <https://doi.org/10.1080/19420862.2015.1008351>
- 136 Wang, Y. et al. 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Scientific Reports* **2015** *5*:1 **5** (2015). <https://doi.org/10.1038/srep16273>
- 137 Sharma, P. et al. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Research* **40** (2012). <https://doi.org/10.1093/nar/gkr1176>
- 138 Donnelly, M. L. L. et al. Analysis of the aphthovirus 2A/2B polyprotein ‘cleavage’ mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal ‘skip’. *Journal of General Virology* **82** (2001). <https://doi.org/10.1099/0022-1317-82-5-1013>

- 139 Haryadi, R. *et al.* Optimization of Heavy Chain and Light Chain Signal Peptides for High Level Expression of Therapeutic Antibodies in CHO Cells. *PLOS ONE* **10** (2015). <https://doi.org/10.1371/journal.pone.0116878>
- 140 Novy, R., Yaeger, K., Held, D. & Mierendorf, R. Coexpression of multiple target proteins in *E. coli*. *inNovations* **15** (2002).
- 141 Held, D., Yaeger, K. & Novy, R. New coexpression vectors for expanded compatibilities in *E. coli*. *inNovations* **18** (2003).
- 142 Raines, R. T., McCormick, M., Oosbree, T. R. V. & Mierendorf, R. C. in *Methods in enzymology* Vol. 326 362-376 (2000).
- 143 Snoeck, S. *et al.* “Metabolic burden” explained: stress symptoms and its related responses induced by (over)expression of (heterologous) proteins in *Escherichia coli*. *Microbial Cell Factories* 2024 23:1 **23** (2024).
<https://doi.org/10.1186/s12934-024-02370-9>
- 144 Maina, C. V. *et al.* An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74** (1988). [https://doi.org/10.1016/0378-1119\(88\)90170-9](https://doi.org/10.1016/0378-1119(88)90170-9)
- 145 Shahravan, S. H., Qu, X., Chan, I.-S. & Shin, J. A. Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. *Protein expression and purification* **59** (2008).
<https://doi.org/10.1016/j.pep.2008.02.015>
- 146 Asano, R. *et al.* Efficient Construction of a Diabody Using a Refolding System: Anti- Carcinoembryonic Antigen Recombinant Antibody Fragment. *The Journal of Biochemistry* **132** (2002). <https://doi.org/10.1093/oxfordjournals.jbchem.a003303>
- 147 Thomson, C. A., Olson, M., Jackson, L. M. & Schrader, J. W. A Simplified Method for the Efficient Refolding and Purification of Recombinant Human GM-CSF. *PLOS ONE* **7** (2012). <https://doi.org/10.1371/journal.pone.0049891>
- 148 Escobar, E. E. *et al.* Precision Mapping of O-Linked N-Acetylglucosamine Sites in Proteins Using Ultraviolet Photodissociation Mass Spectrometry. *J Am Chem Soc* **142**, 11569-11577 (2020). <https://doi.org/10.1021/jacs.0c04710>
- 149 Osset, M., Piñol, M., Fallon, M. J. M., Llorens, R. D. & Cuchillo, C. M. Interference of the carbohydrate moiety in Coomassie Brilliant Blue R-250 protein staining. *Electrophoresis* **10** (1989). <https://doi.org/10.1002/elps.1150100412>

- 150 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).
<https://doi.org/10.1126/science.abj8754>
- 151 Geoghegan, K. F. *et al.* Spontaneous α-N-6-Phosphogluconoylation of a “His Tag” inEscherichia coli:The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Analytical Biochemistry* **267** (1999).
<https://doi.org/10.1006/abio.1998.2990>
- 152 Schweida, D. *et al.* The NMR signature of gluconylation: a frequent N-terminal modification of isotope-labeled proteins. *Journal of Biomolecular NMR* **2019** *73*:1-73 (2019). <https://doi.org/10.1007/s10858-019-00228-6>
- 153 Love, C. E. *et al.* Biolayer interferometry as a biological detection platform. *TrAC Trends in Analytical Chemistry* **130***56* (2024). <https://doi.org/10.1117/12.3013538>
- 154 Yang, Y. *et al.* Increase recombinant antibody yields through optimizing vector design and production process in CHO cells. *Appl Microbiol Biotechnol* **106**, 4963-4975 (2022). <https://doi.org/10.1007/s00253-022-12051-5>
- 155 Kolb, A. F. & Siddell, S. G. Expression of a recombinant monoclonal antibody from a bicistronic mRNA. *Hybridoma* **16**, 421-426 (1997).
<https://doi.org/10.1089/hyb.1997.16.421>
- 156 Li, J. *et al.* A comparative study of different vector designs for the mammalian expression of recombinant IgG antibodies. *J Immunol Methods* **318**, 113-124 (2007). <https://doi.org/10.1016/j.jim.2006.10.010>
- 157 Kang, T. H. & Seong, B. L. Solubility, Stability, and Avidity of Recombinant Antibody Fragments Expressed in Microorganisms. *Front Microbiol* **11**, 1927 (2020). <https://doi.org/10.3389/fmicb.2020.01927>
- 158 Ortega, C., Oppezzo, P. & Correa, A. Overcoming the Solubility Problem in E. coli: Available Approaches for Recombinant Protein Production. *Methods Mol Biol* **240***6*, 35-64 (2022). https://doi.org/10.1007/978-1-0716-1859-2_2
- 159 Mahalakshmi, N. *et al.* Molecular evolution of single chain fragment variable (scFv) for diagnosis of lymphatic filariasis. *Mol Biol Rep* **46**, 5409-5418 (2019).
<https://doi.org/10.1007/s11033-019-04995-1>
- 160 Sharma, P. *et al.* Structure-guided engineering of the affinity and specificity of CARs against Tn-glycopeptides. *Proc Natl Acad Sci U S A* **117**, 15148-15159 (2020). <https://doi.org/10.1073/pnas.1920662117>

- 161 Haltiwanger, R. S., Holt, G. D. & Hart, G. W. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *J Biol Chem* **265**, 2563-2568 (1990).
- 162 Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567 (2011). <https://doi.org/10.1038/nature09638>
- 163 Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J Biol Chem* **276**, 9838-9845 (2001). <https://doi.org/10.1074/jbc.M010420200>
- 164 Hurtado-Guerrero, R., Dorfmueller, H. C. & van Aalten, D. M. Molecular mechanisms of O-GlcNAcylation. *Curr Opin Struct Biol* **18**, 551-557 (2008). <https://doi.org/10.1016/j.sbi.2008.09.005>
- 165 Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022 (2007). <https://doi.org/10.1038/nature05815>
- 166 Ramakrishnan, P. et al. Activation of the transcriptional function of the NF-kappaB protein c-Rel by O-GlcNAc glycosylation. *Sci Signal* **6**, ra75 (2013). <https://doi.org/10.1126/scisignal.2004097>
- 167 Comer, F. I. & Hart, G. W. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40**, 7845-7852 (2001). <https://doi.org/10.1021/bi0027480>
- 168 Vosseller, K. et al. O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* **5**, 923-934 (2006). <https://doi.org/10.1074/mcp.T500040-MCP200>
- 169 Dupas, T. et al. An overview of tools to decipher O-GlcNAcylation from historical approaches to new insights. *Int J Biochem Cell Biol* **151**, 106289 (2022). <https://doi.org/10.1016/j.biocel.2022.106289>
- 170 Ferrand, Y. et al. A synthetic lectin for O-linked beta-N-acetylglucosamine. *Angew Chem Int Ed Engl* **48**, 1775-1779 (2009). <https://doi.org/10.1002/anie.200804905>

- 171 Klement, E., Lipinszki, Z., Kupihar, Z., Udvardy, A. & Medzihradszky, K. F. Enrichment of O-GlcNAc modified proteins by the periodate oxidation-hydrazide resin capture approach. *J Proteome Res* **9**, 2200-2206 (2010).
<https://doi.org/10.1021/pr900984h>
- 172 Yu, S. H. et al. Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. *Proc Natl Acad Sci U S A* **109**, 4834-4839 (2012). <https://doi.org/10.1073/pnas.1114356109>
- 173 Xu, S., Sun, F. & Wu, R. A Chemoenzymatic Method Based on Easily Accessible Enzymes for Profiling Protein O-GlcNAcylation. *Anal Chem* **92**, 9807-9814 (2020). <https://doi.org/10.1021/acs.analchem.0c01284>
- 174 Chen, Y. et al. Highly Efficient Enrichment of O-GlcNAc Glycopeptides Based on Chemical Oxidation and Reversible Hydrazide Chemistry. *Anal Chem* **93**, 16618-16627 (2021). <https://doi.org/10.1021/acs.analchem.1c04031>
- 175 Butler, T. et al. Chemoenzymatic synthesis of biotinylated nucleotide sugars as substrates for glycosyltransferases. *Chembiochem* **2**, 884-894 (2001).
[https://doi.org/10.1002/1439-7633\(20011203\)2:12<884::AID-CBIC884>3.0.CO;2-2](https://doi.org/10.1002/1439-7633(20011203)2:12<884::AID-CBIC884>3.0.CO;2-2)
- 176 Gastinel, L. N., Cambillau, C. & Bourne, Y. Crystal structures of the bovine beta4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J* **18**, 3546-3557 (1999).
<https://doi.org/10.1093/emboj/18.13.3546>
- 177 Ramakrishnan, B., Moncrief, A. J., Davis, T. A., Holland, L. A. & Qasba, P. K. Investigations on beta1,4-galactosyltransferase I using 6-sulfo-GlcNAc as an acceptor sugar substrate. *Glycoconj J* **30**, 835-842 (2013).
<https://doi.org/10.1007/s10719-013-9488-4>
- 178 Williams, S. J., Mark, B. L., Vocadlo, D. J., James, M. N. & Withers, S. G. Aspartate 313 in the Streptomyces plicatus hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. *J Biol Chem* **277**, 40055-40065 (2002).
<https://doi.org/10.1074/jbc.M206481200>
- 179 Heins, R. A. et al. Phylogenomically guided identification of industrially relevant GH1 beta-glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry. *ACS Chem Biol* **9**, 2082-2091 (2014).
<https://doi.org/10.1021/cb500244v>

- 180 Coines, J., Raich, L. & Rovira, C. Modeling catalytic reaction mechanisms in glycoside hydrolases. *Curr Opin Chem Biol* **53**, 183-191 (2019).
<https://doi.org/10.1016/j.cbpa.2019.09.007>
- 181 Lu, L., Li, Z. & Xiao, M. Recent progress on galacto-oligosaccharides synthesis by microbial beta-galactosidase--a review. *Wei Sheng Wu Xue Bao* **48**, 980-985 (2008).
- 182 Bultema, J. B., Kuipers, B. J. & Dijkhuizen, L. Biochemical characterization of mutants in the active site residues of the beta-galactosidase enzyme of *Bacillus circulans* ATCC 31382. *FEBS Open Bio* **4**, 1015-1020 (2014).
<https://doi.org/10.1016/j.fob.2014.11.002>
- 183 Lu, L., Guo, L., Wang, K., Liu, Y. & Xiao, M. beta-Galactosidases: A great tool for synthesizing galactose-containing carbohydrates. *Biotechnol Adv* **39**, 107465 (2020). <https://doi.org/10.1016/j.biotechadv.2019.107465>
- 184 Oliveira, C., Guimaraes, P. M. & Domingues, L. Recombinant microbial systems for improved beta-galactosidase production and biotechnological applications. *Biotechnol Adv* **29**, 600-609 (2011).
<https://doi.org/10.1016/j.biotechadv.2011.03.008>
- 185 Henze, M. et al. Rational design of a glycosynthase by the crystal structure of beta-galactosidase from *Bacillus circulans* (BgaC) and its use for the synthesis of N-acetyllactosamine type 1 glycan structures. *J Biotechnol* **191**, 78-85 (2014).
<https://doi.org/10.1016/j.jbiotec.2014.07.003>
- 186 Chatham, J. C., Zhang, J. & Wende, A. R. Role of O-Linked N-Acetylglucosamine Protein Modification in Cellular (Patho)Physiology. *Physiol Rev* **101**, 427-493 (2021). <https://doi.org/10.1152/physrev.00043.2019>
- 187 Pratt, M. R. & Vocadlo, D. J. Understanding and exploiting the roles of O-GlcNAc in neurodegenerative diseases. *J Biol Chem* **299**, 105411 (2023).
<https://doi.org/10.1016/j.jbc.2023.105411>
- 188 Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**, 545-600 (1997).
<https://doi.org/10.1146/annurev.physchem.48.1.545>
- 189 Kapon, R., Nevo, R. & Reich, Z. Protein energy landscape roughness. *Biochem Soc Trans* **36**, 1404-1408 (2008). <https://doi.org/10.1042/BST0361404>

- 190 Neelamraju, S., Wales, D. J. & Gosavi, S. Protein energy landscape exploration
with structure-based models. *Curr Opin Struct Biol* **64**, 145-151 (2020).
<https://doi.org/10.1016/j.sbi.2020.07.003>
- 191 Varela, A. E., England, K. A. & Cavagnero, S. Kinetic trapping in protein folding.
Protein Eng Des Sel **32**, 103-108 (2019). <https://doi.org/10.1093/protein/gzz018>
- 192 Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: a close
look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* **105**, 8256-8261
(2008). <https://doi.org/10.1073/pnas.0801340105>
- 193 Culyba, E. K. et al. Protein native-state stabilization by placing aromatic side
chains in N-glycosylated reverse turns. *Science* **331**, 571-575 (2011).
<https://doi.org/10.1126/science.1198461>
- 194 Carroll, E. C., Greene, E. R., Martin, A. & Marqusee, S. Site-specific
ubiquitination affects protein energetics and proteasomal degradation. *Nat Chem
Biol* **16**, 866-875 (2020). <https://doi.org/10.1038/s41589-020-0556-3>
- 195 Zhu, Y., Shan, X., Yuzwa, S. A. & Vocadlo, D. J. The emerging link between O-
GlcNAc and Alzheimer disease. *J Biol Chem* **289**, 34472-34481 (2014).
<https://doi.org/10.1074/jbc.R114.601351>
- 196 Bartolome-Nebreda, J. M., Trabanco, A. A., Velter, A. I. & Buijnsters, P. O-
GlcNAcase inhibitors as potential therapeutics for the treatment of Alzheimer's
disease and related tauopathies: analysis of the patent literature. *Expert Opin
Ther Pat* **31**, 1117-1154 (2021). <https://doi.org/10.1080/13543776.2021.1947242>
- 197 Marotta, N. P. et al. O-GlcNAc modification blocks the aggregation and toxicity of
the protein α -synuclein associated with Parkinson's disease. *Nature Chemistry*
2015 **7:11** 7 (2015). <https://doi.org/10.1038/nchem.2361>
- 198 Levine, P. M. et al. alpha-Synuclein O-GlcNAcylation alters aggregation and
toxicity, revealing certain residues as potential inhibitors of Parkinson's disease.
Proc Natl Acad Sci U S A **116**, 1511-1519 (2019).
<https://doi.org/10.1073/pnas.1808845116>
- 199 DiFiglia, M. et al. Aggregation of huntingtin in neuronal intranuclear inclusions
and dystrophic neurites in brain. *Science* **277**, 1990-1993 (1997).
<https://doi.org/10.1126/science.277.5334.1990>
- 200 Kumar, A. et al. Decreased O-linked GlcNAcylation protects from cytotoxicity
mediated by huntingtin exon1 protein fragment. *J Biol Chem* **289**, 13543-13553
(2014). <https://doi.org/10.1074/jbc.M114.553321>

- 201 Du, P., Zhang, X., Lian, X., Holscher, C. & Xue, G. O-GlcNAcylation and Its Roles in Neurodegenerative Diseases. *J Alzheimers Dis* **97**, 1051-1068 (2024). <https://doi.org/10.3233/JAD-230955>
- 202 Yang, W. H. *et al.* Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat Cell Biol* **8**, 1074-1083 (2006). <https://doi.org/10.1038/ncb1470>
- 203 Meier, K. & Brehm, A. Chromatin regulation: how complex does it get? *Epigenetics* **9**, 1485-1495 (2014). <https://doi.org/10.4161/15592294.2014.971580>
- 204 Medina, I., Calleja, M. & Morata, G. Tumorigenesis and cell competition in Drosophila in the absence of polyhomeotic function. *Proc Natl Acad Sci U S A* **118** (2021). <https://doi.org/10.1073/pnas.2110062118>
- 205 Gambetta, M. C. & Muller, J. A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma* **124**, 429-442 (2015). <https://doi.org/10.1007/s00412-015-0513-1>
- 206 Yuzwa, S. A. *et al.* Mapping O-GlcNAc modification sites on tau and generation of a site-specific O-GlcNAc tau antibody. *Amino Acids* **40**, 857-868 (2011). <https://doi.org/10.1007/s00726-010-0705-1>
- 207 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- 208 UniProt, C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531 (2023). <https://doi.org/10.1093/nar/gkac1052>
- 209 Bludau, I. *et al.* The structural context of posttranslational modifications at a proteome-wide scale. *PLoS Biol* **20**, e3001636 (2022). <https://doi.org/10.1371/journal.pbio.3001636>
- 210 Macauley, M. S., Stubbs, K. A. & Vocadlo, D. J. O-GlcNAcase catalyzes cleavage of thioglycosides without general acid catalysis. *J Am Chem Soc* **127**, 17202-17203 (2005). <https://doi.org/10.1021/ja0567687>
- 211 Kim, C. A., Gingery, M., Pilpa, R. M. & Bowie, J. U. The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* **9**, 453-457 (2002). <https://doi.org/10.1038/nsb802>
- 212 Regelsberger, G. *et al.* The role of distal tryptophan in the bifunctional activity of catalase-peroxidases. *Biochem Soc Trans* **29**, 99-105 (2001). <https://doi.org/10.1042/0300-5127:0290099>

- 213 Dragovich, P. S. *et al.* Identification of substituted 2-thio-6-oxo-1,6-dihdropyrimidines as inhibitors of human lactate dehydrogenase. *Bioorg Med Chem Lett* **23**, 3186-3194 (2013). <https://doi.org/10.1016/j.bmcl.2013.04.001>
- 214 Kotaka, M. *et al.* Structural studies of glucose-6-phosphate and NADP⁺ binding to human glucose-6-phosphate dehydrogenase. *urn:issn:0907-4449* **61** (2005). <https://doi.org/10.1107/S0907444905002350>
- 215 Yuzwa, S. A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nature Chemical Biology* **2012** *8:4* **8** (2012). <https://doi.org/10.1038/nchembio.797>
- 216 Narayan, A., Bhattacharjee, K. & Naganathan, A. N. Thermally versus Chemically Denatured Protein States. *Biochemistry* **58**, 2519-2523 (2019). <https://doi.org/10.1021/acs.biochem.9b00089>
- 217 Xu, S., Yin, K. & Wu, R. Combining Selective Enrichment and a Boosting Approach to Globally and Site-Specifically Characterize Protein Co-translational O-GlcNAcylation. *Analytical Chemistry* **95** (2023). <https://doi.org/10.1021/acs.analchem.2c04779>

APPENDIX

FULL GELS AND IMMUNOBLOTS

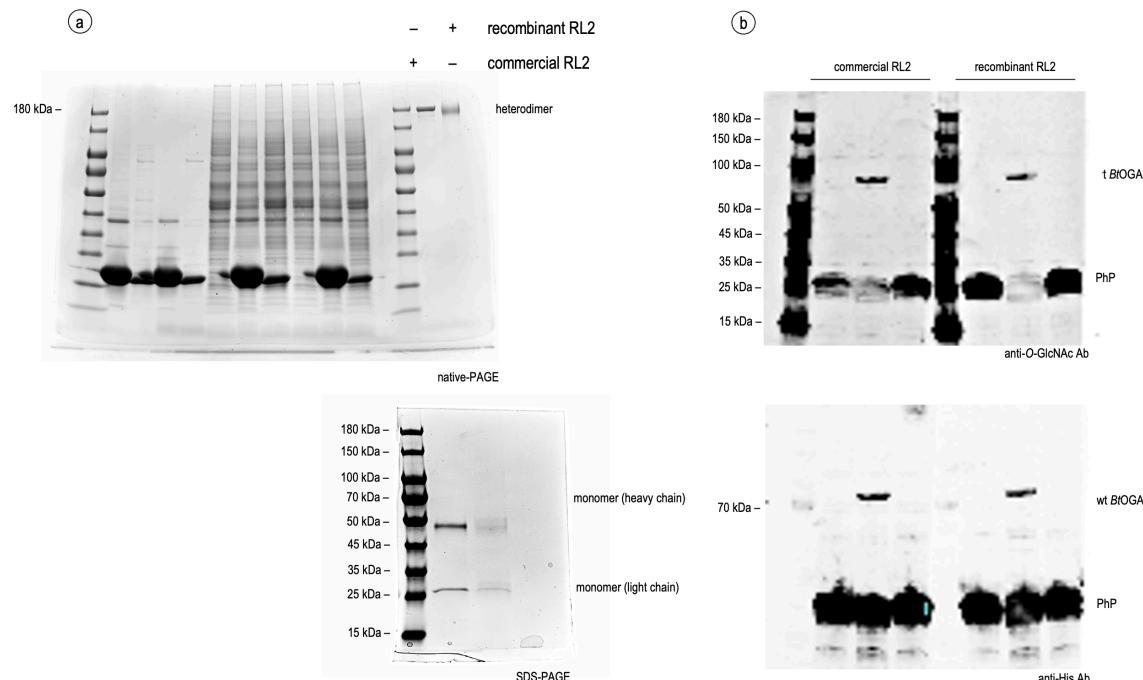


Figure A.1 PAGEs and immunoblots for RL2 characterization

Uncropped SDS-PAGE and Native PAGE gels with molecular weight markers corresponding to Figure 2.8. (a) Full SDS-PAGE and native-PAGE gels comparing commercial RL2 with recombinantly expressed RL2. The native-PAGE (top) shows the mobility of both antibodies, while the SDS-PAGE (bottom) under reducing conditions displays the heavy and light chain bands for both versions. Molecular weight markers are included for reference. (b) Full Western blot using recombinant RL2 to detect O-GlcNAc-modified and unmodified proteins. The top section shows a direct comparison of commercial RL2 and recombinant RL2 binding, while the lower section contains bands corresponding to the His-tagged Ph-P1397-1589 construct. wt. BtOGA was used to enzymatically remove O-GlcNAc, while a catalytically inactive mutant served as a negative control. Markers and full gel images are provided for transparency and reproducibility.

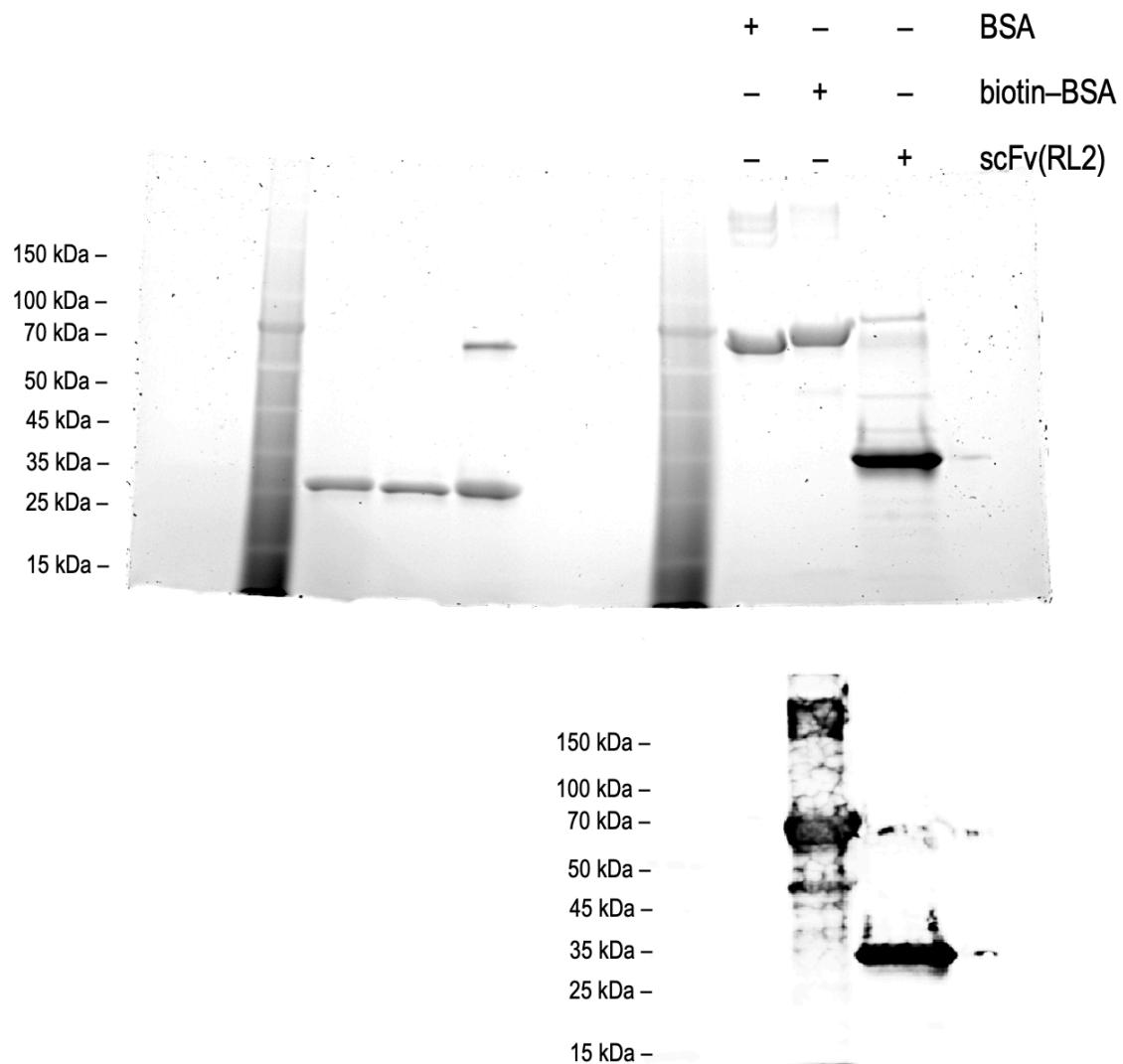


Figure A.2 PAGE and blot analysis of scFv(RL2)

Uncropped SDS-PAGE and streptavidin blot images with molecular weight markers corresponding to Figure 2.9. Full SDS-PAGE gel and streptavidin blot confirming the expression and biotinylation of scFv(RL2). The SDS-PAGE gel displays the migration pattern of scFv(RL2), while the streptavidin blot demonstrates successful biotinylation through a strong signal comparable to the positive control (biotinylated-BSA). Molecular weight markers are included for reference. Full gel images are provided to ensure transparency and reproducibility of the results.

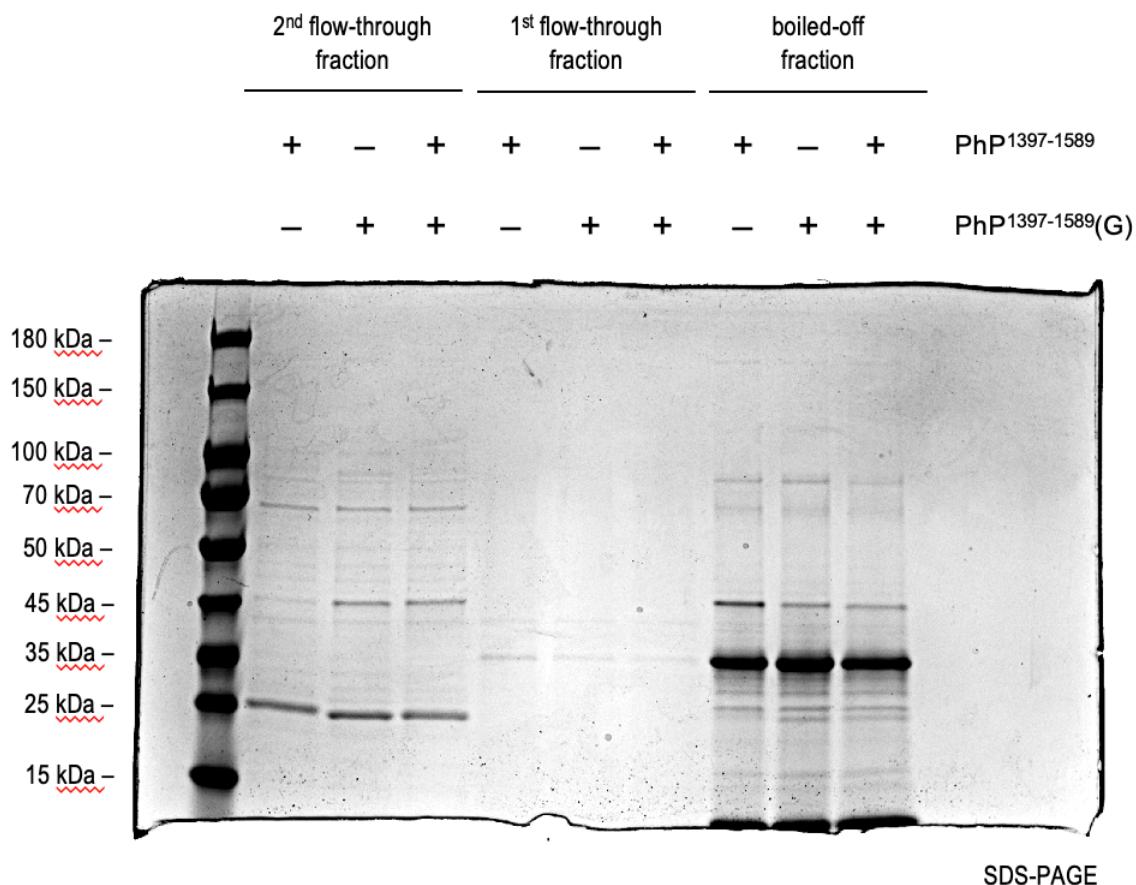


Figure A.3 Pull-down assay of O-GlcNAc modified Ph-P using scFv(RL2)
 SDS-PAGE gels with molecular weight markers corresponding to Figure 2.15, showing the full pull-down assay using scFv(RL2) immobilized on streptavidin-coated magnetic beads. The top panel presents the first flow-through, indicating the amount of scFv bound after 30 minutes of incubation. The middle panel displays the second flow-through after a 1-hour incubation with modified and non-modified Ph-P. The bottom panel shows the remaining Ph-P retained on the beads after washing and boiling. Molecular weight markers are included for reference, ensuring transparency and reproducibility.

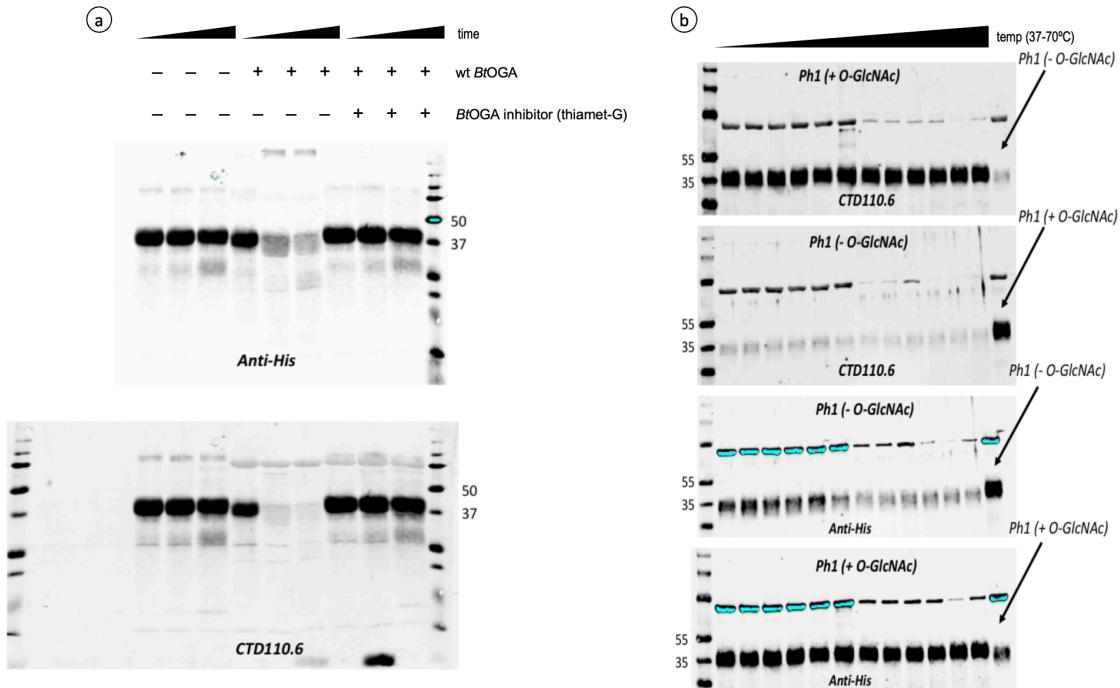


Figure A.4 Full immunoblots for Ph-P¹²²⁷⁻¹⁵⁹⁸ (Figure 4.5)

Full immunoblots corresponding to Figure 4.5, verifying O-GlcNAc modification of Ph-P¹²²⁷⁻¹⁵⁹⁸ and assessing its thermal stability (37–70°C). The top panel detects total protein (anti-His), while the bottom panel monitors glycosylation (anti-O-GlcNAc). O-GlcNAcylation enhances thermal stability, as shown by increased solubility at higher temperatures.

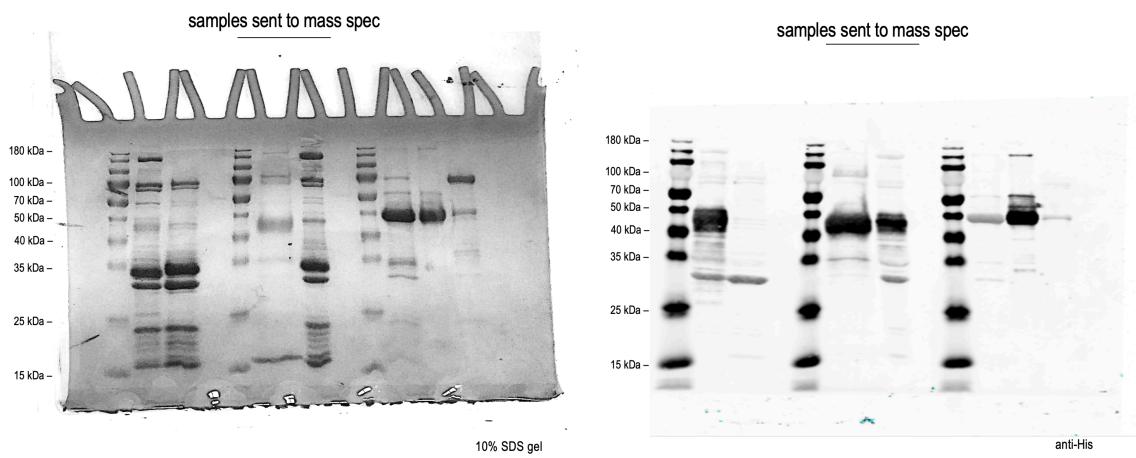


Figure A.5 Purification analysis of recombinant Ph-P¹²²⁷⁻¹⁵⁹⁸

Full SDS-PAGE gel and immunoblot corresponding to Figure 4.6, assessing the purity of Ph-P¹²²⁷⁻¹⁵⁹⁸ after recombinant expression. (Left) The SDS-PAGE gel displays multiple protein bands, with a faint band corresponding to Ph-P¹²²⁷⁻¹⁵⁹⁸, indicating the presence of significant contaminants. Protein bands were excised and analyzed by MS to identify the contaminants. (Right) The anti-His blot confirms Ph-P¹²²⁷⁻¹⁵⁹⁸ expression, though its low yield and smaller size compared to contaminating proteins highlight the challenges in obtaining a highly purified sample. Molecular weight markers are included for reference.

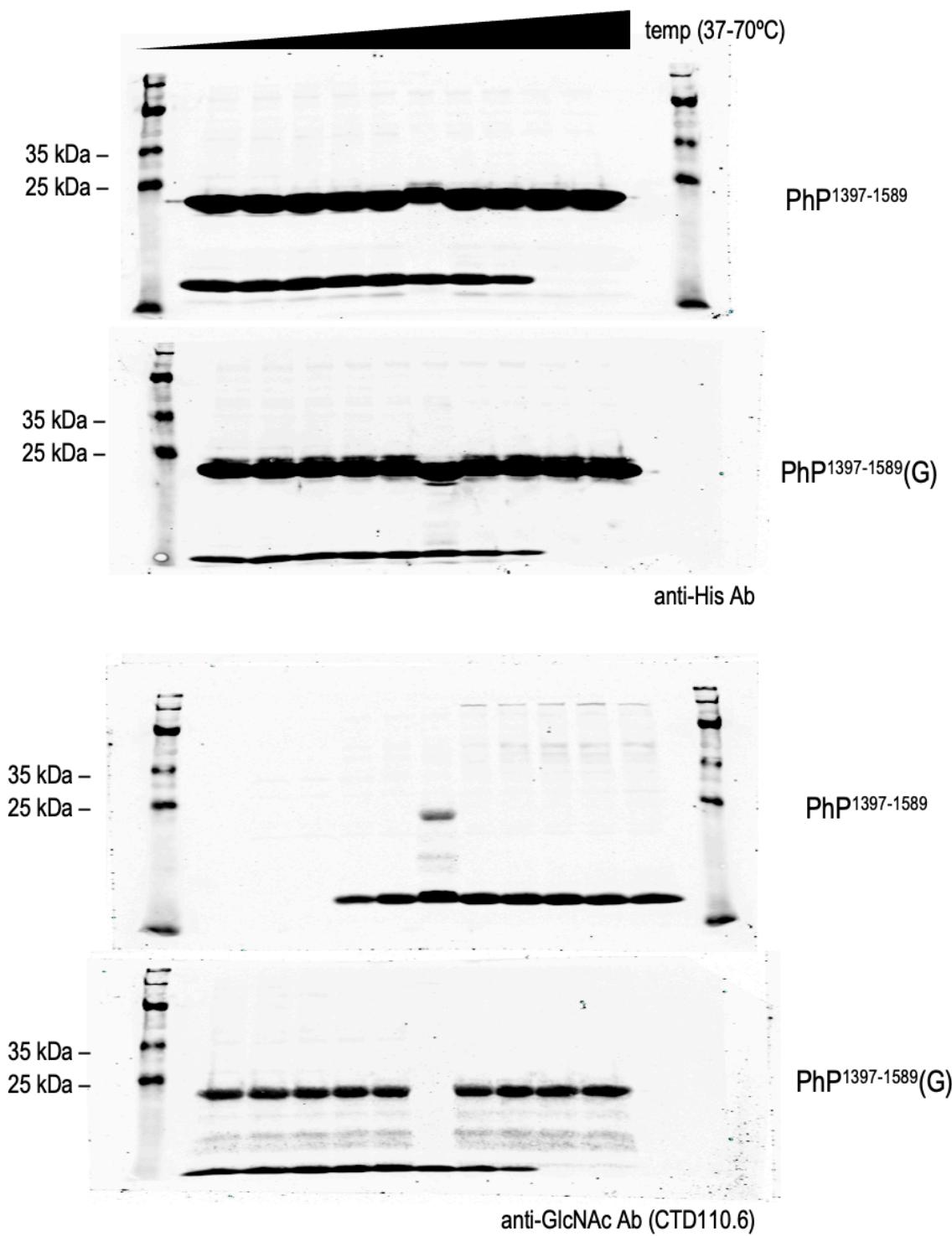


Figure A.6 Full immunoblots for Ph-P¹³⁹⁷⁻¹⁵⁸⁹ (Figure 4.7)

Full immunoblots corresponding to Figure 4.7, evaluating the thermal stability of Ph-P¹³⁹⁷⁻¹⁵⁸⁹ (37–70°C). The top panel detects total protein (anti-His), while the bottom panel monitors glycosylation (anti-O-GlcNAc). Results show Ph-P¹³⁹⁷⁻¹⁵⁸⁹ remains stable regardless of glycosylation. (NOTE: In Figure 4.7, Lane 6 had an inadvertent sample swap.)