
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Biochemical Characterization of Interactions Between O-GlcNAc Transferase and Its Protein Binding Partners

Citation

Hammel, Forrest. 2025. Biochemical Characterization of Interactions Between O-GlcNAc Transferase and Its Protein Binding Partners. Doctoral Dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://dash.harvard.edu/handle/1/42719422

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

THESIS ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Doctoral Program of Chemical Biology

have examined a dissertation entitled

Biochemical Characterization of Interactions Between
O-GlcNAc Transferase and Its Protein Binding Partners

presented by Forrest A. Hammel

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Philip A. Cole

Signature _____

Typed name: Prof. Ralph Mazitschek

Signature _____

Typed name: Prof. Susan Shao

Signature _____

Typed name: Prof. Suzanne Walker

Signature _____

Typed name: Prof.

Date: February 4, 2025

Biochemical Characterization of Interactions Between *O*-GlcNAc Transferase and Its Protein Binding Partners

A dissertation presented

by

Forrest A. Hammel

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

February 2025

© 2025 Forrest A. Hammel

All rights reserved.

Dissertation Advisor: Professor Suzanne Walker Forrest A. Hammel

Biochemical Characterization of Interactions Between *O*-GlcNAc Transferase and Its Protein Binding Partners

Abstract

O-GlcNAc transferase (OGT) is an essential mammalian enzyme that is solely responsible for the post-translational modification of *O*-GlcNAcylation, the addition of *N*-acetyl-D-glucosamine (GlcNAc) onto serine and threonine sidechains on thousands of proteins in the nucleus and cytoplasm. OGT catalyzes another post-translational modification: the cleavage of host cell factor 1 (HCF-1), which it does via *O*-GlcNAcylation of a glutamate residue in one of six highly conserved proteolytic repeats. OGT also engages in a wide variety of protein complexes, facilitated primarily by its superhelical tetratricopeptide repeat (TPR) domain. Previous work from the lab has demonstrated that OGT's *O*-GlcNAcylation and noncatalytic scaffolding functions are essential for cellular proliferation.

In this thesis, I build on previous notions of interactions between OGT and both its protein substrates and non-substrate interactors, defining features that drive these essential functions. Chapter 2 details a series of work performed in collaboration with other members of the Walker lab, studying the effects of mutations and truncations within the TPR domain on OGT's functions and its ability to support cell proliferation; we show that truncation of OGT's N-terminal TPRs impacts viability and all three functions, while preventing OGT homodimerization leads to increased cell growth. Additionally, in that chapter, I show that changes to protein-interacting features in the TPR lumen do not impact all of OGT's substrates equally. In Chapter 3, I harness TR-FRET technology to develop a quantitative binding assay for OGT-substrate interactions. Using that assay, I identify a polypeptide derived from HCF-1's proteolytic repeats, HCF3R, that has picomolar affinity and inhibits OGT in vitro and in cells. I then modify this probe to create a cellular OGT inhibitor that can be switched on and off by addition of two small molecules. In

Chapter 4, I use a combination of biochemical and structural tools to interrogate the nature of the OGT-HCF3R interaction, finding that the medial TPR region (TPRs 6-8) contains features that are essential for the observed phenotypes. Taken together, this work demonstrates methods that can be used to further probe OGT's functions, and insights by which new understanding of OGT's protein-protein interactions can be derived.

Contents

Title Page i
Copyright ii
Abstract iii
Table of Contents v
List of Tables viii
List of Figures ix
Acknowledgments xii
Dedication xv

1 Introduction **1**
1.1 Post-translational modification of proteins in eukaryotes 1
1.2 Eukaryotic protein *O*-GlcNAcylation is catalyzed by *O*-GlcNAc transferase (OGT) . 2
1.3 OGT cleaves the transcriptional coregulator host cell factor 1 (HCF-1) in vertebrates 4
1.4 OGT's structure enables its function as a scaffolding binding partner in protein complexes 6
1.5 The gene encoding OGT is essential and highly conserved in mammals 9
1.6 Bespoke tools to study *O*-GlcNAc transferase *in vitro* and in cells 10
1.6.1 Recombinant OGT expression vectors 11
1.6.2 OGT small molecule inhibitors (OSMIs) 11
1.6.3 Endogenous *OGT* deletion and replacement 13
1.7 Not all of OGT's biochemical functions are essential in mammalian cells 15
1.8 OGT is an attractive therapeutic target 17
1.9 Conclusion 18

2 OGT's tetratricopeptide repeat domain is a key controller of its substrate and interactor selection **19**
2.1 There is no consensus sequence for OGT's glycosylation substrates 20
2.2 The role of homodimerization in OGT's conserved activities and cellular proliferation 21
2.2.1 OGT homodimerizes with a dissociation constant of 39 nM 21
2.2.2 The ability to homodimerize is crucial for glycosylation of some substrates *in vitro* 23

2.2.3 Preventing homodimerization influences OGT's catalytic and noncatalytic functions in cells, but does not negatively affect cellular proliferation 24
2.3 Lumenal residues in OGT's tetratricopeptide repeat (TPR) domain contact protein substrates 29
2.4 The TPR asparagine ladder generally promotes substrate glycosylation *in vitro* and in cells 31
2.5 The TPR aspartate ladder offers a counterbalancing restriction on substrate glycosylation 33
2.6 The TPR macro-structure is important for maintaining OGT's conserved activities and promoting cellular proliferation 35
2.6.1 Truncation of the TPR domain successively reduces OGT's catalytic activity 35
2.6.2 The TPRs drive cellular proliferation and play a role in proper nucleocytoplasmic localization of OGT 38
2.6.3 Truncation of the TPR domain greatly alters OGT's interactome in cells . . . 40
2.7 Global trends of glycosylation and feature-driven OGT regulation are not necessarily recapitulated on a per-substrate level 42
2.8 Conclusion 46
2.9 Methods 48

3 Development of an OGT binding assay and discovery of a polypeptide inhibitor with picomolar affinity 65
3.1 The HCF-1 proteolytic motif is highly conserved across *Chordata* 66
3.2 In mammals, the full HCF-1 proteolytic processing domain is also conserved . . . 69
3.3 HCF-1 cleavage can be monitored *in vitro* using substrates consisting of the first three HCF-1 repeats 71
3.4 Existing tools used to profile OGT-substrate and OGT-ligand interactions lacked quantitative capabilities 72
3.5 Development of a quantitative TR-FRET binding assay to measure the affinity between OGT and its substrates at equilibrium 74
3.6 A single HCF-1 repeat binds OGT with high-nanomolar affinity in the presence of a UDP-containing cofactor 77
3.7 3.7. HCF3R binds wildtype OGT with picomolar affinity in the presence of UDP . . 79
3.8 HCF3R binding and cleavage is highly dependent on interactions with the TPR asparagines closest to the OGT active site 80
3.9 Loss-of-binding OGT asparagine mutants promote glycosylation of HCF3R, but not all OGT substrates 82
3.10 HCF3R is a potent OGT inhibitor 84
3.11 Cellular expression of a degron-tagged HCF3R under an inducible promoter enables rapid inhibition and recovery of OGT activity using two small molecules 85
3.12 (*TetON*)-HCF3R specifically interacts with OGT in HCT116 cells 89

3.13 Conclusion . . . 89
3.14 Methods . . . 92

4 Biochemical and structural dissection of the OGT-HCF3R interaction 106
4.1 Structural solution of proteins by cryo-electron microscopy (cryo-EM) . . . 107
4.2 Advances in OGT structural biology enabled by cryo-EM . . . 108
4.3 Software packages used in cryo-EM data processing and analysis . . . 110
4.3.1 CryoSPARC . . . 110
4.3.2 RELION . . . 111
4.4 Isolation and purification of the OGT-HCF3R-UDP heterotrimer . . . 113
4.5 Initial data processing in CryoSPARC . . . 115
4.5.1 2D class averages reconstitute the OGT homodimer . . . 116
4.5.2 3D structures converged on the OGT homodimer with one well-ordered monomeric catalytic domain . . . 117
4.5.3 Attempts to improve resolution at OGT's N-terminus by 3D classification were unsuccessful . . . 117
4.6 Re-processing using new job types and parameters in CryoSPARC . . . 119
4.7 Motion in the complex hinders structural solution of the OGT N-termini and the second monomer . . . 123
4.8 Local refinement marginally improves density at OGT's N-terminus . . . 125
4.9 Biochemical determination of the OGT-HCF3R binding interface . . . 129
4.9.1 The first two repeats in HCF3R control its picomolar-affinity interaction with OGT . . . 129
4.9.2 OGT's oligomerization state does not significantly affect HCF3R binding . . . 130
4.9.3 AlphaFold2 predicts interactions between OGT and HCF3R in the N-terminal TPRs . . . 132
4.9.4 N-terminal truncations of OGT can still bind and cleave HCF3R . . . 133
4.9.5 The conserved cysteines in HCF3R contribute collectively toward binding OGT . . . 138
4.10 Re-analysis of cryo-EM data to extract meaning in light of biochemical findings . . . 138
4.10.1 Local refinement of the medial TPRs reveals additional density . . . 139
4.10.2 Multibody refinement in RELION improved density of marginal regions . . . 139
4.11 Conclusion . . . 140
4.12 Methods . . . 142

5 Summary and future directions 148
5.1 Summary of findings . . . 148
5.2 Recommendations for technical improvements . . . 150
5.3 Discussion and future scientific directions . . . 151

Appendix A Protein constructs described in this dissertation **154**
A.1 OGT protein constructs 154
A.1.1 OGT bacterial expression constructs 154
A.1.2 OGT mammalian expression constructs 178
A.2 HCF-1 repeat protein constructs 185
A.2.1 HCF-1 repeat bacterial expression constructs 185
A.2.2 HCF-1 repeat mammalian expression constructs 193
A.3 OGT substrate bacterial expression constructs 195

6 References **198**

List of Tables

4.1 Data collection parameters for cryo-EM analysis of the OGT-HCF3R-UDP complex 115

List of Figures

1.1 The sequence hypothesis governs biological processes 2
1.2 Protein *O*-GlcNAc is cycled by two enzymes 3
1.3 The hexosamine biosynthetic pathway produces UDP-GlcNAc from glucose 5
1.4 HCF-1 is cleaved by OGT 5
1.5 Structure of monomeric OGT 7
1.6 Alignment of OGT's TPRs 8
1.7 Structure of dimeric OGT 9
1.8 OGT is highly conserved and essential in mammals 10
1.9 Evolution of OGT Small Molecule Inhibitors (OSMIs) 12
1.10 Strategies to complement the OGT-encoding gene in MEFs 14
1.11 Separation-of-function OGT mutants 16
1.12 Growth curves of MEFs expressing separation-of-function OGT mutants 17

2.1 Sequence alignment of OGT's substrates 21
2.2 OGT homodimerizes with a K_D of 39 nM 22
2.3 Monomeric OGT cleaves and glycosylates HCF3R substrates effectively 23
2.4 Growth curve of Cre-MEFs expressing monomeric OGT 25
2.5 Monomeric OGT is catalytically deficient in Cre-MEFs 25
2.6 Co-immunoprecipitation of monomeric OGT interactors 27
2.7 Expression of monomeric OGT alters the proteome 28
2.8 Residues in OGT's TPR lumen contact substrate proteins 30
2.9 Glycosylation of HeLa S3 cell extracts by OGT-5N5A mutants 32
2.10 Glycosylation of HeLa S3 cell extracts by OGT-D2A mutants 34
2.11 Glycosylation of HeLa S3 cell extracts by TPR truncation OGT mutants 36
2.12 Activity of TPR truncation OGT mutants in Cre- and dTAG-MEFs 37
2.13 Nuclear localization of OGT does not drive proliferation 39
2.14 OGT's N-terminal TPRs control its interactome 41
2.15 Activity of TPR variant OGTs against purified substrates 43
2.16 D2A mutations increase turnover of TAB1 by OGT 45

3.1 Domain structure of HCF-1 homologs 66
3.2 Conservation of the HCF-1 repeat 68
3.3 Conservation of the HCF-1 proteolytic processing domain 70

3.4 HCF3R polypeptides are model substrates for OGT . . . 72
3.5 Development of a TR-FRET binding assay for OGT-ligand interactions . . . 75
3.6 Binding of HCF1R and HCF3R probes **1** and **2** to OGT-WT . . . 78
3.7 Loss of proximal TPR asparagines prevents cleavage and binding of HCF3R . . . 81
3.8 Glycosylation activity of single-asparagine OGT mutants . . . 83
3.9 HCF3R is a potent OGT inhibitor . . . 84
3.10 An on/off switchable HCF3R inhibits OGT in cells . . . 87
3.11 Doxycycline and dTAGV-1 turn HCF3R expression on and off . . . 88
3.12 HCF3R specifically interacts with OGT in cells . . . 90

4.1 OGT structures through the years . . . 109
4.2 Standard data processing pipeline in CryoSPARC . . . 111
4.3 Standard data processing pipeline in RELION . . . 112
4.4 Purification of the OGT-HCF3R-UDP complex . . . 114
4.5 Representative grid image from cryo-EM data collection . . . 115
4.6 2D class averages resemble dimerized OGT . . . 116
4.7 3D reconstruction yielded an OGT homodimer density map . . . 118
4.8 3D classification did not improve density . . . 120
4.9 2D class averages from TOPAZ indicate more sample heterogeneity . . . 121
4.10 Re-processing yielded better density . . . 122
4.11 Reference-based motion correction slightly improved density and resolution . . . 123
4.12 3DFlex reconstituted continuous motion across the particle stack . . . 124
4.13 3D classification using 3DFlex frames did not improve N-terminal density . . . 126
4.14 Local refinement of the catalytic domain improved density and resolution . . . 127
4.15 Local refinement of the N-terminal TPRs improved density and resolution . . . 128
4.16 Competition binding of HCF2R polypeptides to OGT . . . 129
4.17 Affinity of HCF3R truncation polypeptides for OGT . . . 130
4.18 OGT dimerization and HCF3R binding are independent of one another . . . 131
4.19 AlphaFold2 prediction of OGT-HCF3R interactions . . . 133
4.20 AlphaFold2 prediction of OGT-HCF2R-1/2 interactions . . . 134
4.21 Schematic for OGT TPR truncation variants . . . 134
4.22 Binding and cleavage of HCF3R by OGT TPR truncation variants . . . 136
4.23 Binding and cleavage of HCF3R by OGT N-terminal point mutants . . . 137
4.24 HCF3R cysteines contribute collectively towards binding . . . 139
4.25 Local refinement improved density in the medial TPRs . . . 140
4.26 Multibody refinement improved density at distal regions of the complex . . . 141

Acknowledgments

A doctoral degree is not earned in a vacuum, nor is it a solitary effort exerted by one person alone. Many people contributed to this process in many ways, through scientific and personal mentorship, friendship, and companionship.

I joined the Walker lab primarily for two reasons. First, although less important, I enjoyed the science and I could see myself learning about OGT for the length of a full Ph.D. Second, I felt like I belonged there, and that I was in good hands in terms of mentorship and community. To Suzanne, thank you for cultivating that atmosphere. Thank you for challenging me to apply myself to the project, and to continuously grow my ability to conduct experiments and think about the results and what they mean. Thank you for reminding me to consider the question of "why" before all else, and for teaching me that good data means nothing if you can't communicate it effectively.

To my committee members, Phil Cole, Ralph Mazitschek, and Susan Shao, thank you for finding the balance between criticism and praise. Thank you for guiding me along throughout the years as I solved the problems of my project. I greatly appreciate all of the advice you have provided, personally, professionally, and scientifically.

To the many collaborators and scientists who helped facilitate work I conducted during my time here: Connor Payne, Morgan Gilman, Richard Walsh, Megan Mayer, Shaun Rawson, Joao Paulo, Richard Siu, Paula Montero Llopis, Praju Vikas Anekal, and Adrienne Wells. I could not have done this work without you all. Thank you for sharing your expertise with me and helping me to grow to be a better scientist.

To the Microbiology Graduate Student Journal Club, thank you for free lunch on (most) Fridays, and for making me occasionally read articles from outside of my field of study.

To the members of the Walker lab, past and present, thank you for making me want to come to lab, even when my science was failing and my sanity was wearing thin. Thank you for teaching me how to act like a scientist, and how to not act like a fool (most of the time). To the OGTeam, especially Cassie, Sarah, Bettine, and Victoria, thank you for teaching me about OGT, for helping with my experiments, and for enforcing a collaborative environment that I believe was instrumental to my own success. To Cassie and Sarah, thank you for being my mentors, both during my rotation and after I joined the lab. My success is due to the training I received from

you. To Bailey and Eric, thank you for being sounding boards for the finer points of biochemistry and structural biology, especially when it seemed that we were alone in our endeavors in the lab.

To everyone whose instruction and training allowed me to get to Harvard in the first place, thank you. To Professor Steve Rokita, thank you for letting me join your lab as a sophomore, thank you for not giving up on me when I made mistakes, and thank you for encouraging me to think and act like a graduate student as an undergraduate. Thank you for making me apply to Harvard's Chemical Biology program, and thank you for recommending that I talk to Dan and Suzanne during my interview weekend. To Professor Craig Townsend, thank you for letting me get a crash course in enzymology in your lab, even if it was just for six months. To Anton and Dani, thank you for being there for me during difficult times in the lab. To Erica, thank you for holding me to a high standard and teaching me how to think about proteins as a chemist.

To the friends that I met before graduate school and the friends I made while I was here, thank you for getting me through it all. To Brooks and Vik, thank you for the group chat rants, the endless banter, and for being general voices of reason in a chaotic time. To Soham, Nick, Elias, Aaron, Bryan, and David, thank you for always (usually) being down for a late-night gaming session. Those kept my spirits up in many trying times. To my cohort-mates and the associated crew, especially Sebastian, Gerry, Isabel, Sam, Emma, Jenny, thank you for the fun times at retreats, the hiking/ski trips, the late-night JackBox games, and for reminding me that graduate students can be people too (sometimes). To Dayna and Porter, thank you for the conversations and game nights, and thank you for pulling me out of my academic echo chamber.

To HGSU-UAW Local 5118 and the friends and colleagues that I've made through my membership in it, thank you for tirelessly advocating and fighting for student workers' rights. Harvard does not work without you.

To Opal, who is a cat and will sadly never be able to read or understand this. Thank you for being an absolute goober and making my days brighter.

To my siblings, Nick and Liz, thank you for being there for me as we grew up into our own individual selves. To my mom, thank you for the unconditional support you've always shown me. This dissertation is partially your accomplishment, too. To Bob, thank you for being a part of our lives as well. To my extended family, Grandpa Chuck, Grandma Karen (rest in peace), Grandma Pat, Grandma Suzi, and Kent, thank you for cheering me on and supporting me. To the Dittemore

family, thank you for welcoming me into your home and your lives so warmly and graciously.

Finally, to Georgia. Thank you for being my biggest supporter and my best friend. Your presence in my life, and throughout this thesis-writing process, has enriched it so greatly. Thank you for listening to me ramble about science that made no sense, and for cheering me on when things were working. Thank you for listening to my practice talks, even when I thought that I could teach everyone math. Thank you for bringing me into your interests, and for caring about mine. I can't describe how happy and grateful I am to have you in my life.

To those who led me here.

Chapter 1

Introduction

1.1 Post-translational modification of proteins in eukaryotes

The sequence hypothesis offers a very simple view of biological processes: within the cell, DNA sequences encode various genes that are transcribed into messenger RNA strands, which are in turn translated by the ribosome into linear polypeptides comprised of amino acids linked together by amide bonds, which through various biophysically-governed processes adopt secondary and tertiary structures, folding into proteins (Figure 1.1).[1,2] In short, the original DNA sequence for each gene encodes for a specific sequence of amino acids that comprises a specific protein, and all of the downstream information can be (theoretically) determined a priori from the initial sequence. This concept is also referred to as the central dogma of molecular biology, and has been confirmed over decades of studies.[3,4]

This oversimplified model can be elaborated on by RNA splicing to produce alternative gene products,[9] as well as the addition of post-translational modifications (PTMs) to proteins.[10–12] In eukaryotic organisms, a wide variety of PTMs occur, including the addition of small chemical groups such as phosphorylation,[13] methylation,[14] and O-GlcNAcylation;[15–17] covalent addition of small protein tags as seen in polyubiquitination and neddylation;[18–20] other processing steps such as proteolytic cleavage or the formation or breakage of cysteine-cysteine disulfide bonds;[21,22] and numerous other chemical modifications that can occur enzymatically or non-enzymatically. PTMs have a wide range of functions. Some are well-defined and function narrowly – for example, polyubiquitination is almost always a signal that targets proteins for degradation in

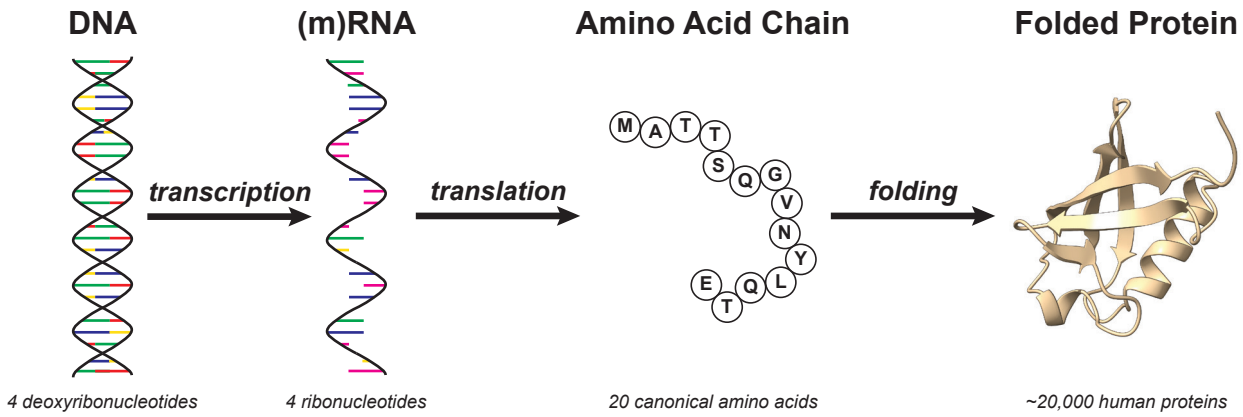


Figure 1.1: The sequence hypothesis governs biological processes. Displayed protein is human ubiquitin (PDB: 1UBQ).[5] Structure visualized using UCSF ChimeraX.[6–8]

the proteasome.[18] Others play more diverse roles, like phosphorylation, which introduces a hydrophilic group with negative charge onto various amino acid sidechains (commonly serine, threonine, histidine, and tyrosine; less frequently, lysine and arginine). This modification can induce steric clashes or shift the hydrophobicity of a region of the target protein, resulting in gross or subtle conformational changes. Moreover, the ease by which the hundreds of unique kinases and phosphatases in the human proteome work to install and remove phosphate groups from proteins makes this process quite dynamic, allowing the cell to use protein phosphorylation to sense and transduce external signals into cellular responses.[23]

1.2 Eukaryotic protein *O*-GlcNAcylation is catalyzed by *O*-GlcNAc transferase (OGT)

This dissertation will focus primarily on the actions of one eukaryotic enzyme conserved in mammals, *O*-GlcNAc transferase (OGT), which catalyzes the attachment of the monosaccharide *N*-acetyl-D-glucosamine (GlcNAc) onto proteins at serine and threonine sidechains, a PTM termed *O*-GlcNAcylation (Figure 1.2).[24–27] This modification is distinct from other mechanisms of protein glycosylation found in the secretory pathway, which decorate proteins with long, branched polysaccharide chains termed glycans.[28] Whereas the secretory pathway glycosyltransferases operate in the endoplasmic reticulum (ER) and Golgi apparatus and act on substrate proteins on

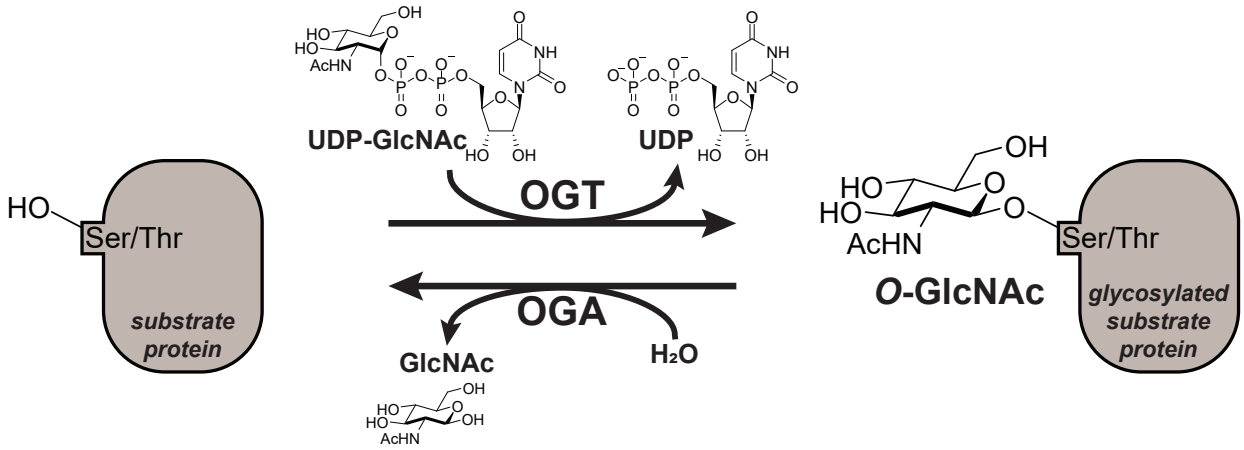


Figure 1.2: Protein *O*-GlcNAc is cycled by two enzymes, *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA). OGT transfers GlcNAc onto substrate serine/threonine residues from UDP-GlcNAc, and OGA removes it by hydrolysis.

the topological "exterior" of plasma membranes (i.e., inside membrane-bound organelles or on the cell surface), OGT acts within the topological "interior" on nuclear and cytoplasmic proteins.[25,26] While there are other eukaryotic glycosyltransferases that act in the nucleus and the cytoplasm, such as the Arabidopsis thaliana *O*-fucosyltransferase SPINDLY,[29] OGT remains the only known glycosyltransferase that is nucleocytoplasmic-localized in mammals.

In contrast with paradigms established for other PTMs, *O*-GlcNAcylation is carried out by one enzyme in eukaryotes – *O*-GlcNAc transferase (OGT) – and its removal is conducted hydrolytically by protein *O*-GlcNAcase (OGA).[30] OGT transfers GlcNAc from the cofactor UDP-GlcNAc onto serine and threonine sidechain hydroxyl groups on its substrate proteins. This reaction is inverting with respect to the sugar's anomeric carbon, and it proceeds via an ordered bi-bi mechanism in which UDP-GlcNAc binds OGT's active site prior to the substrate protein, and following catalysis the *O*-GlcNAcylated protein unbinds before the spent UDP cofactor is released.[31,32]

O-GlcNAc was first identified in the 1980s as a monosaccharide tag distinct from *O*-linked glycans, initially on nuclear pore proteins and then on many others in the nucleus and cytoplasm.[24] In the 1990s, through work primarily in Gerald Hart's group at Johns Hopkins University, OGT was isolated and purified, and its activity was confirmed *in vitro* by detecting glycosylation of a synthetic peptide using radiolabeled UDP-GlcNAc as a cofactor.[26,27] The primary OGT isoform, nuclear/cytoplasmic OGT (ncOGT, hereafter referred to simply as OGT), clocks in at around 115

kilodaltons (kDa), and two other isoforms named mitochondrial (mOGT, ~100 kDa) and short (sOGT, ~75 kDa) have been identified in the literature, however their veracity and function remain unanswered.[33,34]

OGT glycosylates myriad substrates in the mammalian nucleus and cytoplasm – thousands have been identified in the human proteome, with wide-ranging functions including transcription factors, chromatin remodelers, protein kinases, cytoskeletal scaffolding proteins, and nucleoporins, among others.[35] As with phosphorylation, the effects of *O*-GlcNAcylation on its protein substrates are also myriad, and include structural modification,[36] occlusion of substrates for other PTM-installing enzymes,[37] thermal stabilization or destabilization,[38] and recruitment of neo-interactors.[39] Furthermore, because OGT's sugar donor UDP-GlcNAc is the end product of the hexosamine biosynthetic pathway (HBP; Figure 1.3),[40] a highly conserved metabolic pathway whose function has been tied to numerous cancers, some have postulated that *O*-GlcNAcylation acts as a means for cells to sequester or store excess sugar as *O*-GlcNAc, which could then be rapidly recovered by OGA's hydrolase activity.[41] While these claims are not without merit, the roles of *O*-GlcNAc clearly extend beyond serving as a sugar sink, as many groups have demonstrated the specific impacts of *O*-GlcNAcylation in a variety of conditions on a variety of protein substrates.

1.3 OGT cleaves the transcriptional coregulator host cell factor 1 (HCF-1) in vertebrates

Host cell factor 1 (HCF-1) is a mammalian protein that was initially identified as part of a complex with the transcription factor Oct-1 and the herpes simplex virus (HSV) protein VP-16.[42,43] Together, this complex activates transcription of early genes during HSV infection. Since this discovery, HCF-1 has been identified as a key transcriptional coregulator in mammals, where it has been shown to associate with chromatin to regulate cell cycle progression.[44]

Mammalian HCF-1 is a ~2000-residue protein consisting of an N-terminal beta-propeller Kelch domain and a series of C-terminal fibronectin-like repeats, connected by lengthy basic and acidic regions that flank a central core of six highly conserved 26-residue repeats (Figure 1.4A). Within each of these repeats, OGT glycosylates a conserved glutamate residue (E10), producing a

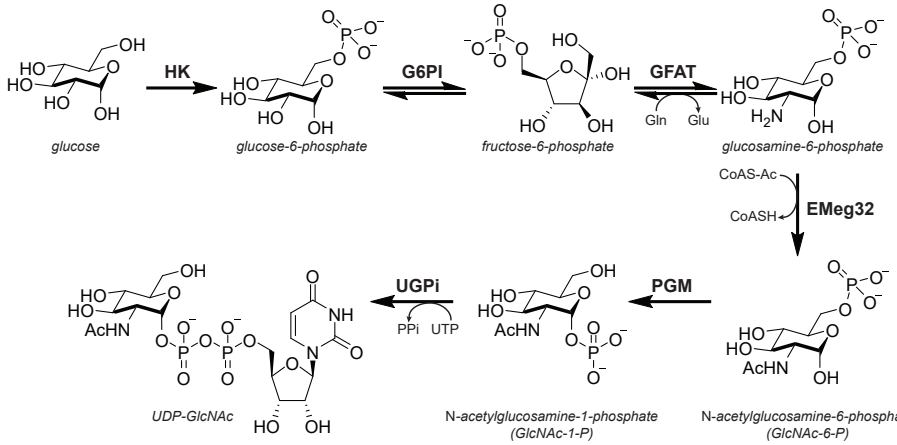


Figure 1.3: In the hexosamine biosynthetic pathway, glucose is phosphorylated to glucose-6-phosphate by hexokinase (HK), which is then isomerized by glucose-6-phosphate isomerase (G6PI) into fructose-6-phosphate. Fructose-6-phosphate is converted to glucosamine-6-phosphate by glutamine:fructose-6-phosphate aminotransferase (GFAT), which is the first committed step in the HBP. Glucosamine-6-phosphate is first acetylated by EMeg32 using the cofactor acetyl-CoA as a donor, forming *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P), then isomerized to *N*-acetylglucosamine-1-phosphate (GlcNAc-1-P) by phospho-GlcNAc mutase (PGM). Finally, the end product UDP-GlcNAc is produced from GlcNAc-1-P and uridine triphosphate (UTP) by UDP-GlcNAc pyrophosphorylase (UGPi)

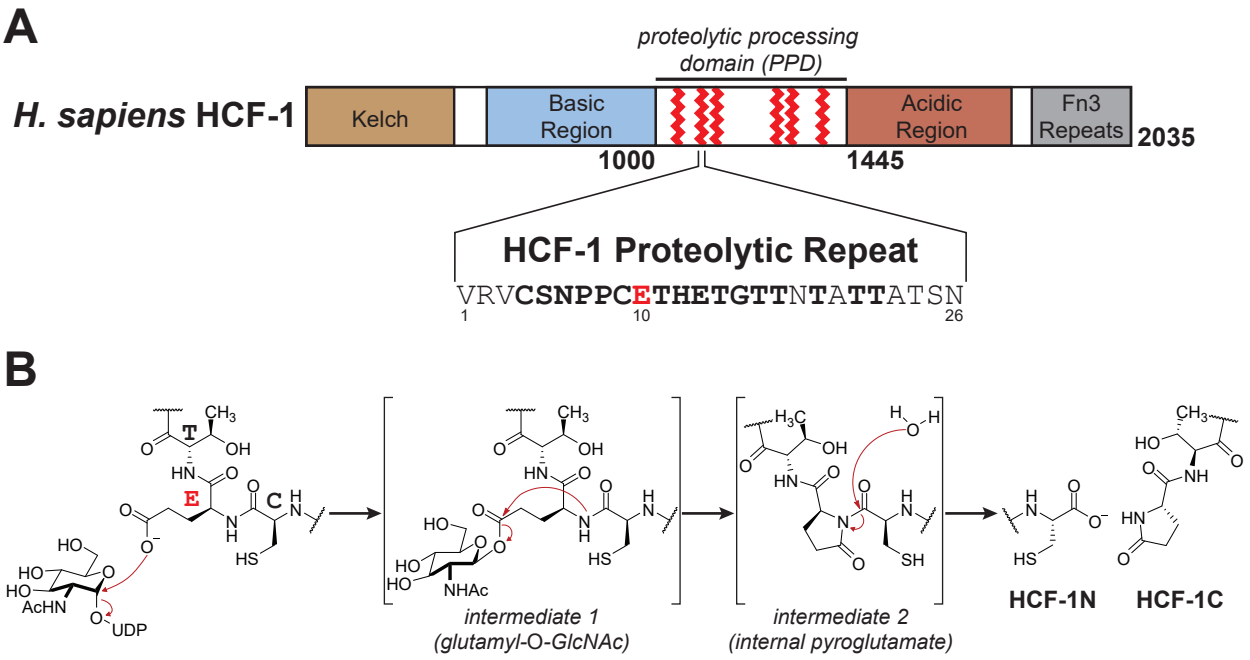


Figure 1.4: HCF-1 is cleaved by OGT. (A) Domain structure of HCF-1, with the sequence for the conserved proteolytic repeat (red zig-zags) shown. Residues conserved across all repeats are bolded, and the glycosite glutamate is shown in red. (B) Mechanism for OGT-mediated HCF-1 cleavage.[45]

glutamyl-*O*-GlcNAc species that cyclizes into a five-membered pyroglutamate ring on the peptide strand due to the stability of *O*-GlcNAc as a leaving group. This pyroglutamatyl peptide is then released from the enzyme, where it can be hydrolyzed into two pieces, resulting in proteolytic maturation of HCF-1 (Figure 1.4B).[45,46] This process can be repeated at all of the repeats within the proteolytic processing domain (PPD), resulting in net cleavage of the 200 kilodalton full-length HCF-1 into a set of 100-150 kilodalton matured polypeptides and 5-50 kilodalton cleavage fragments, depending on how many cleavage reactions occurred.[47,48]

The discovery that HCF-1 is cleaved by OGT came in multiple stages. Knowledge that HCF-1 was proteolytically processed within its PPD existed in the 1990s,[48] and by the early 2000s it was well understood that the N- and C-terminal HCF-1 fragments remained associated with each other following cleavage,[49,50] although it took until 2012 for structural studies to elucidate the mechanism by which this self-association occurs.[51] In 2003, Winship Herr's group found that HCF-1 depletion resulted in cell cycle arrest in the G_1 phase and induced cytokinesis defects marked by multinucleate cells.[52] The cytokinesis defects were not rescued by complementation with an HCF-1 that lacked the PPD, showing that cleavage of HCF-1 does play a specific cellular role. Later work from both the Herr and Walker labs demonstrated that HCF-1 was specifically cleaved by OGT in mammals via recognition of the proteolytic repeats and glycosylation at the conserved E10 glutamate.[46,53]

To date, no other proteolytic substrates of OGT have been reported, although it is theoretically possible that more exist in the proteome. This is likely due to the nature of how HCF-1's cleavage by OGT was identified and the lack of top-down searches for new OGT cleavage substrates in the literature.

1.4 OGT's structure enables its function as a scaffolding binding partner in protein complexes

Structurally, OGT is composed of a C-terminal catalytic domain and an N-terminal tetratricopeptide repeat (TPR) domain (Figure 1.5). Its catalytic domain contains two lobes that each form a Rossmann fold, separated by an intervening domain (IntD) of unknown function.[32] The two lobes come together to form OGT's active site in a typical GT-B glycosyltransferase motif, which binds

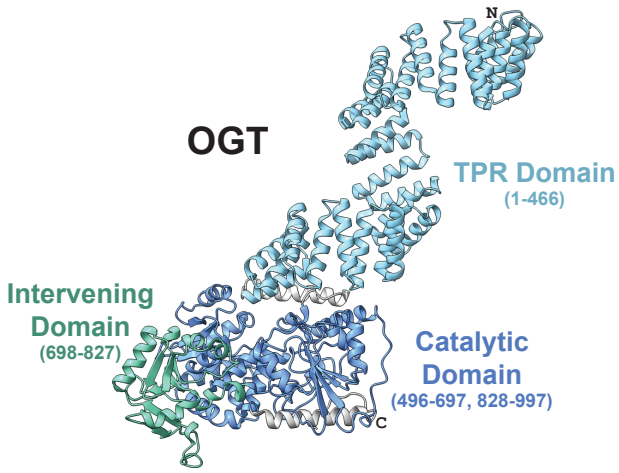


Figure 1.5: Structure of monomeric OGT. OGT consists of an N-terminal TPR domain (light blue, res. 1-466) and a C-terminal catalytic domain with two Rossmann-folded lobes (blue, res. 496-697/828-997) connected by an intervening domain of unknown function (teal, res. 698-827). Helices connecting domains are shaded in grey. PDB: 7YEH.[57] Structure visualized using UCSF ChimeraX.[6–8]

UDP-GlcNAc and the substrate protein at its glycosite.[54] The biochemical and cellular functions of IntD are largely unknown, however recent work out of both the Vocadlo and Jiang groups identified short peptides that bind IntD and inhibit OGT's glycosyltransferase activity, a surprising find given that the active site does not directly interface with IntD.[55,56]

OGT's TPR domain extends out from the catalytic domain and consists of thirteen 34-residue helix-turn-helix motifs that form a right-handed superhelix, which creates a tunnel-like lumen that extends out from OGT's active site.[54,58,59] Beyond the canonically conserved residues that define the TPR fold, OGT's TPR domain possesses additional points of conservation, particularly of aspartates at position 2, asparagines at position 6 and 10, and lysines at position 12 (Figure 1.6); notably, these additional conserved positions are all oriented facing the TPR lumen, and thus could contact substrate proteins exiting the active site.[58] The asparagine and aspartate ladders are of particular interest in understanding OGT's protein-protein interactions,[60–62] and will be discussed further in depth in Chapter 2.

Through a parallel and antisymmetric interface on the outside of its sixth and seventh TPRs, OGT is also capable of homodimerization (Figure 1.7). This was first identified in a crystal structure of the 11 most N-terminal TPRs solved in the early 2000s,[58] but has since been independently confirmed by cryo-electron microscopy structures of full-length OGT alone and in complex with a

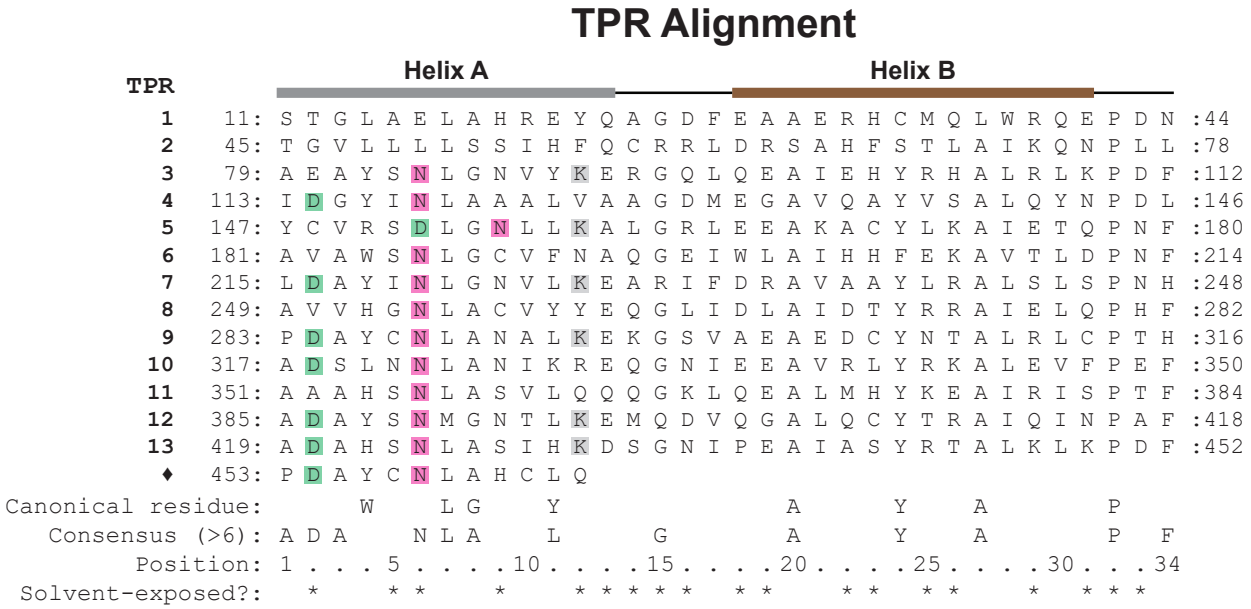


Figure 1.6: Alignment of OGT's thirteen TPRs (diamond indicates additional TPR-like residues after the 13th TPR). Consensus residues (in >6 TPRs), canonical residues, and whether the residue is solvent-exposed are indicated. Figure reproduced from Jinek, et al. (2004).[58]

substrate protein.[57,59] It remains unclear whether OGT homodimerizes in cells, and if so, what proportion of cellular OGT exists in a dimeric state. Furthermore, prior to the work described in this thesis (see Chapter 2.2), it was unclear whether the capacity for OGT homodimerization was essential for cell proliferation.

OGT also plays a role in many well-conserved protein complexes in the cell, particularly in the nucleus.[63] These are facilitated primarily by its TPR domain. Canonically, TPR domains in other proteins enforce protein-protein interactions including in host defense and transcriptional processes,[64] and OGT's TPRs are no different. For example, through interactions with proteins such as the TET family of DNA demethylases or ZNF263,[65–67] OGT associates with chromatin, recruiting various other transcriptional regulation components and driving transcription of specific genes through both stable complex formation and *O*-GlcNAcylation.

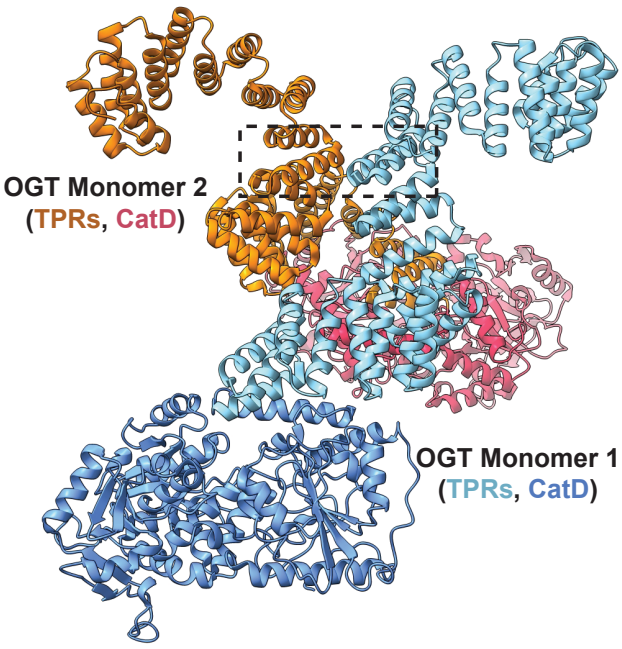


Figure 1.7: Structure of dimeric OGT. OGT dimerizes at an interface on the outside of its 6th and 7th TPRs (black dashed box). PDB: 7YEH[57] Structure visualized using UCSF ChimeraX.[6–8]

1.5 The gene encoding OGT is essential and highly conserved in mammals

Among glycosyltransferase enzymes encoded in the human genome, OGT is the most conserved (Figure 1.8A).[68,69] As such, occurrence of mutations is rare in the population. Additionally, OGT's genetic locus is on the X-chromosome, so loss-of-function mutations are typically only enriched in individuals lacking a second copy of the X-chromosome. While rare, the majority of documented OGT missense mutations occur in its TPR domain, with the majority of clinical outcomes linked to X-linked intellectual deficiency and presentation of other so-called OGT-CDG (Congenital Disorder of Glycosylation) phenotypes.[70–74]

Per data curated by the Cancer Dependency Map Project (DepMap) at the Broad Institute, 1174 out of 1178 (99.7%) tested human cell lines were ranked as dependent on presence of the OGT-encoding gene for proliferation (Figure 1.8B).[75,76] This is consistent with knowledge that loss of OGT induces lethality in early stages of development in mice and in cells; knockout of the *Ogt* allele did not result in viable embryonic stem cells in cultured lines, and all attempts to produce *Ogt*-null mice failed.[77,78] This phenomenon is not conserved across all organisms

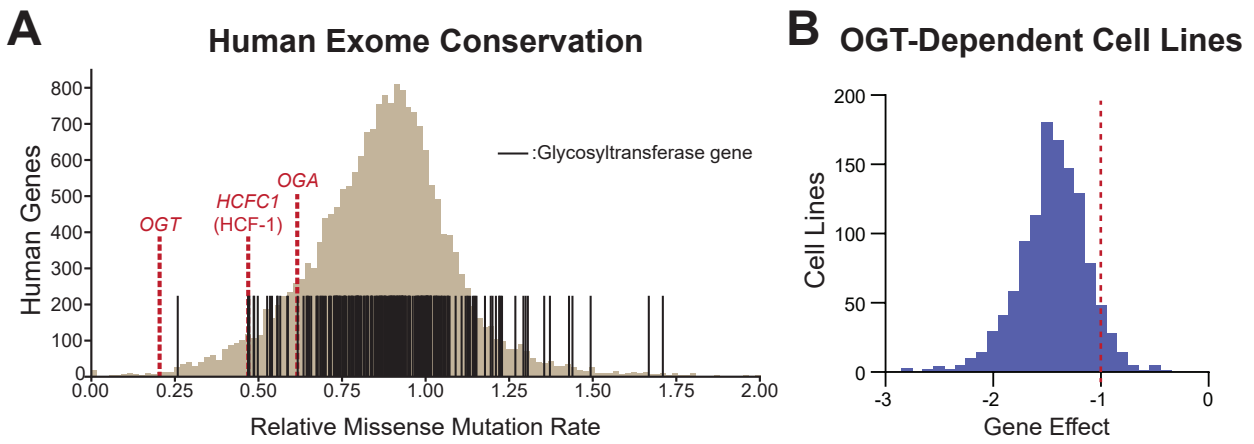


Figure 1.8: OGT is highly conserved and essential in mammals. (A) Genome-wide amino acid conservation data shows that out of all human genes encoding glycosyltransferases, *OGT* is the most conserved.[68] Genes encoding OGA and HCF-1 (*HCFC1*) are also notated. Black lines indicate glycosyltransferase genes. Figure reproduced from Levine, et al. (2021).[69] (B) CRISPR knockout studies curated by DepMap show that 1174/1178 tested cell lines are dependent on *OGT* for proliferation.[75,76] Red dashed line is set at -1, the average gene effect for all essential genes; a more negative gene effect indicates greater essentiality.

with OGT homologues. Knockout of the *C. elegans* OGT homologue (OGT-1) does not affect its development, although it does result in perturbations in insulin signaling, innate immunity, and stress response pathways.[79–82] However, in *Drosophila melanogaster*, OGT retains some of the essentiality observed in mammals. The *Drosophila* OGT homologue, *sxc* (*super sex combs*) is a member of the polycomb family,[83,84] which regulates transcription of genes through chromatin remodeling. Flies without a functional *sxc* allele die before the larval stage and do not complete embryogenesis; however, supplementation of *sxc*-null embryos with functional maternal OGT enables survival through the pupal stage.[85]

Notably, OGT is not strictly essential in post-mitotic, differentiated mammalian cells such as cardiomyocytes and neurons.[86,87] While removal of OGT in both of these cell types leads to physiological defects and increased rates of apoptosis, they are still capable of survival without a functional copy of OGT.

1.6 Bespoke tools to study *O*-GlcNAc transferase *in vitro* and in cells

In addition to a wealth of field-standard tools, including specific antibodies against OGT and protein *O*-GlcNAc, and commercially available assay kits to kinetically characterize *O*-GlcNAcylation,

the Walker lab has developed and utilized various novel and custom approaches to profiling OGT, both *in vitro* and in cells. Except for those developed specifically as a part of this dissertation and discussed in later chapters, these will be described here.

1.6.1 Recombinant OGT expression vectors

Prior to 2005, existing methods for OGT purification from mammalian tissue,[26] or recombinantly in mammalian cells,[88,89] baculovirus,[90] and *E. coli* were poor;[91] protein yields were too low to enable high-throughput screening or structural biology efforts. As part of an effort in the early 2000s to identify small molecule inhibitors of OGT via a ligand displacement fluorescence polarization (FP) assay, the Walker lab cloned an *E. coli* codon-optimized gene encoding human ncOGT into a pET24b vector with a C-terminal polyhistidine affinity tag for purification.[92] Later on, the corresponding N-terminal polyhistidine-tagged variant was constructed; this version served as the foundation for all of the OGT bacterial expression constructs used in this thesis.

1.6.2 OGT small molecule inhibitors (OSMIs)

The FP assay developed by Dr. Benjamin Gross during his time in the Walker lab identified a set of quinoline-6-sulfonamide (Q6S) compounds with mid-micromolar IC_{50}s against sOGT (Figure 1.9A).[92–94] To elaborate on the Q6S moiety, a biased library of Q6S compounds was screened to identify preferred building blocks that would later be incorporated into the first of the OSMI (**O**GT **S**mall **M**olecule **I**nhibitor) series of compounds, OSMI-1 (Figure 1.9BC).[94] This work was heavily aided by Craig Thomas and Damien Duveau at the NIH, who carried out an extensive medicinal chemistry campaign on the scaffold to improve its inhibitory properties. Later, the crystal structure of OGT bound to an OSMI-1 analog with the furan group replaced by a formate was solved, demonstrating that the Q6S group mimics the binding of uridine in the active site, occluding both the substrate cofactor and peptide.[95] This structure demonstrated that addition of a small group at the quinoline position *ortho* to the sulfonamide could improve binding in the active site; addition of a chlorine at this position yielded OSMI-4a, which binds OGT with low nanomolar affinity ($\sim$8 nM by microscale thermophoresis; methyl ester OSMI-4 shown in Figure 1.9F).

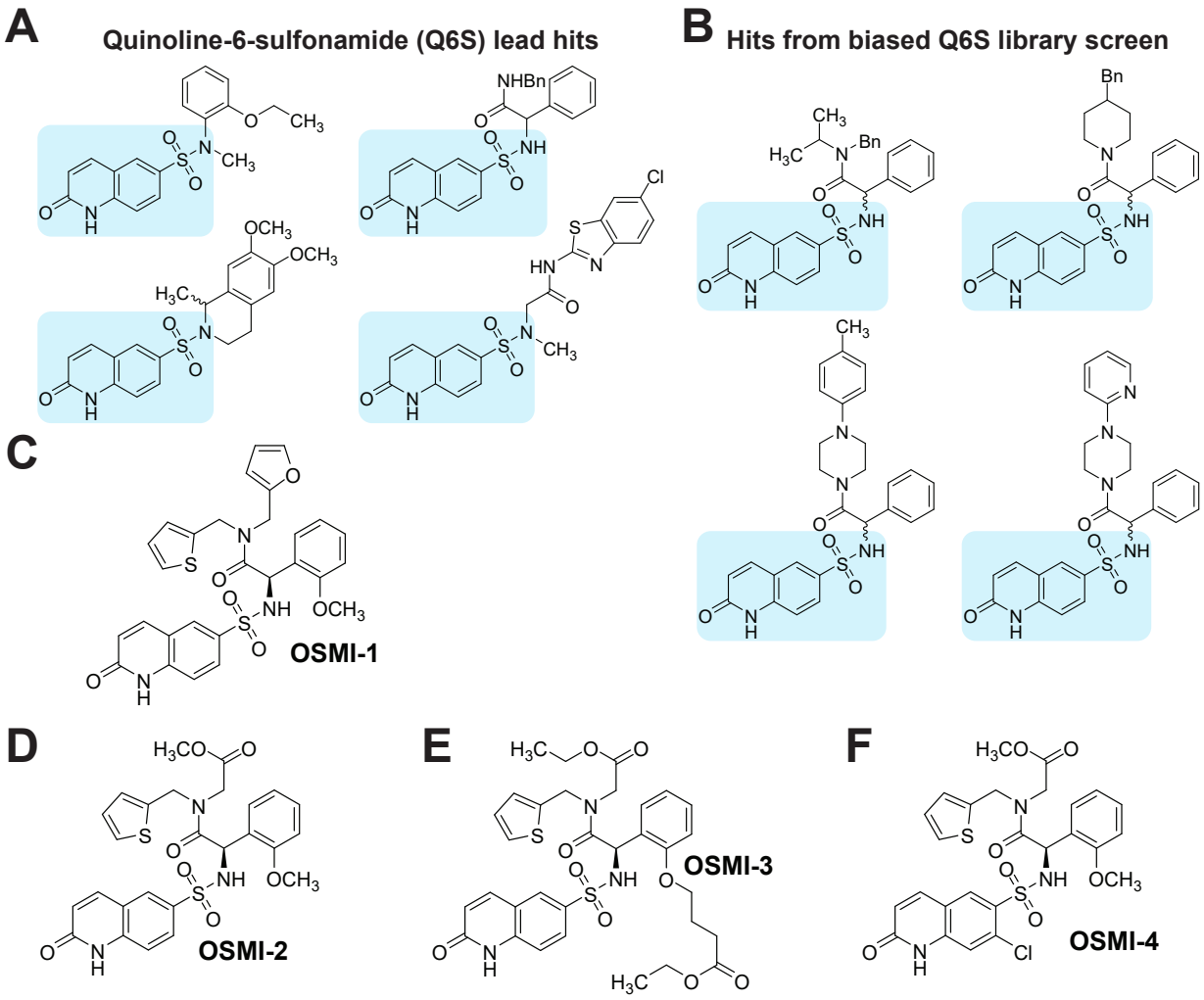


Figure 1.9: Evolution of OGT Small Molecule Inhibitors (OSMIs). (A) Hits from initial FP small molecule library screen containing quinoline-6-sulfonamide (Q6S) moiety, shaded in blue.[92] (B) Secondary hits from screening of a biased library containing Q6S compounds.[94] (C) Chemical structures of cell-permeable analogs OSMI-1, (D) OSMI-2, (E) OSMI-3, and (F) OSMI-4.[95]

Despite the efficacy of OSMI-4 *in vitro*, it suffers from poor cellular permeability. While OSMI-4a had a K_D around 8 nM *in vitro*, it took 5 μM of its ethyl ester derivative OSMI-4b to inhibit OGT in HEK293T cells.[95] Consistent with this notion, we typically use OSMI-4b at concentrations between 10 and 25 μM in cells.

1.6.3 Endogenous *OGT* deletion and replacement

Because OGT is ubiquitous and essential in mammalian tissues and its deletion hinders embryonic viability, a means had to be developed to induce removal of the endogenous copy of the gene after development to study functional variants in cells or *in vivo*. This was done by Natasha Zachara and coworkers in the 2000s,[78,96] building on previous work from Gerald Hart's group.[77] Briefly, transgenic mouse lineages were developed with LoxP sites on either side of exon 6 of *Ogt* in a background where an estrogen receptor-Cre recombinase fusion was also expressed (Figure 1.10, top left).[77,78] By addition of the small molecule 4-hydroxytamoxifen (4-HT), the ER-Cre fusion translocates to the nucleus where it excises the exon between the LoxP sites, resulting in production of a nonsensical transcript and functional deletion of the *Ogt* gene (Figure 1.10, top right). In mice, this system was used to directly show that dividing mammalian cells require a functional copy of OGT. Additionally, through harvesting and immortalization of embryonic fibroblasts (MEFs) pre-knockout, stable cell lines were generated for cellular studies.[96]

In the late 2010s, the Walker lab received these MEFs from the Zachara lab. Using lentiviral vectors, previous graduate student Dr. Zebulon Levine developed two orthogonal cellular systems to replace the endogenous OGT with functional variants, the Cre-MEFs and the dTAG-MEFs.[69] In the Cre-MEF system, lentivirus was used to introduce an ectopic *OGT* transgene fused to a fluorophore protein (mKate2) and an affinity tag (2xFLAG) at its N-terminus, and cells were sorted for expression based on fluorescence. Later, the endogenous *Ogt* could be knocked out by addition of 4-HT over a period of 2-3 days, generating cells expressing only the ectopic OGT variant (Figure 1.10, top). The dTAG-MEF system modified this by first lentivirally incorporating an ectopic wildtype *OGT* transgene fused to the degron tag $FKBP12^{F36V}$ (Figure 1.10, bottom left), enabling its rapid degradation upon addition of a bifunctional proteolysis-targeting chimera (PROTAC) molecule via polyubiquitination and recruitment to the proteasome. Dr. Levine sorted

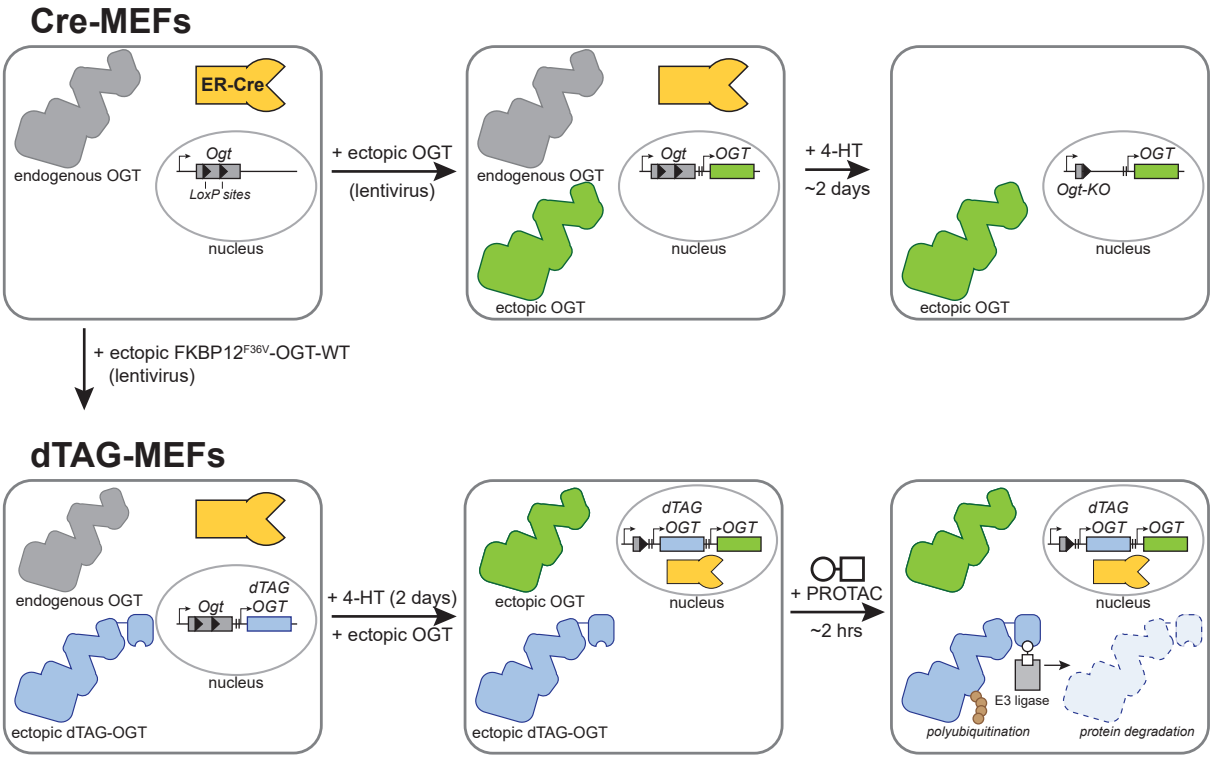


Figure 1.10: Strategies to complement the OGT-encoding gene in MEFs. In the Cre-MEF system (top), ectopic OGTs are introduced to MEFs engineered with LoxP sites flanking a key exon in the *Ogt* gene and expressing an ER-Cre recombinase fusion protein. Upon addition of the small molecule 4-hydroxytamoxifen (4-HT), ER-Cre translocates to the nucleus and excises the *Ogt* gene in between the LoxP sites, effectively knocking it out over a period of 2 days. The dTAG-MEF system (bottom) changes this by first introducing an ectopic wildtype OGT fused to the degron FKBP12^{F36V}. Then, endogenous *Ogt* is knocked out, and the complementary ectopic OGT of interest is introduced. The degron-tagged OGT-WT can be degraded efficiently by introduction of a small-molecule PROTAC degrader, which recruits the E3 ubiquitin ligase Cereblon to FKBP12^{F36V}-OGT, polyubiquitinating it and marking it for proteasomal degradation.

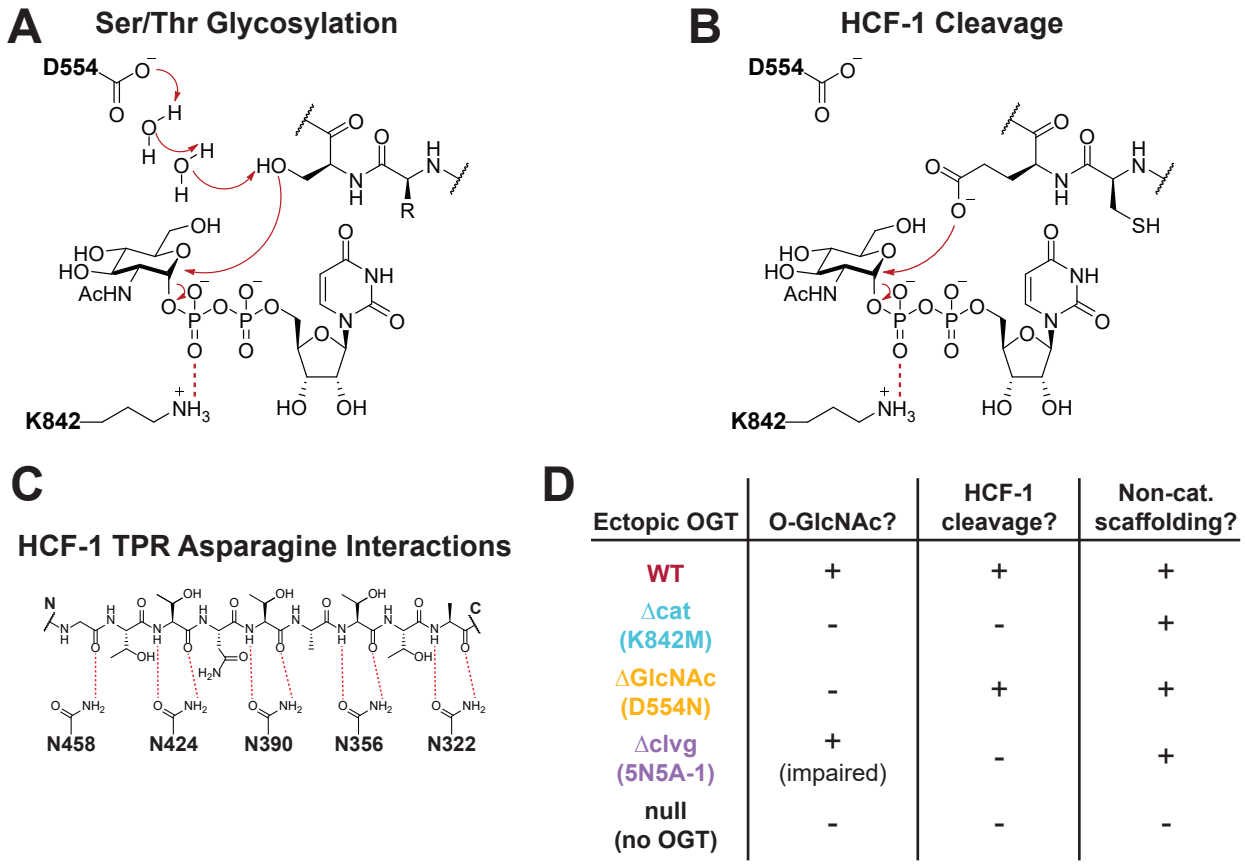
these cells into clonal populations, selected a clone that had endogenous-like OGT expression, and knocked out the endogenous copy using 4-HT. The resultant dTAG-MEFs could then be infected with mKate2-2xFLAG-OGT transgenes using lentivirus as before (Figure 1.10, bottom).

Importantly, in both the Cre- and dTAG-MEF systems, ectopic OGTs are inserted randomly into the genome and may be inserted multiply. Moreover, loss of ectopic OGT in the Cre-MEF system leads to cell death, while degradation of the pseudo-ectopic dTAG-OGT in the dTAG-MEF system only leads to growth inhibition instead of death, making these two systems useful in different cases to monitor essential functions of OGT and how they impact cellular proliferation.

1.7 Not all of OGT's biochemical functions are essential in mammalian cells

Over a series of biochemical and structural studies, the Walker lab established a series of OGT mutants lacking each of OGT's conserved catalytic functions (Figure 1.11). Through dissection of the structure of the substrate-bound active site, the mutant OGT-K842M (OGT-Δcat) was identified as catalytically inactive, because it is unable to bind the cofactor UDP-GlcNAc.[54] Similarly, the mutant OGT-D554N (OGT-ΔGlcNAc) was developed to be inactive towards serine/threonine glycosylation, but still be able to cleave HCF-1, a feat achieved by noticing that the wildtype aspartate D554 participates in proton abstraction from the glycosite hydroxyl to prime it for attack at the GlcNAc anomeric carbon, a process that is unnecessary for HCF-1 cleavage because the glycosite is already deprotonated.[45,46,54] Likewise, a mutant that could carry out serine/threonine glycosylation but not HCF-1 cleavage was developed by mutating the five TPR asparagine ladder residues closest to the active site to alanines; this mutant was named OGT-5N5A-1 (OGT-Δclvg).[60]

Dr. Zebulon Levine, with the assistance of Dr. Sarah Potter, another former Walker lab graduate student, incorporated mKate2-2xFLAG fusions of each of these OGT variants into both the Cre- and dTAG-MEF systems, and also generated 'null' cell lines with no ectopic OGT copy and 'wildtype' cell lines expressing ectopic mKate2-2xFLAG-OGT-WT.[69] After validating by western blotting that the separation-of-function mutants performed properly, they conducted growth curves of all of the Cre-MEFs after knockout of the endogenous *Ogt* (Figure 1.12A), and dTAG-MEFs after degradation of the pseudo-endogenous FKBP12^{F36V}-OGT (Figure 1.12B). In



Ectopic OGT	O-GlcNAc?	HCF-1 cleavage?	Non-cat. scaffolding?
WT	+	+	+
Δcat (K842M)	-	-	+
ΔGlcNAc (D554N)	-	+	+
Δclvg (5N5A-1)	+ (impaired)	-	+
null (no OGT)	-	-	-

Figure 1.11: Separation-of-function OGT variants. (A) The mechanism for serine/threonine *O*-GlcNAcylation requires K842 to stabilize the UDP β-phosphate and D554 to initiate deprotonation of the glycosite sidechain. (B) HCF-1 cleavage only requires K842, not D554, due to the glutamate sidechain being deprotonated at physiological pH. (C) HCF-1 cleavage is dependent on a series of bidentate interactions with asparagine sidechains in the TPRs. (D) Summary of OGT separation-of-function mutants.[69]

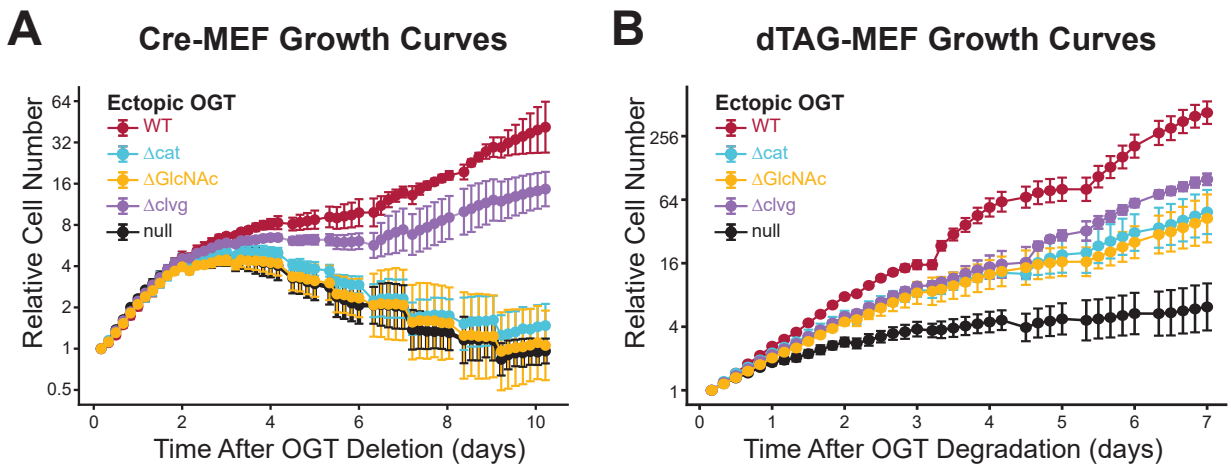


Figure 1.12: Growth curves of MEFs expressing separation-of-function OGT mutants. (A) In Cre-MEFs, serine/threonine glycosylation was an essential function. (B) In dTAG-MEFs, noncatalytic functions of OGT boosted cellular proliferation. Data were obtained by Incucyte live cell imaging. Figures reproduced from Levine, et al. (2021).[69]

the Cre-MEFs, they found that loss of serine/threonine *O*-GlcNAcylation, either through the Δcat or ΔGlcNAc variant, led to cell death along the same timeframe as loss of OGT entirely, while loss of HCF-1 cleavage alone (Δclvg) resulted in cell growth, albeit at a lower rate than cells complemented with OGT-WT.[69] This suggested that OGT's serine/threonine glycosylation functions were strictly essential. However, in the dTAG-MEF system, where the 'null' cells did not die and instead underwent growth arrest, a different pattern emerged: complementation with any OGT variant, including OGT-Δcat, resulted in an intermediate growth phenotype, showing that some noncatalytic functions of OGT drive cell proliferation.

1.8 OGT is an attractive therapeutic target

As mentioned at various points in this chapter, OGT's activity is intrinsically tied to cellular metabolism and homeostasis, both through its direct cellular roles as well as the fact that UDP-GlcNAc is the end-product of the hexosamine biosynthetic pathway. While clinical OGT mutations are rare, dysregulation of *O*-GlcNAcylation has been implicated in many disease phenotypes, and indeed may play a causative role in conditions like diabetes and cardiovascular disease.[97–99] In particular, many aggressive cancers such as pancreatic and prostate cancer are characterized by high *O*-GlcNAc,[100–105] and OGT inhibition has been viewed as a potential therapeutic avenue. To

contrast this, neurodegenerative disorders are typically characterized by having low *O*-GlcNAc levels, which mechanistically could lead to increased protein aggregation characteristic of these conditions.[87,106,107]

1.9 Conclusion

O-GlcNAc transferase is an enigmatic protein. Rarely does one find in the proteome a single enzyme that participates in so many unique, conserved activities – both catalytic and noncatalytic. Indeed, this makes studying OGT both challenging and rewarding, and an enticing prospect. Many carefully crafted global studies have elucidated OGT's broad cellular functions, and many more granular studies have determined more specific biochemical mechanisms that OGT participates in. However, throughout the years leading up to the start of my doctoral research, many questions remained surrounding the question of how OGT maintains activity against such a large repertoire of glycosylation substrates while still enforcing specificity in its substrate selection.

In this thesis, I hope to move towards answering these questions through a variety of means. First, I will discuss work performed in collaboration with Dr. Sarah Potter, a former graduate student, and Dr. Cassandra (Cassie) Joiner, a former postdoctoral fellow in the lab, describing the roles that different features in OGT's TPR domain play in substrate and interactor recognition. In the following chapter, I describe the development of a quantitative binding assay for OGT's substrates, work I performed with the assistance of Professor Ralph Mazitschek's lab and one of his former graduate students, Dr. Connor Payne. In that chapter, I also detail the striking finding that catalytically inactive mutants of HCF-1's PPD bind and inhibit OGT with picomolar affinity. In the third and final data chapter, I employ cryo-electron microscopy and the binding assay developed in the second chapter to dissect the nature of the interaction between OGT and the HCF-1 PPD.

Chapter 2

OGT's tetratricopeptide repeat domain is a key controller of its substrate and interactor selection

Some of the work described in this chapter has been published in two papers on which I am a co-author, and one paper on which I am the first author:

Potter, S.C.; Gibbs, B.E.; Hammel, F.A.; Joiner, C.M.; Paulo, J.A.; Janetzko, J.; Levine, Z.G.; Fei, G.Q.; Haggarty, S.J.; Walker, S. "Dissecting OGT's TPR domain to identify determinants of cellular function." *Proc. Natl Acad. Sci. USA* **2024**, *121* (22):e2401729121.[108]

Joiner, C.M.; Hammel, F.A.; Janetzko, J.; Walker, S. "Protein Substrates Engage the Lumen of *O*-GlcNAc Transferase's Tetratricopeptide Repeat Domain in Different Ways." *Biochemistry* **2021**, *60* (11):847-853.[62]

Hammel, F.A.; Payne, N.C.; Marando, V.M.; Mazitschek, R.; Walker, S. "Identification of a Polypeptide Inhibitor of *O*-GlcNAc Transferase with Picomolar Affinity." *Journal of the American Chemical Society* **2024**, *146* (38):26320-26330.[109]

The first data chapter of this dissertation centers on the growing notion that features of OGT's TPR domain contribute to its ability to glycosylate substrates and bind interactors. While this concept was not new when I joined the lab, it was still in its infancy and much characterization

was needed to flesh out our understanding. Here, I describe work led primarily by two close collaborators and mentors of mine in the Walker lab, former graduate student Dr. Sarah Potter and former postdoctoral fellow Dr. Cassandra (Cassie) Joiner, on which I assisted by performing *in vitro* purification and experimentation while gaining familiarity with the tools available at my disposal to study OGT. I am greatly indebted to both of them for their willingness to take me on as a rotation student and to continue mentoring me after I joined the lab. I additionally describe work on monomeric OGT that Dr. Potter and I conducted jointly with Dr. Victoria Marando, another postdoctoral fellow in the lab.

As alluded to above, this chapter contains multiple sections. In the first part, I describe the role of OGT homodimerization in promoting OGT's catalytic and noncatalytic functions in cells. In the second part, with Dr. Cassie Joiner and Dr. Sarah Potter, I first characterize the global activity of TPR mutants of OGT, then probe the functions of OGT's N-terminal TPRs in OGT's essential cellular activities. And finally, in the last part, working with Dr. Joiner I characterize the role of TPR features on glycosylation of purified substrates *in vitro*, finding that global trends do not always recapitulate on a per-substrate basis.

2.1 There is no consensus sequence for OGT's glycosylation substrates

Many PTM-installing enzymes utilize consensus sequences for facile recognition of their substrates. These consensus sequences bind key regions on the enzyme or its cofactors, facilitating formation of ternary complexes that, in turn, lead to enzymatic catalysis and post-translational substrate modification.

Early work in the OGT field sought to uncover a similar pattern in OGT's substrates, but no strict consensus sequence was found. The field then pivoted to identifying preferential residues by position, and in 2007, Carolyn Bertozzi's group reported a high-throughput ELISA method using single-residue substitutions of a substrate peptide derived from α-A crystallin.[110] They observed loss of activity upon replacement of residues at the -1, and +1 positions relative to the glycosite, and striking gain of activity ($\tilde{5}$-fold increase) with alanine or proline replacement of the glutamate residue at +2. This gave the first hint at any sequence preference for OGT substrate recognition.

Later efforts to characterize OGT's substrate scope and sequence preferences from the

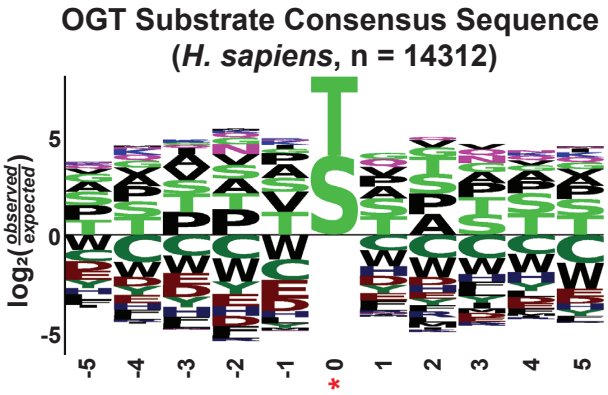


Figure 2.1: Sequence alignment of OGT's substrates. Data obtained from the *O*-GlcNAc Database.[35,111]

Walker,[60,61] Woo,[112] Van Aalten,[113] and Zachara groups,[114] among others, utilized protein microarrays and chemoproteomics approaches, detecting or otherwise enriching for *O*-GlcNAcylated proteins in complex mixtures and plotting amino acid frequencies relative to glycosites. In all of these studies and subsequent meta-analyses,[35] a few patterns emerged: strong preference for prolines at the -2 and/or -3 positions, enrichment of proline and alanines C-terminal to the glycosite, and enrichment of serine and threonine residues upstream and downstream from the glycosite (Figure 2.1). These findings hinted at the uniqueness of OGT's TPR domain in facilitating a broad range of protein-protein interactions, which the Walker lab began exploring in earnest over a decade ago.

2.2 The role of homodimerization in OGT's conserved activities and cellular proliferation

2.2.1 OGT homodimerizes with a dissociation constant of 39 nM

As mentioned previously (see Section 1.4), OGT homodimerizes via a set of symmetric and antiparallel hydrogen-bonding residues on the external face of its 6th and 7th TPRs (Figure 2.2A, 2.2B). This feature was first identified in a crystal structure of OGT's TPR domain deposited in 2004,[58] then later confirmed by unbound and substrate-bound cryo-EM structures published by multiple groups.[57,59] These structural studies additionally identified key mutants that disrupt dimerization, from alanine mutants of key hydrogen-bonding residues (*e.g.*, L199A, H202A)

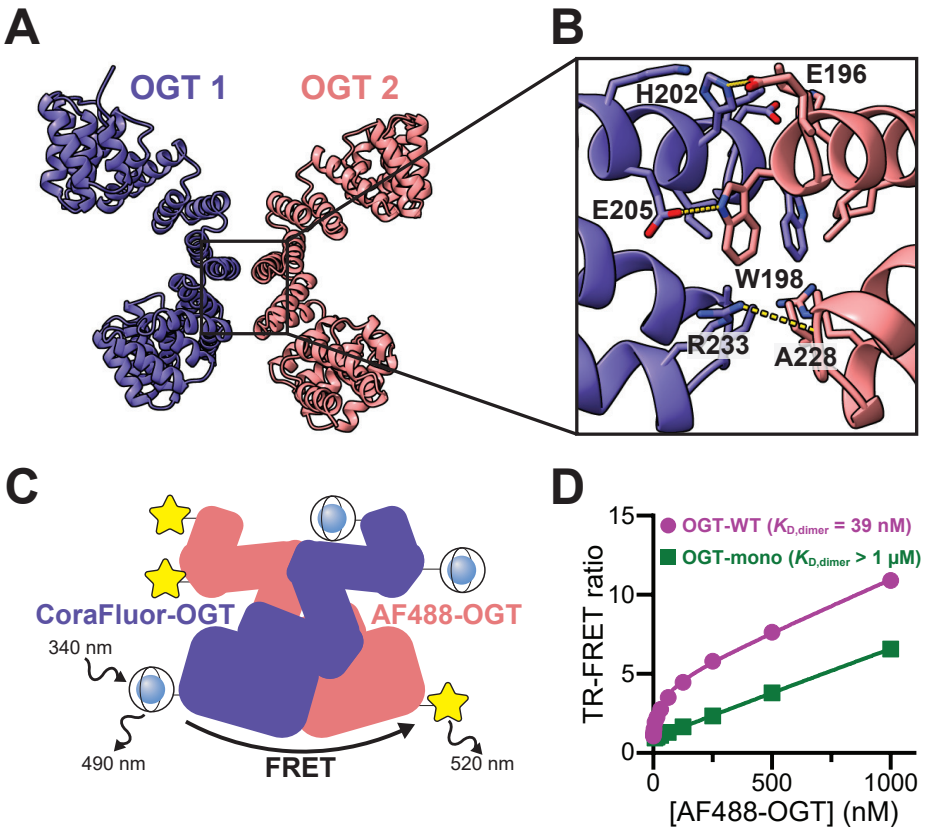


Figure 2.2: OGT homodimerizes with a K_D of 39 nM. (A) Structure of dimerized TPR domains from OGT. Black box is shown in inset (B) demonstrating symmetric and antiparallel interactions at the interface. PDB:1W3B.[58] (C) Schematic for an OGT dimerization binding assay. (D) Uncorrected binding curves show that OGT-WT (purple circles) dimerizes with a 39 nM $K_{D,dimer}$, while OGT-mono (green squares) does not dimerize within the assay parameters. Data are mean ± s.d of two replicates. Structures were visualized using UCSF ChimeraX.[6–8]

to targeted mutations that introduce ionic repulsion in the dimerization interface such as the W198E/I201D double mutant, which we have named OGT-mono because multiple studies have demonstrated that it exists solely in the monomeric state.

While previous studies used size exclusion chromatography with multi-angle light scattering (SEC-MALS) to determine the oligomeric state of OGT-WT and various dimerization-domain mutants,[59] they lacked quantitation that would help determine whether OGT exists in a primarily dimeric or monomeric state in cells or in our *in vitro* glycosylation reactions. Using a time-resolved FRET platform that is elaborated upon further in section 3.5, I developed a simple assay to read out binding of acceptor-tagged OGT-WT to donor-tagged OGT-WT using the TR-FRET donor/acceptor pair of CoraFluor-1 and Alexa Fluor 488 (Figure 2.2C). With this assay, I determined that OGT-WT

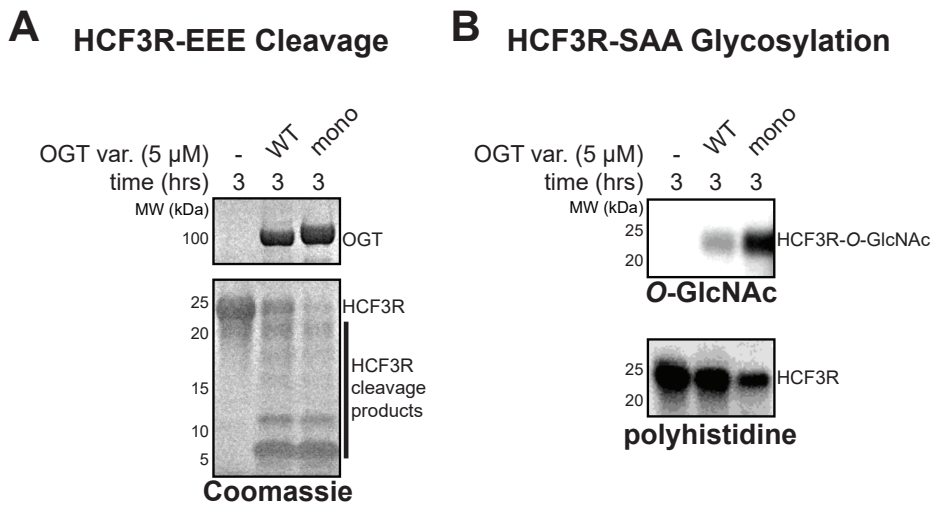


Figure 2.3: Monomeric OGT cleaves and glycosylates HCF3R substrates effectively. (A) Cleavage of HCF3R-EEE by OGT-WT and OGT-mono, visualized by SDS-PAGE and Coomassie staining. (B) Glycosylation of HCF3R-SAA by OGT-WT and OGT-mono, visualized by SDS-PAGE and western blotting.

has a dimerization dissociation constant ($K_{D,dimer}$) of 39 nM, while OGT-mono (W198E/I201D) has a $K_{D,dimer}$ greater than 1 μM, the highest concentration of OGT tested (Figure 2.2D). The wildtype $K_{D,dimer}$ of 39 nM suggests that cellular OGT, whose concentration has been estimated at 100 nM and likely varies by cell type and state,[115] primarily exists as a homodimer but that there is likely a small monomeric population that may be relevant to OGT's biology. More work will need to be done in the future to characterize OGT's oligomeric state in cells.

2.2.2 The ability to homodimerize is crucial for glycosylation of some substrates *in vitro*

After identifying that certain OGT mutants (L199A, H202A, W198E/I201D) existed solely in the monomeric form, Gideon Davies, et al. assayed their ability to glycosylate two substrates *in vitro*: the serine glycosite mutant of a single HCF-1 proteolytic repeat, HCF1R-E10S, and the full-length protein TAB1, whose glycosylation kinetics and glycosites had previously been characterized in the literature.[59,116] They found that while the rate of HCF1R glycosylation was unaffected by any of the mutations, the rate of TAB1 glycosylation was negatively affected by all mutants that abolished OGT homodimerization. This suggested that engagement and glycosylation of TAB1 are in some way dependent on the presence of multiple OGT monomers, however no specific

mechanistic rationale was provided. Later work in this chapter (see Section 2.7) demonstrates that TAB1 glycosylation is dependent on contacts in the N-terminal TPR region near the dimerization interface, while previous work showed that HCF1R does not contact that part of the TPRs, which may explain the observed dependence on dimerization. A more recent study confirmed that monomeric OGT does not glycosylate TAB1 and additionally showed that the substrate OGA, which engages the full TPRs, is dependent on OGT dimerization for its glycosylation.[57]

I was curious if dimerization would affect OGT's catalytic activity against HCF3R substrates derived from the first three proteolytic repeats in HCF-1. With both the wildtype cleavable HCF3R-EEE substrate (Figure 2.3A), and the glycosylation substrate HCF3R-SAA (Figure 2.3B), removal of OGT's ability to dimerize does not affect its catalytic activity towards HCF3R. In fact, OGT-mono appeared to be more effective at catalyzing both reactions compared to OGT-WT. Taken together with the previous work in the literature, this demonstrates that OGT homodimerization is not strictly essential for all of OGT's conserved functions, and much like other features of OGT, its impact must be assessed on a more granular basis.

2.2.3 Preventing homodimerization influences OGT's catalytic and noncatalytic functions in cells, but does not negatively affect cellular proliferation

We wanted to know more about the role that OGT dimerization plays in the cell – if it is essential, and how it affects OGT's cellular functions. Dr. Sarah Potter, a former graduate student in the Walker lab, incorporated an ectopic FLAG-tagged copy of OGT-mono into both Cre-MEFs and dTAG-MEFs (see Section 1.6.3 for an overview of the Cre-MEF and dTAG-MEF systems). By treating these cells with 4-hydroxytamoxifen alongside cells with no ectopic OGT, and cells with an ectopic OGT-WT copy, then tracking growth via live cell imaging over a period of one week, she showed that OGT-mono facilitates cellular growth at greater-than-wildtype levels compared to OGT-WT and no OGT, which induces cell death (Figure 2.4). Thus, we reasoned that whatever role OGT homodimerization plays, it is not one that is required for cell proliferation. However, given the observed increase in cell number compared with wildtype, dimerization might play a role in maintaining normal levels of growth.

Using the established OGT-mono Cre-MEF cell line, we then assessed OGT-mono's expression

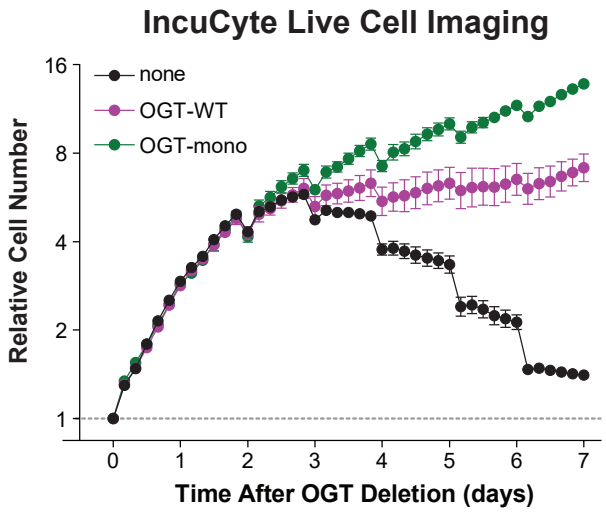


Figure 2.4: Growth curve of Cre-MEFs expressing monomeric OGT. Data are from Incucyte live cell imaging.

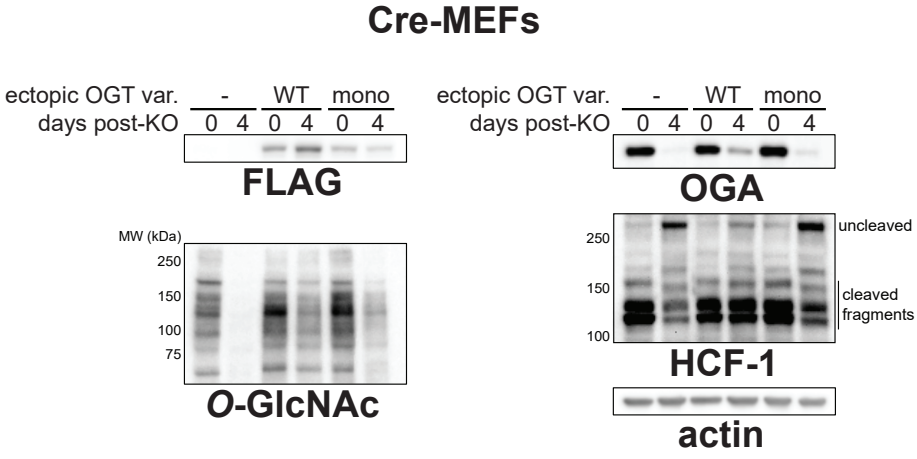


Figure 2.5: Monomeric OGT is catalytically deficient in Cre-MEFs, both in terms of *O*-GlcNAcylation and HCF-1 cleavage. Cells expressing no ectopic OGT, OGT-WT, or OGT-mono were lysed 0 or 4 days after knockout of the endogenous *Ogt*. Lysates were assessed by SDS-PAGE and western blotting.

level and activity by western blot at 0- and 4-days post-knockout of the endogenous *Ogt* gene (Figure 2.5). We found that in comparison to OGT-WT Cre-MEFs, expression of OGT-mono was much lower despite a similar population being collected during cell sorting in construction of the two stable cell lines. Consistent with this notion, we noticed that the overall level of *O*-GlcNAcylation was much lower, with bands at 100 and 75 kilodaltons (kDa) appearing more diminished compared to the corresponding OGT-WT lane 4-days post-knockout. We also noticed that HCF-1 cleavage was greatly reduced compared to WT, with accumulation of uncleaved HCF-1 reaching levels similar to that of the 'null' Cre-MEFs expressing no OGT at day 4.

While reduction in *O*-GlcNAc levels was consistent with lower OGT expression in the OGT-mono Cre-MEFs, the accumulation of full-length HCF-1 suggested that OGT-mono was in some way not facilitating HCF-1 cleavage to the same extent as OGT-WT (Figure 2.5, HCF-1 blot). Because we knew that OGT-mono was capable of robust cleavage of HCF-1 substrates *in vitro* (Figure 2.3A), we asked if this phenotype in cells was due to a change in OGT's interactome caused by preventing OGT homodimerization. With Dr. Victoria Marando, a postdoctoral fellow in the lab, I performed co-immunoprecipitation of FLAG-tagged OGT-WT and OGT-mono from their corresponding dTAG-MEF cell lines following 8 hours of treatment with the degrader compound dTAG-13. We followed this treatment regimen from Dr. Sarah Potter's protocols, used with great effect to study the effects of TPR truncation on OGT's interactome (see Section 2.6.2), as it prevents accumulation of proteome-wide changes in protein expression levels while eliminating confounding effects from the endogenous OGT-WT copy.[108] Dr. Marando and I prepared four replicates for each cell line and submitted FLAG-enriched elution samples to Dr. Joao Paulo, an Instructor in the Gygi Lab, who performed the trypsinization, cleanup, and label-free proteomics. Using the resultant data, we computed Pearson correlation coefficients for each dataset and excluded the most distant replicate for each cell line in further processing (WT 3, mono 3, null 1; Figure 2.6A). We then plotted the enrichment over null and significance for each identified protein in the remaining OGT-WT samples and used this to set cutoffs of $\geq$4-fold enrichment and p-value $\leq$0.05 (Figure 2.6B); these cutoffs yielded a hit list for OGT-WT that contained primarily known OGT interactors such as the Tet and Trak family proteins.[65,117] We performed the same analysis for OGT-mono and noticed that while many known OGT interactors such as HCF-1, Trak2,[118] Kansl3,[119,120] and Tet3 remained as hits,[66] many other interactors present in

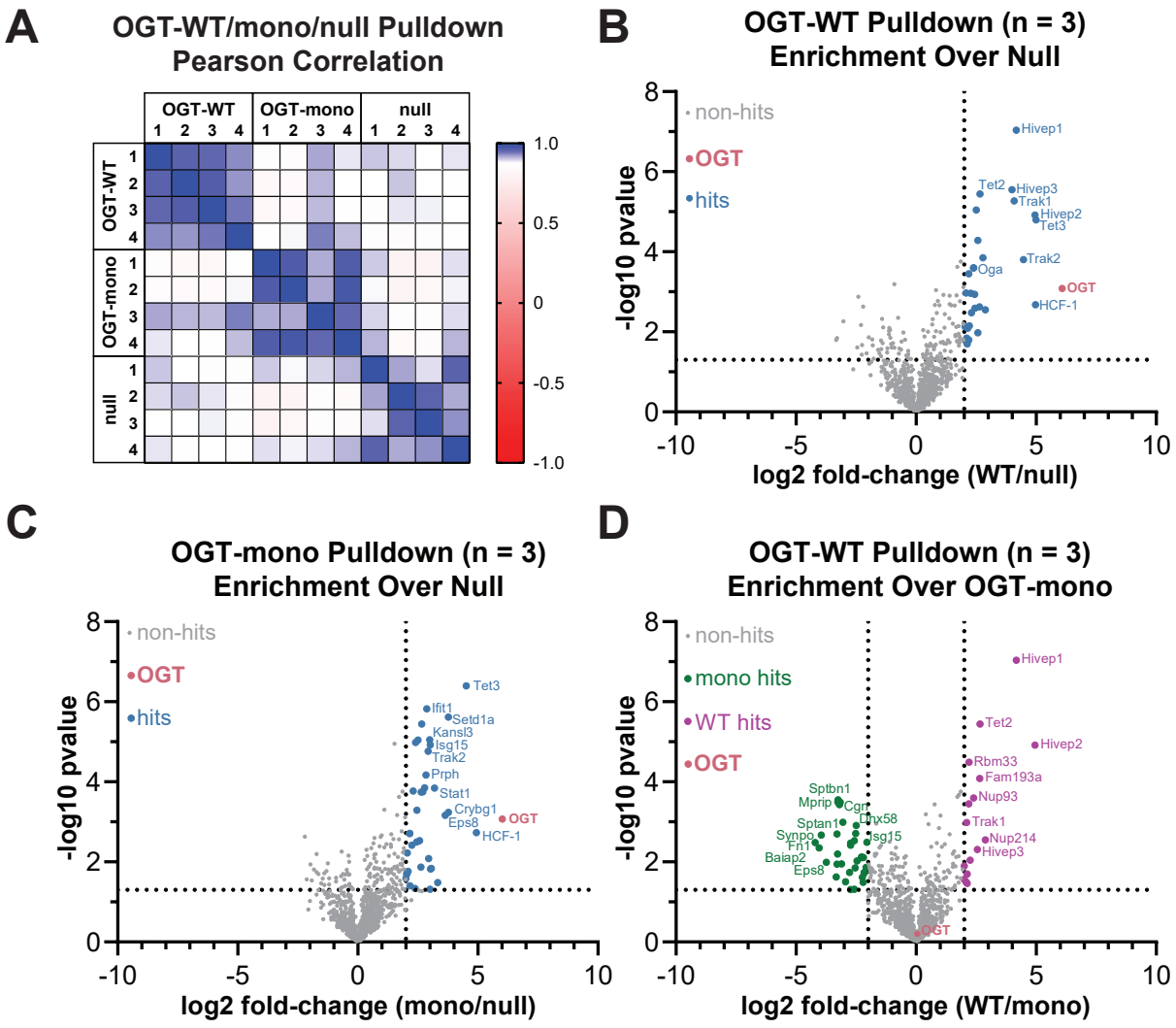


Figure 2.6: Co-immunoprecipitation of monomeric OGT interactors. (A) Pearson correlation of data from all four replicate datasets for each cell line. (B,C,D) Volcano plots showing enrichment and significance of co-immunoprecipitated proteins from (B) OGT-WT over null, (C) OGT-mono over null, and (D) OGT-WT over OGT-mono. Data are averaged from triplicate datasets. In (B) and (C), blue circles indicate hits and red circles indicate the OGT bait protein. In (D), purple circles indicate OGT-WT-specific hits, while green circles indicate OGT-mono-specific hits.

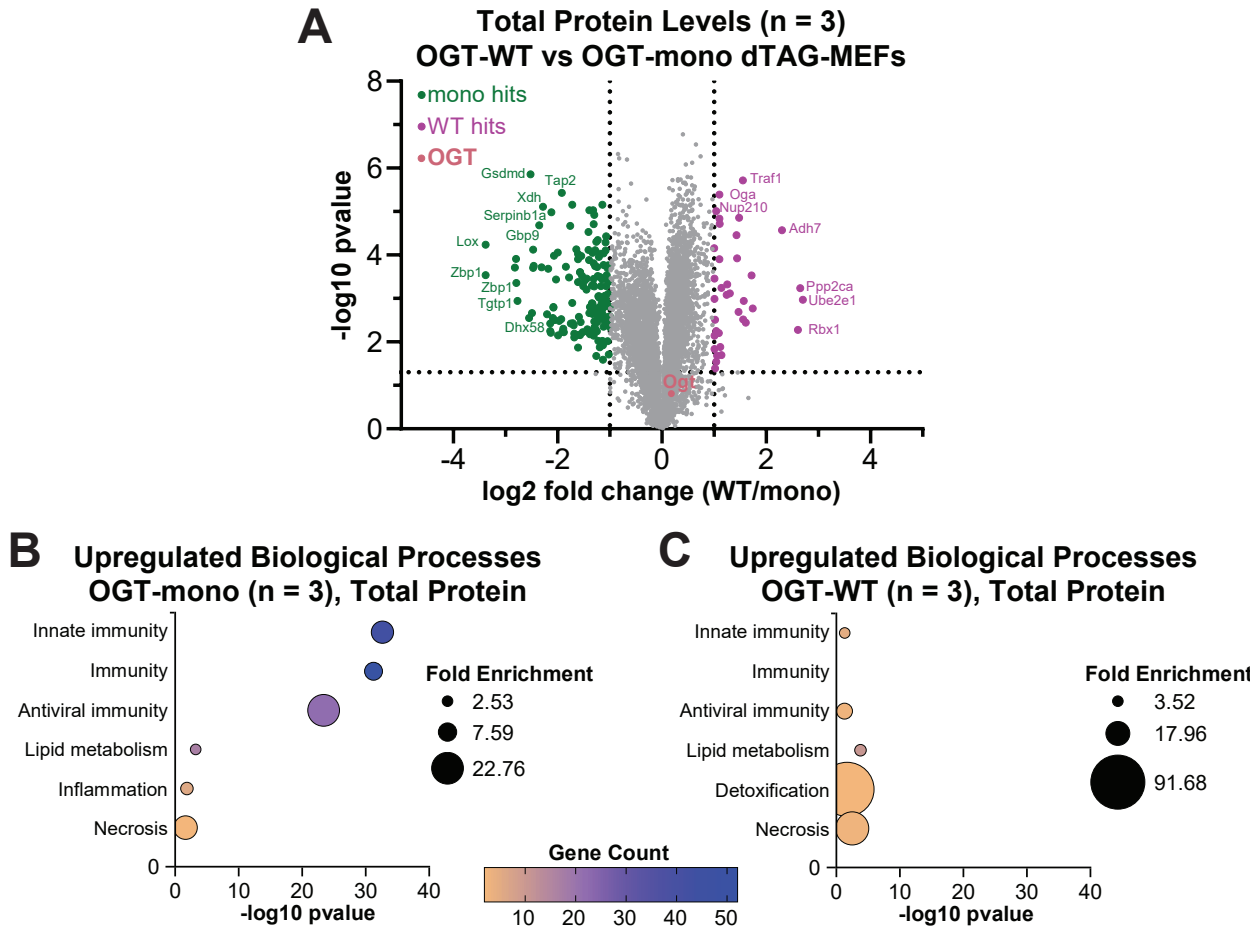


Figure 2.7: Expression of monomeric OGT alters the proteome. (A) Volcano plot showing changes in total protein levels from dTAG-MEFs over three replicates. WT-specific hits are shown in purple, and mono-specific hits are shown in green. Hit cutoffs were set at 2-fold enrichment and p-value < 0.05. (B,C) Bubble plots showing upregulation of proteins involved in immune-related biological processes in (B) OGT-mono and (C) OGT-WT dTAG-MEFs.

the OGT-WT hit list disappeared (Figure 2.6C). Furthermore, a number of new interactors from immune signaling pathways appeared, including STAT1, ISG15, and IFIT1. We also looked at the subcellular compartments the hits were found in and noticed that, while hits for OGT-WT were primarily nuclear, cytoplasmic, or mitochondrial, hits for OGT-mono were excluded from both the nucleus and the mitochondria, instead clustering in the cytoplasm with cytoskeletal and plasma membrane elements.

The emergence of immune signaling interactors for OGT-mono was corroborated by tandem mass tagged mass spectrometry (TMT), also performed by Dr. Joao Paulo, which was used to determine if any proteins were differentially expressed across our experiment. We generated a

volcano plot of the resultant quantitative proteomics data and observed, using cutoffs of 2-fold enrichment and a p-value of 0.05, a far greater number of upregulated proteins in the OGT-mono samples compared to the OGT-WT samples (Figure 2.7A). We then took the hits for both samples and performed DAVID analysis, which provides information on functional annotations for proteins, and found that the upregulated proteins in OGT-mono cells were strongly associated with immune signaling pathways, including innate immunity and antiviral immunity (Figure 2.7B). In contrast, the same was not observed with OGT-WT (Figure 2.7C), demonstrating that this was not a result of the general manipulation steps performed on all of our cell lines (infection, sorting, selection, dTAG-13 treatment, fusion to mKate2-2xFLAG, etc.).

Overall, these results provide key information on the role of OGT's homodimerization interface. While it is not essential for all cellular roles that OGT plays (*e.g.*, cell proliferation) it does appear to influence OGT's substrate repertoire as well as its interactome. The specific role of immune signaling proteins in the OGT-mono interactome is of great interest to us moving forward, as OGT has been implicated in proper immune system function, and *O*-GlcNAc dysregulation appears to be immunomodulatory.[121,122]

2.3 Lumenal residues in OGT's tetratricopeptide repeat (TPR) domain contact protein substrates

OGT's TPR domain, which consists of thirteen full 34-residue tetratricopeptide repeats in a right-handed superhelix (Figure 2.8AB), forms a central lumen, or cavity, extending out from OGT's active site.[54,58] To facilitate binding and promote glycosylation of OGT's substrates, a series of sidechains extend into the TPR lumen (Figure 2.8C). These can generally be separated into three broad classes: a ladder of asparagine residues that each occupy a conserved position within the TPR they are in (Figure 2.8, pink residues), a ladder of aspartate residues that each occupy a conserved position within the TPR they are in (Figure 2.8, green residues), and other assorted residues that are not part of any conserved ladders. The asparagine ladder is canonically found at position 6 within the TPRs, and is present on all TPRs except for 1, 2, and 5. Sidechains from this ladder protrude into the TPR lumen and canonically make sequence-nonspecific bidentate contacts with the peptide backbone of substrate proteins.[46,54,60] This is best illustrated by the

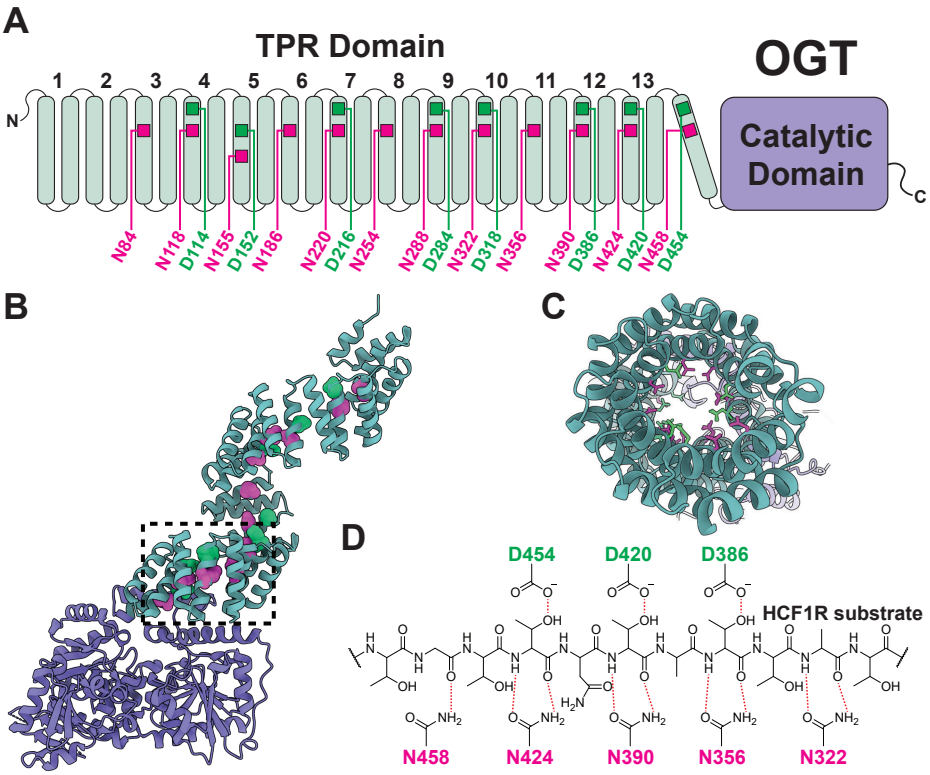


Figure 2.8: Residues in OGT's TPR lumen contact substrate proteins. (A) Schematic of OGT, with residues in the conserved TPR asparagine ladder (pink) and aspartate ladder (green) labeled. (B) Structure of OGT, with asparagines (pink) and aspartates (green) indicated. Black box indicates region shown in D. PDB: 7YEH.[57] (C) Top-down view of the TPR lumen, with asparagines and aspartates shown as atoms. (D) Contacts between the TPR asparagines/aspartates and HCF1R.[45] Structures were visualized with UCSF ChimeraX.[6–8]

substrate peptide HCF1R, which forms bidentate hydrogen bonds with asparagines N322, N356, N390, and N424, as well as a single hydrogen bond with N458 (Figure 2.8D, pink residues). The aspartate ladder, found at the second position within the TPRs (in repeats 4, 7, 9, 10, 12, 13), can form sequence-specific hydrogen bonds between its sidechains and serine or threonine sidechain hydroxyls on substrate proteins, as they do with the threonines in HCF1R (Figure 2.8D, green residues).[61] In addition to these two ladders, other residues extend into the TPR lumen and can facilitate protein-protein interactions with substrates, including non-ladder asparagines N233, N321, and N393; non-ladder aspartate 431; K396; Q362; and residues in the N-terminal TPRs such as H29.[57]

2.4 The TPR asparagine ladder generally promotes substrate glycosylation *in vitro* and in cells

Previously, the Walker lab has investigated the role of the TPR asparagines in reactions with purified proteins and protein microarrays on the benchtop. To do this, an OGT-5N5A-1 mutant, so named because it contained alanine mutations to five sequential asparagines (N322/N356/N390/N424/N458A), was constructed (Figure 2.9A). This OGT-5N5A-1 had impaired glycosylation activity on the vast majority of protein microarray glycosylation hits (736 out of 739), as well as the serine mutant of the HCF-1 proteolytic repeat, HCF1R-E10S.[60] OGT-5N5A-1 was also incorporated into both the Cre-MEF and dTAG-MEF cells by a former graduate student in the lab, Dr. Zebulon Levine.[69] In these cell lines, it was observed that this OGT mutant had much lower *O*-GlcNAcylation activity (although still detectable) but was incapable of cleaving HCF-1, thus it was used as a "Δcleavage" construct.[69] From these results, it was reasoned that the proximal TPR asparagines promote substrate binding and are an indispensable anchor for OGT-substrate interactions that lead to protein *O*-GlcNAcylation and HCF-1 cleavage.

With a postdoctoral fellow in the lab, Dr. Cassie Joiner, I worked on expression and purification of two other multi-asparagine ladder mutants to span the TPR asparagine ladder and supplement OGT-5N5A-1, which only contained mutations of the asparagines closest to OGT's active site. OGT-5N5A-2 contained mutations in five asparagines in the central TPRs (N186/220/254/288/322A) and OGT-5N5A-1contained mutations in the five most N-terminal ladder asparagines (N84/118/155/186/220A; Figure 2.9A). The size exclusion chromatography profiles of these OGTs resembled that of OGT-WT, suggesting that substitution of multiple asparagine residues with alanines does not subtantially impact OGT's three-dimensional conformation.[60,62]

We tested these mutants first in HeLa S3 cell extracts, which are metabolically active cell lysates capable of ATP regeneration that are suitable proxies for global substrate protein pools.[123] We dosed purified OGT variants into extracts at 1 μM for 1, 5, or 15 minutes before quenching and assessing glycosylation by western blot (Figure 2.9B). We found that all three 5N5A mutants had reduced *O*-GlcNAcylation activity compared to OGT-WT, with 5N5A-1 having the lowest activity (almost as low as the no OGT control, which only has background activity from the OGT present in the HeLa cell extracts), and 5N5A-3 having the most activity, approaching the levels of

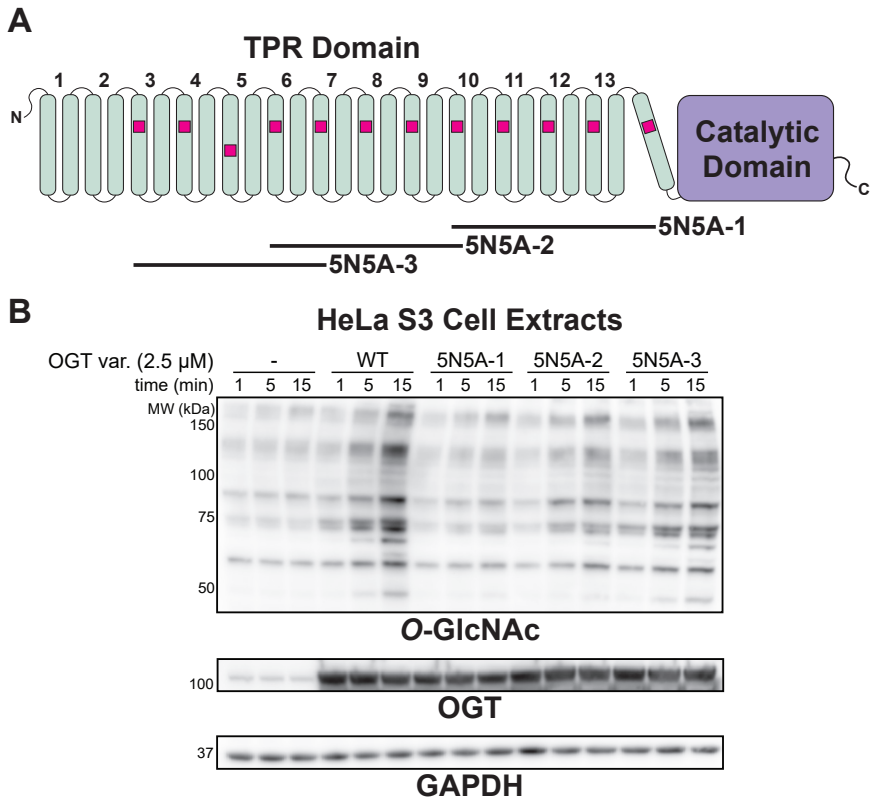


Figure 2.9: Glycosylation of HeLa S3 cell extracts by OGT-5N5A mutants. (A) Schematic of OGT-5N5A mutants, each of which contains five asparagine-to-alanine mutations. (B) Glycosylation activity of OGT-5N5A mutants in HeLa S3 cell extracts after 1, 5, or 15 minutes, assessed by SDS-PAGE and western blotting.

OGT-WT. The observed gradient in activity as asparagines from proximal to distal regions of the TPRs are removed suggests that the most proximal TPR asparagines are the most important for substrate recognition prior to *O*-GlcNAcylation, an observation that makes sense when thinking about how OGT binds its substrates. Since OGT's substrates extend from the active site into the TPRs, many more substrates make contacts with the proximal TPRs than they do with the distal TPRs, and indeed all distally-contacting substrates must also make proximal contacts. Thus, the asparagine ladder residues closer to the active site are likely responsible for contacting more substrates than those farther from the active site, and their loss affects a greater proportion of the overall *O*-GlcNAcome.

2.5 The TPR aspartate ladder offers a counterbalancing restriction on substrate glycosylation

Previous work in the lab, led by Dr. Cassie Joiner, established that alanine mutation of the TPR aspartates D386, D420, and D454 by themselves, in groups of two (D2A-2: D386/420A, D2A-1: D420/454A; Figure 2.10A), or as a triple mutant (D3A: D386/420/454A) led to increased glycosylation in HeLa cell extracts, with additivity evident as the number of mutated aspartates increased.[61] In that work, Dr. Joiner also used chemoproteomic methods to map the glycosites hit by OGT-WT and OGT-D2A-2 (D386/420A). She found that OGT-D2A-2 glycosylated more overall proteins and had an altered substrate specificity compared to OGT-WT. While OGT-WT preferentially glycosylated proteins containing arginine and lysine residues in the +7 to +11 region from the glycosite, corresponding to the area that can be engaged by aspartate residues D386 and D420, OGT-D2A-2 did not glycosylate these substrates, and TPR-binding lysines and arginines were not enriched.

I assisted Dr. Joiner in extending our coverage of the TPR aspartate ladder by making three additional D2A mutants: D2A-3 (D284/318A), D2A-4 (D152/216A), and D2A-5 (D114/152A; Figure 2.10A). We then assayed these mutants' glycosylation activities against HeLa cell extracts by western blotting and found that all were hyperactive compared to OGT-WT (Figure 2.10B), with the biggest fold-change observed for D2A-3 ($\tilde{3}$.5x WT activity). As previously observed, the different D2A mutants hyperglycosylated different substrates (seen by eye on a western blot),

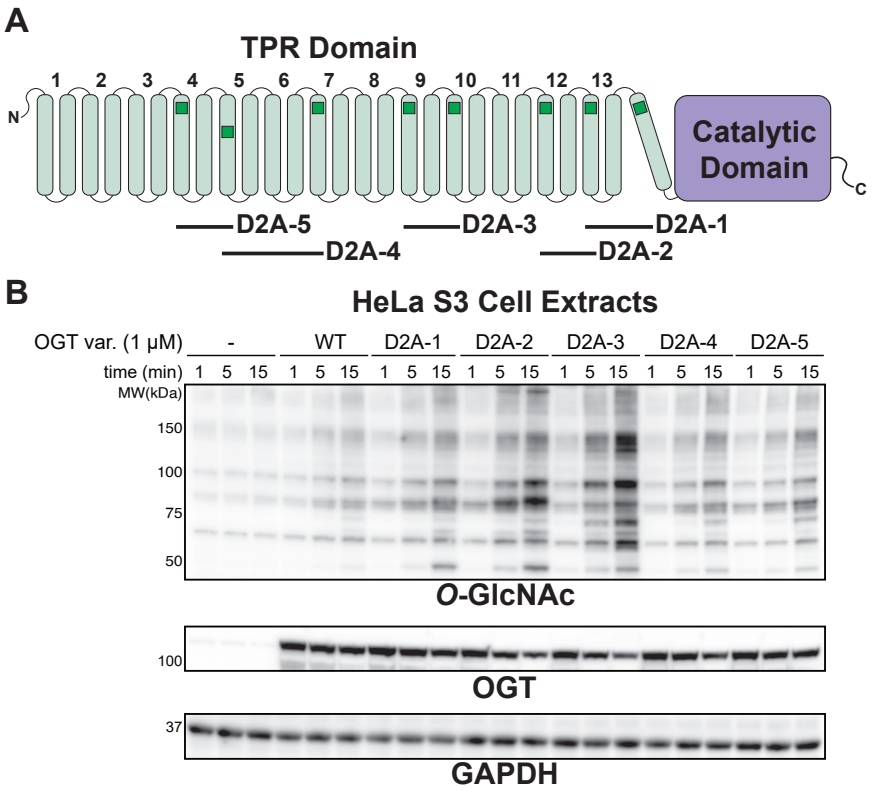


Figure 2.10: Glycosylation of HeLa S3 cell extracts by OGT-D2A mutants. (A) Schematic of OGT-D2A mutants, each of which contains two aspartate-to-alanine mutations. (B) Glycosylation activity of OGT-D2A mutants in HeLa S3 cell extracts after 1, 5, or 15 minutes, assessed by SDS-PAGE and western blotting.

suggesting that different aspartates are responsible for binding or placing restrictions on binding of different groups of potential substrates. Overall, these results show that the aspartate ladder acts in contrast to the asparagine ladder. While its removal does not hinder OGT's ability to glycosylate substrates, it does significantly affect the substrate pool and thus can be thought of as more of a gatekeeper for OGT's "wildtype" substrate specificity.

2.6 The TPR macro-structure is important for maintaining OGT's conserved activities and promoting cellular proliferation

2.6.1 Truncation of the TPR domain successively reduces OGT's catalytic activity

A previous graduate student in the Walker lab, Dr. John Janetzko, developed a series of OGT constructs containing N-terminally-truncated TPR domains for heterologous expression and purification out of *E. coli*.[124] This panel contained six OGTs (OGT-4, -5, -6, -7, -7*, and -8; OGT-7* was truncated in the middle of the sixth TPR), each having an intact active site and between 4 and 8 TPRs (Figure 2.11A); other attempted constructs did not fold correctly and primarily formed aggregates during purification. These truncations were initially made to answer questions concerning cleavage of HCF3R constructs, and Dr. Janetzko found that constructs containing fewer than 8 full TPRs were unable to proficiently cleave HCF3R-EEE, although all constructs were able to glycosylate HCF3R-SAA.[124]

Later, in the course of assisting Dr. Cassie Joiner and Dr. Sarah Potter with their studies, I expressed and purified the same OGT truncations to test their activity against HeLa cell extracts (Figure 2.11B). Here, we chose to exclude OGT-7*, as concurrent work in cells showed that it tended to form aggregates. We found that OGT-8's activity was only slightly lower than WT, excluding a band around 90 kDa that was strongly glycosylated by OGT-WT but not by OGT-8. We also noted that the further truncations (OGT-4, -5, -6, and -7) all had very low activity, only marginally above the background *O*-GlcNAc present in the HeLa cell extracts themselves.

As noted above, Dr. Joiner and Dr. Potter also worked to incorporate the five well-folding constructs into both the Cre- and dTAG-MEFs as mKate2-2xFLAG fusion proteins, in line with other efforts in the lab at the time. I assisted in this process during my rotation, but the bulk of the

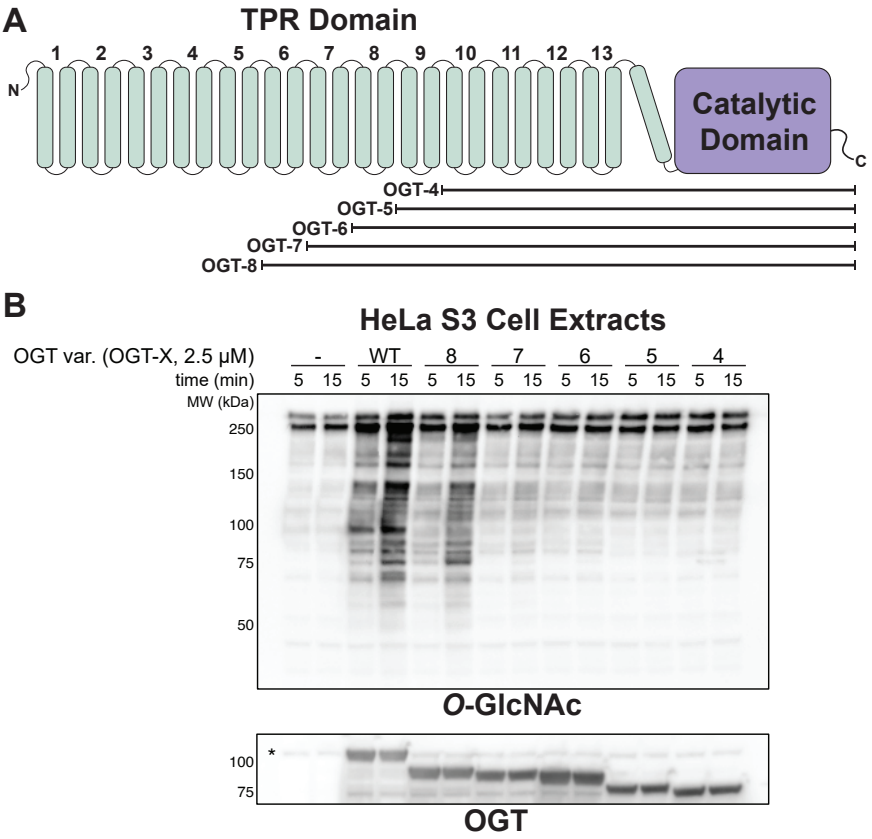


Figure 2.11: Glycosylation of HeLa S3 cell extracts by TPR truncation OGT mutants. (A) Schematic of TPR truncation OGT mutants. (B) Glycosylation activity of TPR truncation OGT mutants in HeLa S3 cell extracts after 5 or 15 minutes, assessed by SDS-PAGE and western blotting.

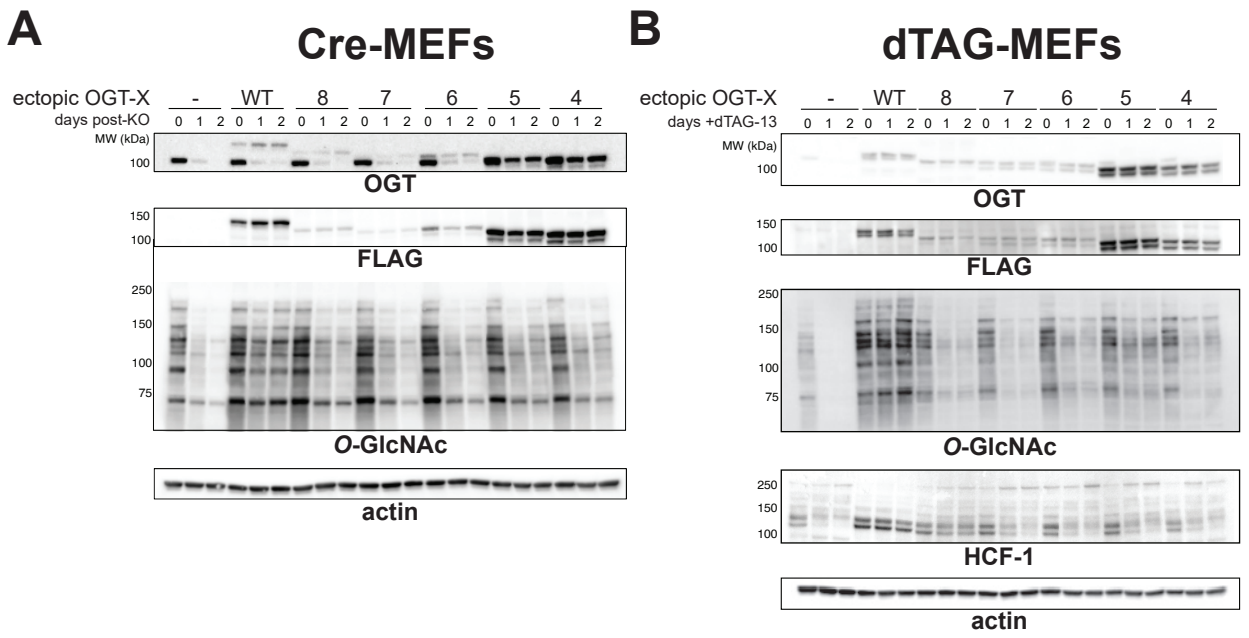


Figure 2.12: Activity of TPR truncation OGT mutants in (A) Cre-MEFs and (B) dTAG-MEFs, assessed by SDS-PAGE and western blotting.

work would be finished by Dr. Potter, who conducted multiple re-infections and re-sorts of some of the truncations in an attempt to normalize expression levels and had to contend with an errant stop codon in the largest truncation construct. However, we eventually obtained five Cre-MEF and five dTAG-MEF cell lines expressing the truncations, which we then tested for expression and activity by western blot following knockout of the endogenous *Ogt* in the Cre-MEFs and degradation of the "endogenous" FKBP12^{F36V}-OGT-WT in the dTAG-MEFs, with corresponding OGT-WT and null cells as controls (Figure 2.12AB). In both systems, we found that truncations OGT-8, OGT-7, and OGT-6 had a lower expression level than OGT-WT, while OGT-5 and OGT-4 were comparatively overexpressed. After two days of knockout or degradation, we saw that *O*-GlcNAc levels in all truncations in both systems were greatly reduced compared to those for OGT-WT, however still above the null background. The same trend was observed with HCF-1 cleavage in the dTAG-MEF cell lines: all truncations were unable to fully cleave HCF-1, and the shortest truncations appeared to be fully incompetent at fulfilling this biochemical function of OGT (Figure 2.12B). We think that the loss of *O*-GlcNAc in the intermediate truncations (OGT-8, -7, -6) compared to the shorter truncations (OGT-5, -4) is due to lower expression; however, the fact that OGT-5 and -4 do not catalyze any HCF-1 cleavage indicates that these, in addition to

OGT-6 and -7, are more deficient at performing that activity due to their reduced TPRs.

2.6.2 The TPRs drive cellular proliferation and play a role in proper nucleocytoplasmic localization of OGT

With the knowledge that truncation of the TPR domain alters OGT's biochemical functions in cells, we were also interested in finding out if it impacted cell proliferation as the sole OGT construct. To do this, we knocked out endogenous *Ogt* and monitored growth using both IncuCyte Zoom live cell imaging as done with OGT-mono (Figure 2.13A), as well as CellTiter-Glo, an endpoint luminescence-based assay that reads out ATP levels as a proxy for metabolically active (*i.e.*, dividing) cells (Figure 2.13B). In both assays, we found that while OGT-WT supported healthy proliferation, the OGT-4, -5, -6, and -7 truncations led to cell death after 6-8 days, consistent with the pattern seen in null cells. However, the OGT-8 truncation supported an intermediate viability in both assays, suggesting it was more capable of supporting cell proliferation than the other truncations.

We then wondered if the N-terminal TPRs played a role in OGT's subcellular localization. Typically, OGT adopts a diffuse nuclear and cytoplasmic localization, with a higher concentration observed in the nucleus.[27] Prior to our work, the mechanistic determinant of OGT's nuclear localization was not strictly known, as OGT itself lacks a nuclear localization signal. One group posited a three-residue Asp-Phe-Pro ('DFP') motif at the C-terminus of the 13th TPR was the controller of OGT's localization;[125] however, in Dr. Potter's hands in the dTAG-MEF system, a triple-alanine mutation of this motif still translocated to the nucleus.[108] Another group working with TPR truncation OGT variants, the Woo lab in Harvard's CCB department, showed in a transient transfection system that truncations containing five or fewer TPRs were excluded from the nucleus.[126] Dr. Potter performed live-cell imaging using the mKate2 fluorescent protein fusion attached to all of our OGT variants and found that all of our truncations, including OGT-8, mislocalized and were nuclear-excluded (Figure 2.13C). This mislocalization was not dependent on whether the endogenous FKBP12^{F36V}-tagged OGT-WT was present. Since all of our truncations contained the 'DFP' motif, here we showed that it was not the sole driver of OGT's nuclear localization. Moreover, since OGT-8 was still mislocalized, we have determined that at least a

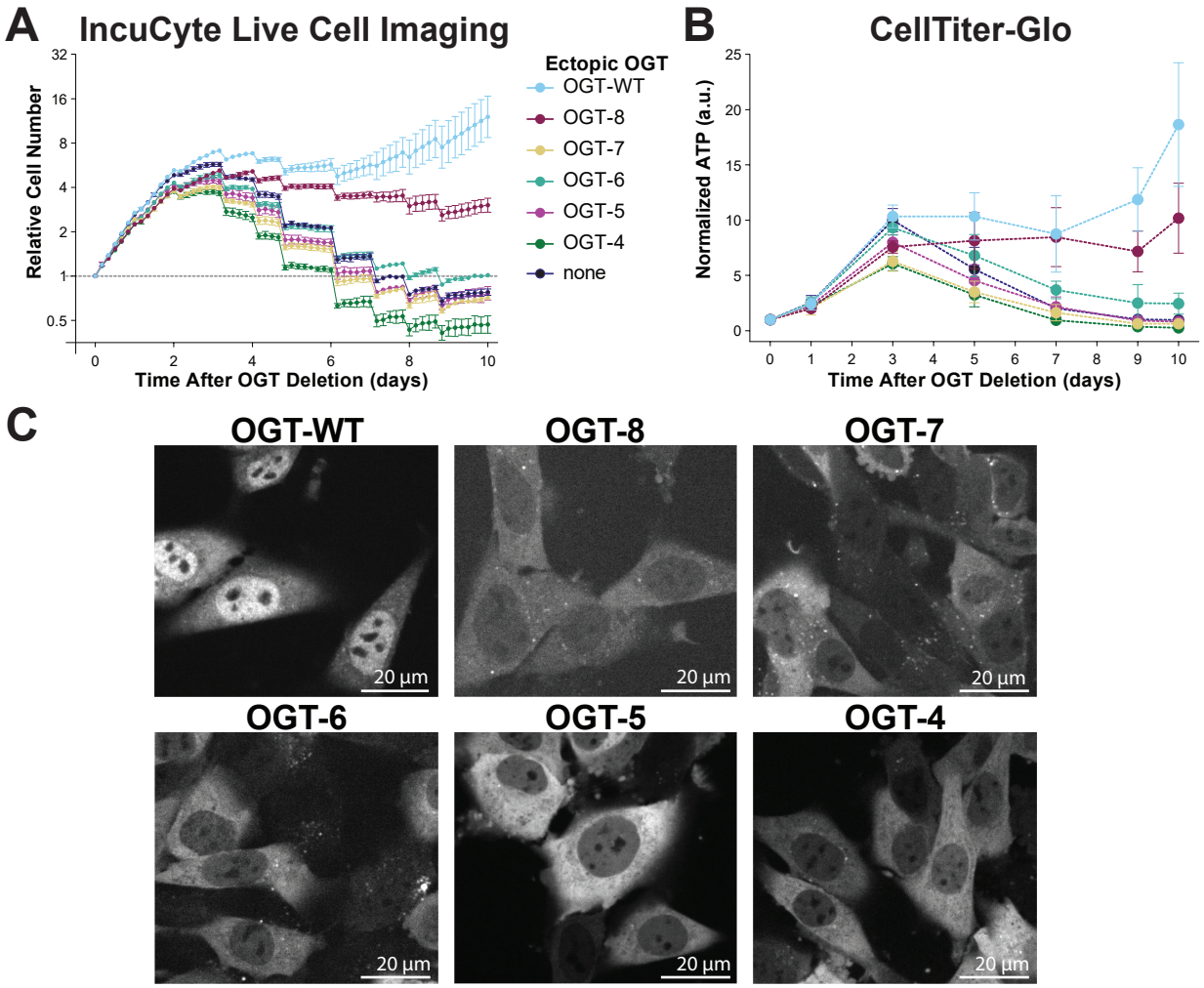


Figure 2.13: Nuclear localization of OGT does not drive proliferation. (A) Incucyte live cell imaging and (B) CellTiter-Glo growth curves of Cre-MEFs expressing TPR truncation OGT variants. (C) Live cell confocal fluorescence microscopy of dTAG-MEFs expressing TPR truncation OGT variants without degradation of dTAG-OGT-WT shows altered localization of TPR truncations. Figure reproduced from Potter, et al. (2024).[108]

partial determinant of OGT's localization lies in the five most N-terminal TPRs. Further work will need to be done to fully characterize the nature of OGT's nuclear localization, whether that be due to some feature of OGT, an external chaperone-like protein shuttling OGT in as cargo, or some other cellular factor.

2.6.3 Truncation of the TPR domain greatly alters OGT's interactome in cells

Our data on the truncations' catalytic activities in cells did not suggest anything that might explain the increased proliferation of cells expressing OGT-8 compared to cells expressing shorter truncations, so Dr. Potter and I moved to assess the ability of the truncated OGTs to bind interacting proteins. In triplicate, we grew up dTAG-MEFs expressing the TPR truncation variants OGT-8, OGT-7, and OGT-4, as well as OGT-WT and no additional OGT as controls, treated with the degrader compound dTAG-13 for 8 hours, harvested cells and lysed under native (non-denaturing) conditions, then pulled down on the FLAG tag present on all of the ectopic OGT constructs using magnetic anti-FLAG beads to isolate the OGT interactome for each of the cell lines. We normalized input by total lysate mass and sent samples to Dr. Joao Paulo, who performed mass spectrometry to enable semi-quantitative proteomics analysis.

We processed data by comparing each dataset to the "null" dataset (which contained no FLAG-tagged OGT) and computing the resultant fold enrichment and p-value. We then set a cutoff of 5-fold enrichment (provided that the average of total peptides found was ≥ 2) and analyzed the resultant hit list by UpSet plot (Figure 2.14A). Using this analysis, we found 113 total hits with significant overlap between different samples. OGT-WT pulled down 42 proteins, of which 35 only interacted with OGT-WT and 31 were identified as OGT interactors in other studies. Of the 113 hits across all samples, 78 were hits for at least one TPR truncation, the majority of which were not found as hits for OGT-WT in our work or in previous studies. These results were not generally confounded by changes in total protein abundance, as total protein levels were generally similar across all cell lines with the exception of *O*-GlcNAcase (OGA), which was identified as a hit in OGT-WT-expressing cells and was also overexpressed in that cell line. This higher expression is likely due to the higher levels of protein *O*-GlcNAcylation in that cell line compared to the truncation cell lines, corroborating results observed in other studies with

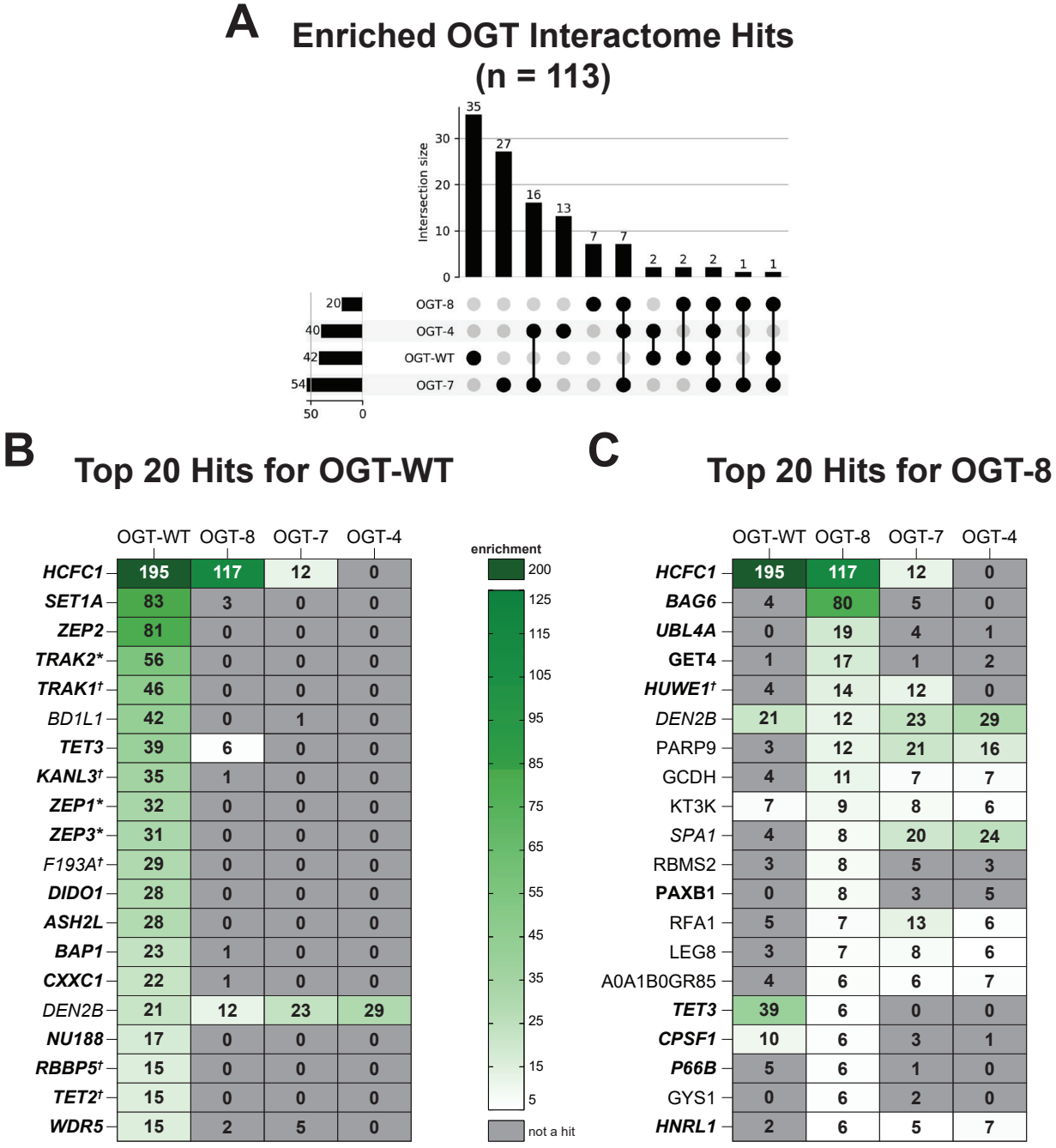


Figure 2.14: OGT's N-terminal TPRs control its interactome. (A) UpSet plot showing overlap of pulldown hits from OGT-WT and TPR truncation variants OGT-4, OGT-7, and OGT-8. (B,C) Top 20 most abundant pulldown hits for (B) OGT-WT and (C) OGT-8, sorted by peptide count. Italics indicate known OGT substrates,[35,111] bolding indicates known OGT interactors,[63] asterisks indicate proteins with elevated expression in the hit cell lines, and crosses indicate proteins for which no expression data was obtained. Figure reproduced from Potter, et al. (2024).[108]

increased *O*-GlcNAcylation.[127,128]

The hit list for OGT-WT greatly resembled previous studies of the OGT interactome (Figure 2.14B); of the 20 most abundant interactors, 17 had previously been identified. HCF-1 was the most abundant interactor, validating previous studies that suggested that upwards of 50% of cellular OGT was bound to HCF-1.[119,129] Moreover, the "usual suspect" interactors in the TET, TRAK, and ZEP/HIVEP protein families were identified, suggesting that our cutoffs and methodology were sound and sufficient to isolate interaction partners of OGT while still removing junk. Interestingly, of the top 20 OGT-WT interactors, only three proteins were found as hits for the truncations: HCF-1, which was a hit in OGT-8 and OGT-7; TET3, which was a hit in OGT-8; and DEN2B (previously unknown as an interactor, but identified as a glycosylation substrate),[130] which was identified as a hit in all tested OGT variants.

The list of hits for OGT-8 tells a different story entirely (Figure 2.14C): of the top 20 hits for OGT-8, only 10 were previously identified as OGT interactors. While the top hit was still HCF-1, the next three hits (BAG6, UBL4A, and GET4) represent components of a conserved complex that functions in synthesis of tail-anchored proteins as well as DNA damage response pathways when translocated to the nucleus.[131–133] This is of special note, as previous work out of the Mariappan lab at Yale found that the OGT-BAG6 interaction is mediated by a small region of BAG6 that bears striking homology to the HCF-1 proteolytic repeat's threonine-rich region (GTTNTATT, res. 598-606).[134] Further work will be necessary to determine if OGT-WT is engaging with BAG6 complex members functionally, and if HCF-1 and BAG6 compete to bind OGT's TPRs.

2.7 Global trends of glycosylation and feature-driven OGT regulation are not necessarily recapitulated on a per-substrate level

The majority of this chapter has focused on the effects of various OGT features – the ability to homodimerize, the TPR asparagine and aspartate ladders, and the macro-structure of the TPRs themselves – on global glycosylation and scaffolding. And while these global analyses are useful, we know from previous studies on OGT's substrate binding roles that conclusions about global trends do not recapitulate on a per-substrate basis.[60,61] To demonstrate and elaborate on this phenomenon, Dr. Cassie Joiner selected three OGT substrates with similar architecture for

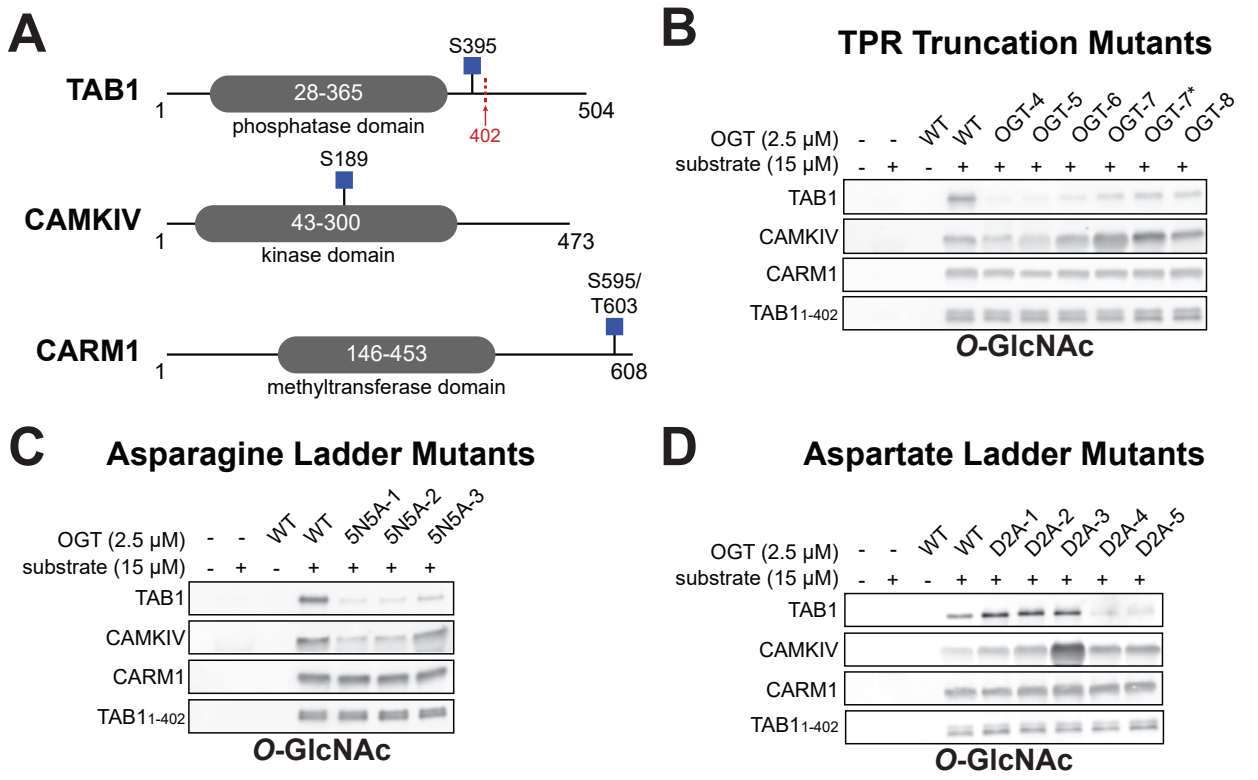


Figure 2.15: Activity of TPR variant OGTs against purified substrates. (A) Schematics for the OGT substrates TAB1, CAMKIV, and CARM1. Blue squares indicate primary glycosites. Red dashed line indicates a cryptic proteolytic site. (B,C,D) Western blots for *O*-GlcNAcylation of TAB1, CAMKIV, and CARM1 by (B) TPR truncation OGTs, (C) 5N5A OGTs, and (D) D2A OGTs.

rigorous *in vitro* characterization: TAK1 binding protein 1 (TAB1), calcium/calmodulin-dependent protein kinase IV (CAMKIV), and coactivator-associated arginine methyltransferase 1 (CARM1; Figure 2.15A). *O*-GlcNAcylation kinetics for all three of these OGT substrates had previously been determined,[116] and plasmids for expression of all three as polyhistidine tag fusions in *E. coli* were obtained as a generous gift from David Vocadlo's laboratory at Simon Fraser University. Dr. Joiner picked these substrates for two reasons: the aforementioned existence of stable expression constructs, and the differential location of glycosites in each substrate. While CARM1's primary glycosites (S595/T603) are at its C-terminus, TAB1's (S395) is located at the N-terminus of its lengthy (>100 residues) C-terminal unstructured tail, and CAMKIV's primary glycosite (S189) is in the center of the protein on a small loop between two folded kinase domain lobes. She wondered if the difference in sub-protein location of the glycosites would lead to different responses to TPR variation.

Dr. Joiner and I worked to purify all three substrates with N-terminal polyhistidine tags using a combination of nickel affinity chromatography and size exclusion chromatography, as were used to purify all of the OGT variants described in this chapter. CARM1 purified very nicely as a monodisperse peak with no clear contaminants. CAMKIV was messier, but we were able to obtain relatively clean fractions. TAB1, however, purified as a mixture of the full-length protein and a degradation product that we hypothesized contained residues 1-402 ($TAB1_{1\text{-}402}$).

Together, we subjected the three substrates to glycosylation reactions using the full panel of OGT truncation variants (including the OGT-7* construct; Figure 2.15B), asparagine ladder mutants (5N5As; Figure 2.15C), and aspartate ladder mutants (D2As; Figure 2.15D). What we found generally confirmed our assumptions about the per-substrate variability of responses to perturbations in OGT's TPR domain.

CARM1 and $TAB1_{1\text{-}402}$ were unaffected by any OGT truncation or mutation to the asparagine or aspartate ladders, consistent with the hypothesis that their C-termini do not extend into the TPR domain and thus make no contacts in its lumen.

Full-length TAB1 glycosylation was negatively affected by all assessed TPR truncations and asparagine mutants, with greater effects observed with mutation/truncation closer to OGT's active site (*e.g.*, OGT-5N5A-1, OGT-4, OGT-5). It also bears repeating that a previous study found that TAB1 glycosylation kinetics were poorer with monomeric OGT, although no further characterization of this effect was documented.[59] Notably, all of the truncations excluding OGT-8 lack the homodimerization motif. It is unclear whether poor glycosylation of Tab1 by these truncations reflects their (presumed) inability to dimerize, their shorter length, or both. In a general validation of effects observed with aspartate mutants on global glycosylation, full-length TAB1 was glycosylated more by OGT-D2A variants 1, 2, and 3, which contained aspartate-to-alanine mutations closest to the active site. I conducted an additional kinetic analysis of TAB1 glycosylation by OGT-WT and OGT-D2A-2, the most highly active OGT variant against TAB1, by running reactions using 1 μM OGT and 20 μM TAB1, then quenching at timepoints between 0 and 20 minutes (Figure 2.16A,C). I then immunoblotted against *O*-GlcNAc and imaged using a fluorophore-conjugated secondary antibody, enabling quantitative correlation of band density to *O*-GlcNAc levels. By plotting band density as a function of time, I was able to calculate a metric for substrate turnover from the slope of the linear range of the graph. Using this analysis, I found

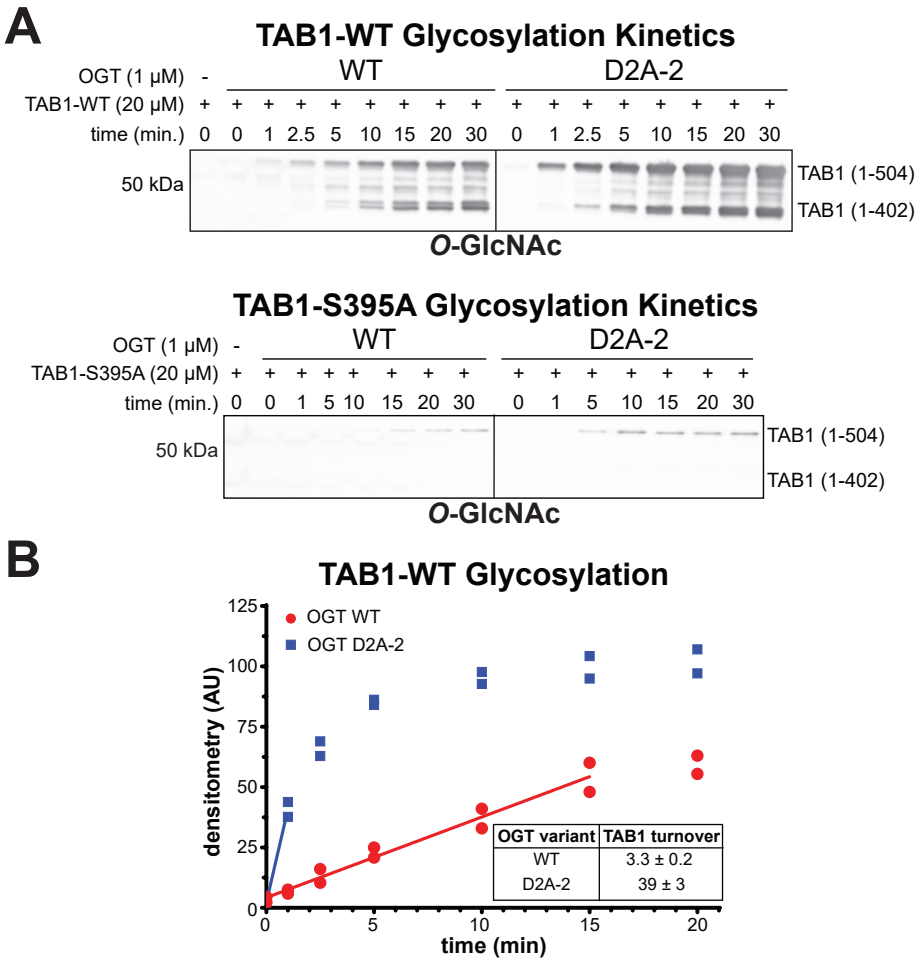


Figure 2.16: D2A mutations increase turnover of TAB1 by OGT. (A,B) Western blots for time-course *O*-GlcNAcylation of (A) TAB1-WT and (B) TAB1-S395A by OGT-WT and OGT-D2A-2. One replicate is shown per experiment. (C) Plotted data for TAB1-WT glycosylation kinetics shows that OGT-D2A-2 turns over TAB1-WT ∼10x faster than OGT-WT.

that OGT-D2A-2 turns over TAB1 around 10 times faster than OGT-WT does. A similar effect was observed with the glycosite alanine mutant TAB1-S395A (Figure 2.16B), although the total level of *O*-GlcNAcylation was significantly lower due to loss of the primary glycosite.

Lastly, glycosylation of CAMKIV was most strongly affected by alteration of the medial region of the TPRs (between TPRs 6 and 10). The extent of glycosylation with OGT-5N5A-1 and -5N5A-2 was lower than for OGT-WT, while glycosylation was greatly increased with OGT-D2A-3 compared with WT. In contrast, other D2A variants showed only a slight increase in glycosylation over WT. However, contrary to the results obtained with TAB1, the other TPR-sensitive substrate, truncation of the TPRs did not generally reduce glycosylation of CAMKIV. While the shortest truncations (OGT-4 and OGT-5) did have less activity against CAMKIV, the longer truncations (OGT-6, -7, -7*, and -8) were more active. This suggests that CAMKIV interacts with OGT differently from the other substrates because the presence of OGT's N-terminal TPRs seems to prevent or limit CAMKIV glycosylation.

Overall, this work showed that global trends about the effects of different TPR features cannot be applied to each substrate in the *O*-GlcNAcome. Therefore, OGT-substrate interactions must be treated on a case-by-case basis. While this work was not comprehensive because OGT glycosylates thousands of substrates and we only dissected its interactions with three, our findings have provided insight into how different types of *O*-GlcNAcylation substrates depend on OGT's TPR domain.

2.8 Conclusion

The work presented in this chapter strengthens previous findings that OGT's TPR domain is a master controller of its protein-protein interactions, which it enforces through conserved binding interfaces spread throughout the repeats.

The finding that OGT's ability to homodimerize is not essential for cellular viability is one that is particularly surprising, given the observed $K_{D,dimer}$ for OGT-WT. Cellular OGT concentrations are estimated to be around 100 nM, so a $K_{D,dimer}$ of 39 nM would enable cellular OGT to exist primarily as a dimer, but also as a monomer in some circumstances, depending on its local concentration. Additionally, if tighter-binding interactors exist at that interface, those could

preferentially bind in place of a second OGT monomer. All of this is to say that some monomeric OGT likely exists in cells and, given the results described here, may play physiological roles distinct from dimeric OGT. It is also interesting that, despite reduced cellular activity against *O*-GlcNAcylation substrates and HCF-1, OGT-mono supports greater-than-wildtype growth in the Cre-MEF system. While we generally interpret cell growth as a good thing in the context of these loss-of-function OGT variants, this could indicate that preventing OGT dimerization actually does induce some increased growth phenotype, which is not always preferential (see: cancer). Further work will be needed to characterize OGT's oligomeric state in cells, and to determine if it is dependent on subcellular localization, cell cycle state, or some other cellular or environmental factor.

The role of the TPR domain in OGT's essential cellular functions had not been directly probed prior to the work described in this chapter. The striking finding that OGT-8 supports cellular proliferation while shorter truncations do not, even though all truncations were excluded from the nucleus, suggests some key role for the 6^{th} TPR. Unfortunately, the activity-based analyses we performed did not reveal what that role could be, and whether it is shared with OGT-WT or is some novel compensatory function that only the truncation can perform. The co-immunoprecipitation experiment comparing interactors of the different TPR truncations did shed some light on how the TPRs regulate OGT's protein-protein interactions, especially with known interactors such as HCF-1, the Tet family proteins, and TRAK1/2; however, they did not provide mechanistic insights. Future studies will need to elucidate the role of the N-terminal TPRs in these interactions. Additionally, the finding that OGT-8 interacts with the BAG6/UBL4A/GET4 complex members is of interest, especially due to the striking resemblance a region of BAG6 bears to the HCF-1 proteolytic repeat.[134]

Lastly, the finding that the roles of TPR features on global substrate pools are not recapitulated on a per-substrate basis *in vitro* is not wholly surprising, but it had not been previously described. Our work with the substrates TAB1, CAMKIV, and CARM1 has allowed us to speculate on mechanisms of target engagement that may be present among different classes of OGT's substrates. The first of these classes contains substrates where the glycosite exists on a short, unstructured region close to its C-terminus, such as CARM1. For these substrates, engagement is simple: the tail sticks into the active site to facilitate glycosylation without substantial interactions in the TPRs.

The second class consists of substrates with glycosites on longer, unstructured tails, including full-length TAB1. With these substrates, interactions in the TPR domain are crucial for locking the glycosite in OGT's active site. The last class of substrates for which we characterized a founding example contain a glycosite on an internal loop between two folded domains, such as CAMKIV. Some TPR interactions are necessary to drive CAMKIV glycosylation, but the N-terminal TPRs inhibit the process. Although structural characterization of different OGT-substrate complexes is necessary to fully understand how substrates bind to OGT, this work provides working models for differential engagement of substrates by the TPR domain.

2.9 Methods

General Vector Construction and Cloning

See Appendix A for information on the plasmids used in this chapter, as well as elsewhere in this dissertation. The OGT heterologous expression plasmids used in this chapter were originally developed by Dr. John Janetzko; the below information pertains to plasmids developed as part of the work presented in this chapter.

All primers and gene blocks were purchased from Integrated DNA Technologies (IDT). Isolated plasmids were verified by sequencing by Quintara Biosciences (Cambridge, MA) or the Harvard Biopolymers Facility (Boston, MA). All PCRs were run using Q5 High-Fidelity Polymerase 2x Master Mix (NEB), according to manufacturer's suggestions. PCR products with installed point mutations, small insertions, or deletions were treated with KLD enzyme mix (NEB) for 1 hour at room temperature. PCR products destined for multi-piece ligation were treated with 1 μL Dpn1 for 1 hour at 37°C, heat-inactivated for 20 minutes at 80°C, then gel-purified and extracted. Purified DNA was then treated with In-Fusion HD 5x mix (Takara) according to manufacturer's guidelines. KLD- or InFusion-treated DNA was then transformed into E. coli STELLAR competent cells (Takara) and plated on Luria-Bertani (LB) agar supplemented with the corresponding antibiotic(s) for 16 hours at 37°C. Following incubation, single colonies were picked and subcultured in 5 mL LB broth supplemented with the corresponding antibiotic(s) for 16 hours at 37°C, then miniprepped using the QIAprep spin miniprep kit (Qiagen) to isolate plasmid DNA. Sequences were confirmed prior to plasmid storage and usage.

Destination vectors for lentivirus production in HEK293T cells were produced from pENTR-1A entry vectors containing the desired gene products via Gateway cloning. Briefly, pENTR-1A vector containing the gene insert in-frame (75 ng) and empty desti-nation vector (75 ng) were mixed with Gateway LR Clonase II enzyme mix (Thermo-Fisher Scientific) in 5 μL volumes, then incubated for 16 hours at room temperature. Proteinase K (0.5 μL, NEB) was added, and the reaction was incubated for 10 minutes at 37°C. The full mixture was transformed into NEB Stable E. coli competent cells (NEB) and plated on LB agar supplemented with corresponding antibiotic(s) for at least 16 hours at 30°C (note: growth at 37°C should be avoided to prevent undesired recombination events). Following transformation, single colonies were subcultured in 5 mL LB broth supplemented with corresponding antibiotic(s) for 16 hours at 30°C, miniprepped using the QIAprep spin miniprep kit (Qiagen), and verified by sequencing (Quintara Biosciences). Following sequence confirmation, NEB Stable E. coli cell cultures containing the desired plasmids were subcultured in 50 mL volumes with corresponding antibiotics at 30°C, then midiprepped (Nucleobond Xtra Midi EF kit, Macherey-Nagel) to yield sufficient DNA for lentivirus production.

Expression and purification of His_8-tagged OGT from *E. coli*

Vectors for heterologous OGT expression were freshly transformed into chemically competent *E. coli* BL21(DE3) cells (NEB) and plated on LB agar with appropriate antibiotics for 16 hours at 37°C. 5-10 colonies were used to seed a 50 mL LB broth + antibiotic subculture, which was incubated for 16 hours at 37°C with 225 rpm shaking. Expression cultures containing 1.5L LB broth + antibiotic were then inoculated at 1% (v/v) from this subculture and grown at 37°C with 225 rpm shaking until the OD_{600} reached 1.0, at which point the cultures were cooled on ice to 4°C. Once cold, IPTG was added to 0.2 mM to induce protein expression, and the cultures were incubated for 16 hours at 16°C with 225 rpm.

Following induction, cells were harvested by centrifugation (4200 rpm, 20 minutes, 4°C) and resuspended in 45 mL HBS on ice, supplemented with 0.1 mg/mL recombinant lysozyme (GoldBio), 250 units Benzonase (Millipore-Sigma) and one Roche cOm-plete EDTA-free protease inhibitor tablet (Millipore-Sigma). Cells were lysed by cell disruption on an Avestin Emulsiflex C3 cell disruptor at 15,000 psi three times. The lysate was clarified by centrifugation (15000g, 30 minutes, 4°C), and the clarified lysate was removed to a new tube, supplemented with 40 mM

imidazole, and batch-bound for 1 hour with rotation at 4°C to 1 mL Nickel-NTA agarose resin (Qiagen), pre-equilibrated in HBS (20 mM HEPES, pH 7.4, 150 mM NaCl) supplemented with 40 mM imidazole. Bound resin was subjected to gravity filtration at 4°C: flowthrough was collected, followed by a 10 CV (column volume) wash in HBS with 50 mM imidazole, then a 6 CV elution in HBS with 250 mM imidazole. Dithiothreitol (DTT) was added to 1 mM in elution fractions to prevent formation of disulfides and aggregates. Elution fractions were concentrated to 300 μL in 100K MWCO Amicon protein concentrators (note that TPR truncation OGTs were concentrated on 50K MWCO Amicon concentrators), then purified by size exclusion chromatography on an AKTA Pure 25 FPLC (Cytiva) fitted with a Superdex 200 Increase 10/300 GL column using an isocratic method into sterile-filtered HBS with 1 mM DTT. 0.5 mL peak fractions were collected and analyzed by 4-20% Tris/Glycine SDS-PAGE using InstantBlue protein stain (Abcam/Novus Biologicals). Suitable fractions were combined, concentrated to $\geq$ 25 μM, aliquoted, frozen in liquid nitrogen, and stored at -80°C until further usage.

Expression and purification of His_8-tagged HCF3R substrates

Polyhistidine-tagged HCF3R constructs were prepared similarly to His_8-OGT variants, as described above, with the following modifications: (1) proteins were concentrated on 10K MWCO Amicon concentrators; (2) proteins were purified by size exclusion chromatography using a Superdex 75 Increase 10/300 GL column, using the same isocratic method and buffer as above; and (3) proteins were concentrated to >200 μM prior to aliquoting, freezing in liquid nitrogen, and storing at -80°.

Expression and purification of His_6-tagged OGT substrates TAB1, CARM1, and CAMKIV

Polyhistidine-tagged OGT substrate proteins TAB1, CARM1, and CAMKIV were purified using their respective pET28a plasmids, which were obtained as a gift from Dr. David Vocadlo.[116] These were purified according to the protocols above with the following modifications: (1) *E. coli* LOBSTR(BL21-DE3) cells were used; (2) cells were grown to an OD_{600} of 0.6 prior to cooling and induction; (3) Tris-buffered saline (TBS, pH 7.4) was used in the place of HBS for all isolation and purification steps; (4) 30K MWCO Amicon concentrators were used to concentrate protein; and (5) proteins were purified by size exclusion chromatography using a Superdex 75 Increase 10/300 GL

column, using an isocratic method and purifying into TBS supplemented with DTT.

CoraFluor-1 and Alexa Fluor 488 (AF488) labeling of OGT

Protein labeling with CoraFluor-1-PFP ester was performed as previously described.[135] Protein was thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates prior to buffer exchanging into 100 mM sodium carbonate, pH 8.5 with 0.05% (v/v) Tween-20 using Zeba 3K MWCO spin columns according to the manufacturer's guidelines (Thermo-Fisher). Protein was then concentrated to $\geq$ 1 mg/mL if necessary. 5 equivalents of CoraFluor-1-PFP ester or AF488-TFP ester were added directly (from a 25 mM stock in dimethylacetamide) and pipetted to mix. The reaction was incubated for 1 hour at room temperature in the dark and buffer-exchanged back into HBS with 0.05% Tween-20 (and 1 mM DTT for OGT variants).

For CoraFluor-1 labeled OGT, concentration and degree of labeling (DOL) were calculated as previously reported, by taking A_{280} and A_{340} by Nanodrop, using an A_{280} correction factor (cf_{340}) of 0.1571, E_{340} of 22000 $M^{-1}cm^{-1}$, and an E_{280} of 210000 $M^{-1}cm^{-1}$ for the antibody, 117580 $M^{-1}cm^{-1}$ for OGT-WT, or 112080 $M^{-1}cm^{-1}$ for OGT-mono, with the following formulae:

$$[CoraFluor] = \frac{A_{340}}{E_{340}} \qquad [protein] = \frac{A_{280} - cf_{280}A_{340}}{E_{protein,280}} \qquad DOL = \frac{[CoraFluor]}{[protein]}$$

Aliquots were diluted 1:1 in glycerol (50% v/v final), frozen in liquid nitrogen, and stored at -80°. Once thawed, aliquots were kept at -20° for use, remaining stable for at least one month under these conditions.

For AF488-labeled OGT, the same calculations were performed using A_{495} in the place of A_{340}, an A_{280} correction factor (cf_{495}) of 0.11, and a molar extinction coefficient for AF488 at 495 nm (E_{495}) of 71000 $M^{-1}cm^{-1}$.

General staining and immunoblotting methods

Except where specified otherwise, all proteins were separated by SDS-PAGE on 4-20% TGX Tris/Glycine gels (Bio-Rad), using a constant voltage of 180V. Protein molecular weight was assessed by comparison to Precision Plus Protein Dual Xtra Prestained ladder (Bio-Rad).

Total protein content was analyzed by staining either with fixative Coomassie Blue R-250 dye or InstantBlue protein stain. Coomassie Blue staining was performed by first fixing the gel in 50% methanol and 10% acetic acid for 1 hour at RT, then staining in 30% methanol and 10% acetic

acid supplemented with 0.1% (w/v) Coomassie Blue R-250 for 1 hour at RT, and destaining for 8-24 hours in 30% methanol and 10% acetic acid at RT. InstantBlue staining was performed by incubating the gel in InstantBlue for 5-15 minutes at RT, then briefly washing in deionized water to remove excess stain. Stained gels were imaged using the 'visible' setting on an Azure C600 imager using auto-determined exposure times. Resultant images were contrast- and brightness-adjusted in FIJI for visibility.

Following SDS-PAGE, samples destined for immunoblotting were transferred either to nitrocellulose membranes or PVDF membranes (pre-soaked in methanol for 2 minutes), then blocked in TBS T-20 Superblock buffer at RT for 1 hour. Membranes were generally incubated in 1:1000 primary antibody (including primary-HRP conjugates, *e.g.*, His-HRP antibody) for 16 hours at 4° with gentle rocking, then warmed to RT, washed three times for 10 minutes in TBST, then (if needed) incubated for 1 hour in 1:2000 dilutions of the corresponding HRP-conjugated secondary antibody, washed three more times with TBST, then imaged using Pierce ECL substrate or Super-Signal West Pico PLUS substrate on a Bio-Rad ChemiDoc MP imager using the chemiluminescent blot setting. Ladder images were taken simultaneously using the colorimetric setting. Images were contrast- and brightness-adjusted in FIJI for visibility. In cases where multiple antibodies were used to probe the same membrane, following the first imaging, membranes were stripped in Restore PLUS stripping buffer (Thermo-Fisher) for 10 minutes at RT, washed three times in TBST, then re-blocked and re-probed as described above.

General tissue culture methods

Unmodified HEK293T cells used in this study were previously ordered from ATCC. Mouse embryonic fibroblasts (MEFs) expressing the Cre-ER fusion with floxed *Ogt* (Cre-MEFs) were previously developed and characterized, and were obtained as a gift from Dr. Natasha Zachara's group.[69,96] MEFs expressing Cre-ER and $FKBP12^{F36V}$-tagged OGT-WT with the endogenous *Ogt* knocked out (dTAG-MEFs) were previously developed and characterized by Dr. Zebulon Levine.[69] All cells were utilized at low passage numbers (<10) throughout this study and were routinely tested to confirm the absence of mycoplasma contamination. All cells were cultured at 37°C with 5% CO_2. MEFs were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (Cytiva) and 1% penicillin/streptomycin solution

(Corning), termed "complete DMEM."

Cells were passaged and split by trypsinization. Briefly, media was removed by aspiration and cells were washed with Dulbecco's phosphate buffered saline (dPBS, Gibco), which was then removed by aspiration. TrypLE Express (Gibco) was added, and cells were incubated at 37°C for 10 minutes to detach from the culture plate. Trypsinized cells were resuspended to 10% final TrypLE in complete medium and centrifuged at 200 g for 5 minutes to pellet. Media was aspirated, then the cell pellet was resuspended in 10 mL complete medium, which was used to seed new plates at the desired cell density.

Cells were harvested by scraping. Briefly, cell plates were placed on ice and media was removed. Cells were washed with 1-5 mL dPBS, which was also removed. 1 mL dPBS was then added to each plate, and cells were scraped into buffer. Cells were removed by pipetting, then centrifuged at 500 g for 5 minutes at 4°C to pellet. Buffer was aspirated and cells were lysed in a minimal amount of RIPA buffer supplemented with 0.05 mM PUGNAc (OGA inhibitor) and 1x Halt proteinase/phosphatase inhibitor (Thermo-Fisher) by pipetting up and down ~20 times and incubating on ice for 10 minutes. Lysates were centrifuged at 21130 rpm for 15 minutes at 4°C to clarify, then moved to new tubes. Protein amounts were quantified by BCA assay (Thermo) against a series of bovine serum albumin standards.

Lentivirus production for cellular integration of OGT variants

10-cm dishes were treated with poly-D-lysine (PDL, 0.01% w/v solution in dPBS) for 1 hour at 37°C. PDL was then removed, and each plate was seeded with ~5,000,000 HEK293T cells in complete DMEM. The following day, media was removed and replaced with serum-free DMEM. For each viral construct, solutions were prepared of 6 μg destination vector, 2.75 μg psPAX-2 plasmid, and 1.25 μg VSV-G plasmid in 40 μL serum-free DMEM. 60 μL polyethyleneimine was then added, and solutions were vortexed briefly and incubated for 15 minutes at room temperature before adding to a plate of HEK293T cells drop-wise. Plates were incubated overnight. The following day, media was aspirated and replaced with 5 mL complete DMEM. On each of the next two days, media containing virus was removed into conical tubes, centrifuged to pellet any cells (200 g, 5 minutes), transferred to a new tube, frozen in liquid nitrogen, and stored at -80°C until needed. Media was also replaced with 5 mL complete DMEM on the first day of harvesting. Following the

second harvest, cells were bleached and aspirated.

Viral infection, puromycin selection, and sorting of MEFs expressing OGT variants

Low-passage Cre- or dTAG-MEFs were thawed from frozen into complete DMEM and grown to confluence, then used to seed 6-well plates at 50,000 cells/well. The following day, viral aliquots were thawed on ice, and media in wells was removed and replaced with 2 mL virus/well (two replicate wells per virus). On each plate, at least one well was replaced with 2 mL complete DMEM as a control. Plates were then spinfected by centrifugation at 1000 g and 30°C for 90 minutes. Following spinfection, 1 mL media was removed from each well and replaced with 1 mL complete DMEM, and plates were incubated overnight. The following day (day 1), media was removed and replaced with 2 mL complete DMEM. On day 2, media was replaced with complete DMEM supplemented with 2 μg/mL puromycin. From days 3-6, media was changed as necessary to remove cell debris, replacing with complete DMEM and 2 μg/mL puromycin. Cells were also expanded to 10-cm dishes when confluent. From days 7-14, this was increased to 10 μg/mL puromycin and cells were expanded to 15-cm dishes as necessary. Once cells reached confluence in 15-cm dishes after day 14, cells were aliquoted and frozen in cryopreservation media (1,000,000 cells/mL, complete DMEM + 10% DMSO).

Approximately four days prior to fluorescence-activated cell sorting (FACS), one tube per cell line was thawed into 10 mL complete DMEM in a 10-cm dish. Additionally, one tube of unmodified Cre- or dTAG-MEFs was thawed as a negative control. On the day of sorting, cells were trypsinized and passed through a 40 μM cell strainer. Cells were sorted on either a BD FACSAria sorter using a 594 nm laser and a 610/20 bandpass filter, or on a MoFlo Astrios EQ sorter equipped with a 592 nm laser and a 620/29 bandpass filter at the Harvard Medical School Department of Immunology's Flow Cytometry Core facility. Gating was set such that fewer than 0.1% of the cells in the negative control would be captured.

***Ogt* knockout in Cre-MEFs**

One day prior to knockout, Cre-MEFs were plated in complete DMEM, which was supplemented with 1 μM 4-hydroxytamoxifen (4-HT) to begin knockout. After 24 hours, media was replaced with complete DMEM and changed regularly until cells were collected for lysis and

analysis.

OGT degradation in dTAG-MEFs

One day prior to degradation, dTAG-MEFS were plated in complete DMEM, which was supplemented with 500 nM dTAG-13 to begin degradation. Media was replaced daily with complete DMEM + 500 nM dTAG-13 until cells were collected for lysis and analysis.

OGT dimerization binding assay

Aliquoted proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates. Protein concentrations were then re-calculated by A_{280} (and A_{340} or A_{495} for CoraFluor- or AF488-labeled OGT). Master mixes were prepared containing UDP (0.1 mM final), DTT (2.5 mM final), and CoraFluor-1-tagged OGT (1 nM) in HBS FRET buffer (20 mM HEPES, pH 7.4, 20 mM $MgCl_2$, 150 mM NaCl, 0.5 mg/mL bovine serum albumin (BSA), 0.005% (v/v) Tween-20). AF488-tagged OGTs were diluted to the target concentration in HBS supplemented with 1 mM DTT and 0.3% (v/v) Tween-20, then dispensed in 1:2 dilution series across each column using a D300e digital dispenser (Tecan) fitted with a T8+ dispensehead (HP), with the final row having no AF488-OGT. Plates were covered in sealing tape (Perkin-Elmer), briefly centrifuged to eliminate bubbles (1500 g, RT, 1 minute), and incubated in the dark at room temperature for 4 hours. Plates were then moved to 4°C for overnight incubation, after which they were warmed to room temperature over a period of 4 hours prior to data collection.

TR-FRET measurements were acquired on a Perkin-Elmer EnVision plate reader using 340nm excitation with a 407nm/505nm advanced dual mirror and 520nm (25nm bandwidth, FITC) and 495nm (10nm bandwidth, Tb) filters, using a 50 μs delay and 300 μs integration. FRET ratio was taken as the ratio of 520nm/490nm emission on a per-well basis, then averaged across replicates. Data was fitted to a 'one site – total binding' model in GraphPad Prism 10 (Equation below). Due to usage of directly-labeled OGTs, data was not corrected or normalized to a model of fractional binding.

$$Y = \frac{\mathrm{B}_{max} * X}{\mathrm{K}_D + X} + NS * X + Background$$

In vitro cleavage of HCF3R-EEE by OGT variants

Purified proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates. Protein concentrations were re-quantified by A_{280} on Nanodrop. Reactions were prepared on ice in 10 μL volumes containing 5 μM OGT, 50 μM HCF3R-EEE, 1 mM DTT, and 0.01 units/μL Quick CIP (NEB), in HBS supplemented with 20 mM $MgCl_2$. Controls were created that lacked either OGT or HCF3R. UDP-GlcNAc (Promega) was added to 1 mM and samples were incubated at 37°C with 300 rpm mixing for 3 hours. Following incubation, reactions were quenched by addition of 1 volume 2x reducing Laemmli buffer, then boiled at 95°C for 10 minutes. 14 μL of each sample was loaded and run on a 4-20% Tris/Glycine SDS-PAGE gel for 40 minutes at 180V. Gels were fixed in fixation buffer (50% methanol, 10% acetic acid, aqueous) for 1 hour at room temperature, then stained in Coomassie stain (0.1% w/v Coomassie brilliant blue R-250 in fixation buffer) for 1 hour at room temperature and destained fully in destain buffer (30% methanol, 10% acetic acid, aqueous) overnight at room temperature. Stained gels were imaged using the 'visible' setting on an Azure C600 imager using automatically-determined exposure times. Resultant images were contrast- and brightness-adjusted in FIJI for visibility.

In vitro glycosylation of HCF3R-SAA by OGT variants

Purified proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates. Protein concentrations were re-quantified by A_{280} on Nanodrop. Reactions were prepared on ice in 15 μL volumes containing 1 μM OGT, 50 μM HCF3R-SAA, 1 mM DTT, and 0.01 units/μL Quick CIP (NEB), in HBS supplemented with 20 mM $MgCl_2$. Controls were created that lacked either OGT or HCF3R. UDP-GlcNAc (Promega) was added to 1mM and samples were incubated at 37°C with 300 rpm mixing for 1 hour. Following incubation, reactions were quenched by addition of 1 volume 2x reducing Laemmli buffer, then boiled at 95°C for 10 minutes. 14 μL of each sample was loaded onto two separate 4-20% Tris/Glycine SDS-PAGE gels and run at 180V for 60 minutes.

One gel was stained in InstantBlue stain according to the manufacturer's protocols to assess for uniformity of protein loading across samples. The other gel was transferred to a nitrocellulose membrane and blotted using pan-*O*-GlcNAc antibodies as described above.

Incucyte live cell imaging of Cre-MEFs expressing OGT variants

Cells were plated in complete DMEM at 5000 cells/mL in 100 μL aliquots in 6 technical replicates per cell line in a 96-well plate (black, glass bottom), with the outer rim wells filled with 200 μL PBS. 24 hours after plating the media was replaced with complete DMEM supplemented with 1 μM 4-HT, and cells were replaced in the Essen Incucyte Zoom instrument set to image every 4 hours. Every 24 hours thereafter, cells were removed, media was exchanged for fresh complete DMEM, and cells were replaced in the instrument. If any cell line reached 85% confluence at the time of media replacement, those cells were split at a 1:3 ratio and replated.

Cell confluence was obtained using the basic analyzer tool within Incucyte Zoom's software (v.2016b). Confluence was computed on a per-well basis by comparing to that well's initial confluence value at the first data collection. The base-2 log of this value was used to determine cell doublings per well. If cells were split, the confluence of the well post-split was artificially raised to the value pre-split. Analysis was carried out using python scripts and GraphPad Prism.

CellTiter-Glo analysis of Cre-MEFs expressing OGT variants

Cells were plated in complete DMEM at 5000 cells/mL in 100 μL aliquots in 12 technical replicates per cell line in a 96-well plate (white, opaque bottom), with the outer rim wells filled with 200 μL PBS. 24 hours after plating the media was replaced with complete DMEM supplemented with 1 μM 4-HT. Every 24 hours thereafter until analysis, media was replaced with complete DMEM.

At each timepoint (days 0, 1, 3, 5, 7, 9, 10), a plate was removed, cooled to RT, and assayed by the following protocol. PBS was removed from the left-most column of the plate, and 100 μL of each in a series of ATP standards (0, 0.001, 0.01, 0.1, 1 μM in DMEM) was added across the column. CellTiter-Glo reagents (Promega G7573) were prepared according to the manufacturer's protocols and added to each well on the plate. The plate was mixed at 300 rpm for 2 minutes at RT, then incubated at RT for 10 minutes without mixing. Data was collected using the luminescence setting with a 1 second integration time on a GloMax Explorer plate reader.

The ATP standard curves were used to convert luminescence to ATP concentrations on a per-plate basis. Data was corrected based on day 0 values on a per-cell line basis to account for slight differences in seeding density. Relative ATP (arbitrary units) was plotted at each time point.

Live cell confocal fluorescence microscopy of dTAG-MEFs expressing OGT variants

Data shown in Figure 2.13C was acquired by Dr. Sarah Potter during the course of her graduate work; details regarding image collection can be found in her thesis, as well as in her work published in *PNAS* in 2024.[108,136]

Co-immunoprecipitation (Co-IP) of OGT variants from dTAG-MEFs

dTAG-MEFs expressing TPR truncation OGTs (also WT and null) were seeded at 13,000,000 cells/dish in triplicate (quadruplicate for WT/mono/null experiments) 15-cm dishes in complete DMEM. After 12 hours, this media was supplemented with 500 nM dTAG-13. After 8 hours of dTAG treatment, cells were harvested by scraping into cold PBS, then pelleted by centrifugation (500 g, 4°, 5 minutes). Pellets were resuspended in minimal Pierce IP lysis buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% (v/v) IGEPAL CA-630, 5% (v/v) glycerol, 1x Halt protease/phosphatase inhibitor cocktail) and incubated for 10 minutes on ice to lyse, then clarified by centrifugation (21130 rpm, 4°, 10 minutes). Supernatants were transferred to new tubes, and protein concentration was estimated by BCA assay.

For each sample, 55 μL of packed M2 anti-FLAG magnetic beads were prepared by washing three times with 300 μL cold TBS, then loaded with 2,400 μg of total protein lysate. Samples were incubated overnight at 4° with end-over-end rotation. After overnight incubation, bead flowthrough was collected, then beads were washed twice with 400 μL cold TBS, then twice with 400 μL cold HBS, each time for 5 minutes at 4° with rotation. Elution was performed by addition of 200 μL 5% formic acid, rotating for 20 minutes at RT, then repeating. Elution fractions were combined and pushed through to mass spectrometry.

Elutions were dried by vacuum centrifugation, then resuspended in 100 μL HEPES (100 mM, pH 8.5) and digested in 1:100 trypsin at 37° for 16 hours. Trypsinized peptides were desalted using StageTips, dried by vacuum centrifugation, and resuspended in acetonitrile (5%) and formic acid (5%) for LC-MS^2 analysis.

Data were collected using an Exploris 480 mass spectrometer fitted with a Proxeon 1200 Liquid Chromatograph. Peptides were separated on a 100 μm inner-diameter microcapillary column packed with Accucore C18 resin (~25 cm, 2.6 μm/150Å); 1 μg of peptides were loaded onto the column for analysis.

Peptides were separated on an acetonitrile gradient (3-25% in 0.125% formic acid over 90 minutes at 520 nL/minute). Scanning began with an Orbitrap MS^1 spectrum using the following parameters: 60,000 resolution; 350-1350 Th scan range; "standard" automatic gain control (AGC) target; "auto" maximum injection time; 40% RF lens setting; centroid spectrum data type. The top twenty precursors were selected for MS^2 analysis consisting of HCD with the following parameters: 15,000 resolution; "standard" AGC; "auto" maximum injection time; 0.7 Th isolation window; 28 normalized collision energy (NCE); centroid spectrum data type. Unassigned and singly-charged species were excluded from MS^2 analysis and dynamic exclusion was set to 90 s.

Spectra were processed using in-house software based on Sequest. MS spectra were converted to mzXML format via a modified version of ReAdW.exe. The mouse uniport database plus additional bain sequences, concatenated with a reverse database composed of all proteins in reversed order, were used as a database for sequence searching and assignment, which were performed using a 50 ppm precursor ion tolerance. Product ion tolerance was set to 0.03 Th. Cysteine carbamidomethylation (+57.0215 Da) was set as a static modification, while methionine oxidation (+15.9949 Da) was set as a variable modification.

Peptide spectral matches (PSMs) were altered to a 1% false discovery rate (FDR) according to literature protocols,[137,138] and PSM filtering was performed using linear discriminant analysis considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy.[139] PSMs were identified, quantified, first collapsed to 1% FDR, then further collapsed to a protein-level FDR of 1%. Protein assembly was guided by principles of parsimony to produce the smallest set of proteins encapsulating the observed peptides. Spectral counts were extracted for analysis.

Tandem mass tagging (TMT) analysis of total proteins from dTAG-MEFs expressing OGT-WT or -mono

Following native lysis of dTAG-MEFs (see above) and protein quantification by BCA assay, 100 μg total protein from each sample was separately pushed through TMT labeling procedures, described here.

Proteins were reduced by addition of TCEP (Pierce neutralized TCEP) to 5 mM for 15 minutes at RT, alkylated by addition of iodoacetamide to 10 mM for 30 minutes in the dark at RT, and

quenched by addition of DTT to 5 mM for 15 minutes at RT.

Protein was extracted by methanol-chloroform precipitation as follows. Sample was first diluted in lysis buffer to 100 μL. 400 μL methanol was added, then 100 μL chloroform, then 300 μL water, with short (5 second) vortexes after each addition. Sample was centrifuged (14,000 g, RT, 1 minute), and both aqueous and organic layers were removed, yielding a 'pancake' of protein, which was washed with 400 μL methanol and centrifuged again (21,000 g, RT, 2 minutes). This time, the supernatant was removed and the pellet was resuspended in 100 μL 200 mM EPPS buffer (pH 8.5).

Resuspended samples were digested for 16 hours at RT with shaking by addition of 1:50 LysC protease. Then, trypsin was added at a 1:100 ratio and the reaction was incubated for 6 hours at 37°. 25 μL were combined with 7.5 μL anhydrous acetonitrile, then mixed with 5 μL TMT reagent (one of TMT-18plex), vortexed briefly, and incubated at RT for 1 hour. 2 μL of each sample was combined in 100 μL 1% formic acid, desalted by stage tip, and briefly analyzed by LC-MS to check digestion and labeling efficiency.

The remainder of the samples were quenched by addition of hydroxylamine to a 0.3% (v/v) concentration at RT for 15 minutes. Samples were combined, frozen, and evaporated on a SpeedVac, then acidified to a pH of 2-3 by addition of formic acid and loaded onto a SepPak cartridge before being fractionated by high pH reverse-phase HPLC from buffer A (10 mM ammonium bicarbonate, 5% acetonitrile, pH 8.0) using a linear gradient from 13% to 42% buffer B (10 mM ammonium bicarbonate, 90% acetonitrile, pH 8.0) over 60 minutes, on an Agilent 300Extend C18 column fitted onto an Agilent 1200 HPLC at 0.6 mL/minute. Samples were collected and dried via SpeedVac overnight.

Samples were cleaned by StageTip desalting into 100 μL buffer (1% formic acid, 70% acetonitrile), vacuum centrifuged to dryness, resuspended into 10 μL 5% acetonitrile with 5% formic acid, and analyzed by mass spectrometry, briefly summarized as follows. Mass spectrometry data were collected on an Orbitrap Fusion Lumos mass spectrometer fitted with a Proxeon EASY-nLC 1200 LC pump. Peptides were separated on a gradient of 4 to 25% acetonitrile in 0.125% formic acid over 2 hours at 525 nL/minute on a 30-cm long 100 μm Accucore (2.6 μm, 150Å) bead column. MS spectra were acquired at a 120,000 resolution. The 10 highest-intensity MS^1 ions were pushed to MS^2 analysis. Isolation width was set to 0.4 Da, and isolated precursors were fragmented by

CID at an NCE of 35%. After MS^2 spectrum acquisition, SPS MS^3 spectra were collected on the 10 highest-intensity MS^2 ions. SPS-MS^3 precursors were fragmented by HCD at an NCE of 65% and analyzed on the Orbitrap.

Spectra were processed using in-house software based on Sequest. MS spectra were converted to mzXML format via a modified version of ReAdW.exe. The mouse uniport database plus additional bain sequences, concatenated with a reverse database composed of all proteins in reversed order, were used as a database for sequence searching and assignment, which were performed using a 50 ppm precursor ion tolerance. Product ion tolerance was set to 0.03 Th. Cysteine carbamidomethylation (+57.0215 Da) was set as a static modification, while methionine oxidation (+15.9949 Da) was set as a variable modification.

Peptide spectral matches (PSMs) were altered to a 1% false discovery rate (FDR) according to literature protocols,[137,138] and PSM filtering was performed using linear discriminant analysis considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy.[139] PSMs were identified, quantified, first collapsed to 1% FDR, then further collapsed to a protein-level FDR of 1%. Protein assembly was guided by principles of parsimony to produce the smallest set of proteins encapsulating the observed peptides. Spectral counts were extracted for analysis.

Data were processed using spectral counts to determine enrichment over the null control and statistical significance. Pearson correlation was performed on GraphPad Prism to determine R values, which were used to eliminate the most outlying sample out of each group of quadruplicate datasets. From the remaining triplicates, hits were defined as having at least 4-fold enrichment over the null dataset, with a p-value of less than 0.05. DAVID (Database for Annotation, Visualization, and Integrated Discovery) analysis was performed on the hits to describe functional annotations of key words and gene ontology (GO) terms for biological processes.[140]

Data processing for Co-IP from dTAG-MEFs expressing OGT-WT or TPR truncations

To determine hits, average peptide counts across all three replicates were calculated for each identified protein. Enrichment was determined by dividing the average peptide counts for a sample over the corresponding negative control. Peptide counts of 0 in the negative control were artificially set to 0.33 (the smallest possible non-zero peptide average) to enable calculation of

enrichment.

Hit cutoffs were set at ≥3-fold enrichment over negative control, with only proteins with ≥2 average peptides across replicates considered. Hit proteins were identified for each OGT variant used as bait. Gene ontology (GO) terms were identified using Uniprot. Known OGT interactors and substrates were identified using OGT-PIN and the O-GlcNAc Database, respectively.[35,63] The STRING database was used to identify functional relationships between proteins using a medium confidence cutoff of 0.4.[141]

Significance analysis was performed using custom python scripts using the pandas, NumPy, and UpSetPlot libraries in python3, using the non-negative integer peptide counts in triplicate for each condition as the dataset. Analysis was performed on each identified protein for each condition, using the null dataset as a point of comparison. First, an f-test was calculated to determine equality of variances. If the f-test statistic was < 0.05, Welch's t-test was performed without assuming equal population variances; otherwise, an independent 2-sample t-test assuming equal variances was performed. The t-test outputs were set as p-values alongside the enrichment over the null dataset and the base-2 log transform of enrichment. P-value cutoffs were set at < 0.05.

Values of NaN/inf/-inf/zero were individually corrected as follows. If both peptide averages were zero, enrichment and $\log_2$(enrichment) were set to 0, and p-value was set to 1. If one peptide was zero and one was nonzero, enrichment was set as the absolute value of the nonzero peptide count average.

UpSet plots were produced from datasets trimmed by enrichment and p-value cutoffs.[142]

Data processing for Co-IP from dTAG-MEFs expressing OGT-WT or -mono

The above data processing protocol was repeated, using hit cutoffs of greater than 2-fold enrichment and p-value less than 0.05.

Preparation of HeLa S3 cell extracts

HeLa S3 cells were grown to 2,000,000 cells/mL in 1L spinner flasks, then centrifuged (1000 rpm, 4°, 5 minutes) to pellet and flash frozen. Pellets were thawed on ice then resuspended in swelling buffer (40 mM HEPES, pH 7.5, 5 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, with 1 Roche

cOmplete EDTA-free protease inhibitor tablet per 25 mL buffer) to around 80% of the pellet volume. Energy mix (7.5 mM creatine phosphate, 1 mM ATP, 0.1 mM EGTA, 1 mM $MgCl_2$, pH 7.7) was added to 1x from a 50x stock, and the suspended cells were then incubated under nitrogen gas at 1000 psi for 30 minutes at 4°. After pressure was released, lysate was collected, and the process was repeated one more time. Obtained lysates were clarified by centrifugation (15000 g, 4°, 30 minutes), and the resultant supernatant was aliquoted, flash frozen in liquid nitrogen, and stored at -80°. All extracts used in this dissertation were prepared by Dr. Cassie Joiner.

***In vitro* glycosylation of HeLa S3 cell extracts by OGT variants**

Extracts were thawed on ice, then centrifuged at 45000 rpm for 20 minutes at 4°C to pellet cell debris and aggregates. Supernatant was removed and aliquoted into fresh tubes in 24 μL volumes. OGT variants were thawed on ice, then centrifuged at 21130 rpm for 10 minutes at 4°C to pellet aggregated proteins and re-quantified by A_{280} on Nanodrop. 5x concentration mixes of each OGT were constructed on ice in 6 μL volumes containing 5 (D2A vars.) or 12.5 μM (5N5A vars., TPR truncations) OGT, 5 mM DTT, and 5 mM UDP-GlcNAc in HBS supplemented with 20 mM $MgCl_2$.

6 μL of each 5x mix was added to 24 μL cell extract and pipetted briefly to mix. Samples were incubated at 37°C with 300 rpm mixing. At timepoints of 1, 5, and 15 minutes, 8 μL of each sample was removed, combined with 8 μL 2x reducing Laemmli buffer, and boiled at 95°C for 10 minutes. Samples were loaded at 7 μL per well on a 4-20% Tris/Glycine SDS-PAGE gel (Bio-Rad), which was run for 50 minutes at 180V. Protein was transferred to PVDF membranes and blotted using pan-*O*-GlcNAc, polyhistidine-HRP, and GAPDH antibodies as described above.

For glycosylation by TPR truncation variants of OGT, only timepoints at 5 and 15 minutes were collected. GAPDH western blots were not performed for these experiments either. The glycosylation reactions of HeLa S3 cell extracts by OGT 5N5A and D2A variants were performed by Dr. Cassie Joiner.

***In vitro* glycosylation of OGT substrates (TAB1, CAMKIV, CARM1) by OGT variants**

Proteins were thawed from frozen on ice, then centrifuged (21130 rpm, 4°, 10 minutes) to pellet aggregates and re-quantified by A_{280}. Reaction mixtures were constructed containing OGT (2.5 μM), substrate (15 μM), and UDP-GlcNAc (1 mM) in OGT reaction buffer (20 mM Tris, pH 7.4,

150 mM NaCl, 20 mM $MgCl_2$, 1 mM DTT) in 20 μL volumes, then incubated at 37° for 90 minutes with 350 rpm shaking. Following incubation, reactions were quenched by addition of 20 μL 2x reducing Laemmli SDS-PAGE loading buffer, boiled for 10 minutes, then separated by SDS-PAGE as described above in this methods section. Reactions were analyzed by western blot, and total protein was assessed by InstantBlue staining to confirm evenness of loading. Western blots were conducted using anti-mouse IgG-680RD near-IR-conjugated secondary antibodies (LI-COR), which necessitated 3x TBS washes (no Tween-20) following staining.

***In vitro* TAB1 glycosylation kinetic assay**

Proteins were thawed from frozen on ice, then centrifuged (21130 rpm, 4°, 10 minutes) to pellet aggregates and re-quantified by A_{280}. For each reaction, OGT (1 μM), TAB1 (20 μM), and UDP-GlcNAc (1 mM) were incubated in reaction buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 20 mM $MgCl_2$, 1 mM DTT) to a total volume of 90 μL. Reactions were incubated at 37° with 350 rpm shaking. At set timepoints (TAB1-WT: 0/1/2.5/5/10/15/20/30 minutes; TAB1-S395A: 0/1/5/10/15/20/30 minutes), 10 μL of each reaction was removed and quenched by addition of 10 μL 2x reducing Laemmli SDS-PAGE loading buffer, then boiled for 10 minutes. All samples were separated by SDS-PAGE on two separate gels, with one gel being stained by InstantBlue, and the other western blotted for *O*-GlcNAc levels using the CTD antibody and anti-mouse IgG-680RD IRDye secondary. Activity was determined by measuring densitometry of the full-length glycosylated band in FIJI. Substrate turnover metrics were determined by taking the slope of the fitted lines for the linear range of the activity curves, then calculating relative turnover. Data are an average of two replicate experiments.

Chapter 3

Development of an OGT binding assay and discovery of a polypeptide inhibitor with picomolar affinity

The work described in this chapter has been published in the following paper, on which I am the first author:

Hammel, F.A.; Payne, N.C.; Marando, V.M.; Mazitschek, R.; Walker, S. "Identification of a Polypeptide Inhibitor of *O*-GlcNAc Transferase with Picomolar Affinity." *Journal of the American Chemical Society* **2024**, *146* (38):26320-26330.[109]

In this chapter, I harness insights built up over more than a decade of work in the Walker lab to quantify the affinity of interaction between OGT and the HCF-1 proteolytic repeats. Using the newfound knowledge of the picomolar affinity interaction between these binding partners, I develop a cellular OGT inhibitor based on a three-repeat scaffold, which I then modify to enable rapid on-and-off switching of OGT inhibition in cells. Most of the experimentation was performed by myself; however, I was greatly assisted in the design and initial execution of the binding assay by Dr. N. Connor Payne, a former graduate student in Ralph Mazitschek's research group at Massachusetts General Hospital. Additionally, Dr. Victoria Marando, a postdoctoral fellow in the Walker lab, generated and sorted the HCT116 cell line expressing degron-tagged HCF3R under a doxycycline-inducible promoter. I also want to thank Dr. Cassandra (Cassie) Joiner, a former

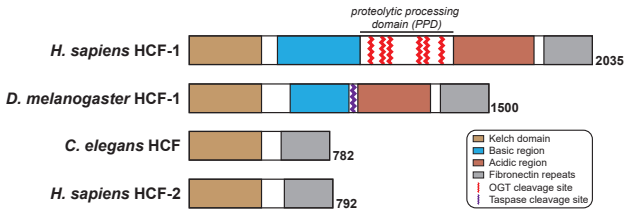


Figure 3.1: Domain structure of HCF-1 homologs from *H. sapiens*, *D. melanogaster*, and *C. elegans*, as well as *H. sapiens* HCF-2.

postdoctoral fellow in the Walker lab, for giving me the original idea to develop a binding assay and for encouraging me to move forward into this project.

Background work supporting the genesis of this project was conducted primarily in the Walker lab by a plethora of former students and postdoctoral fellows, whom I will list here and cite as necessary in the text of this chapter: Dr. Michael Lazarus, whose structural work on the OGT-HCF1R interaction was necessary for understanding the role of TPR residues in binding substrates; Dr. John Janetzko, whose initial design of HCF3R constructs and intense interest in the chemistry of HCF-1 cleavage motivated a lot of this work; and Dr. Zebulon Levine and Dr. Sarah Potter, who developed the initial Cre- and dTAG-MEF cell lines used as controls and background cell lines in some experiments shown in this chapter.

3.1 The HCF-1 proteolytic motif is highly conserved across *Chordata*

OGT-mediated HCF-1 cleavage occurs at a series of proteolytic repeats in the central unstructured proteolytic processing domain (PPD) of the protein.[46,52,53] However, the presence of these repeats is only conserved within *Chordata* (Figure 3.1). In other organisms, the HCF-1 homolog is either not cleaved (*e.g.*, *C. elegans*, whose HCF-1 homolog more closely resembles human HCF-2 and lacks the PPD), or is cleaved by another process (*e.g.*, *D. melanogaster*, whose HCF-1 homolog is cleaved by the protease Taspase).[143–147]

The HCF-1 proteolytic repeat is defined as a 26-residue sequence with the glycosite glutamate residue at position 10.[45,46] Alignment of the six canonical repeats from human HCF-1 revealed two conserved motifs flanking the E10 position: an N-terminal 'CSNPPC' motif (res. 4-9), and a C-terminal threonine-rich region containing six threonine residues (res. 14-22) (Figure 3.2A).[148]

Previous mutagenesis studies on the glycosite region showed that the cysteine at position 9 is not strictly essential – arginine, lysine, and tyrosine mutants were still able to be robustly cleaved by OGT *in vitro*, while other mutants (methionine, alanine, serine) maintained a low amount of activity.[124] However, the threonine-rich region – which engages OGT's aspartate ladder – is required for cleavage.[148] Additionally, while any mutation to the E10 residue prevents cleavage because the ability of the glycosylated glutamate to decompose to an internal pyroglutamate is necessary for that function, mutation to aspartate enables a side reaction, producing an isoaspartyl peptide from decomposition of a succinimide intermediate.[149]

Conservation of the HCF-1 proteolytic repeat motif is high among homologs in *Chordata* (Fig. 3.2B, see Methods for more detailed computational methodology). In HCF-1 isoforms from 400 unique species, I identified 27,201 glutamate residues, of which 2,973 (10.9%) were immediately preceded by a cysteine, and 2,663 (9.8%) were followed by threonines in the +7, +9, and +11 positions, resembling the threonine-rich region. Alignment of both of these groups of potential repeat motifs reconstituted the full HCF-1 repeat, with high conservation of the 'CSNPPC' motif and Thr-rich region, as well as a nearly invariant region surrounding the E10 glycosite ('CETHET'). This suggests that the conserved elements correlate at high rates. Restricting this analysis to just 229 mammalian HCF-1 homologs revealed a similar pattern of conservation (Fig. 3.2C). In both analyses, the cysteines in positions 4 and 9 are highly conserved; this is of particular note, as cysteines with high conservation typically play roles in reactivity or binding (*e.g.*, disulfide formation);[150] however, no such interactions have been reported or theorized for HCF-1.

Upon comparison of the HCF-1 repeats and OGT's glycosylation substrates, a few patterns emerge. For one, while the 'CX_4C' motif preceding the glutamate glycosite is almost universally conserved in HCF-1 proteins, cysteines are rarely observed preceding normal serine/threonine glycosites in OGT's substrates.[60,61,112] Additionally, residues with hydroxyl sidechains (S/T) are found in high abundance C-terminal to the glycosite for OGT's glycosylation substrates; however, they are not as universally enriched as the threonines in the HCF-1 repeats, and their positions are not so strictly conserved. One other key difference is the presence of prolines immediately C-terminal to the glycosite, which HCF-1's repeats are lacking entirely.[151,152] The absence of prolines allows the HCF-1 peptide backbone to maintain optimal contacts with the TPR asparagine sidechains, facilitating the interactions necessary for HCF-1 cleavage.[46,60]

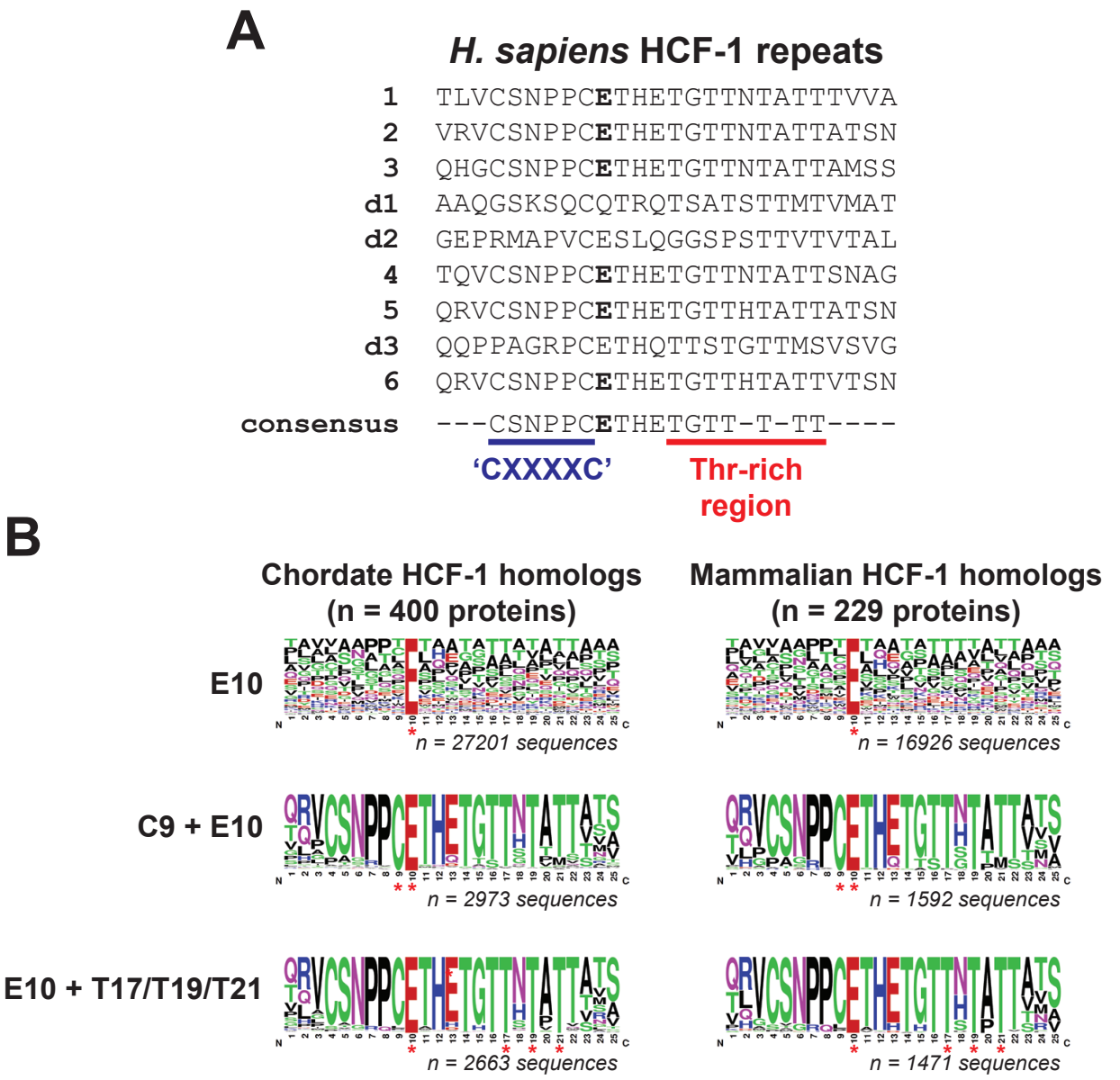


Figure 3.2: Conservation of the HCF-1 repeat. (A) Alignment of all six repeats (1-6) and three pseudorepeats (d1-d3) from *H. sapiens* HCF-1, with consensus CX_4C and threonine-rich regions annotated. (B) Sequence conservation of the 26-mer repeat in chordates (left) and mammals (right) using different criteria: presence of E10 (top), presence of C9 and E10 (middle), and presence of E10 and threonines T17, T19, and T21 (bottom). Red asterisks indicate residues made artificially invariant for analysis.

3.2 In mammals, the full HCF-1 proteolytic processing domain is also conserved

In humans, the HCF-1 PPD consists of six canonical repeats, which are cleaved upon *O*-GlcNAcylation, and three degenerate repeats, which lack either the conserved E10 glycosite or elements of the threonine-rich region and are not cleaved by OGT (Fig. 3.2A, 3.3A).[48,49,53] The six canonical repeats are arranged semi-symmetrically into two groups of three separated by a large unstructured region (Fig. 3.3A), where the two outer repeats in each group (1 & 2, 5 & 6) are separated by a large (>25 residues) spanning region, whereas the two inner repeats in each group (2 & 3, 4 & 5) are separated by three or fewer residues each.

Conservation of this 6-repeat structure is high among mammals and lower among other chordates (Fig. 3.3BC). Using a restrictive definition of a proteolytic repeat (E10 + 'C4-XXXX-C9' + T17 or T19 or T21), 178 out of 400 (44.5%) chordate HCF-1 homologs have six repeats (Fig. 3.3B); removing the C4 and T11 residues from the requirements shifts the histogram right, and by either definition many chordate homologs have >8 repeats. In contrast, restriction to mammalian species condenses the histogram towards a six-repeat structure, with 161/229 mammalian homologs (70.3%) possessing six repeats by the more restrictive definition (Fig. 3.3C). This means that only 17 out of the 171 non-mammals (9.9%) in the initial chordate dataset have six repeats, suggesting that the PPD architecture is much more highly conserved in mammals than in non-mammals. A separate analysis I conducted on 2068 pBLAST hits from the search sequence 'CETHET' in the phylum *Chordata* identified that the 6-repeat architecture is highly conserved across mammals, excluding the order *Monotremata* (platypus and echidnas), whose members have at least 13 repeats. However, in that analysis I found that other classes of chordates including *Aves* (birds), *Actinopteri* (ray-finned fish), and *Reptilia* (reptiles) often have more than 10 well-defined proteolytic repeats in their HCF-1 proteoforms, with one species of bird (*Pseudopodoces humilis*, Tibetan ground-tit) having 49 canonical HCF-1 repeats.[153]

Among the 161 mammalian HCF-1 homologs with six putative repeats, the distance of the spanning regions in between repeats is highly conserved as well, excluding the lengthy (>150 aa.) span between repeats 3 and 4 (Fig. 3.3D). This conservation extends to the sequence level as well, as frequency plots of the spans between repeats 1 and 2, 2 and 3, 4 and 5, and 5 and 6 (which

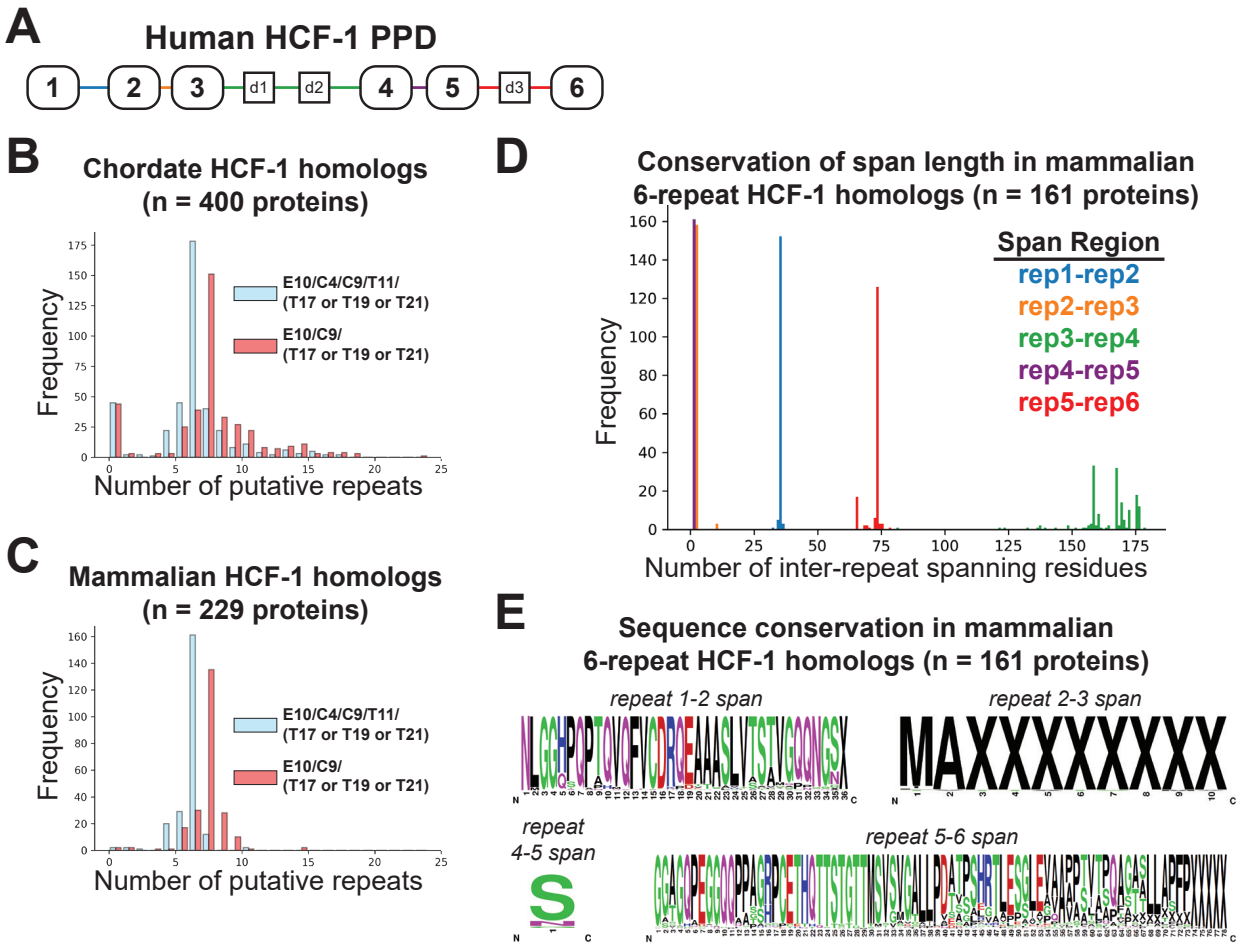


Figure 3.3: Conservation of the HCF-1 proteolytic processing domain (PPD). (A) Schematic of the human HCF-1 PPD. Spanning regions are colored according to the color scheme in Figure 3.3D. (B,C) Histograms of number of predicted repeats in (B) chordates and (C) mammals containing either E10, C4, C9, T11, and one of T17/T19/T21 (blue); or E10, C9, and one of T17/T19/T21 (red). (D) Histogram showing conservation of the spanning lengths between repeats in mammalian HCF-1 homologs with six repeats. (E) Sequence level conservation of spanning regions between repeats in mammalian HCF-1 homologs with six repeats.

contains pseudorepeat d3) indicate a high degree of conservation (Fig. 3.3E).

All of this is to say: despite the non-essentiality of HCF-1 cleavage for cellular proliferation in both our Cre-Lox and dTAG MEF systems,[69] the high degree of conservation of the HCF-1 repeats and of the entire PPD in mammals suggests mechanistic importance. And thus, further biochemical and biophysical investigation of the interaction between OGT and the HCF-1 PPD would be a useful venture to gain insights into the roles that binding and cleavage of the PPD may play.

3.3 HCF-1 cleavage can be monitored *in vitro* using substrates consisting of the first three HCF-1 repeats

As HCF-1 is a lengthy protein (~2000 residues) with significant portions predicted to be unstructured, heterologous purification of full-length HCF-1 from *E. coli* or other organisms would be very challenging. In order to detect and track HCF-1 cleavage *in vitro*, truncations were made to improve stability. Since only the PPD of HCF-1 is required for its cleavage,[45,46,129] it made sense to only include that domain in the construct. Previously, the Walker lab has utilized HCF3R constructs consisting of the HCF-1 sequence containing the first three repeats with an N-terminal polyhistidine tag (Fig. 3.4A).[124] Early on in my rotation, I installed a C-terminal tri-tyrosine (YYY) tag to increase the extinction coefficient at 280 nm, making quantitation possible. Cleavage of the wildtype HCF3R-EEE construct, which contains E10 glycosites in all three repeats, is easily monitored by SDS-PAGE and Coomassie staining (Fig. 3.4B). Furthermore, mutation of the glycosites to serines enables glycosylation without cleavage (observable by western blotting, Fig. 3.4C), and mutations to alanine prevent any reaction catalyzed by OGT. HCF1R constructs consisting of a single repeat have been cloned using repeat 2, however these are not robustly cleaved *in vitro* unless the next five C-terminal residues are added, producing $HCF1_{short}$.[124] By itself, $HCF1_{short}$ is useful for LC/MS-based assays, but if fused to a larger protein (*e.g.*, glutathione S-transferase or maltose-binding protein) it can also provide a similar SDS-PAGE cleavage readout to HCF3R. Similarly, "HCF-1rep1" constructs consisting of the first repeat plus the spanning region between repeats 1 and 2 are easily cleaved *in vitro*,[46,53] suggesting that the 26-mer repeat alone is insufficient for cleavage, and some factors outside of the core repeat serve to promote

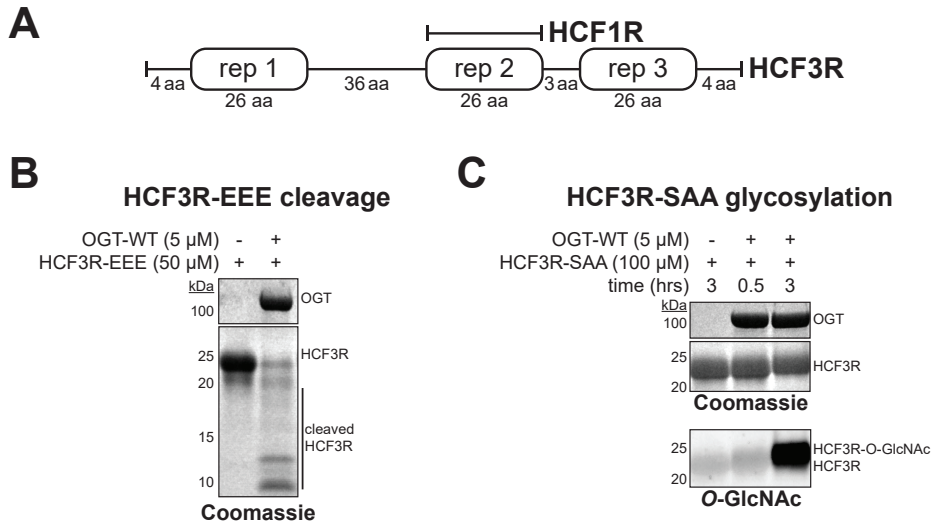


Figure 3.4: HCF3R polypeptides are model substrates for OGT. (A) Schematic of single-repeat (HCF1R) and triple-repeat (HCF3R) substrates. (B) Cleavage of HCF3R-EEE can be tracked by SDS-PAGE and Coomassie staining. (C) Glycosylation of HCF3R-SAA can be tracked by western blotting.

OGT-mediated proteolysis.

3.4 Existing tools used to profile OGT-substrate and OGT-ligand interactions lacked quantitative capabilities

Prior to the start of the work detailed in this chapter, a variety of assays to profile OGT-ligand interactions were described in the literature. In 2005, the Walker lab reported a fluorescence polarization (FP) assay using a fluorescently-derivatized UDP-GlcNAc, which was used to identify small molecule inhibitors that bound OGT's active site.[154] This work ultimately led to the development of the OSMI series of OGT inhibitors (see Chapter 1).[93,94] This assay was expanded upon by screening a biased library of quinolinone-6-sulfonamide compounds to identify preferred building blocks that would later become the first OSMI compounds. Craig Thomas and Damien Duveau at the NIH carried out additional medicinal chemistry on the OSMI scaffold, which led to improved activity. While this fluorescent ligand displacement assay was useful as a tool for screening hits and was capable of outputting quantitative binding data for OGT-ligand interactions, it had a limited dynamic range due to the inherent lack of sensitivity for FP assays. Moreover, because signal relied on displacement of UDP-GlcNAc, it was only useful for screening active site binders

that were competitive with respect to UDP-GlcNAc and could not be used to screen OGT-substrate interactions.

Co-immunoprecipitation (co-IP) and proximity-based labeling methods such as TurboID and BioID have been used to assay OGT's interactions with its substrates. Using a co-IP strategy followed by Western blotting, some information was obtained about the requirements of the HCF-1 repeats for OGT binding; however, the results provided only qualitative data.[46,53] TurboID and BioID assays consisting of modified biotin ligases fused to OGT in cells have also been used to generate OGT interactomes,[155–157] but the results are convoluted by the mixture of substrates and non-substrate interactors in the resultant interactome datasets, and are also only qualitative.

Other groups have drawn conclusions about substrate binding from assaying glycosyltransferase activity. The Vocadlo lab used biotinylated substrates and fluorophore-linked UDP-GlcNAc in a plate-based assay to read out kinetics of glycosylation directly.[158] A ratiometric approach based on a similar design was also taken by Jiaoyang Jiang's research group at the University of Wisconsin: they synthesized GlcNAc electrophilic probes (GEPs) which can either glycosylate substrates that bind in the active site or crosslink directly to OGT in the absence of substrate binding.[159] By clicking on fluorescent probes post-reaction and comparing the ratio of fluorescent substrate to fluorescent OGT, they were able to determine the reactivity of various OGT mutants towards different substrates.

Most recently, after the bulk of the work presented in this chapter had been performed, the Vocadlo and Jiang groups separately developed phage-displayed peptide libraries to enrich for OGT-binding peptides.[55,56] Strikingly, both groups landed upon small peptides with the sequence x-P-x-Y-x (note that the groups found different preferences for flanking residues, likely due to differences in the encoded peptide libraries) which bound OGT's intervening domain (IntD), a cryptic interface in between the two lobes of its catalytic domain. They then used fluorescence polarization, microscale thermophoresis, and biolayer interferometry assays with fluorophore-labeled variants of their initial hit peptides to determine binding affinity, and separately determined that their peptides had the capacity to inhibit OGT, despite binding outside of its active site.

While all these tools were useful, most lacked the ability to quantify the interactions between OGT and its substrates or interacting partners. The IntD-peptide binding assays provided quantitative data, but for bespoke peptides targeted at a specific region of OGT; these assays lack

utility for characterizing the affinities of known OGT substrates or interactors.

3.5 Development of a quantitative TR-FRET binding assay to measure the affinity between OGT and its substrates at equilibrium

At the start of this project, it was clear that the OGT field needed a quantitative substrate binding assay. Too often in the literature, glycosyltransferase activity has been directly linked to binding affinity, a faulty assumption. Moreover, we observed increased glycosyltransferase activity for OGT-D2A mutants (see Chapter 2.5), where loss of hydrogen bonding interactions (*i.e.*, lower affinity for some substrates) resulted in increased glycosylation, a finding that highlights the dangers in assuming that substrate binding affinity correlates with activity.[160]

As I began designing my thesis project, the question of what binding assay technology I should utilize was central. Each proposed assay type had clear benefits and drawbacks: traditional ELISA is relatively simple to set up but would be time-consuming to execute due to the repeated washing steps; SPR and BLI provide robust data with kinetic on- and off-rates but throughput suffers; and TR-FRET provides high sensitivity and medium/high-throughput capabilities, but the necessary lanthanide probes were previously cost-prohibitive. Fortunately, around the beginning of my Ph.D. work, Ralph Mazitschek's group developed CoraFluor probes (Fig. 3.5A).[135] These terbium cryptate species provide a replacement for commercial TR-FRET donors such as Lumi4-Tb, enabling cost-effective TR-FRET assay development in academic environments. Moreover, the route for CoraFluor synthesis enables functionalization with a variety of reactive handles, including pentafluorophenyl (pfp) esters and HaloTag chloroalkanes, allowing for covalent labeling of target proteins.

To set up a basic TR-FRET binding assay, three components are necessary: a donor OGT, an acceptor probe, and a compatible buffer solution. For the donor, I nonspecifically labeled anti-polyhistidine tag antibodies with CoraFluor-1-*pfp* and incubated this with polyhistidine-tagged wildtype OGT. As an initial acceptor, Connor Payne from the Mazitschek lab prepared OSMI-4-FITC, an OSMI-4 derivatized with a $(PEG)_2$-fluorescein moiety at its solvent-exposed carbonyl (Fig. 3.5B). We anticipated from crystal structures of OSMI-bound OGT that this position was amenable to significant modification,[95] and synthesis from the initial OSMI-4-ethyl compound

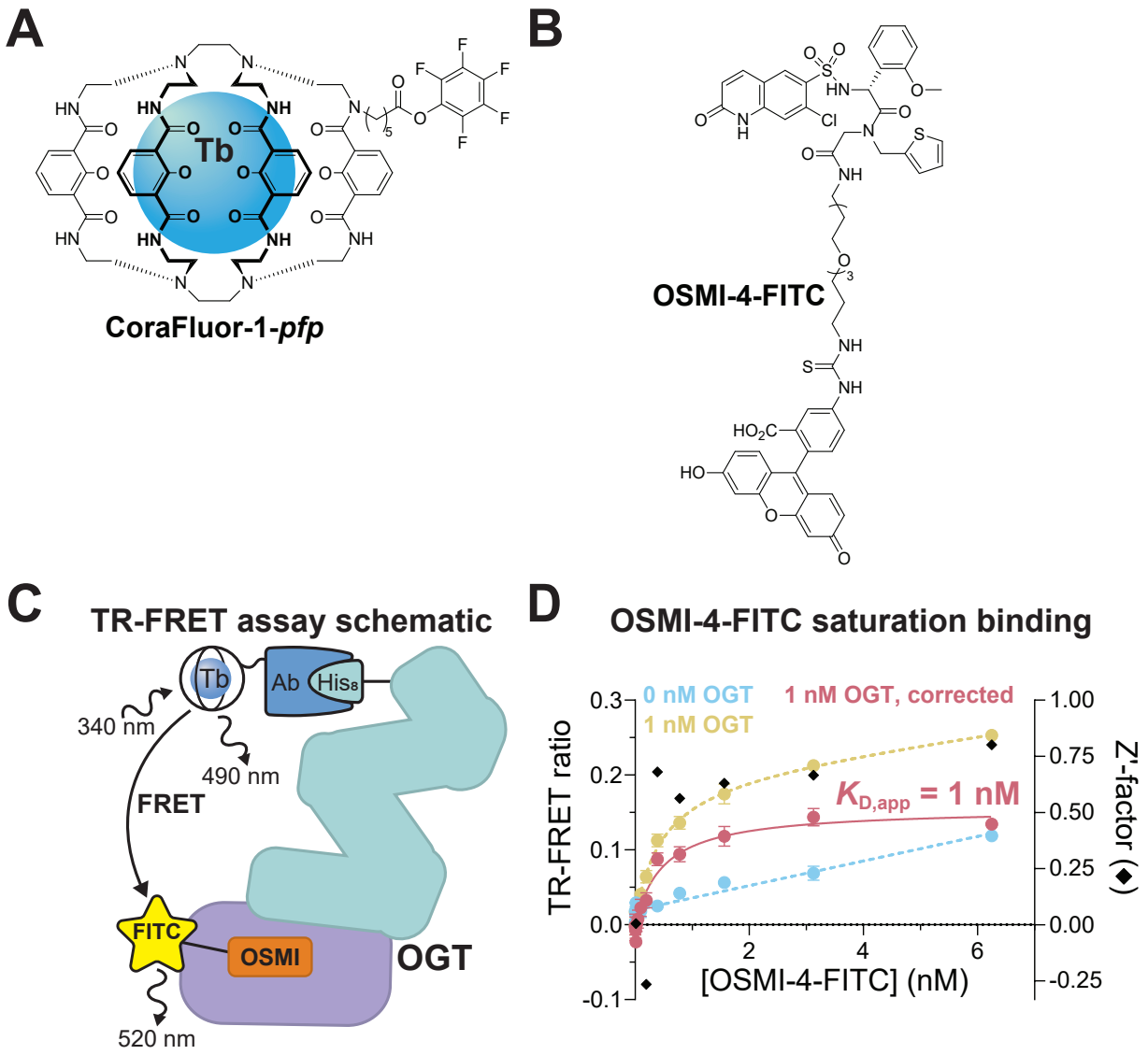


Figure 3.5: Development of a TR-FRET binding assay for OGT-ligand interactions. (A) Structure of the terbium cryptate pentafluorophenyl (PFP) ester CoraFluor-1-*pfp*.[134] (B) Structure of the fluorescent TR-FRET acceptor probe OSMI-4-FITC. (C) Schematic for a TR-FRET assay to measure the affinity of the OGT/OSMI-4-FITC interaction. (D) Results from the OGT/OSMI-4-FITC binding assay. Uncorrected signal (yellow) was corrected by subtraction of background (blue) to yield a corrected curve (red) representing a $K_{D,app}$ of 1 nM. Z′-factor was plotted by concentration in black diamonds. Data are mean ± s.d. of three replicates, and are representative of at least two independent experiments.

was straightforward, following a general peptide coupling scheme. We also anticipated that this would bind with relatively high affinity, as OSMI-4's K_D was estimated to be $\tilde{8}$ nM by microscale thermophoresis.[95] Lastly, I modified the standard OGT reaction buffer (20 mM Tris, pH 7.4, 20 mM $MgCl_2$, 150 mM NaCl) by swapping the buffering salt for HEPES, adding 0.5 mg/mL bovine serum albumin (BSA), and 0.005% (v/v) Tween-20 to improve assay sensitivity and signal-to-noise ratio (SNR). A reducing agent, dithiothreitol (DTT, 2.5 mM), was also added in order to improve protein stability. These changes were mostly made in line with what the Mazitschek lab had previously observed to improve SNR in their assays.

As a trial of the assay, I conducted a dose-response saturation binding assay in a 384-well plate format, holding OGT-WT and the CoraFluor-labeled antibody constant at 1 nM in the buffer solution, then adding OSMI-4-FITC in a 1:2 dilution series down each column in triplicate. As a negative control, I repeated this on wells containing no OGT, allowing background signal to be subtracted on a per-concentration basis. Data was collected per-well by exciting the CoraFluor at 340 nm and reading out non-FRET emission from the CoraFluors at 490 nm and FRET emission from the fluorescein at 520 nm; the ratio of the 520 nm emission to the 490 nm emission was reported as the TR-FRET ratio (Fig. 3.5C). After averaging ratios across each concentration for each sample (data and background), the background ratio was subtracted from the data to produce a hyperbolic curve consistent with specific ligand binding, with the following general formula, where B_{max} is the maximum signal, [L] is ligand concentration, and K_D is the ligand's equilibrium dissociation constant:

$$Y = \frac{B_{max} * [L]}{[L] + K_D}$$

Here, the K_D is the concentration of ligand at which binding is at half of its maximum ($B_{max}/2$). After plotting and curve-fitting using this equation, I extracted a $K_{D,app}$ of 1 nM for OSMI-4-FITC (Fig. 3.5D). Notably, the ligand dissociation constant K_D is only absolute if it is far greater ($\tilde{1}$0x) than the enzyme concentration. Since in this case, the enzyme concentration is around the same as the K_D, we are likely underestimating the affinity of OSMI-4-FITC for OGT, and thus we report the dissociation constant as an apparent K_D, or a $K_{D,app}$.

Of special note is the fact that this assay with the OSMI-4-FITC tracer could be easily adapted to high-throughput screening (HTS) for new OGT active site-directed inhibitors. I computed the Z'-factor, a metric of statistical effect size often used to determine the capabilities of an assay in high-throughput screening,[161] across the concentrations used for the initial binding assay. A Z'-factor of 1 is ideal, and a Z'-factor greater than 0.5 is generally regarded as sufficient for robust separation of hits and non-hits. At OSMI-4-FITC concentrations between 0.39 and 25 nM, this assay had a Z' greater than 0.5 (Figure 3.5D), suggesting a usefulness in screening that could be exploited in the future. Should such screening be considered, however, the high affinity of the OSMI-4-FITC probe could be disadvantageous, as usage of the probe at its $K_{D,app}$ would only yield 1 nM available fluorophore for maximal signal. Since competitive HTS efforts usually rely on high signal-to-noise ratios for the displaced probe, one could maximize available signal by instead fluorescently labeling an earlier-generation OSMI, such as OSMI-1a (K_D = 140 nM) or OSMI-3a (K_D = 45 nM),[95] and using the resultant probe to screen new modalities.

3.6 A single HCF-1 repeat binds OGT with high-nanomolar affinity in the presence of a UDP-containing cofactor

With a functional OGT ligand binding assay in hand, I moved on to adapt it to quantify OGT-substrate interactions. For the substrate acceptor probe, I chose to use the HCF-1 repeat substrate HCF1R because of the wealth of structural and biochemical information regarding its interaction with OGT.[46,53] Initially, I purchased HCF1R as a synthetic peptide with an N-terminal fluorescein label and a glycosite E10S or E10A substitution; however, I later developed a recombinant expression system consisting of an N-terminal StrepTag, a HaloTag fusion (which was covalently labeled with an Alexa Fluor 488 chloroalkane),[162] and the HCF1R peptide bearing the catalytic-dead E10A glycosite mutation, which I termed probe **1** (Fig. 3.6A). At HCF1R concentrations out to 10 μM, binding was not observed (data not shown). However, upon adding UDP, fulfilling requirements of OGT's ordered bi-bi mechanism,[54] full binding was achieved and HCF1R's K_D for OGT was found to be 450 nM in the presence of UDP (Fig. 3.6B).

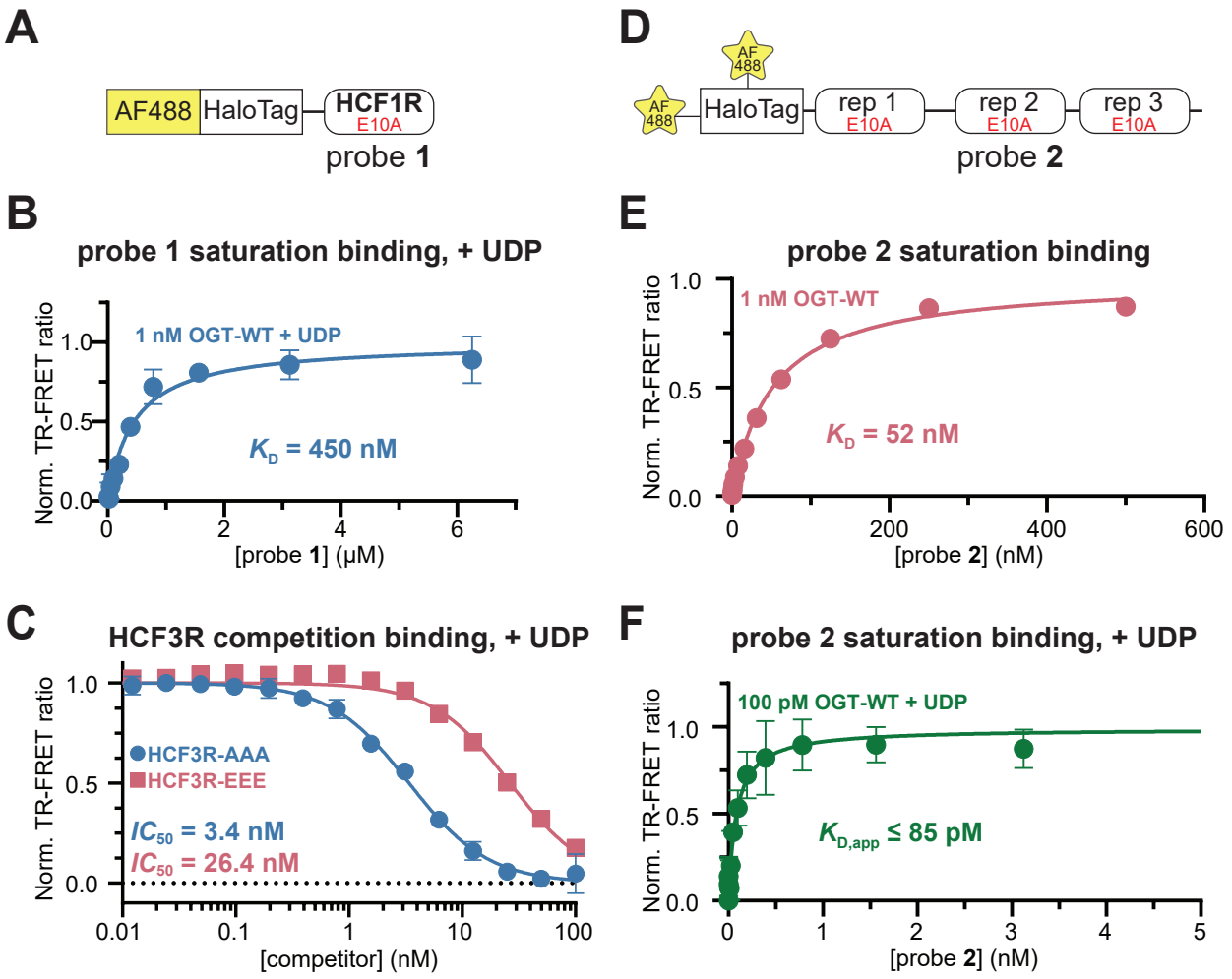


Figure 3.6: Binding of HCF1R and HCF3R probes **1** and **2** to OGT-WT. (A) Schematic for probe **1**. (B) Saturation binding of **1** to OGT-WT in the presence of UDP, with a K_D of 450 nM. (C) Competition binding of HCF3R-AAA (IC_{50} = 3.4 nM) and -EEE (IC_{50} = 26.4 nM) to OGT-WT in the presence of UDP and **1** at its K_D. (D) Schematic for probe **2**. (E,F) Saturation binding of **2** to OGT in the (E) absence (K_D = 52 nM) or (F) presence ($K_{D,app} \leq$ 85 pM) of UDP. Data in B, C, E, F are mean ± s.d. of three (B,C,E) or four (F) replicates, and are representative of at least two independent experiments.

3.7 3.7. HCF3R binds wildtype OGT with picomolar affinity in the presence of UDP

While HCF1R cannot be cleaved by OGT, HCF3R is robustly cleaved. I hypothesized that this would be represented as an increase in binding affinity for HCF3R over HCF1R, so I devised a competition binding assay using probe **1** at its K_D in the presence of UDP, and varying the amount of competitive peptide added in a 1:2 dilution series. I purified tag-free HCF3R-AAA and HCF3R-EEE and performed the assay in triplicate, yielding hyperbolic curves consistent with competition of **1** in the active site. After curve fitting to a variable-slope inhibitor vs. response equation and using the Cheng-Prusoff equation to determine the resultant competitor K_D,[163] I found that HCF3R-EEE binds OGT with a K_D of 13.2 nM, while the triple alanine variant HCF3R-AAA binds with a tighter $K_{D,app}$ of 1.7 nM (Fig. 3.6C). The decreased affinity of the wildtype peptide is likely due to ionic repulsion and steric clashes induced by the glutamate sidechain in OGT's active site, which has been represented in previous structural studies.[46]

Encouraged by this, I modified probe **1** to incorporate the first three HCF-1 repeats and nonspecifically labeled the recombinant protein with fluorescein-NHS ester (achieving a degree of labeling ≥ 1) to generate the HCF3R TR-FRET probe **2** (Fig. 3.6D). I then performed a saturation binding assay with **2**, 1 nM OGT-WT, and no cofactor, and was rewarded with a K_D of 52 nM (Fig. 3.6E), showing that this peptide is capable of interaction even in the absence of the active site cofactor. I repeated the assay with 1 nM in the presence of UDP and obtained a $K_{D,app}$ on the order of ½ the OGT concentration; repetition with 100 pM OGT yielded a $K_{D,app}$ of 85 pM, and further dilution of the OGT concentration decreased SNR significantly; thus, probe **2**'s K_D in the presence of UDP is $\leq$85 pM (Fig. 3.6F).

These data present the first evidence of OGT-substrate binding in the absence of cofactor. We found that probe **2**'s affinity for OGT decreased from 85 pM to 52 nM ($\tilde{1}0^3$-fold) when UDP was not included, which represents a change in the free energy of binding of at least 3.6 kcal/mol. Using this estimate of the contribution of free energy due to peptide-UDP interactions in the active site, probe **1** would have an affinity of >100 μM for OGT in the absence of UDP, which is substantially outside of the concentration range tested in our assay. We believe probe **1** to be more in line with the affinity of most OGT substrates compared to probe **2**, thus we reason that the

typical OGT substrate does not bind OGT in the absence of UDP or UDP-GlcNAc in its active site.

Additionally, we were particularly interested in the fact that HCF3R maintains picomolar affinity for the OGT:UDP complex despite its intrinsic disorder. Proteins with intrinsic disorder have been thought to maintain that disorder to promote binding specificity at the cost of binding strength, as the "induced folding" of the intrinsically-disordered regions presents an entropic penalty to binding.[164–166] However, some suggest that intrinsic disorder skews the spectrum of potential binding affinity towards weaker binding, but can often lead to very high-affinity binding,[167] as seen here in the case of HCF3R. Based on these initial results, I was encouraged to further investigate the interaction between OGT and HCF3R.

3.8 HCF3R binding and cleavage is highly dependent on interactions with the TPR asparagines closest to the OGT active site

HCF-1 cleavage does not occur *in vitro* or in cells if the five most C-terminal TPR asparagines are all mutated to alanine, as in the OGT-5N5A-1 construct.[60,69] To assess the role of each individual asparagine in this phenotype, I cloned and purified the single mutant OGTs N322A, N356A, N390A, N424A, and N458A and assessed HCF3R cleavage by these OGTs (Fig. 3.7B). Surprisingly, even though they constituted loss of only one peptide-engaging amino acid, the N356A, N390A, and N424A mutants were fully incapable of cleavage, while the N322A variant only partially cleaved HCF3R after 3 hours of reaction. Mutation of N458, the only one of these asparagines that only makes a single hydrogen bond to the substrate peptide backbone, did not appreciably affect cleavage.

To compare this activity readout to equilibrium binding, I repeated the saturation binding assay with probe **2** using each of these OGT mutants (Fig. 3.7C). Recapitulating the cleavage assay results, loss of N356, N390, and N424 fully abolished binding out to 1 μM probe **2**, while the N322A mutant only reached $\tilde{2}$0% saturation at that concentration and the N458A mutant maintained low-nanomolar affinity to **2**. The change in free energy of binding due to the N356A, N390A, and N424A mutations approaches $\tilde{7}$ kcal/mol, a striking change for a single amino acid. However, these asparagine residues make bidentate hydrogen bonds, so mutating each removes two H-bonds to the HCF3R peptide. Even so, the $\tilde{7}$ kcal/mol value approaches the intrinsic

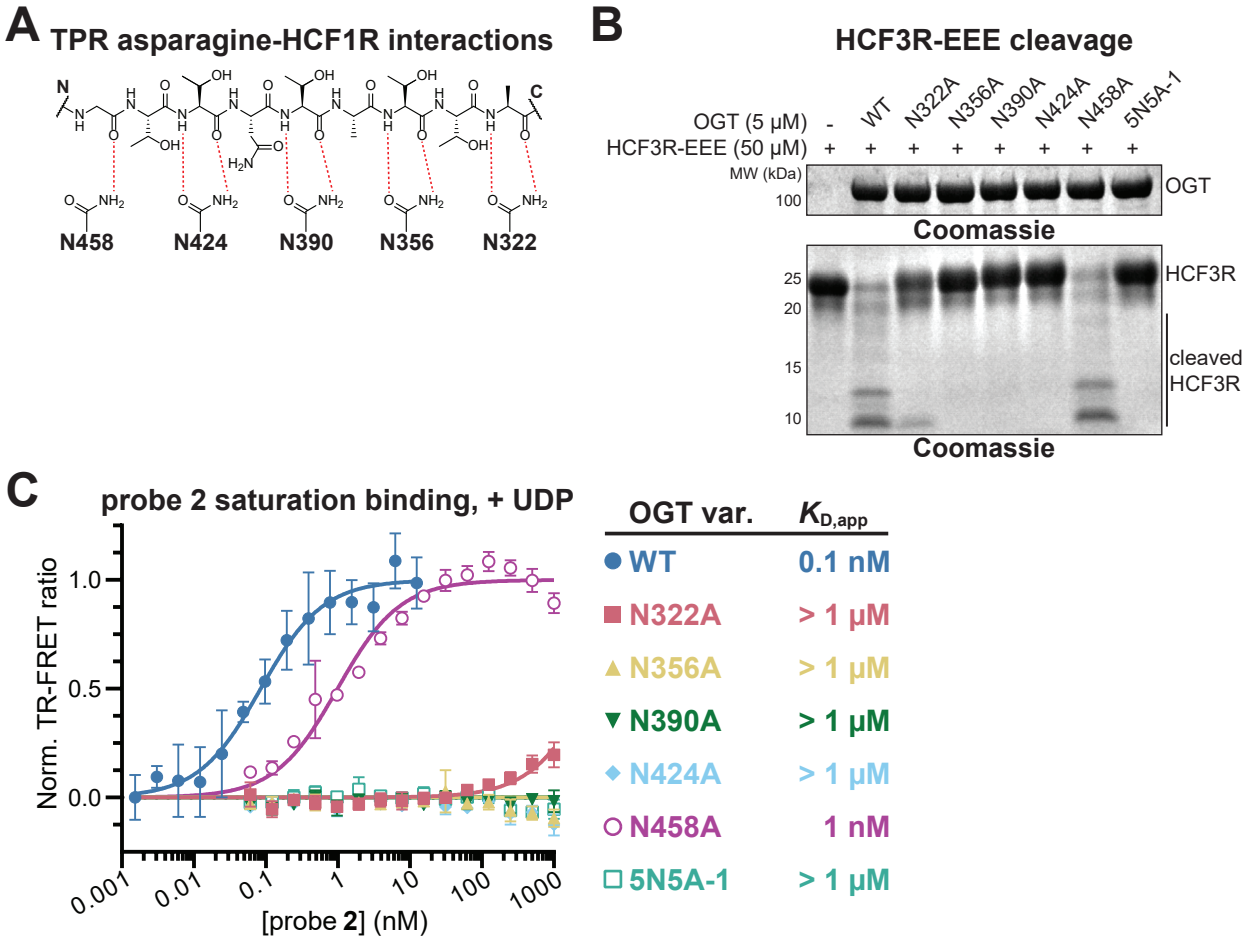


Figure 3.7: Loss of proximal TPR asparagines prevents cleavage and binding of HCF3R. (A) Proximal asparagines N322/356/390/424 form bidentate contacts with the HCF1R peptide backbone. (B) Coomassie stain of HCF3R-EEE cleavage by asparagine mutants shows limited activity by N322A/N356A/N390A/N424A/5N5A-1 mutants. (C) Saturation binding of probe **2** to TPR asparagine mutants shows significant loss of affinity upon mutation. Data are mean ± s.d. of three replicates, and are representative of at least two independent experiments.

energetic worth of two hydrogen bonds.[168–170] We believe the majority of the unfavorable change in binding free energy can be attributed to the loss of the hydrogen bonds themselves, instead of due to positive changes in desolvation entropy. Using analyses postulated by Jencks to explain why the change in free energy due to multiple hydrogen bonds is greater than the sum of the changes in free energy from individual hydrogen bonds, we hypothesize that formation of the first hydrogen bond in each bidentate interaction helps 'pay the cost' of free energy losses due to restricting rotational and translational entropy, which enables the energetic worth of the second hydrogen bond to be fully realized.[168,171] This explanation does not satisfactorily explain the $\tilde{7}$ kcal/mol change because the intrinsic energetic worth of an H-bond in water has been estimated to be $\tilde{4}$ kcal/mol. However, some studies have shown dramatic effects on binding when H-bonds in hydrogen bonded arrays are altered.[172] We think those findings are applicable to understanding the effects of the single Asn mutants on HCF3R binding. That is, the bidentate contacts formed by each Asn affect the energetics of binding of the adjacent residues with the result that any disruption, especially in the region close to OGT's active site, wreaks havoc on the rest of the network.

3.9 Loss-of-binding OGT asparagine mutants promote glycosylation of HCF3R, but not all OGT substrates

I was interested in probing whether the effects of the asparagine mutations on cleavage were the same as they were on glycosylation. Previous work reported that OGT-5N5A-1 had reduced glycosylation activity both *in vitro* and in cells,[60,62,69] but like with HCF3R/HCF-1 cleavage, there were no reports on the activity of the single asparagine mutants. Given the striking impact of a single asparagine-to-alanine mutation on binding and cleavage, I hypothesized a similar effect on glycosylation would be observed. I initially tested this in HeLa cell extracts after 5 and 15 minutes of reaction, using no OGT and OGT-WT as negative and positive controls (Fig. 3.8A). Here, I observed patterns that did not match the HCF3R data: while OGTs N322A, N356A, and 5N5A-1 had sharply reduced glycosylation activity, OGTs N390A, N424A, and N458A were only slightly less active than WT.

I then performed a similar glycosylation assay over a period of 1 hour using these OGTs and

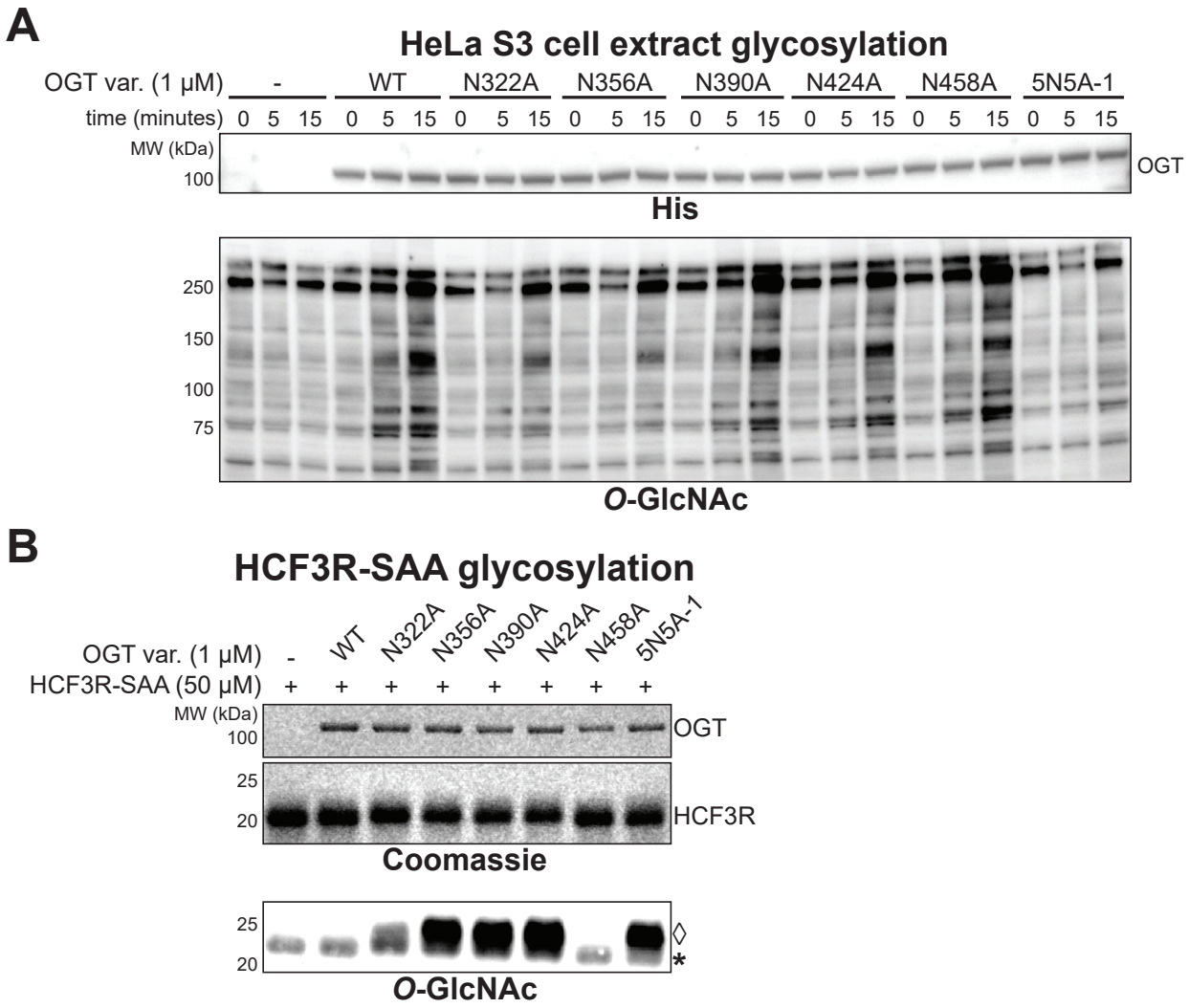


Figure 3.8: Glycosylation activity of single-asparagine OGT mutants. (A) In HeLa S3 cell extracts, the mutants N322A, N356A, and 5N5A-1 were globally deficient in *O*-GlcNAcylation activity. (B) Against the substrate HCF3R-SAA, asparagine mutants that had low binding capability (*e.g.*, N356A, 5N5A-1) displayed hypermorphic activity. Asterisk indicates unglycosylated starting material, while diamond indicates glycosylated substrate.

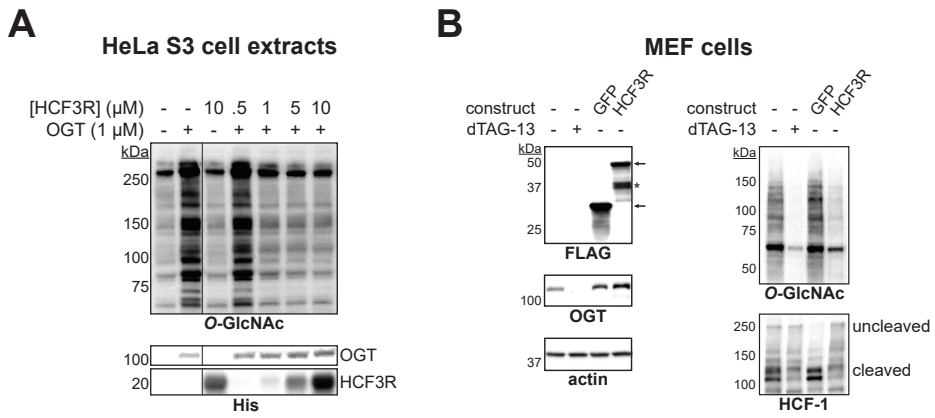


Figure 3.9: HCF3R is a potent OGT inhibitor. (A) Western blotting shows that 1 μM HCF3R fully inhibits 1 μM OGT in HeLa S3 cell extracts. (B) Expression of GFP-tagged HCF3R in dTAG-MEFs fully inhibited *O*-GlcNAcylation and HCF-1 cleavage to levels recapitulated by 24 hours of dTAG-mediated OGT degradation.

recombinant HCF3R-SAA. After separation by SDS-PAGE and immunoblotting against *O*-GlcNAc, a trend emerged that was opposite the trend observed from HCF3R cleavage: decreased binding affinity led to greatly increased *O*-GlcNAcylation (Fig. 3.8B).

3.10 HCF3R is a potent OGT inhibitor

Because HCF3R-AAA bound OGT with picomolar affinity, I became interested in whether it would also inhibit OGT. I initially used HeLa cell extracts to answer this question by dosing 1 μM OGT-WT and varying amounts of HCF3R-AAA from 0.5 to 10 μM (0.5 to 10 equivalents) into the extracts, incubating for 15 minutes, then quenching and analyzing by western blot (Fig. 3.9A). In this assay, even 1 equivalent of HCF3R nearly fully inhibited *O*-GlcNAcylation, which provided the rationale to test HCF3R in cells.

I integrated HCF3R-AAA as an mEGFP-$FLAG_2$ fusion in dTAG-OGT MEFs under the constitutive PGK promoter using lentivirus, then sorted for GFP expression. As a control, I also generated and sorted a cell line expressing mEGFP-$FLAG_2$. I chose this cell line because I wanted a 'positive' control for depleted OGT activity, which could be easily achieved in this cell line by inducing OGT degradation with the small molecule dTAG-13.[69] After sorting, I harvested cells from both cell lines, as well as the parent dTAG-OGT MEF line with and without dTAG-13 treatment, then blotted for the FLAG-tagged constructs, protein *O*-GlcNAcylation, HCF-1 cleavage, OGT expression,

and actin as a loading control. HCF3R expression reduced OGT activity (both *O*-GlcNAcylation and HCF-1 cleavage) to levels comparable to 24 hours of dTAG-13 treatment in this cell line. I also performed live cell fluorescence microscopy, confirming diffuse nuclear and cytoplasmic expression of both the control GFP, and the tagged HCF3R. I kept these cell lines proliferating for future analysis, but I began to observe slower growth in the HCF3R-expressing cells as the passage number increased, suggesting reduced fitness due to continuous OGT inhibition. This was disadvantageous, as I did not want to accumulate or report on gross cellular defects, so I began searching for alternative ways to deliver or induce expression of HCF3R.

3.11 Cellular expression of a degron-tagged HCF3R under an inducible promoter enables rapid inhibition and recovery of OGT activity using two small molecules

One effective means of preventing background expression of a gene is to tie its expression to a conditionally-inducible promoter, instead of a constitutively-active promoter as was previously used. Perhaps the most widely used set of inducible promoters are derived from the bacterial *Tet* operon that provides resistance to the protein translation inhibitor tetracycline, in which transcription of the *tetA* gene encoding a tetracycline efflux pump is controlled by the repressor TetR.[173,174] Natively, TetR binds *tetO* elements upstream of *tetA* and prevents its transcription, however binding of tetracycline induces a conformational shift in TetR, preventing its association to DNA and inducing transcription of the *tetA* gene. Notably, TetR's high affinity for tetracycline enables this de-repression to occur at lower concentrations than are required for protein translation inhibition. This operon was initially modified for use in eukaryotic cells in 1992, when the Tet-OFF system was developed.[175] In this system, the activation domain from the herpes simplex virus protein VP16 was fused to TetR, generating a tetracycline-controlled transcriptional activator, or tTA. By expressing tTA in cells and fusing *tetO* elements to the commonly-used eukaryotic CMV promoter region upstream of a gene of interest, transcription could be deactivated in the presence of tetracycline or its derivative doxycycline. Tet-OFF had the notable drawback that continuous administration of tetracyclines was necessary to prevent expression; to remedy this, random

mutants of TetR were screened for their ability to bind *tetO* in the presence of tetracyclines instead of in their absence. The resultant reverse-TetR was fused to VP16's activation domain, creating an rtTA that initiates transcription of the gene of interest with a *tetO*-fused promoter region only in the presence of tetracyclines (Fig. 3.10A).[176] This Tet-ON system has been further iterated upon to improve its sensitivity to doxycycline, reducing adverse cytotoxic effects.[177]

While Tet-ON could provide an on-switch for HCF3R expression, I also wanted to incorporate an off-switch to shut down expression and stop OGT inhibition. To do this, I turned to the $FKBP12^{F36V}$ degron that the Walker lab had previously used to enable OGT degradation.[69,108,178] Briefly, degron fusion promotes proximity-based polyubiquitination by E3 ubiquitin ligases upon treatment with bifunctional PROTAC molecules, which bind both the degron and the E3 ligase, and the resultant polyubiquitinated protein is degraded in the proteasome.

I cloned the $FKBP12^{F36V}$ degron into to the existing construct in between the GFP and FLAG tags (Fig. 3.10A). I then cloned the full mEGFP-$FKBP12^{F36V}$-$FLAG_2$-HCF3R gene into a lentiviral destination vector under a *tetO*-fused eukaryotic promoter sequence that also encoded the rtTA machinery necessary for the Tet-ON system (pCW57.1) and produced lentivirus containing this genetic material. I chose to transfect unmodified HCT116 cells, an adherent human colorectal carcinoma cell line, with this gene to synchronize our cell lines with other efforts in the lab, including ongoing genetic screens. After selection, I induced expression with addition of doxycycline at 10 μg/mL for 24 hours, then sorted for high GFP signal. I then recovered these cells and set up two assays to profile the ability of this system to turn on and turn off HCF3R expression and OGT inhibition.

In the first experiment, I seeded cells into a 6-well plate and treated with 10 μg/mL doxycycline for variable times (0-24 hours), harvesting and lysing at the requisite stop time. I also seeded a second 6-well plate, whose wells were all treated with 10 μg/mL doxycycline for 24 hours then treated with 250 nM dTAGV-1 from 0-24 hours to induce degradation of HCF3R (note that dTAGV-1, a VHL E3 ligase-targeting PROTAC, was used due to the low expression of CRBN E3 ligase in HCT116 cells).[179] Following the conclusion of all timepoints, I separated these samples by SDS-PAGE and immunoblotted against FLAG (proxy for HCF3R), *O*-GlcNAc, and actin (Fig. 3.10B). I found that HCF3R expression began around 2 hours of doxycycline treatment, and sharply increased by 24 hours, while dTAGV-1 treatment induced rapid degradation of HCF3R

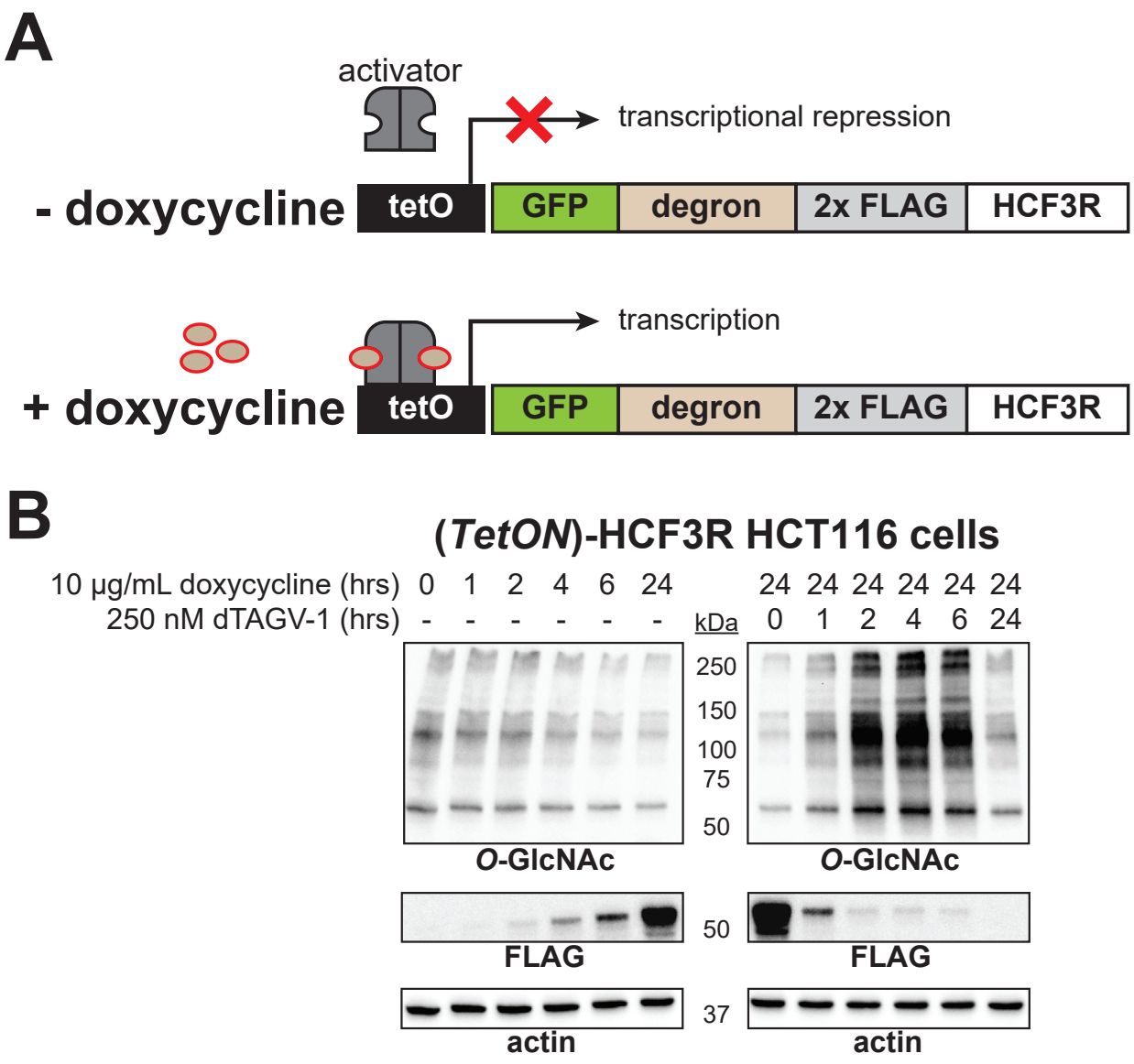


Figure 3.10: An on/off switchable HCF3R inhibits OGT in cells. (A) Schematic for doxycycline-inducible (*TetON*)-GFP-degron-2xFLAG-HCF3R, which can be degraded by addition of dTAGV-1. (B) Western blotting from a time-course of induction (+doxycycline) and degradation (+dTAGV-1) of HCF3R in HCT116 cells.

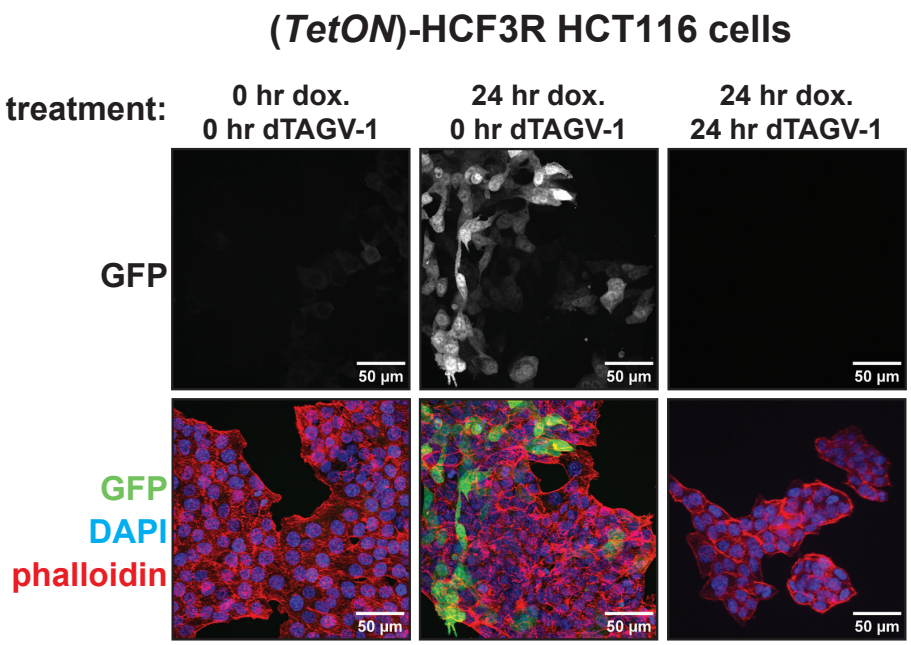


Figure 3.11: Doxycycline and dTAGV-1 turn HCF3R expression on and off, as seen in fixed-cell confocal fluorescence microscopy images. Top is the GFP (HCF3R) channel, which was combined with DAPI (nuclear DNA) and phalloidin (f-actin) on the bottom. Images are of maximum intensity projections, normalized across all three conditions.

to basal levels over a period of just 2 hours. Additionally, and more importantly, the pattern of *O*-GlcNAcylation activity inversely mirrored HCF3R expression: after 24 hours of induction with doxycycline, *O*-GlcNAc was at a minimum, while after 1 hour of degradation it had recovered to wildtype-like levels. Interestingly, the *O*-GlcNAc levels between 2 and 6 hours of dTAGV-1 treatment were substantially above those of untreated cells, suggesting some sort of rebound phenomenon not previously observed due to lack of a rapidly-acting off-switch for other OGT inhibitors described in the literature. While the exact mechanism of this phenomenon leaves much room for speculation, it is clear that our system enables switchability to OGT inhibition that has not been possible before.

In the second experiment, I aimed to visualize these changes in expression by fluorescence microscopy. I seeded cells into 6-well MatTek microscopy dishes and treated with either no compound, doxycycline (10 μg/mL) for 24 hours, or doxycycline for 24 hours then dTAGV-1 (250 nM) for 24 hours. At the conclusion of these timepoints, I fixed the cells and stained using DAPI (cell nuclei) and fluorophore-tagged phalloidin (filamentous actin, a proxy for the full extent of the cytoplasm). I did not specifically stain for the HCF3R construct; instead, I imaged

residual fluorescence from the mEGFP tag that was not abolished by the fixation process. Without compound treatment, HCF3R expression was very low; 24 hours of doxycycline induced robust expression; and addition of dTAGV-1 for 24 hours cut expression to undetectably low levels (Fig. 3.11). This reaffirmed the results observed by western blotting and confirms that our inhibitor's expression is switched on and off by the addition of two commercially-available small molecules.

3.12 (*TetON*)-HCF3R specifically interacts with OGT in HCT116 cells

Because I was developing a cellular OGT inhibitor, I wanted to make sure that HCF3R was not interacting with off-target proteins or complexes that might confound the observed phenotypes. Using principles established in Chapter 1, I performed co-immunoprecipitation on lysates from (*TetON*)-HCF3R HCT116 cells treated with either 10 μg/mL doxycycline or 250 nM dTAGV-1 for 48 hours in triplicate, with total protein loaded onto the anti-FLAG magnetic beads normalized across all samples. I confirmed that the pulldown worked by blotting for the FLAG tag in all samples (flowthroughs, washes 1-3, elutions; Figure 3.12A), then processed all elution samples for in-gel digestion, trypsinization, and proteomics analysis at the Taplin Mass Spectrometry facility (see Methods, 3.14, for additional workflow). I then conducted a simple analysis of the results, averaging the peptides across all replicates and comparing across the doxycycline- and dTAG-treated samples, showing that the only substantially-enriched proteins pulled down in the doxycycline-treated, HCF3R-expressing cells were GFP and FKBP12 (both from the HCF3R construct itself), and OGT (Figure 3.12B). Thus, the HCF3R polypeptide inhibitor construct selectively binds OGT in cells.

3.13 Conclusion

In this chapter, I developed a facile and modular quantitative binding assay for OGT and its ligands that can be adapted towards small molecules or protein substrates. This assay, as initially described using the fluorescent OSMI-4-FITC probe, is well-suited to high-throughput competition-based screening at low nanomolar probe concentrations. While developing an HTS-compatible binding assay was an initial goal of this project, I quickly moved away from the idea of library

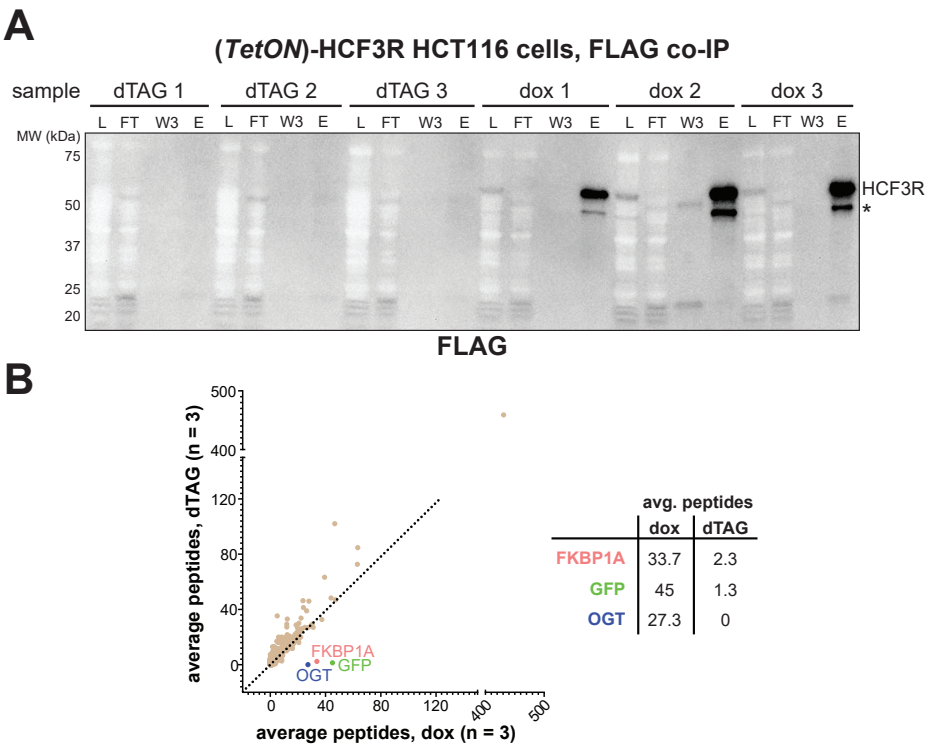


Figure 3.12: HCF3R specifically interacts with OGT in cells. (A) FLAG western blots from pulldowns of FLAG-tagged HCF3R indicate minimal loss of HCF3R upon washing (L = lysate, FT = flowthrough, W3 = wash 3, E = elution). (B) Pulldown of HCF3R from HCT116 cells only enriched the bait and OGT.

screening as I identified the picomolar binding affinity of HCF3R polypeptides. Future work could use either OSMI-4-FITC, or probes **1** and **2** in HTS efforts to identify new modalities that bind either OGT's active site or the TPR lumen. The assay could potentially be used to screen for compounds that displace HCF1R or HCF3R, then counterscreen using OSMI-4-FITC to eliminate compounds that just bind the active site.

The striking observation that HCF3R binds OGT with picomolar affinity in the presence of UDP is of particular importance. The HCF-1 proteolytic repeats are thought to be relatively disordered, but upon forming bidentate backbone contacts with the TPR asparagines, they are entropically frozen and locked into place. Removal of any of the bidentate bond-forming asparagines (N322, N356, N390, N424) results in significant loss of binding affinity as the backbone is not oriented properly for the other asparagine residues to interface with it. The lack of proline residues in the region C-terminal to the HCF3R glycosites additionally contributes to this, as prolines distort the peptide backbone in ways that may prevent sequential association of bidentate contacts in the TPRs;[57] additionally, the proline amide nitrogen is not protonated at physiological pH, and so it cannot be a hydrogen bond donor.

That HCF3R is a potent and selective inhibitor of OGT should not come as a surprise, given its picomolar affinity for the enzyme *in vitro*. Its status as a lengthy (1̃00 amino acid) peptide makes it likely beyond the realm of deliverable therapeutics; however, its potency and amenability to modification make it a valuable contribution to the field. I have demonstrated that fusion to the degron $FKBP12^{F36V}$ enables rapid degradation, providing an off-ramp for OGT inhibition that is not possible with the current suite of small molecule inhibitors, and further modification to enable cell penetrance could also be possible. However, to pursue that would likely require shortening of the construct, as cell-penetrating peptides generally benefit from shorter length.[180] Additionally, HCF3R could have particular use in animal models, where the existing small molecule OGT inhibitors are limited due to poor permeability and solubility. Delivery mechanisms for peptides are also much more diverse and could be used to target specific tissues, but this would require significant engineering to optimize delivery and activity. The work described in the following chapter of this dissertation will present some evidence of minimal tightly-binding constructs derived from the HCF3R polypeptide. Future work will use that information to develop improved inhibitors based on this scaffold.

3.14 Methods

General Vector Construction and Cloning

See Appendix A for information on the plasmids used in this chapter, as well as elsewhere in this dissertation. OGT heterologous expression plasmids were originally developed by Dr. John Janetzko. His_8-HCF3R expression plasmids were initially developed by Dr. John Janetzko, then later modified to include a C-terminal tri-tyrosine (YYY) tag by Dr. Cassie Joiner.

See Chapter 2 Methods for details on cloning protocols used throughout this chapter.

General staining and immunoblotting methods

See Chapter 2 Methods for detailed procedures on SDS-PAGE electrophoresis, Coomassie/InstantBlue staining, and western blotting.

General tissue culture methods

Unmodified HEK293T and HCT116 cells used in this study were previously ordered from ATCC. MEFs expressing Cre-ER and $FKBP12^{F36V}$-tagged OGT-WT with endogenous *Ogt*-KO (dTAG-MEFs) were previously developed and characterized by Dr. Zebulon Levine.[69] All cells were utilized at low passage numbers (<10) throughout this study and were routinely tested to confirm the absence of mycoplasma contamination. All cells were cultured at 37° with 5% CO_2. HEK293T cells and MEFs were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco), which unless otherwise specified was supplemented to 10% with fetal bovine serum (FBS, Cytiva) and 1% with penicillin/streptomycin solution (Corning), termed "complete DMEM." HCT116 cells were similarly cultured in McCoy's 5A medium (Gibco), made "complete" by addition of 10% FBS and 1% Pen/Strep solution.

Cells were passaged and split by trypsinization. Briefly, media was removed by aspiration and cells were washed with Dulbecco's phosphate buffered saline (dPBS, Gibco), which was then removed by aspiration. TrypLE Express (Gibco) was added, and cells were incubated at 37°C for 10 minutes to detach from the culture plate. Trypsinized cells were resuspended to 10% final TrypLE in complete medium and centrifuged at 200 g for 5 minutes to pellet. Media was aspirated, then the cell pellet was resuspended in 10 mL complete medium, which was used to seed new

plates at the desired cell density.

Cells were harvested by scraping. Briefly, cell plates were placed on ice and media was removed. Cells were washed with 1-5 mL dPBS, which was also removed. 1 mL dPBS was then added to each plate, and cells were scraped into buffer. Cells were removed by pipetting, then centrifuged at 500 g for 5 minutes at 4°C to pellet. Buffer was aspirated and cells were lysed in a minimal amount of RIPA buffer supplemented with 0.05 mM PUGNAc (OGA inhibitor) and 1x Halt proteinase/phosphatase inhibitor (Thermo-Fisher) by pipetting up and down ~20 times and incubating on ice for 10 minutes. Lysates were centrifuged at 21130 rpm for 15 minutes at 4°C to clarify, then moved to new tubes. Protein amounts were quantified by BCA assay (Thermo) against a series of bovine serum albumin standards.

Lentivirus production for cellular integration of HCF3R constructs

Lentivirus was produced as described in Chapter 2 Methods.

Viral infection, puromycin selection, and sorting of dTAG-MEFs expressing GFP and GFP-HCF3R

Low-passage dTAG-MEFs were thawed from frozen into complete DMEM and grown to confluence, then used to seed 6-well plates at 50,000 cells/well. The following day, viral aliquots were thawed on ice, and media in wells was removed and replaced with 2 mL virus/well (two replicate wells per virus). On each plate, at least one well was replaced with 2 mL complete DMEM as a control. Plates were then spinfected by centrifugation at 1000 g and 30°C for 90 minutes. Following spinfection, 1 mL media was removed from each well and replaced with 1 mL complete DMEM, and plates were incubated overnight. The following day (day 1), media was removed and replaced with 2 mL complete DMEM. On day 2, media was replaced with complete DMEM supplemented with 2 μg/mL puromycin. From days 3-6, media was changed as necessary to remove cell debris, replacing with complete DMEM and 2 μg/mL puromycin. Cells were also expanded to 10-cm dishes when confluent. From days 7-14, this was increased to 10 μg/mL puromycin and cells were expanded to 15-cm dishes as necessary. Once cells reached confluence in 15-cm dishes after day 14, cells were aliquoted and frozen in cryopreservation media (1,000,000 cells/mL, complete DMEM + 10% DMSO).

Approximately four days prior to fluorescence-activated cell sorting (FACS), one tube per cell line was thawed into 10 mL complete DMEM in a 10-cm dish. Additionally, one tube of unmodified dTAG-MEFs was thawed as a negative control. On the day of sorting, cells were trypsinized, resuspended in sorting buffer (dPBS with 1% FBS and 2 mM EDTA) and passed through a 35 μM cell strainer. FACS was conducted on a Beckman-Coulter MoFlo Astrios EQ sell sorter using a 488 nm laser. The top 50% of GFP-expressing cells were sorted into complete DMEM. Gating was set to minimize collection of cells not expressing GFP (<1% of the control). Half of the collected cells were immediately plated and recovered, while the other half were cryopreserved.

Viral infection, puromycin selection, and sorting of HCT116 cells expressing GFP-HCF3R

Low-passage HCT116 cells were thawed from frozen into complete McCoys 5A medium and grown to confluence, then used to seed 6-well plates at 50,000 cells/well. The following day, viral aliquots were thawed on ice, and media in wells was removed and replaced with 2 mL virus/well (two replicate wells per virus). On each plate, at least one well was replaced with 2 mL complete McCoys 5A medium as a control. Plates were then spinfected by centrifugation at 1000 g and 30°C for 90 minutes. Following spinfection, 1 mL media was removed from each well and replaced with 1 mL complete McCoys 5A medium, and plates were incubated overnight. The following day (day 1), media was removed and replaced with 2 mL complete McCoys 5A medium. On day 2, media was replaced with complete McCoys 5A medium supplemented with 2 μg/mL puromycin. From days 3-6, media was changed as necessary to remove cell debris, replacing with complete McCoys 5A medium and 2 μg/mL puromycin. Cells were also expanded to 10-cm dishes when confluent. From days 7-14, this was increased to 10 μg/mL puromycin and cells were expanded to 15-cm dishes as necessary. Once cells reached confluence in 15-cm dishes after day 14, cells were aliquoted and frozen in cryopreservation media (1,000,000 cells/mL, complete McCoys 5A medium + 5% DMSO).

Approximately four days prior to FACS, one tube of cells per cell line was thawed into 10 mL complete McCoys 5A medium in a 10-cm dish. Additionally, one tube of unmodified HCT116 cells was thawed as a negative control. 48 hours before sorting, all plates were supplemented with 10 μg/mL doxycycline to induce expression of GFP-tagged HCF3R constructs. On the day

of sorting, cells were trypsinized, resuspended in sorting buffer (dPBS with 1% FBS and 2 mM EDTA) and passed through a 35 μM cell strainer. FACS was conducted on a Beckman-Coulter MoFlo Astrios EQ sell sorter using a 488 nm laser. The top 50% of GFP-expressing cells were sorted into complete DMEM. Gating was set to minimize collection of cells not expressing GFP (<1% of the control). Half of the collected cells were immediately plated and recovered, while the other half were cryopreserved.

Analysis of HCF-1 proteolytic repeat conservation across taxa

All analysis was performed using custom python3 scripts using the pandas, csv, ete3, NumPy, and MatPlotLib packages.

HCF-1 protein sequences (n=2050) were extracted from NCBI using the longest human HCFC1 protein isoform as a query sequence. Sequences were pruned to include only the largest protein isoform from each species, and to remove any without species identifiers (n=493). Sequences were separated by taxonomic information, with chordate (n=400) and mammalian (n=229) species pushed to further analysis. Frequency plots were constructed using WebLogo centering on all glutamate residues not preceded by a glutamate three residues upstream (E10 with no E7, eliminating erroneous centering of the conserved E13 glutamate in the proteolytic repeat), glutamates preceded immediately by a cysteine (E10 + C9 with no E7), and glutamates followed by threonines in the +7/+9/+11 positions (E10 + T17/19/21 with no E7).[181]

Sequences in both chordate and mammalian categories were also assessed for number of repeats. Mammalian HCF-1 homologs with 6 repeats were assessed for the conservation of their spanning residues by counting the number of amino acids in between the repeats, and plotting frequency maps of the spanning sequences using WebLogo.[181] Positions in frequency maps that were empty were represented with the symbol 'X'.

Expression and purification of polyhistidine-tagged OGT variants

See Chapter 2 Methods for detailed methodology.

Expression and purification of StrepTag-HaloTagged HCF1R and HCF3R variants

Plasmids for heterologous expression of StrepTag-HaloTag-HCF1R and -HCF3R in *E. coli* were freshly transformed into *E. coli* BL21(DE3) chemically competent cells and plated on LB agar with

corresponding antibiotics overnight at 37°. 5-10 colonies were used to seed overnight cultures of 50 mL LB with corresponding antibiotic, which were grown at 37° with 250 rpm shaking for 16 hours. Large cultures (1L LB) with corresponding antibiotic were inoculated from overnight cultures at 1% (v/v) and grown at 37° with 225 rpm shaking until the OD_{600} reached 0.7, at which point IPTG was added to 0.3 mM and cultures were incubated at 37° with 225 rpm shaking for 3 hours to induce protein expression.

Cells were pelleted by centrifugation, resuspended in 30 mL Streptactin XT Buffer W (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) supplemented with 0.1 mg/mL lysozyme, 0.1 mg/mL DNAse I, and 1 mM PMSF, and lysed by cell disruption as described in Chapter 2 Methods. Lysate was clarified by centrifugation (15,000 g, 4°, 20 minutes), and clarified lysate was applied directly to 1 mL Streptactin-XT 4FLOW high-capacity resin (Iba Life Sciences). Flowthrough, 3x 2 column volume (CV) washes in Buffer W, and 3x 1 CV elution fractions in Buffer BXT (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 50 mM biotin) were collected by gravity filtration. All fractions were supplemented with 1 mM DTT.

Elution fractions were combined, buffer exchanged into HBS with 1 mM DTT by serial concentration and dilution in 30K MWCO Amicon concentrators, then finally concentrated to < 0.5 mL. Elution was then purified by size exclusion chromatography on an AKTA Pure 25 FPLC fitted with a Superdex 200 Increase 10/300 GL column (Cytiva), using an isocratic method, into HBS with 1 mM DTT. Peak fractions were assessed by SDS-PAGE and InstantBlue staining, and pure fractions were combined, concentrated to > 200 μM, aliquoted, flash frozen in liquid nitrogen, and stored at -80°.

Expression and purification of tag-free HCF3R variants

Tag-free HCF3R variants were purified from polyhistidine-tagged HCF3Rs according to the same protocol up to and including the nickel affinity chromatography step (see Chapter 2 Methods for more detailed methodology). Following nickel affinity chromatography, elutions were concentrated to 2.5 mL and desalted into 3.5 mL HBS using a PD-10 desalting column via the gravity method (Cytiva); desalted elutions were supplemented with 1 mM DTT and 1 mM EDTA and treated with 1:50 mass ratio HRV 3C protease (purified recombinantly with a polyhistidine tag). Protease reactions were incubated for 16 hours at 4° with rotation, diluted to 10 mL in HBS

with 1 mM DTT and 10 mM imidazole, and batch bound to 1 mL Nickel-INDIGO Purecube resin (Purecube; note that this resin is not stripped of metals by addition of 1 mM EDTA) for 1 hour at 4° with gentle rotation. Flowthrough was collected and combined with a 5 mL wash (HBS + 1 mM DTT + 10 mM imidazole), concentrated to < 0.5 mL, then injected onto an AKTA 25 FPLC fitted with a Superdex 75 Increase 10/300 GL column, pre-equilibrated in HBS with 1 mM DTT. Peak fractions were assessed by SDS-PAGE and InstantBlue staining, and pure fractions were combined, concentrated to > 200 μM, aliquoted, flash frozen in liquid nitrogen, and stored at -80°.

CoraFluor-1-*pfp* synthesis

CoraFluor-1-pentafluorophenyl (PFP) ester was synthesized by Dr. N. Connor Payne according to synthetic schema developed during his time in the Mazitschek lab,[135] and was provided in solution in anhydrous dimethylacetamide.

CoraFluor-1 labeling of anti-polyhistidine antibody

CoraFluor-1 labeling of anti-polyhistidine antibodies (Abcam a18184) was performed as previously described.[135] Protein was thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates prior to buffer exchanging into 100 mM sodium carbonate, pH 8.5 with 0.05% (v/v) Tween-20 using Zeba 3K MWCO spin columns according to the manufacturer's guidelines (Thermo-Fisher). Protein was then concentrated to $\geq$ 1 mg/mL if necessary. 10 equivalents of CoraFluor-1-PFP ester were added directly (from a 25 mM stock in dimethylacetamide) and pipetted to mix. The reaction was incubated for 1 hour at room temperature in the dark and buffer-exchanged back into HBS with 0.05% Tween-20. Protein concentration and degree of labeling (DOL) were calculated as previously reported, by taking A_{280} and A_{340} by Nanodrop, using an A_{280} correction factor (cf_{340}) of 0.1571, E_{340} of 22000 $M^{-1}cm^{-1}$, and an E_{280} of 210000 $M^{-1}cm^{-1}$ for the antibody with the following formulae:

$$[AF488] = \frac{A_{494}}{E_{494}} \qquad [protein] = \frac{A_{280} - cf_{494} * A_{494}}{E_{protein,280}} \qquad DOL = \frac{[AF488]}{[protein]}$$

Aliquots were diluted 1:1 in glycerol (50% v/v final), frozen in liquid nitrogen, and stored at -80°. Once thawed, aliquots were kept at -20° for use, remaining stable for at least one month under these conditions.

Synthesis of OSMI-4-FITC

OSMI-4-FITC was prepared from the carboxylate of OSMI-4, OSMI-4a, by Dr. N. Connor Payne. 3.6 mg OSMI-4a (1 eq.) were combined with 8.1 mg *N*,*N*-diisopropylethylamine (10 eq.) in anhydrous dimethylformamide, to which the trifluoroacetate salt of FITC-$(PEG)_3$-propylamine (5.4 mg, 1.2 eq.) and PyBOP (benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate, 3.4 mg, 1.05 eq.) were added. Mixture was stirred for 15 minutes, then diluted into dichloromethane, washed twice with 0.2 N HCl, washed once with saturated brine, dried over sodium sulfate, filtered, and concentrated by rotary evaporation to dryness. Crude product was purified by reverse phase flash chromatography into aqueous acetonitrile with 0.1% formic acid, then concentrated to yield the final product OSMI-4-FITC (NCP451, 2.95 mg, 40% yield) as a yellow film. Product was dissolved to 5 mM in deuterated DMSO and confirmed by LCMS (expected MW 1167.71 Da).

Synthesis of probe 1 (AF488-StrepTag-HaloTag-HCF1R)

Probe **1** was produced by covalent labeling of the HaloTag fusion protein using an Alexa Fluor 488 (AF488) HaloTag-reactive chloroalkane ligand (Promega).[162] Protein was thawed on ice, centrifuged for 10 minutes at 21130 rpm and 4° to pellet aggregates, then diluted to $\leq$10 μM and $\geq$100 μL in HBS with 1 mM DTT. AF488-chloroalkane ligand was added to a 1:1 final stoichiometry from a 1 mM stock in DMSO, the mixture was briefly mixed by pipetting, and the reaction was incubated at RT in the dark for 1 hour. Reaction was serially concentrated and diluted in HBS with 1 mM DTT five times using a 10K MWCO Amicon Ultra-Cel concentrator to remove unreacted ligand. Degree of labeling was then calculated by taking A_{280} and A_{494} measurements by Nanodrop (Thermo-Fisher), using an A_{280} correction factor (cf_{494}) of 0.11, E_{494} of 75000 $M^{-1}cm^{-1}$ and E_{280} determined by the protein sequence with the following formulae:

Labeled probe was aliquoted, flash frozen in liquid nitrogen, and stored at -80° until usage.

Synthesis of probe 2 (AF488-StrepTag-HaloTag-HCF3R)

Probe **2** was produced by direct labeling of StrepTag-HaloTag-HCF3R using an AF488-tetrafluorophenyl (TFP) ester (Thermo-Fisher), using the same protocol as CoraFluor-1-*pfp* labeling with the following modifications: (1) 5 equivalents of AF488-TFP ester were used; (2) protein was not diluted to 50% glycerol following labeling; and (3) concentration and degree of labeling were taken by absorbance at 280 and 495 nm, using a correction factor of 0.11 and a molar extinction

coefficient for AF488 at 495 nm of 71,000 $M^{-1}cm^{-1}$. Labeled probe was aliquoted, flash frozen in liquid nitrogen, and stored at -80° until usage.

Saturation binding of OSMI-4-FITC to OGT

Aliquoted proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates. Protein concentrations were then re-calculated by A_{280} (and A_{494} or A_{495} for labeled protein). Master mixes were prepared containing CoraFluor-1-labeled anti-His_6 antibody (1 nM final), DTT (2.5 mM final), and His_8-tagged OGT (1 nM final) in HBS supplemented with 20 mM $MgCl_2$, 0.5 mg/mL bovine serum albumin (BSA, Sigma-Aldrich) and 0.005% (v/v) Tween-20. Negative control mixes were prepared identically with the addition of no OGT. Each mix was pipetted into one column per replicate of a 384-well, low-volume plate (Corning #3824) at 20 μL/well using an electronic repeating multichannel pipette. OSMI-4-FITC was diluted to the target concentration in anhydrous DMSO, then dispensed in 1:2 dilution series across each column using a D300e digital dispenser (Tecan) fitted with a T8+ dispensehead (HP), with the final row having no probe. Plates were covered in sealing tape (Perkin-Elmer), briefly centrifuged to eliminate bubbles (1500 g, 1 minute), and incubated in the dark at room temperature for at least 4 hours. Plates were then moved to 4°C for overnight incubation, after which they were warmed to room temperature over a period of at least 4 hours before data collection (note: $\geq$ 24 hours of total incubation).

TR-FRET measurements were acquired on a Perkin-Elmer EnVision plate reader using 340nm excitation with a 407nm/505nm advanced dual mirror and 520nm (25nm bandwidth, FITC) and 495nm (10nm bandwidth, Tb) filters, using a 50 μs delay and 300 μs integration. FRET ratio was taken as the ratio of 520nm/490nm emission on a per-well basis, then averaged across replicates. Data was corrected by subtracting the average of wells containing no OGT (corrected error determined by $\sigma_{corr} = (\sigma_1{}^2 + \sigma_2{}^2)^{1/2}$), then fitted to a 'one site – specific binding' model in GraphPad Prism 10. Data were normalized to a fractional binding model using built-in normalization tools in GraphPad Prism to scale the B_{max} to 1.0.

Saturation binding of probe 1 or probe 2 to OGT variants

Aliquoted proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and

4°C to pellet aggregates. Protein concentrations were then re-calculated by A_{280} (and A_{494} or A_{495} for labeled protein). Master mixes were prepared containing CoraFluor-1-labeled anti-His_6 antibody (1 nM final), UDP (0.1 mM final, when required), DTT (2.5 mM final), and His_8-tagged OGT (100 pM for experiments with **2** excluding asparagine mutants; 1 nM for all others) in HBS supplemented with 20 mM $MgCl_2$, 0.5 mg/mL bovine serum albumin (BSA, Sigma-Aldrich) and 0.005% (v/v) Tween-20. Negative control mixes were prepared identically with the addition of no OGT. Each mix was pipetted into one column per replicate of a 384-well, low-volume plate (Corning #3824) at 20 μL/well using an electronic repeating multichannel pipette. Fluorescent probes were diluted to the target concentration in HBS supplemented with 1 mM DTT and 0.3% (v/v) Tween-20, then dispensed in 1:2 dilution series across each column using a D300e digital dispenser (Tecan) fitted with a T8+ dispensehead (HP), with the final row having no probe. Plates were covered in sealing tape (Perkin-Elmer), briefly centrifuged to eliminate bubbles (1500 g, 1 minute), and incubated in the dark at room temperature for at least 4 hours. Plates were then moved to 4°C for overnight incubation, after which they were warmed to room temperature over a period of at least 4 hours before data collection (note: $\geq$ 24 hours of total incubation).

TR-FRET measurements were acquired on a Perkin-Elmer EnVision plate reader using 340nm excitation with a 407nm/505nm advanced dual mirror and 520nm (25nm bandwidth, FITC) and 495nm (10nm bandwidth, Tb) filters, using a 50 μs delay and 300 μs integration. FRET ratio was taken as the ratio of 520nm/490nm emission on a per-well basis, then averaged across replicates. Data was corrected by subtracting the average of wells containing no OGT (corrected error determined by $\sigma_{corr} = ({\sigma_1}^2 + {\sigma_2}^2)^{1/2}$), then fitted to a 'one site – specific binding' model in GraphPad Prism 10. When required, data were normalized to a fractional binding model using built-in normalization tools in GraphPad Prism to scale the B_{max} to 1.0.

Competition binding of competitive peptides to OGT variants

Aliquoted proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and 4°C to pellet aggregates. Protein concentrations were then re-calculated by A_{280} (and A_{494} or A_{495} for labeled protein). Master mixes were prepared containing CoraFluor-1-labeled anti-His_6 antibody (1 nM final), UDP (0.1 mM final), DTT (2.5 mM final), His_8-tagged OGT-WT (1 nM), and probe **1** at its K_D (450 nM) in HBS FRET buffer. Untagged competitor peptides were diluted to the target

concentration in HBS supplemented with 1 mM DTT and 0.3% (v/v) Tween-20, then dispensed in 1:2 dilution series across each column using a D300e digital dispenser (Tecan) fitted with a T8+ dispensehead (HP), with the final row having no competitor. Plates were covered in sealing tape (Perkin-Elmer), briefly centrifuged to eliminate bubbles (1500 g, 1 minute), and incubated in the dark at room temperature for 4 hours. Plates were then moved to 4°C for overnight incubation, after which they were warmed to room temperature over a period of 4 hours before data collection.

TR-FRET measurements were acquired on a Perkin-Elmer EnVision plate reader using 340nm excitation with a 407nm/505nm advanced dual mirror and 520nm (25nm bandwidth, FITC) and 495nm (10nm bandwidth, Tb) filters, using a 50 μs delay and 300 μs integration. FRET ratio was taken as the ratio of 520nm/490nm emission on a per-well basis, then averaged across replicates. Top and bottom measurements were determined by the average of the "no competitor" wells and the highest dosage wells, respectively, which were used to normalize data to represent fractional binding. Data were fitted to an '[inhibitor] vs. response – variable slope (four parameters)' model in GraphPad Prism 10 and normalized to a fractional binding model by using built-in normalization tools in GraphPad Prism to scale the 'bottom' to 0.0 and the 'top' to 1.0.

***In vitro* cleavage of HCF3R-EEE by OGT variants**

See Chapter 2 Methods for detailed methodology.

***In vitro* glycosylation of HCF3R-SAA by OGT variants**

See Chapter 2 Methods for detailed methodology.

***In vitro* glycosylation of HeLa S3 cell extracts by OGT variants**

See Chapter 2 Methods for detailed methodology.

***In vitro* inhibition of OGT by HCF3R in HeLa S3 cell extracts**

Extracts were thawed on ice, then centrifuged at 45000 rpm for 20 minutes at 4°C to pellet cell debris and aggregates. Supernatant was removed and aliquoted into fresh tubes in 24 μL volumes. OGT and HCF3R aliquots were thawed on ice, then centrifuged at 21130 rpm for 10 minutes at 4° to pellet aggregated proteins and re-quantified by A_{280} on Nanodrop. 5x concentration mixes of each OGT were constructed on ice in 6 μL volumes containing 5 μM OGT (or 0 in

negative controls), 5 mM DTT, 5 mM UDP-GlcNAc, and between 0 and 50 μM HCF3R in HBS supplemented with 20 mM $MgCl_2$.

6 μL of each 5x mix was added to 24 μL cell extract and pipetted briefly to mix. Samples were incubated at 37°C with 300 rpm mixing. After 15 minutes, each sampled was removed, combined with 30 μL 2x reducing Laemmli buffer, and boiled at 95°C for 10 minutes. Samples were loaded at 7 μL per well on two separate 4-20% Tris/Glycine SDS-PAGE gel (Bio-Rad), which were run for 50 minutes at 180V. Protein was transferred to PVDF membranes and blotted using pan-*O*-GlcNAc and polyhistidine-HRP antibodies as previously described.

Time course of HCF3R expression and degradation in (*TetON*)-GFP-degron-2xFLAG-HCF3R HCT116 cells

Sorted HCT116 cells containing HCF3R genes under doxycycline-inducible promoters were thawed into complete McCoy's 5A medium. Once confluent, cells were seeded into two 6-well plates at 500,000 cells/well. The time course started the following day. At 0 hours, doxycycline was added to 10 μg/mL in well 6 of plate 1 and wells 1-6 of plate 2. Doxycycline was added to 10 μg/mL in plate 1 wells 5, 4, 3, and 2 at 18, 20, 22, and 23 hours. At 24 hours, all cells from plate 1 were harvested by trypsinization and lysed, and lysates were frozen and stored at -80°C. At 24, 42, 44, 46, and 47 hours, media was removed from plate 2 wells 6, 5, 4, 3, and 2, respectively, and replaced with complete McCoy's 5A medium supplemented with 250 nM dTAGV-1. At 48 hours, all cells in plate 2 were harvested by trypsinization and lysed. Plate 1 lysates were thawed on ice and all protein concentrations were determined by BCA assay. Western blotting was conducted as previously described, and images were color- and brightness-adjusted in FIJI.

Live cell confocal fluorescence microscopy of dTAG-MEFs expressing GFP-2xFLAG-HCF3R

dTAG-MEFs constitutively expressing GFP-2xFLAG or mEGFP-2xFLAG-HCF3R were grown to confluence in 10-cm dishes, then seeded into MatTek #1.5 6-well plates at 50,000 cells/well in complete DMEM. Once confluent, media was exchanged for complete DMEM lacking phenol red for imaging.

Images were acquired in an OKO Lab heated enclosure with CO_2 control at 37°C and 5% CO_2 using a Yokogawa CSU-W1 spinning disk scan head fitted with a 50 μm, 4000 rpm pinhole

disk, mounted on a Nikon Ti inverted microscope. A 488 nm laser was used to excite the GFP fluorophores on both constructs (1x1 binning, 100ms exposure time, rolling shutter at 12-bit dual gain 1/4, readout 200 MHz, 25% power), and emission was read out using an Andor Zyla 4.2 Plus sCMOS monochrome camera fitted with a Chroma ET 525/36 filter. All images were acquired using a Plan Fluor 40x/1.3 Oil DIC H/N2 objective and NIS Elements AR 5.02 software. Image processing was performed using FIJI.

Fixed cell confocal fluorescence microscopy of (*TetON*)-GFP-degron-2xFLAG-HCF3R HCT116 cells

Three days prior to fixation, HCT116 cells expressing GFP-FKBP12^{F36V}-2xFLAG-HCF3R under the *TetON* promoter were seeded into three wells of an 8-well #1.5 glass-bottomed chambered slide (Nunc Lab-Tek II) at a confluence of 25,000 cells/well in 400 μL complete McCoy's 5A medium. 48 hours prior to fixation, one well was treated with 10 μg/mL doxycycline to induce expression; 24 hours prior to fixation, media was removed from this well and replaced with complete McCoy's 5A medium supplemented with 250 nM dTAGV-1. To another well, doxycycline was added to 10 μg/mL 24 hours prior to fixation.

To fix cells, media was aspirated, and cells were washed twice with dPBS then incubated in a solution of 8% (v/v) formaldehyde (Millipore-Sigma) in dPBS for 45 minutes at room temperature in a covered box. Cells were washed four times in dPBS, then permeabilized overnight in 0.1% (v/v) TritonX-100 in dPBS at 4°. Cells were stored in this solution until imaging, at which point they were warmed to room temperature and stained in 200 μL dPBS supplemented with 0.1% (v/v) TritonX-100, 1:5000 DAPI dilactate, and 1:40 phalloidin-Alexa Fluor 568 conjugate (Thermo Fisher) for 1 hour at room temperature. Stain was removed and cells were washed four times in dPBS, storing in 400 μL dPBS for imaging.

Images were acquired using a Yokogawa CSU-W1 spinning disk scan head fitted with a 50 μm, 4000 rpm pinhole disk, mounted on a Nikon Ti inverted microscope. A 405 nm laser was used to excite DAPI (1x1 binning, 200ms exposure, rolling shutter at 12-bit Gain 4, readout 200 MHz, 10% power), a 488 nm laser was used to excite GFP (1x1 binning, 400ms exposure, rolling shutter at 12-bit Gain 4, readout 200 MHz, 25% power), and a 561 nm laser was used to excite AF568-labeled phalloidin (1x1 binning, 200ms exposure, rolling shutter at 12-bit Gain 1/4, readout 200 MHz,

10% power). Emission was read out using an Andor Zyla 4.2 Plus sCMOS monochrome camera fitted with Chroma ET 455/50 (DAPI), 525/36 (GFP), and 605/52 (AF568-phalloidin) filters. All images were acquired using a Plan Fluor 40x/1.3 Oil DIC H/N2 objective and NIS Elements AR 5.02 software. Image processing was performed using FIJI. GFP levels were normalized across all images in order to enable direct comparison of intensity. Representative images were chosen for each construct. Maximum intensity projections calculated from a z-stack were displayed.

FLAG co-immunoprecipitation from (*TetON*)-GFP-degron-2xFLAG-HCF3R HCT116 cells and data analysis

Prior to the experiment, six 10-cm plates were seeded equally with HCT116 cells expressing GFP-degron-2xFLAG-HCF3R under the TetON promoter, in complete McCoy's 5A medium. 48 hours before these plates reached confluence, media on three plates was replaced with complete McCoy's 5A medium supplemented with 10 μg/mL doxycycline, while the media on the other three plates was replaced with complete McCoy's 5A medium supplemented with 250 nM dTAGV-1.

48 hours later, cells were placed on ice, media was aspirated, plates were washed in 1 mL dPBS, and cells were harvested by scraping. Cells were pelleted by centrifugation (500 g, 4°, 5 minutes), then lysed in Pierce native lysis buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% (v/v) IGEPAL CA-630, 5% (v/v) glycerol, 1x Halt protease/phosphatase inhibitor cocktail) on ice for 10 minutes. Lysates were clarified by centrifugation (max speed, 4°, 15 minutes), and lysate protein concentrations were estimated by BCA assay.

For each lysate (3x +dox replicates, 3x +dTAG replicates), 90 μL of M2 anti-FLAG magnetic bead slurry was combined with 1.9 mg protein and diluted to 200 μL in cold HBS. Bead-bound lysates were incubated at 4° for 16 hours with gentle rotation. Flowthrough was removed, and beads were washed three times in 150 μL cold HBS for 10 minutes at 4° with gentle rotation. Bound proteins were eluted by addition of 150 μL 3xFLAG peptide (150 ng/μL in HBS; Genscript) at RT for 1 hour with gentle rotation.

Bound protein was assessed by SDS-PAGE and western blotting to confirm retention of the FLAG-tagged species through flowthrough and wash steps. Each elution was diluted to 25 μL in 1x reducing Laemmli SDS-PAGE loading buffer, then separated on 10% Tris/Glycine SDS-PAGE

for 7 minutes (just until the protein exited the stacking region). Gels were Coomassie stained in glass dishes, then destained overnight in 30% methanol and 10% acetic acid. Gels were then rinsed in water, imaged, and protein bands were cut out into separate microcentrifuge tubes and covered with 100 μL milliQ-water. Samples were submitted to the Taplin Mass Spectrometry Facility at Harvard Medical School for proteomics analysis on an Orbitrap mass spectrometer.

Data were assessed very simply by averaging peptide counts from all three +dox samples and all three +dTAG samples and comparing averaged counts on a scatter plot. Due to the not-strictly-quantitative nature of the sample preparation and proteomics readouts, statistical significance metrics were not calculated.

Chapter 4

Biochemical and structural dissection of the OGT-HCF3R interaction

Some of the work described in this chapter has been published in the following paper, on which I am the first author:

Hammel, F.A.; Payne, N.C.; Marando, V.M.; Mazitschek, R.; Walker, S. "Identification of a Polypeptide Inhibitor of *O*-GlcNAc Transferase with Picomolar Affinity." *Journal of the American Chemical Society* **2024**, *146* (38):26320-26330.[109]

In this chapter, I investigate the means by which OGT and HCF3R bind with picomolar affinity using a combination of cryo-electron microscopy (cryo-EM) and the binding assay toolkit developed in Chapter 2. The work in this chapter was enabled by a variety of people and resources, most notably Dr. Morgan Gilman, a postdoctoral fellow in Andrew Kruse's lab who helped me get started processing my cryo-EM data and remained an amazing sounding board for structural ideas; Dr. Richard Walsh and Megan Mayer at the Harvard Cryo-Electron Microscopy Center for Structural Biology, who taught me how to prepare sample grids and collect data for cryo-EM and were overall an invaluable resource; and the SBGrid Consortium,[182] whose computational resources and staff were a necessary part of my attempts at structural biology during my doctoral studies. While most of the background supporting this chapter was part of my thesis project (see Chapter 3), some of the original research was done previously. In particular,

Dr. John Janetzko's Ph.D. project and dissertation provided a wealth of thoughts and ideas surrounding the OGT-HCF3R interaction, and his work with Dr. Michael Lazarus on the structure of OGT bound to HCF1R and UDP-GlcNAc derivatives has been an amazing starting point for my own structural modeling. The work done here will be presented in two parts, following an introduction concerning principles of cryo-EM and structural solution. In the first part, I will describe the advances I have made towards generating atomic-resolution density maps of the OGT-UDP-HCF3R complex. In the second part, I will attempt to define the key features of the complex by biochemical characterization. Unfortunately, as will become evident throughout this chapter, this was a painstakingly difficult endeavor, and I did not ultimately resolve the structural features that fully explain the picomolar affinity of HCF3R for OGT. However, this was still a valuable experience, and I learned a great deal about structural biology tools in the process. As I walk through the steps I took in trying to solve this structure, I will attempt to explain my rationale and, in the case that my methods did not work, I will try to explain what went wrong. In doing so, I hope to provide a clear depiction of my efforts in this project.

4.1 Structural solution of proteins by cryo-electron microscopy (cryo-EM)

For decades, structural solution of proteins and protein complexes was dominated by two methodologies: X-ray crystallography, which shoots beams of X-rays at protein crystals and analyzes diffraction patterns produced at different rotations of the crystals to generate electron density maps, and protein NMR, which typically involves a combination of 2-dimensional NMR methodologies applied on samples in solution. While the union of these methods is generally seen as robust, they are not without their general drawbacks. For example, X-ray crystallography requires that protein samples crystallize, which is not always possible, especially for large, flexible, or otherwise conformationally dynamic samples.[183–185] Likewise, protein NMR eliminates the necessity for rigidity, but is really only feasible on very small samples whose residues can be readily assigned.[186]

In cryo-EM, a beam of electrons is focused at and transmitted through protein samples encased in vitreous (*i.e.*, non-crystalline) ice, and the resultant projections are collected and used

for analysis. While cryo-electron microscopy (cryo-EM) is not a *new* technology – researchers demonstrated its ability to solve structures of various virus particles in the 1980s[187] – a series of technical advancements made over the 2000s and 2010s have made it much more accessible to the average structural biologist.[188] As a result, it is gaining widespread popularity; currently, cryo-EM depositions to the PDB are outpacing crystallographic depositions, and that trend appears to be permanent.

Cryo-EM offers many benefits over other methods for structural characterization. For one, its sole requirement is that protein can be obtained from some source, whether that be from heterologous purification or homogenization of organismal tissue. No crystallization is required, and most buffer systems are compatible with the vitrification techniques available. Moreover, cryo-EM is well suited towards solution of large macromolecular structures, including entire virus particles. Furthermore, because the software packages used to analyze cryo-EM data (see Section 4.3) excel at separating out different conformational states, conformational heterogeneity is tolerated, so complexes do not have to be rigorously purified to homogeneity prior to data collection.

Cryo-EM is not without its drawbacks, particularly in terms of molecular size. Currently, structural solution of complexes less than 50 kDa in size is seen as difficult-to-impossible,[189,190] although new advancements to hardware and software may lower this limit in the future. Additional advances in fusion protein strategies to increase particle size, as seen in the GPCR field with BRIL fusions, may artificially lower this limit.[191] Furthermore, optimization of grid preparation conditions can be time-consuming and empirical, as the only means to assess grid viability is to take images on the electron microscope itself. This often leads projects to stall at the stage of "screening by microscope" due to lack of microscope access. Despite these flaws, cryo-EM appears to be the principal methodology by which new structures are being solved.

4.2 Advances in OGT structural biology enabled by cryo-EM

Prior to 2021, there were no published structures of full-length OGT in the protein databank (PDB). Moreover, all structures published prior to that point had been solved by X-ray crystallography. The first structure containing any significant portion of OGT was published in 2004,[58] and consisted

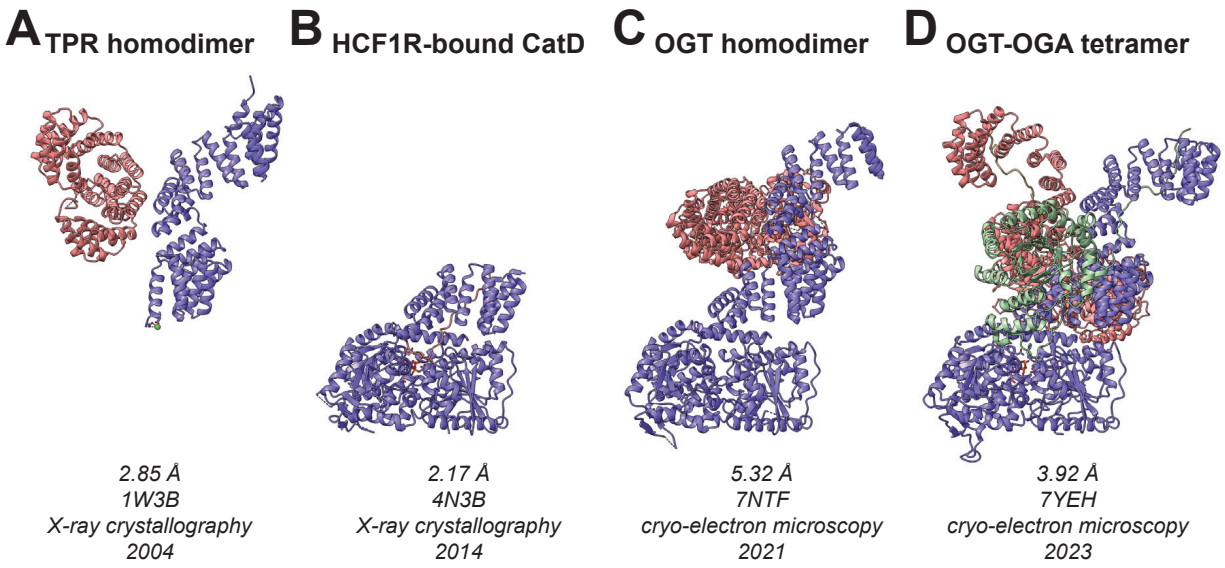


Figure 4.1: OGT structures through the years. (A) Structure of dimerized TPR domains, determined by X-ray crystallography in 2004.[58] (B) Structure of OGT-4 bound to UDP-GlcNAc and HCF1R, determined by X-ray crystallography in 2013.[46] (C) Structure of homodimeric OGT, determined by cryo-EM in 2021.[59] (D) Structure of homodimerized OGT-OGA heterodimers, determined by cryo-EM in 2023.[57] Figure produced using UCSF ChimeraX.[6–8]

of a homodimer of OGT's TPR domain (residues 23-410; Figure 4.1A). This structure provided the basis for our biochemical understanding of the TPR domain for the better part of a decade, until various groups including the Walker lab published structures of the OGT-4 truncation bound to small molecule inhibitors or peptide substrates (Figure 4.1B).[31,54] It was hypothesized, based on the TPR-only crystal structure and the nature of the TPRs, that the flexibility of the TPR superhelix would thwart any attempts at crystallizing full-length OGT; indeed, full-length structures were only made possible through usage of cryo-electron microscopy. In 2021, Gideon Davies's group published an apostructure of full-length OGT solved by cryo-EM (Figure 4.1C).[59] This structure was solved at relatively low resolution, and did not fully encapsulate the N-terminal TPRs; however, it did confirm that full-length OGT homodimerizes *in vitro*. More recently, a group from Westlake University in China published two higher-resolution cryo-EM structures of full-length dimerized OGT, one unbound and one bound as a dimer of dimers to the substrate protein OGA (Figure 4.1D).[57] These structures provided a much better view of OGT's N-terminus, and offered insights into how non-HCF-1 substrates might engage the TPR lumen through via direct hydrogen-bonding interactions with residues throughout the entire TPR domain.

With this buildup of knowledge in the literature, it was very clear from the start that if I wanted to solve the structure of the OGT-HCF3R complex, I would need to use cryo-EM.

4.3 Software packages used in cryo-EM data processing and analysis

4.3.1 CryoSPARC

CryoSPARC (**Cryo**-EM **S**ingle **P**article **A**b-Initio **R**econstruction and **C**lassification) is a software package developed and managed by Structura Biotechnology, Inc., that is primarily used to process and analyze cryo-EM data, reconstructing density maps from single- and multi-frame grid movies taken from the microscope. It can also be used to process cryo-electron tomography (cryo-ET) and negative stain data, although those functionalities are less well-used and will not be described here. Importantly, CryoSPARC is constantly being updated and improved, so some of the information here will likely become obsolete at some point; the background and methodology described here is current as of CryoSPARC v4.4.1. Information on current and in-development CryoSPARC tools, as well as in-depth discussion on usage, can be found at the CryoSPARC Discuss forums (https://discuss.cryosparc.com/), which have been indispensable for my understanding of how to use this toolkit.

On the back end, CryoSPARC's primary initial achievement was the combinatorial usage of two algorithms: one enabling unsupervised *ab initio* classification of 3-dimensional sample states without input of an initial structure, and the other speeding up structural refinement by eliminating redundancies in image alignment steps. The first advancement was enabled by usage of stochastic gradient descent starting from arbitrary initial structures, which allows for the algorithm to sample a wide variety of potential structures in 3D space before converging on a probable structure. The second advancement uses a 'branch-and-bound' approach to set a lower bound for alignment error and eliminate potential 3D poses that are above this bound, reducing computational costs and vastly improving speed.

CryoSPARC's front end also represents great improvements over other structural solution packages. First, CryoSPARC is accessible by a web browser-based graphical user interface (GUI) that enables simple access and setup of all jobs within the software package. This GUI can also be operated on a high-performance cluster (HPC) virtually, lowering the barrier to entry for many

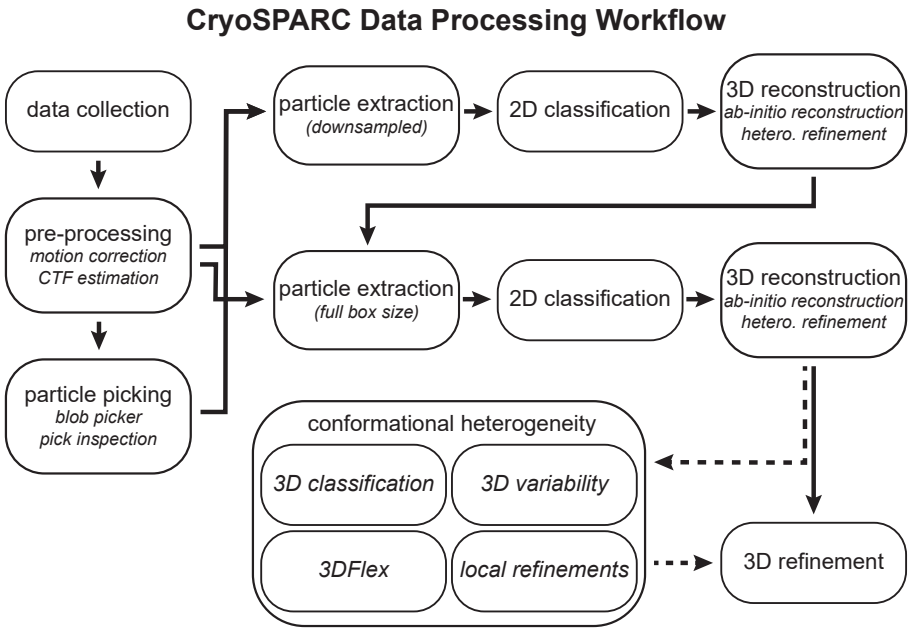


Figure 4.2: Standard data processing pipeline in CryoSPARC.

research groups. Within the GUI, jobs can be queued in sequence, parameters are easily modified to tune various jobs, and all job outputs are easily visualized without the need for downloading files or using additional visualization software.

In CryoSPARC (Figure 4.2), the standard pipeline for data processing consists of pre-processing steps (movie motion correction and contrast transfer function (CTF) estimation), particle picking and extraction, 2D classification, *ab initio* 3D reconstruction from classes representing particles, then various 3D classification and refinement jobs until a high-resolution map is obtained. Depending on the nature of the dataset, various additional jobs can be performed (*e.g.*, symmetry expansion for symmetric complexes), but the bulk of the data processing route remains constant.

4.3.2 RELION

Whereas CryoSPARC is a sleek and shiny plug-and-play package that sometimes lacks in functionality and manipulability, RELION (**RE**gularized **LI**kelihood **O**ptimizatio**N**) is its no-frills, no-guiderails, more-customizable counterpart. RELION was developed as an open-source toolkit for cryo-EM structural solution and analysis in the early 2010's,[192] and it maintains that open-source status even in its current iteration, RELION-5. It is accessible via command line and has a

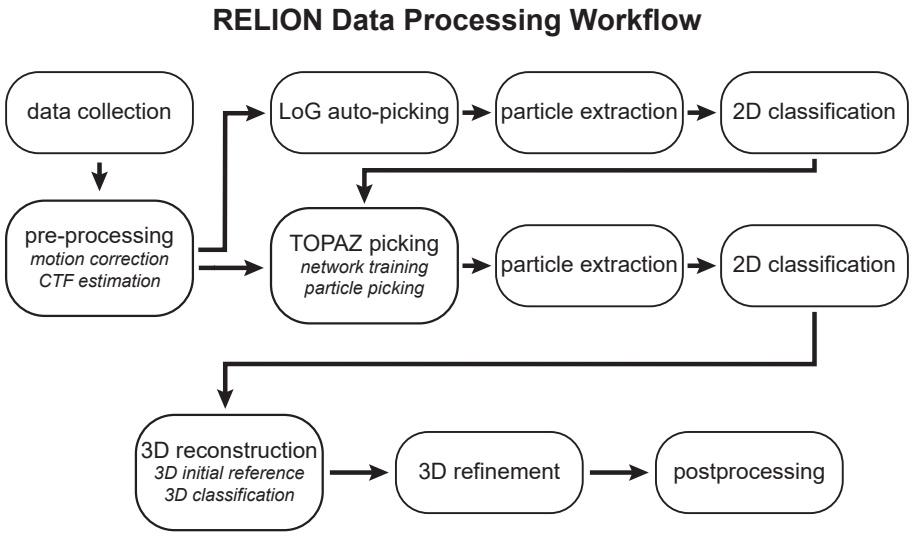


Figure 4.3: Standard data processing pipeline in RELION.

barebones GUI that enables parameter tuning and data access.

RELION was developed prior to CryoSPARC, and thus its computational advancements were more significant compared to its predecessors. It implements a Bayesian approach to determination of structures, which its initial developer, Sjors H.W. Scheres, described as a reformulation of the "reconstruction problem as… the optimization of a single target function" to find a model that fits both the data and available prior information concerning the dataset.[192] At the time, this was a huge advancement in democratizing access to cryo-EM data processing, as the existing tools were only usable by experts in the field and had parameters that needed to be manually tuned over the course of data processing, whereas RELION's Bayesian approach self-tuned parameters using the data itself. Still though, RELION relies on a far greater amount of user input compared to CryoSPARC, and determination of core parameters can only be done empirically in many cases, making RELION much more finicky and difficult to operate effectively as a non-expert user.

A typical data processing scheme in RELION looks similar to CryoSPARC (Figure 4.3). First, movies are preprocessed to correct for motion and estimate the CTF, then particles are picked and 2D class averages are generated. From the 2D classes, RELION can generate *de novo* 3D references, which are then used for 3D classification. Good 3D classes can be refined in high resolution and then passed to end-stage processing jobs, including per-movie CTF and aberration refinement, per-particle motion-correction ("Bayesian polishing"), and mask-focused local classification (essentially,

running 3D classification using a focus mask). The key difference is that because RELION is so highly tunable, many groups have found ways to trick it into doing specific tasks that it was not necessarily hard-coded into doing. This includes, for example, "focused classification with signal subtraction," a method that classifies subtracted particles within a focus mask without realigning to enable reproduction of the full particle after de-subtraction.[193] Moreover, where CryoSPARC often codes in the most likely usage case for certain jobs, RELION requires much more user input, which can result in misuse when not properly monitored.

4.4 Isolation and purification of the OGT-HCF3R-UDP heterotrimer

Because OGT and HCF3R associate with a picomolar $K_{D,app}$ in the presence of UDP, I sought to use a simple size exclusion chromatography (SEC) step to isolate the complex. I first ran loading control samples consisting of only His_8-OGT-WT or His_8-HCF3R-SAA on a Superdex 200 Increase column to determine their elution volumes. I then combined the two proteins at a 5:1 ratio of HCF3R to OGT in the presence of saturating UDP and re-purified them by SEC; complexation induced a slight leftward shift in elution time compared to OGT-WT alone (Figure 4.4). I then concentrated these samples to 0.5 mg/mL in buffer containing additional saturating UDP (0.1 mM), dithiothreitol as a reducing agent, and glycerol (0.5% v/v), and prepared grids for sample collection. Importantly, because the nominal OGT concentration of 0.5 mg/mL (18.2 μM) is significantly greater than OGT's $K_{D,dimer}$ of 39 nM, the purified complex was expected to contain fully homodimerized OGT.

I took the purified complex and vitrified several grids of two varieties (C-flat holey carbon on copper 1.2/1.3 and UltrAUfoil holey gold 1.2/1.3) at different concentrations diluted from the stock solution, using a variety of vitrification conditions (varied blot force and blot time), screening for optimal grid conditions on the Talos Arctica microscope at the Harvard Cryo-EM Center with the assistance of Megan Mayer and Dr. Richard Walsh. We ultimately found the following conditions to be optimal, yielding grids with dense but separated complex particles: 5 blot force, 7 second blot time, 10 second wait time, and 0.5 mg/mL complex on the UltrAUfoil grids. We then proceeded to data collection using a single optimized grid on the Titan Krios II microscope, where we ended up with ~15,000 total movies (representative microscope image in

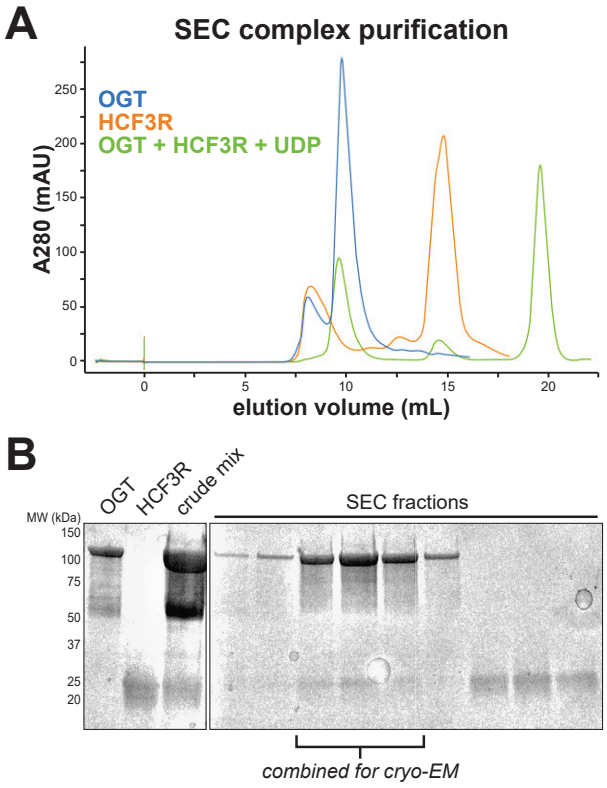


Figure 4.4: Purification of the OGT-HCF3R-UDP complex. (A) Size exclusion chromatography traces of OGT (blue), HCF3R (orange), and the OGT-HCF3R-UDP complex (green). (B) InstantBlue stains of fractions from size exclusion chromatography.

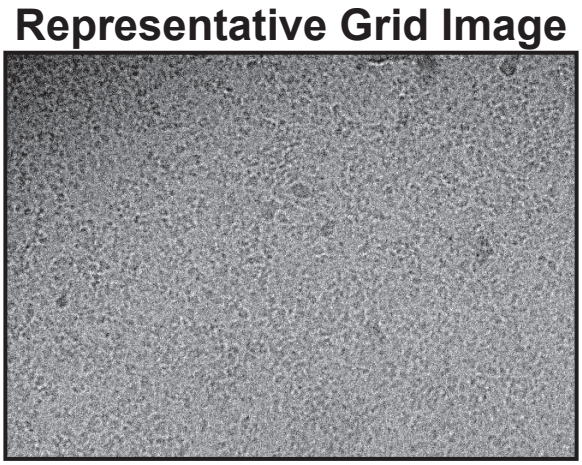


Figure 4.5: Representative grid image from cryo-EM data collection.

Data Collection Parameters	
Magnification (X)	105,000
Voltage (kV)	300
Detector	Gatan K3
Exposure time (s)	2.984
Total electron exposure (e/Å^2)	53.934
Number of frames	54
Dose per frame (e/Å^2)	1
Defocus range (μm)	-0.8, 2.2
Pixel size (Å)	0.825

Table 4.1: Data collection parameters for cryo-EM analysis of the OGT-HCF3R-UDP complex.

Figure 4.5). Full data collection parameters can be found in Table 4.1.

4.5 Initial data processing in CryoSPARC

My first iteration of data processing was performed in CryoSPARC, initially with the supervision and support of Dr. Morgan Gilman, who performed a rudimentary processing pipeline separately to help me follow along. After following standard parameters for importing the raw data, per-micrograph motion correction, and CTF estimation, I picked particles within a range of 75 and 225 Å based on the estimated smallest and largest diameters of the OGT homodimeric complex. I computationally eliminated noisy particles (often due to presence of ice or the edge of the grid), then extracted all remaining particles at a box size of 384 pixels, downsampled by a factor

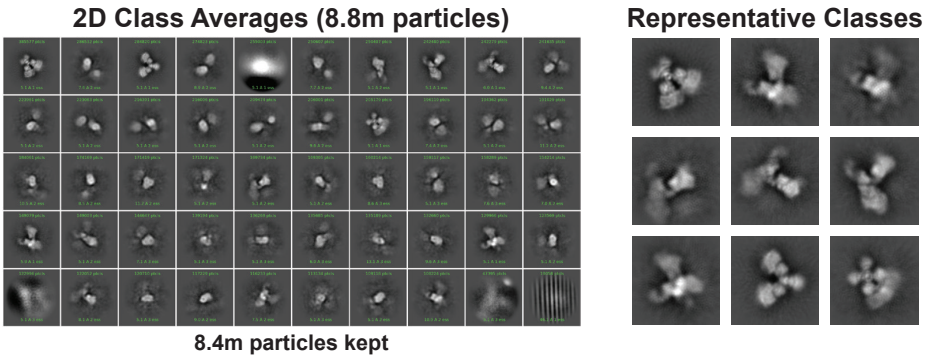


Figure 4.6: 2D class averages resemble dimerized OGT. Selected representative classes are reproduced on the right.

of two to 192 pixels. This box size was chosen based on a protocol from Wen Jiang's Lab at Purdue University,[194] where the optimal box size is generally around 150 Å larger than the largest dimension of the anticipated structure. Since OGT's largest dimension is around 170 Å (the length of one monomer), I needed a box size of ∼320 Å, which corresponds to ∼390 pixels at 0.825 Å/px. The closest optimal box sizes were 384 and 392 pixels, so I chose the lower value at first.

4.5.1 2D class averages reconstitute the OGT homodimer

After extraction of 8.8 million downsampled particles, I performed 2D classification to acquire a set of 50 2D class averages. Of note, this job had to be run using a single GPU, and GPU parallelization resulted in many seemingly empty classes. Classification yielded many well-structured classes that resembled homodimeric OGT, which was highly reassuring (Figure 4.6). I noticed that various poses were captured in classification, including ones where the complex was oriented with the N-termini facing out of the grid, where the complex was viewed from the side, and seemingly where the complex was upside down on the grid. From this initial classification, I selected all classes that weren't completely junk (46 classes, 8.39m particles, 95%) and proceeded to *ab initio* 3D reconstruction.

This process was a major success, as it confirmed that not only did I have millions of workable particles, but also that I would not likely have to return to the microscope for additional data collection, as many unique poses were captured by a simple 2D classification. However, the following sections will detail that, despite the appearance of success, I was still very far away

from a full, workable density map.

4.5.2 3D structures converged on the OGT homodimer with one well-ordered monomeric catalytic domain

Following 2D classification, I performed *ab initio* 3D reconstruction of three volumes using a randomly-selected subset of the good particles from 2D classification ("good" here meaning that they resembled a pose of OGT and were not totally junk; Figure 4.7, top left). From this, I obtained one volume that resembled the OGT homodimer. I pushed all three volumes into a Heterogeneous Refinement job using all non-junk particles, and found that 42% (3.5m) sorted into the OGT dimer class, with the others being split between the two junky classes (Figure 4.7, middle left). As a general confirmation that the refinement converged on an accurate sorting, I took the particles from the best class and re-sorted them into all three classes in a new Heterogeneous Refinement job (Figure 4.7, bottom left). This yielded minimal re-sorting of good particles into junk classes, suggesting that my particles were initially well-sorted.

I took the best class from the initial heterogeneous refinement job (still downsampled by a factor of 2) and conducted non-uniform refinement. I used non-uniform instead of homogeneous refinement because the adaptive marginalization algorithms built into non-uniform refinement aids in resolving flexible or variably-occupied regions,[195] and based on existing structures of OGT, I was anticipating such problems.[57,59] Non-uniform refinement yielded a density map with 3.39Å resolution (Figure 4.7, right), which at high contours (transparent shaded region) lacks density outside of the in-focus monomer's catalytic domain.

4.5.3 Attempts to improve resolution at OGT's N-terminus by 3D classification were unsuccessful

While the density map obtained from non-uniform refinement possessed good local resolution at the in-focus monomer's catalytic domain, density completely fell off at the N-terminus and the out-of-focus monomer. To me, this indicated potential heterogeneity at those regions, an assertion backed up by previous observations of flexibility in the TPRs from both crystallography and cryo-EM data.[58,59]

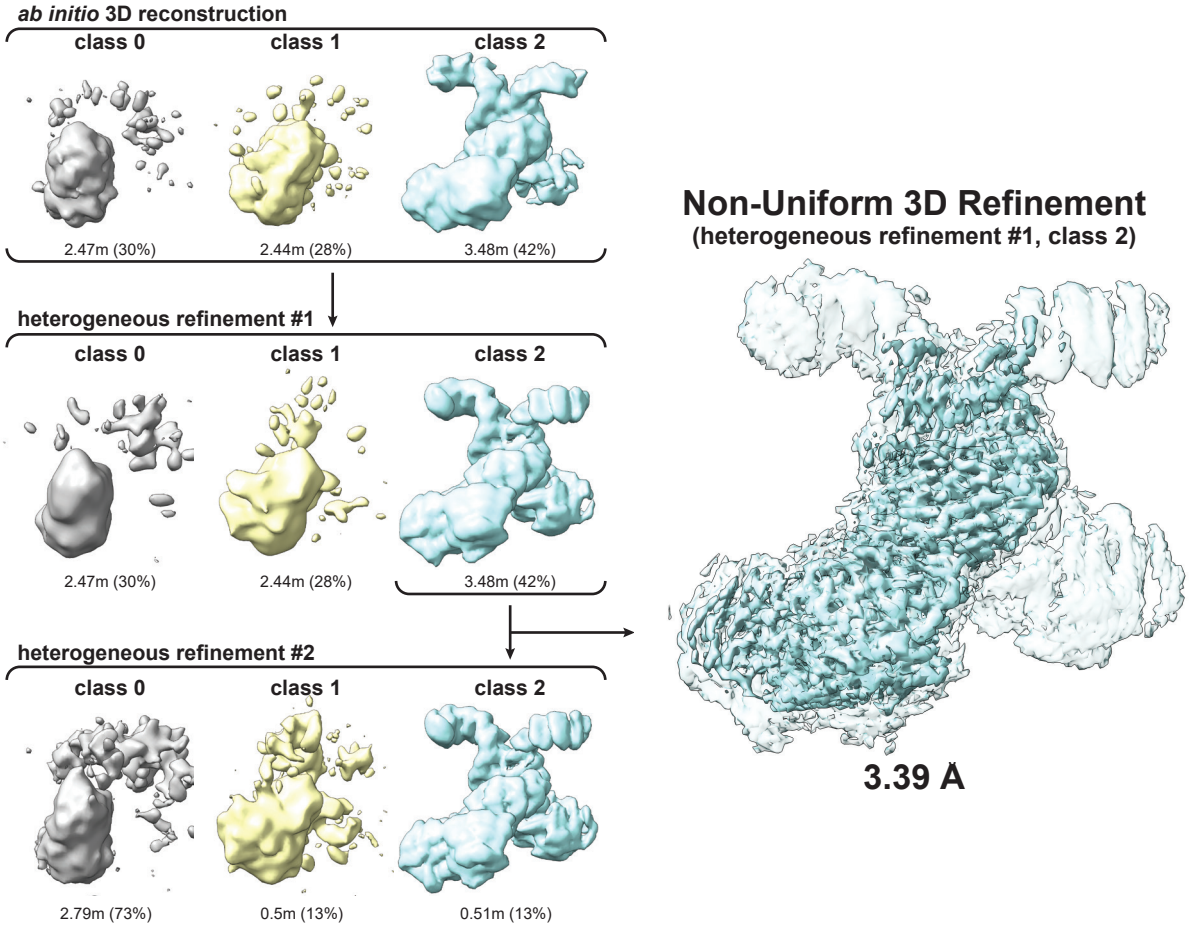


Figure 4.7: 3D reconstruction yielded an OGT homodimer density map. (*left*) Pipeline for 3D reconstruction and heterogeneous refinement to remove junk particles. (*right*) Density map produced by non-uniform refinement of class 2 from heterogeneous refinement 1. Opaque region represents density at a high-resolution contour, while transparent region represents density at a low-resolution contour. Figure produced using UCSF ChimeraX.[6–8]

Conformational heterogeneity in cryo-EM data can generally be dealt with in one of three ways without returning to the microscope: employing a careful "masking" and local refinement strategy (see section 4.8 for further discussion), determining modes of protein motion and fitting particles to states of these modes (see section 4.7 for further discussion), or 3D classification.[196,197] Of these strategies, the third is both the simplest to conceptualize and the easiest to employ, and since I was working with 3.8 million particles, I had more than enough data to populate a diverse variety of class volumes.

I ran a variety of 3D classifications in CryoSPARC from the starting high-resolution refinement, varying parameters (number of classes, target resolution, initialization mode, soft vs. hard classification, etc.) to try to improve the density at the low-resolution regions of the map (example 3D classification results in Figure 4.8A). Additionally, I attempted to make a focus mask to restrict classification to just the in-focus monomer's N-terminus. However, while some outputs from 3D classification were convincing, running non-uniform refinements using the output volumes and particles at their full box size (*i.e.*, no longer downsampled) resulted in more of the same: maps with a well-oriented primary catalytic domain, but lacking continuous density everywhere else (Figure 4.8B).

4.6 Re-processing using new job types and parameters in CryoSPARC

At this point in the project, I hit a bit of a wall. Nothing I was doing was particularly working, and while structural solution is often an empirical endeavor that requires meticulous parameter testing and optimization, I wasn't confident that I had set myself up with a good launching point for such a task. So, I went back to the drawing board and started back at the beginning of the processing scheme, making a few key modifications to the processing pipeline to improve my initial data. Many of the steps described below were adapted from an end-to-end processing guide for an inactive GPCR complex,[198] in which members of the CryoSPARC team describe full processing of a dataset containing multiple complex conformations.

First, I increased the full extraction box size from 392 to 440 pixels. While this may seem frivolous, because my particles seemed to be centered on the in-focus OGT monomer's catalytic domain and not the actual particle center-of-mass (which would be somewhere in the space

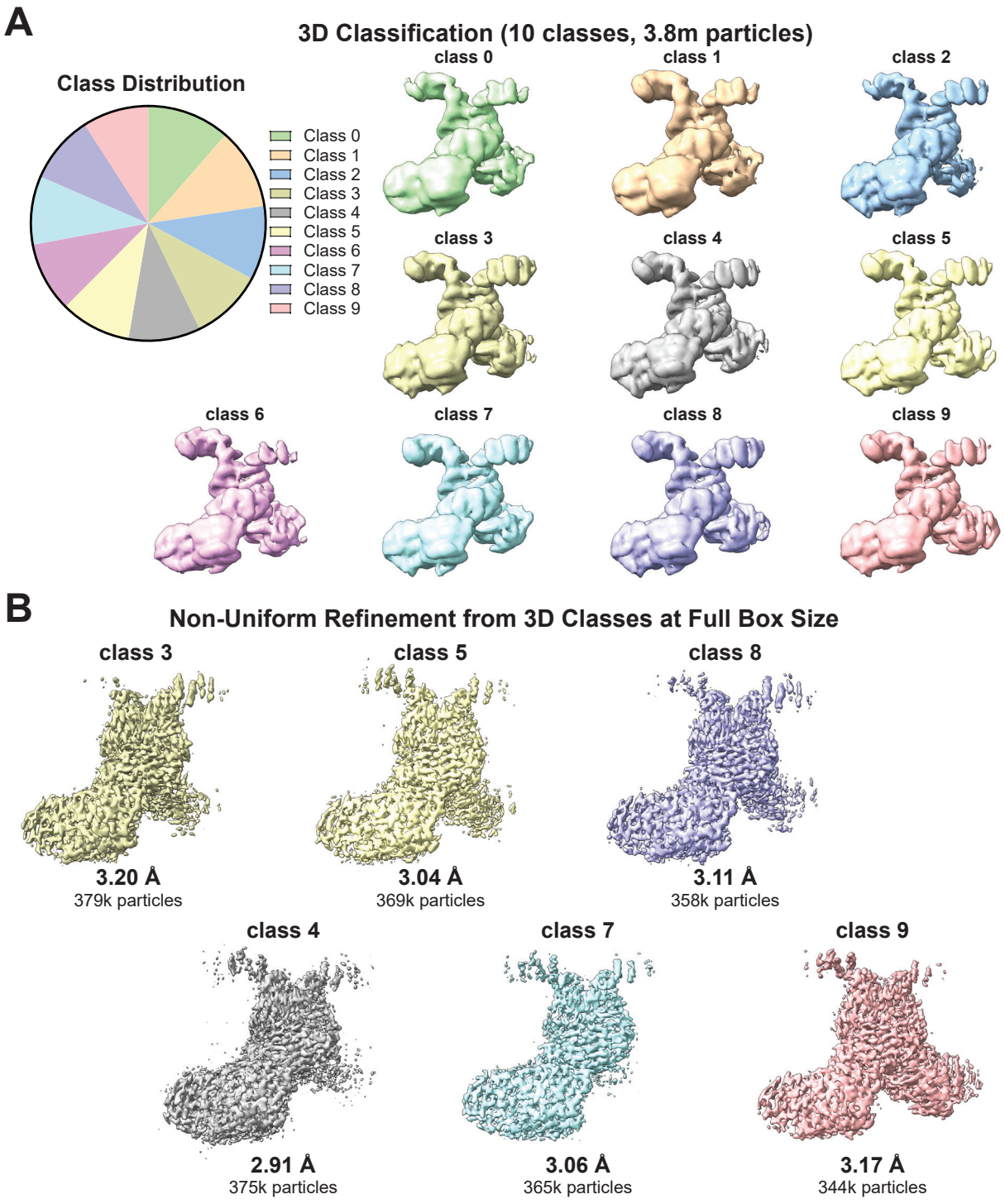


Figure 4.8: 3D classification did not improve density. (A) Class distribution and resultant density maps from a 10-class 3D classification, with soft classification. (B) Density maps from non-uniform refinement of select classes from 3D classification at the un-binned box size. Figure produced using UCSF ChimeraX.[6–8]

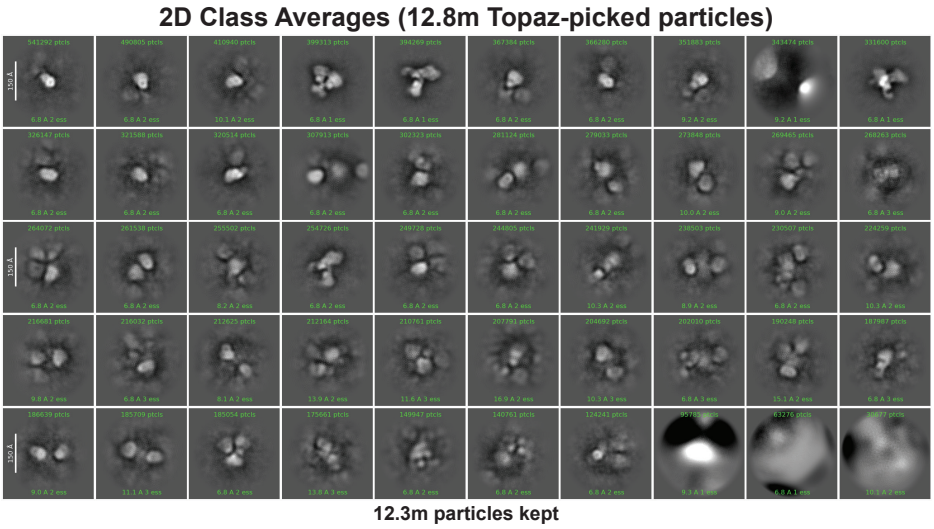


Figure 4.9: 2D class averages from TOPAZ indicate more sample heterogeneity.

between the two medial TPR regions), the box size required to keep those particles fully within the boundary is likely larger than would generally be necessary.[194]

Second, once I obtained new 2D class averages, I fed the "good" classes into TOPAZ, a neural network-based particle picker that excels at picking particles that recapitulate rare poses.[199] This resulted in 12.8 million particles (52% increase from previous count) from good 2D classes (Figure 4.9). I then proceeded through the 3D reconstruction steps, adding CTF refinement steps into the existing non-uniform refinement script to help correct for anisotropy, which may have contributed to the "streaking" phenomena observed in the N-terminus of the previous density maps. I also turned off dynamic masking during the final refinement to prevent the algorithm from locking in on artifactual density. This led to a map corresponding to 2.1 million particles with a nominal resolution of 2.46Å (Figure 4.10A). Importantly, local resolution estimation demonstrated that the new particle stack yielded superior resolution at the N-terminus. Additionally, the conical FSC area ratio (cFAR) and sampling compensation factor (SCF),[200–202] metrics that demonstrate completeness of sampled particle orientations, were well within ranges that would indicate a good dataset (Figure 4.10BC). Moreover, visual inspection of the resultant density map demonstrated improved density at the N-termini, consistent with an improved set of captured poses in picked particles.

With a higher-quality density map in hand, I wanted to see if any further improvements could

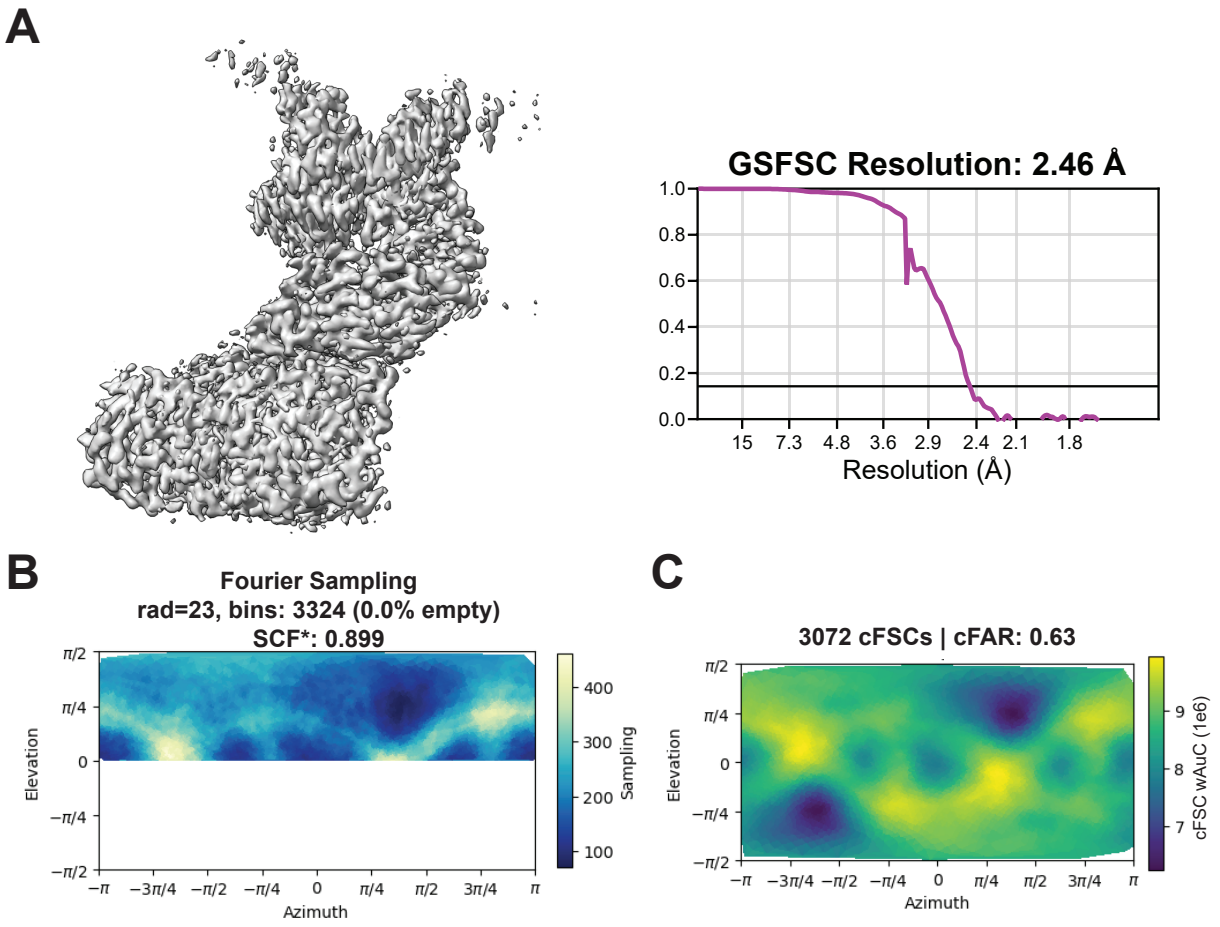


Figure 4.10: Re-processing yielded better density. (A) Autosharpened density map from non-uniform refinement with CTF refinement on and no masking, with a nominal resolution of 2.46Å. (B) Fourier sampling indicated an SCF of 0.899, suggesting good orientation distribution. (C) The dataset had a conical FSC area ratio (cFAR) of 0.63, which also indicates no preferred orientation within the sample. Figure produced using UCSF ChimeraX.[6–8]

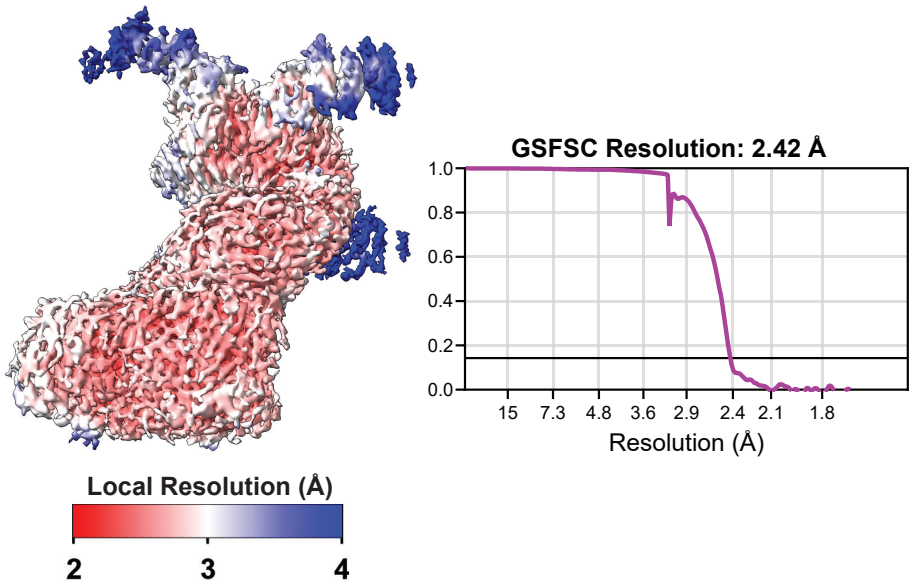


Figure 4.11: Reference-based motion correction slightly improved density and resolution. Figure produced using UCSF ChimeraX.[6–8]

be made before resolving the problems at the N-terminus and the second monomer directly. One way to do so is to utilize reference-based motion correction (RBMC), which is the CryoSPARC equivalent of RELION's Bayesian polishing.[203] RBMC essentially determines the motion of each particle across the frames of the initially collected movies to calculate precise trajectories, which are then used to perform motion correction on the particle. It also determines radiation damage incurred over the course of the movie and performs dose weighting to correct this. Overall, RBMC can be a powerful tool to improve the quality of an already-workable density map. I performed RBMC on the previous non-uniform refinement volume and particles, then re-ran non-uniform refinement, yielding a map with a nominal resolution of 2.42Å, a marginal improvement of 0.04Å from the previous refinement with the same particle stack (Figure 4.11).

4.7 Motion in the complex hinders structural solution of the OGT N-termini and the second monomer

At this stage, I was rather disappointed that I had not achieved a full density map corresponding to homodimeric OGT bound to HCF3R, so I decided to look into what could be causing the loss

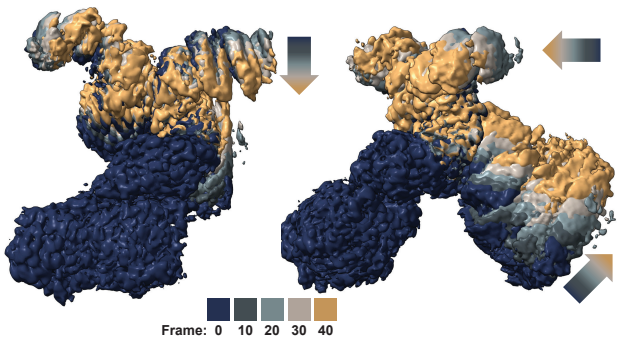


Figure 4.12: 3DFlex reconstituted continuous motion across the particle stack. Arrows indicate movement across frames. Figure produced using UCSF ChimeraX.[6–8]

of density at the N-terminus of the primary monomer and the entire second monomer. Previous structural solution of OGT's TPRs (both by themselves and as part of the full protein) identified that the TPRs are capable of a large degree of flexibility, seen both as lateral movement of the N-terminal TPRs and as 'relaxation' or local unwinding of parts of the TPRs.[57–59] Given this, I assumed the problem to be one of flexibility, or in cryo-EM jargon, "conformational heterogeneity." Unfortunately, the problem of conformational heterogeneity is not one that has been explicitly solved.[196,204]

The first step in resolving this problem is fully characterizing it. To do this, I utilized 3D Flexible Refinement (3DFlex), a tool that is built into CryoSPARC. 3DFlex uses a neural network to model continuous conformational heterogeneity as deformations to a tetrahedral mesh corresponding to the complex volume, and can output motion trajectories along principal components of molecular motion as volume series.[205] I generated a mesh corresponding to the OGT homodimer at a low contour level, trained the 3DFlex model, and used the trained model to generate density maps corresponding to 41 frames of continuous motion along the first principal component (Figure 4.12). The results indicated a large degree of flexibility in the N-terminal TPRs, both in terms of rotation around the lumenal axis, and differential translation of that axis as it extends towards the N-terminus. Similar motion was observed with the second monomer.

One common strategy employed to resolve continuous motion, once it is defined, is 3D classification using densities from continuous motion frames as input volumes.[198] I attempted this with 11 evenly-spaced frames, using a focus mask encapsulating the N-terminus. Sadly, I was met with 3D classes that were incomplete, lacking in the same way as the starting volume (Figure

4.13). While 3D classification can be performed exhaustively, often to reasonable success, this is a very empirical endeavor, and one that I thought would not be worthwhile to pursue unless all other options failed.

4.8 Local refinement marginally improves density at OGT's N-terminus

After confirming that Topaz improved the quality of my picked particles, I also attempted to locally refine the different regions of the primary monomer – the catalytic domain and the flexible N-terminal TPRs. In CryoSPARC, local refinement takes aligned particles from a full refinement job (*e.g.*, non-uniform or homogeneous refinement) and refines a subvolume determined by a padded focus mask around the region of interest. This is particularly helpful for flexible or otherwise variable regions outside of the particle's center-of-mass, as it can overcome shifts and rotations in these regions on a per-particle basis while ignoring the relative rigidity of the center-of-mass. Based on these general principles, I thought this would be a very good starting point for improving the quality of the density map at the N-terminus in the main monomer.

I began this protocol using the non-uniform refined particles after reference-based motion correction. I first downloaded the map, applied a Gaussian filter to smooth it, and created two separate subvolumes containing the primary catalytic domain (CatD-1) and the primary N-terminus (NTerm-1), which I then imported back into CryoSPARC, extended, and padded to create smooth and generous focus masks for local refinement (CatD, green volume in Figure 4.14A; NTerm, pink volume in Figure 4.15A). I then ran local refinement on the aligned particles using each of these focus masks, which improved the resolution at both regions. For the catalytic domain, I noticed better coverage of the region encompassing the intervening domain, and more uniformity of resolution across the region (Figure 4.14B). In the N-terminus, I observed better resolution of the dimerization interface, but poor coverage of the N-terminal TPRs (Figure 4.15B). Overall, this seemed like a good strategy moving forward, however I was still unable to model OGT's N-terminus and the bound peptide using local refinement around the N-terminal TPRs.

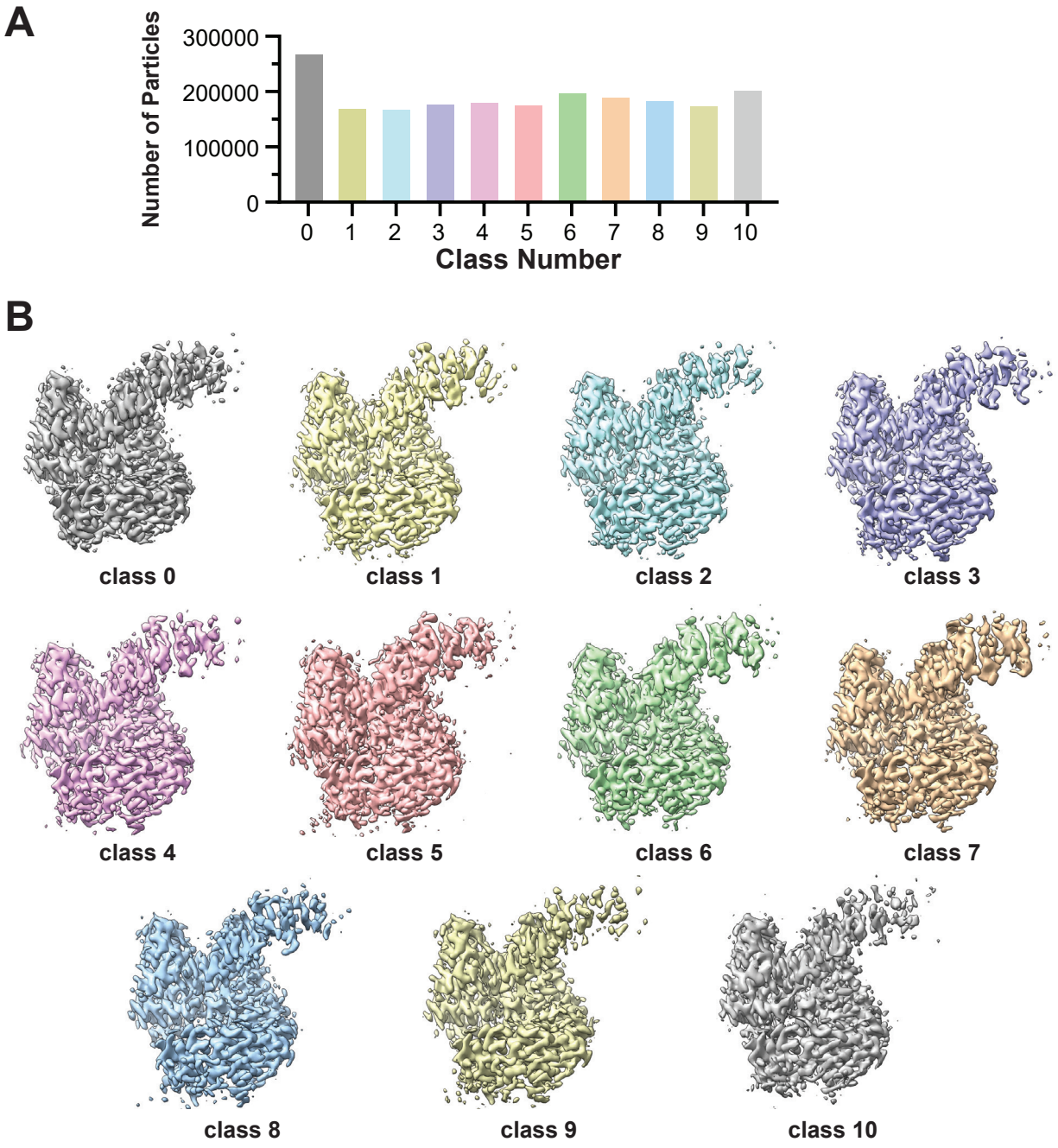


Figure 4.13: 3D classification using 3DFlex frames did not improve N-terminal density. (A) Class distribution from 11-class hard 3D classification. (B) Heterogeneous reconstruction of all 11 3D classes resulted in little observed improvement to the N-terminal TPR density. Figure produced using UCSF ChimeraX.[6–8]

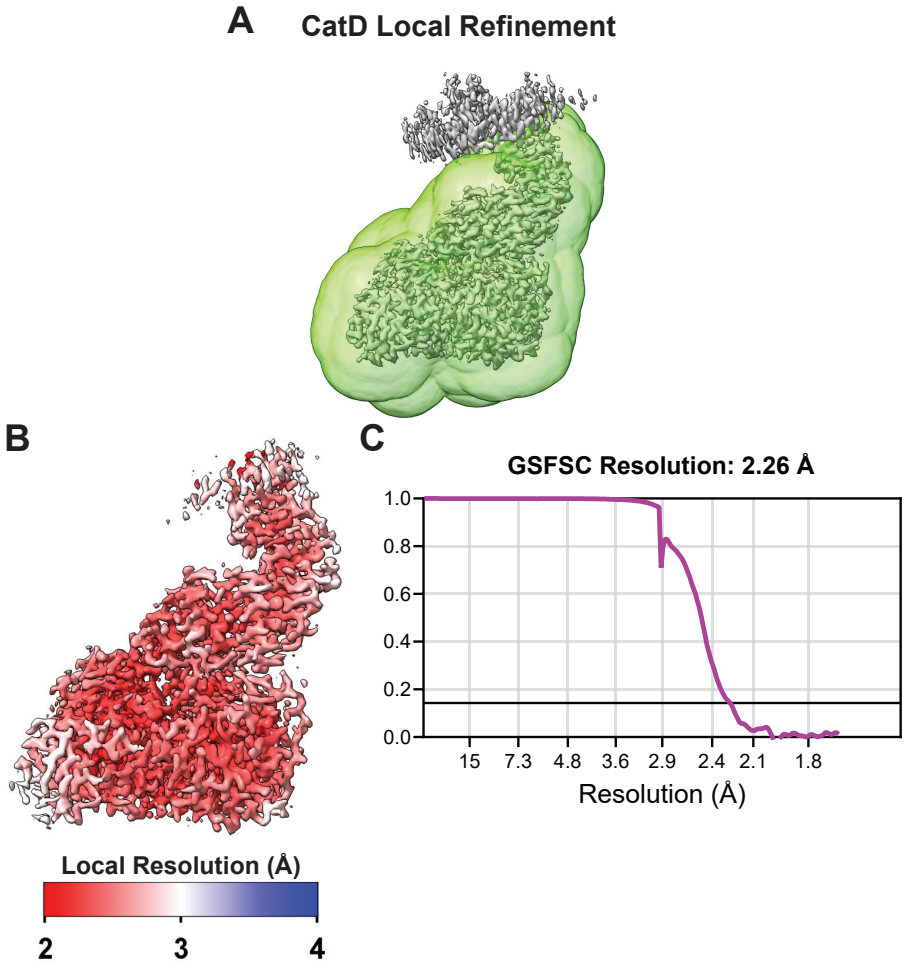


Figure 4.14: Local refinement of the catalytic domain improved density and resolution. (A) Mask surrounding the catalytic domain (transparent green). (B) Density map from local refinement using the catalytic domain mask, colored by local resolution. (C) GSFSC resolution plot of the local refinement. Figure produced using UCSF ChimeraX.[6–8]

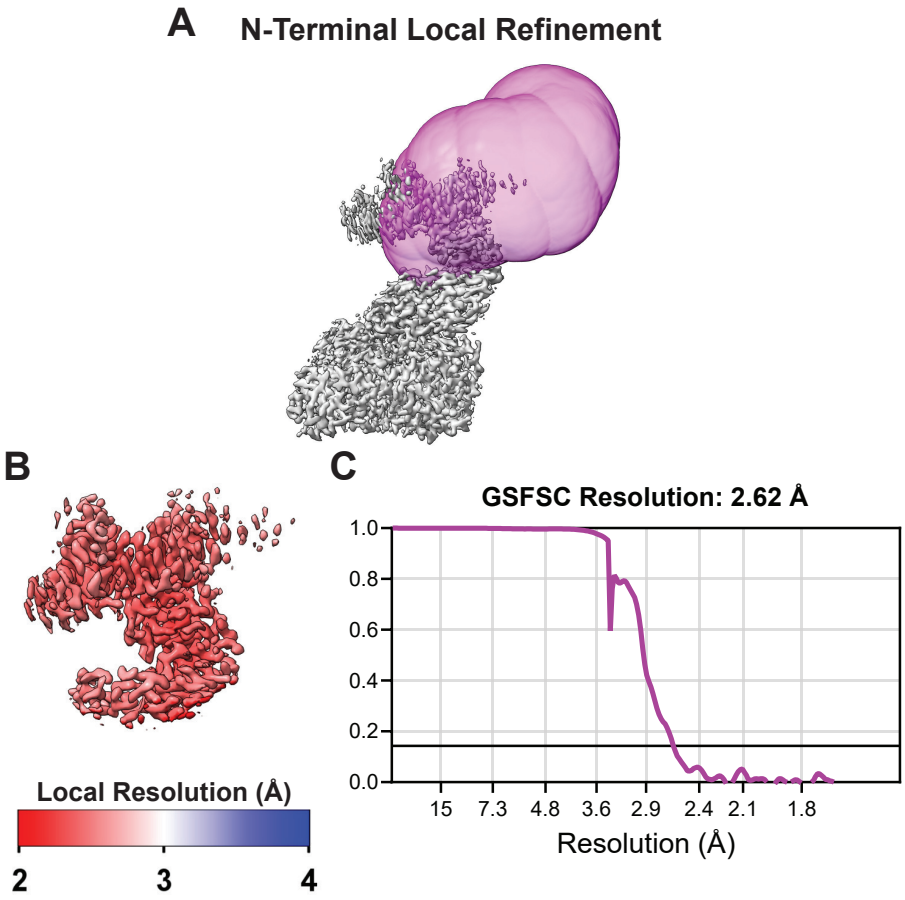


Figure 4.15: Local refinement of the N-terminal TPRs improved density and resolution. (A) Mask surrounding the N-terminal TPRs (transparent pink). (B) Density map from local refinement using the N-terminal TPR mask, colored by local resolution. (C) GSFSC resolution plot of the local refinement. Figure produced using UCSF ChimeraX.[6–8]

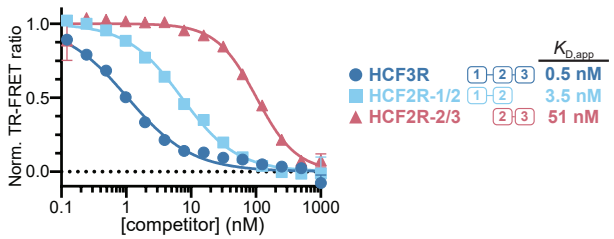


Figure 4.16: Competition binding of HCF2R polypeptides to OGT. Data shown are mean ± s.d. of three replicates and were fitted to an '[inhibitor] vs response – four parameters' binding curve in GraphPad Prism. Data were normalized to fractional binding.

4.9 Biochemical determination of the OGT-HCF3R binding interface

Faced with the problem of apparent continuous flexibility and hindered by lack of structural clarity in the medial and N-terminal TPRs, I decided to take a step back from playing around with CryoSPARC. In its place, I chose to use a combination of *in vitro* HCF3R cleavage assays and the TR-FRET binding assay described in Chapter 2. My hope was that I would be able to elucidate mechanisms by which HCF3R binds OGT, which could then guide refinement of specific features in CryoSPARC later.

4.9.1 The first two repeats in HCF3R control its picomolar-affinity interaction with OGT

HCF3R contains three nearly-identical proteolytic repeats. Therefore, the question of which repeats bind is of the utmost importance: does just one repeat bind to one OGT monomer, or is there a multiplicity of binding? Does one HCF3R bind to both monomers within an OGT homodimer? How far up the TPRs does HCF3R bind? To answer all of these questions, I first had to determine which components of HCF3R were required for binding.

To do this, I employed a competition binding assay using UDP, fluorescent-labeled HCF1R at its K_D, and purified derivatives of HCF3R lacking any polyhistidine tags. I purified two "HCF2R" substrates – one containing the first two repeats ("HCF2R-1/2"), and the other containing the last two repeats ("HCF2R-2/3"), both with alanine mutations of their E10 glycosites. I ran the binding assay with these competitor peptides and found that, while HCF2R-1/2 maintained a low-nanomolar affinity of 3.5 nM, HCF2R-2/3's K_D was 51 nM (Figure 4.16). Thus, the minimal

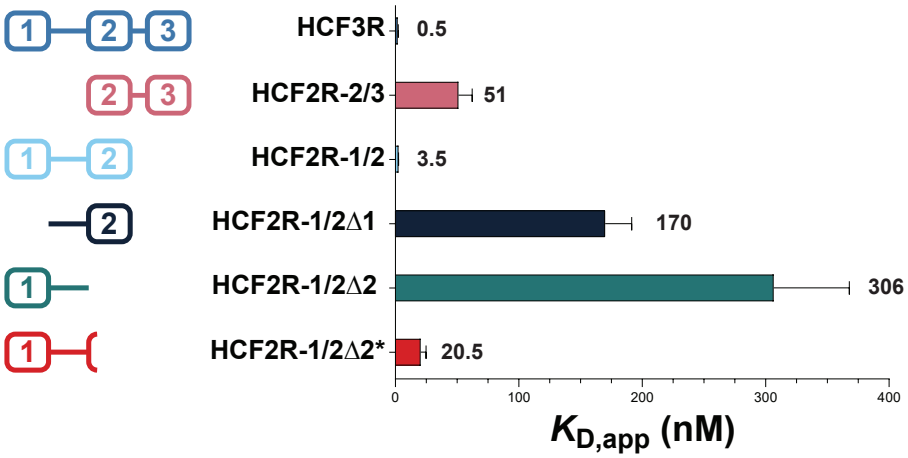


Figure 4.17: Affinity of HCF3R truncation polypeptides for OGT. Data shown are extracted $K_{D,app}$ values in nM from triplicate data fitted to an '[inhibitor] vs response – four parameters' binding curve in GraphPad Prism. Error bars represent the 95% confidence interval.

tight-binding construct is likely contained within the first two repeats, and not within the third repeat.

To further elaborate on this, I constructed a series of truncations of HCF2R-1/2 containing one full repeat and some of the spanning region connecting repeats 1 and 2. I assessed the binding of these mutants by competition assay as done above, and found that removal of either the full 26-mer repeat 1 (HCF2R-1/2Δ1) or 2 (HCF2R-1/2Δ2) leads to loss of low-nanomolar affinity binding; however, addition of the span plus the first 9 residues of the second repeat ('VRVCSNPPC'; HCF2R-1/2Δ2*) recovers a $K_{D,app}$ in the low nanomolar range (Figure 4.17).

4.9.2 OGT's oligomerization state does not significantly affect HCF3R binding

In Chapter 2.2, I determined that OGT dimerization occurs at a dissociation constant of 39 nM. Therefore, under the conditions in all of the binding assays described in this thesis ([OGT] $\leq$ 1 nM), OGT exists primarily in a monomeric state, unless some other factor (*e.g.*, HCF3R) is shifting the equilibrium towards dimerization. To answer the question of whether HCF3R binding influences OGT dimerization, or whether OGT dimerization could influence HCF3R binding, I took a three-pronged approach.

First, I tested OGT-mono's ability to cleave HCF3R-EEE as a proxy for its ability to bind it. After three hours, OGT-mono and OGT-WT cleaved the HCF3R substrate to a similar degree.

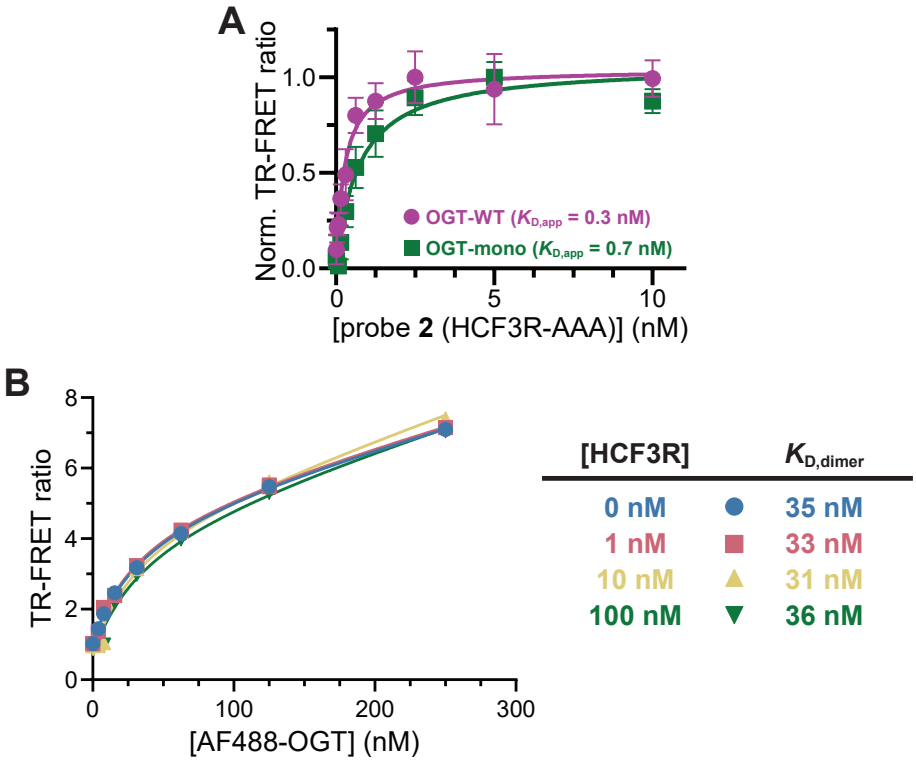


Figure 4.18: OGT dimerization and HCF3R binding are independent of one another. (A) Saturation binding of probe **2** to 0.5 nM OGT-WT and OGT-mono. Data are mean ± s.d. of corrected triplicates. (B) Saturation binding of AF488-OGT-WT to CoraFluor-1-OGT-WT in the presence of 0.1 mM UDP and varying amounts of HCF3R. Data are mean ± s.d. of duplicates.

I also tested OGT-mono's ability to bind HCF3R in a saturation binding assay and found that HCF3R still maintains a sub-nanomolar apparent $K_{D,app}$ in the presence of UDP (0.7 nM, Figure 4.18A). From this data, I reasoned that OGT's ability to oligomerize did not play a role in HCF3R binding; however, I also wanted to prove the converse, that HCF3R binding did not influence OGT's oligomeric state. To do this, I conducted an abbreviated saturation binding curve of CoraFluor-1-labeled OGT-WT to FITC-labeled OGT-WT from 0 to 100 nM, encapsulating its 39 nM $K_{D,dimer}$ four times, each time adding saturating UDP and either 0, 1, 10, or 100 nM HCF3R (Figure 4.18B). Because the binding curve did not shift upon addition of HCF3R and UDP, I knew that HCF3R did not promote or hinder OGT homodimerization.

4.9.3 AlphaFold2 predicts interactions between OGT and HCF3R in the N-terminal TPRs

The knowledge that OGT-OGT and OGT-HCF3R interactions do not significantly influence one another is crucial for understanding how OGT and HCF3R interact. If there was some multiplicity of binding (*e.g.*, one HCF3R engaging an OGT homodimer), this would likely exert some shift of the OGT-OGT binding equilibrium, which would have been observed in the experiments described in section 4.9.2. Because this was not true, however, we can reasonably assume that each OGT monomer binds one HCF3R, and the complex is symmetric about the OGT homodimerization domain.

To better understand how OGT binds HCF3R, I used AlphaFold2 (AF2) to predict binding of monomeric OGT to HCF3R.[206,207] AF2 predicted generally that the second repeat binds OGT's active site, while the third repeat binds in the N-terminal TPRs and the first repeat floats around the catalytic domain (Figure 4.19). However, it did so with generally poor pLDDT values for all regions outside of the second repeat, indicating poor local confidence.[206]

I also performed AF2 prediction of the interaction between an OGT monomer and HCF2R-1/2 (Figure 4.20), since I knew that construct was still capable of low-nanomolar affinity interaction with OGT. This resulted in placement of repeat 2 in the active site as before,[46,54] however repeat 1 was predicted to interact in the N-terminal TPR lumen. While some predictions (ranks 0 and 3) had repeat 1 oriented in the same direction as repeat 2, others (ranks 1 and 2) predicted it to be

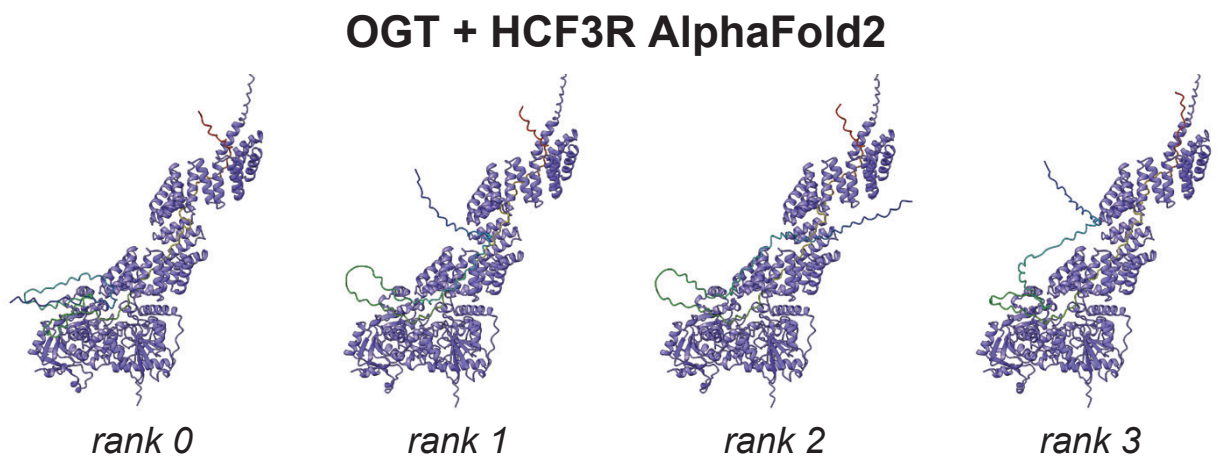


Figure 4.19: AlphaFold2 prediction of OGT-HCF3R interactions. Top four ranked models were shown (out of 25 total). HCF3R peptide was colored by rainbow from N- to C-termini. Figure produced using UCSF ChimeraX.[6–8]

oriented *anti* to the second repeat, with a possible interface in the medial TPRs.

4.9.4 N-terminal truncations of OGT can still bind and cleave HCF3R

Dr. John Janetzko, a former graduate student in the Walker lab, created a series of N-terminal truncations of OGT (Figure 4.21).[62,108,124] Using these, he showed that removal of the five most N-terminal TPRs resulted in an OGT (OGT-8) that could still cleave HCF3R-EAA (one cleavage site in the first repeat), while further truncation resulted in negligible cleavage activity.[124] This was reinforced by data from dTAG-MEFs expressing the same OGT truncations: OGT-8 retained robust HCF-1 cleavage activity, but shorter truncations resulted in accumulation of uncleaved HCF-1 (see Chapter 2.6.1, Figure 2.12B).[108]

I first sought to independently validate and extend this finding using HCF3R-EEE as an *in vitro* substrate. Some initial work on the TR-FRET binding assay (not shown) demonstrated that, while HCF3R-EEE bound substantially worse than HCF3R-AAA, HCF3R-EAA maintained picomolar affinity for OGT. This suggested that repeat 2 is likely bound in the active site under equilibrium conditions. Therefore, HCF3R-EAA may not be an optimal substrate to model the activity of cleavage-deficient OGT variants, as the most stably-bound glycosite (in repeat 2) cannot be glycosylated as efficiently. Using HCF3R-EEE, I found that the truncations OGT-8, OGT-7*, and OGT-7 had wildtype-like activity at 1 and 3 hours of reaction, while OGT-6 had minimal activity

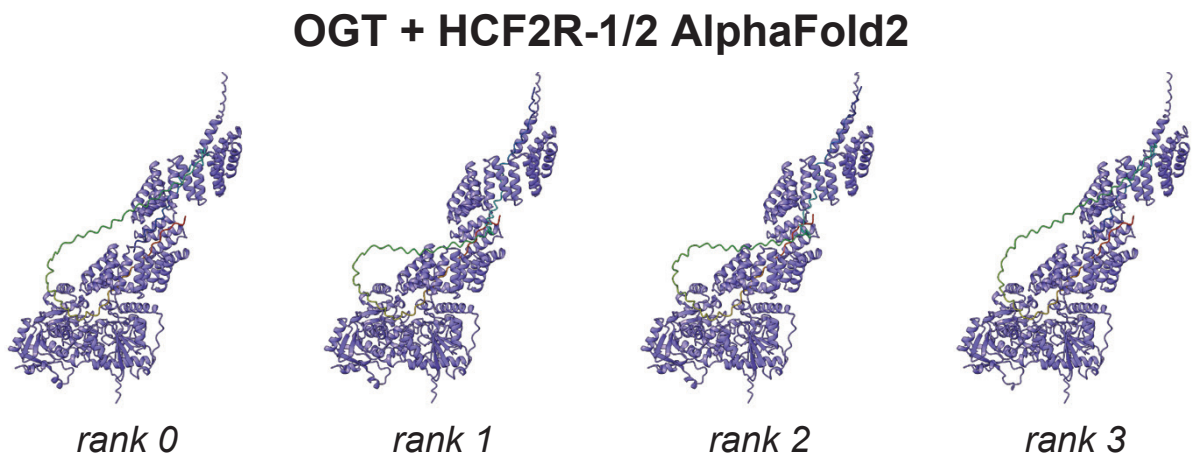


Figure 4.20: AlphaFold2 prediction of OGT-HCF2R-1/2 interactions. Top four ranked models were shown (out of 25 total). HCF2R peptide was colored by rainbow from N- to C-termini. Figure produced using UCSF ChimeraX.[6–8]

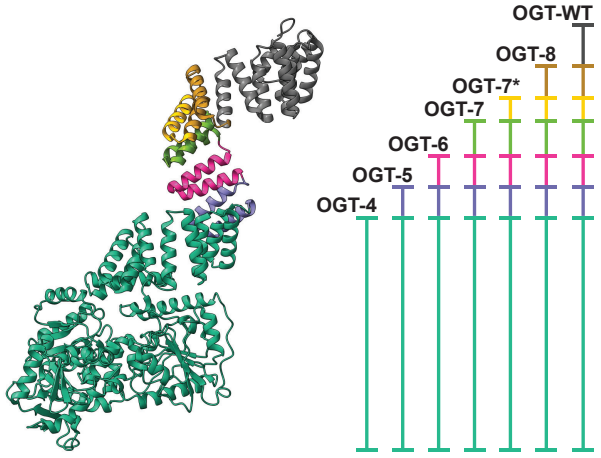


Figure 4.21: Schematic for OGT TPR truncation variants. Figure was produced using UCSF ChimeraX.[6–8]

at both timepoints, and both OGT-5 and OGT-4 had no detectable activity (Figure 4.22A).

In the case of the TPR asparagine mutants, cleavage and binding of HCF3R were directly correlated (*i.e.*, more cleavage == tighter binding). Thus, I wanted to test whether the TPR truncations with HCF3R cleavage activity were still capable of binding HCF3R tightly. Using the same binding assay described in Chapter 3.7, I measured the affinity of probe **2** (fluorescently-labeled HCF3R-AAA) for the TPR truncation variant OGTs in the presence of saturating UDP (Figure 4.22B). Under the assay conditions (1 nM OGT), the cleavage-active truncation mutants OGT-8, OGT-7*, and OGT-7 maintained sub-nanomolar $K_{D,app}$ values consistent with tight binding, while the other OGTs were incapable of detectably binding HCF3R. Notably, the B_{max} values observed were different across these TPR truncations. However, the tightly-binding OGT variants still fit well to a hyperbolic curve. I speculate that the shifting B_{max} values are due to differences in placement of the CoraFluor species bound via an anti-polyhistidine antibody to the OGTs' N-termini. The findings from the binding assay were in line with the observed cleavage activity of the truncation variant OGTs, although given the observation from CryoSPARC refinement that there was some N-terminal peptide binding visible in the TPR lumen, I was still somewhat surprised with these results.

To validate that the N-terminal TPRs do not play a key role in controlling binding of HCF3R, I made a series of alanine point mutations to potentially key residues in the N-terminal TPR lumen, including residues in the TPR asparagine and aspartate ladders (N84A, D114A, N118A, N220A), cysteines proximal to the region of observed peptide density (C148A, C189A), and other residues observed to interact with the substrate protein OGA in an existing crystal structure (H19A, E80A). I assessed the binding affinity of these OGT mutants by competition binding using probe **1** as the tracer at its K_D and unlabeled HCF3R-AAA as the competitor ligand (Figure 4.23A), and found that all of these mutants retained sub-nanomolar $K_{D,app}$ values, providing another point of confirmation that the N-terminal TPRs are not crucial for HCF3R binding. I also assessed the ability of a subset of these mutants to cleave HCF3R-EEE, and found no deviation from wildtype activity across the mutants used (Figure 4.23B).

These results suggested that any observed interactions between OGT and HCF3R in OGT's N-terminus are superfluous to binding, and that the key interactions span TPRs 7-13, as removal of TPR 7 (OGT-7 to OGT-6) leads to loss of HCF3R cleavage and binding. Additionally, these

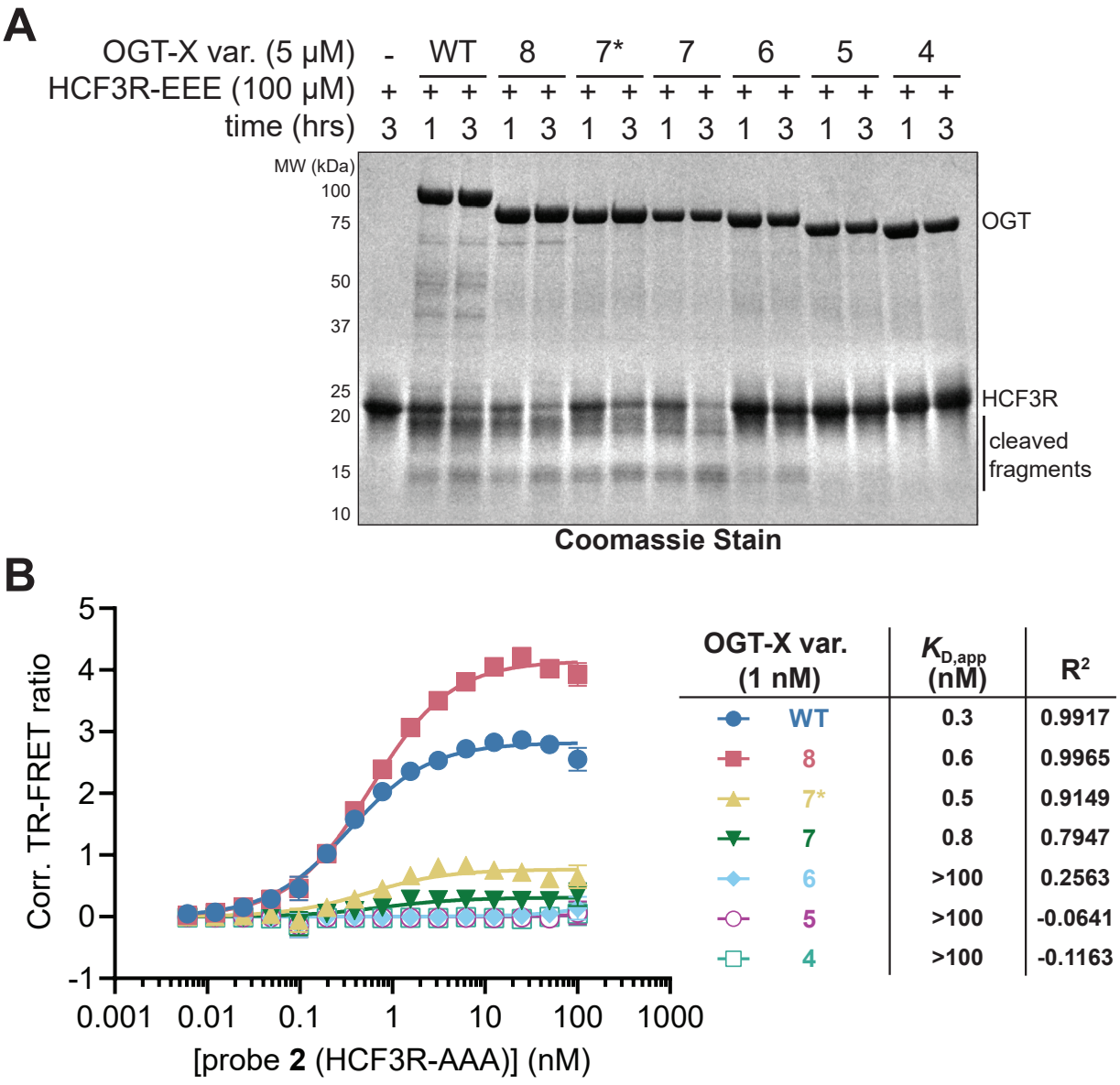


Figure 4.22: Binding and cleavage of HCF3R by OGT TPR truncation variants. (A) Coomassie stained SDS-PAGE readout from HCF3R-EEE cleavage by OGT truncations after 3 hours of reaction. (B) Saturation binding of probe **2** to OGT truncations in the presence of UDP. Data are mean ± s.d. of corrected triplicates, fitted to a 'one site – specific binding' model in GraphPad Prism.

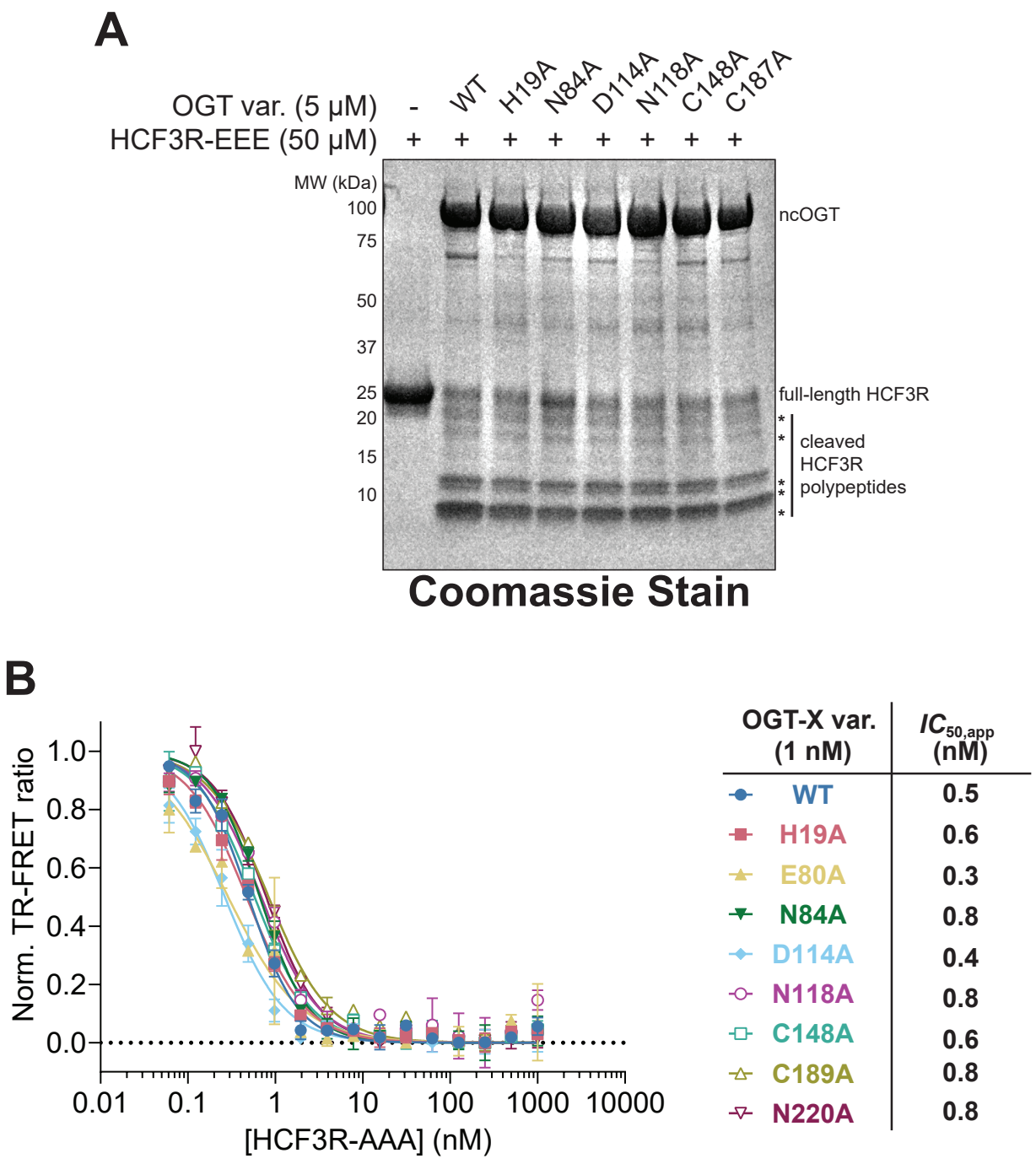


Figure 4.23: Binding and cleavage of HCF3R by OGT N-terminal point mutants. (A) Coomassie stained SDS-PAGE readout from HCF3R-EEE cleavage by OGT N-terminal point mutants after 3 hours of reaction. (B) Competition binding of HCF3R to OGT N-terminal point mutants in the presence of UDP. Data are mean ± s.d. of triplicates, fitted to a '[inhibitor] vs response – four parameters' model in GraphPad Prism. $IC_{50,app}$ values are shown in nM.

refute the AlphaFold2 predictions. Thus, moving forward I will focus on refinement of the medial TPR region, as it seems to be significantly contributing to HCF3R binding. It is possible that the HCF3R peptide resides in the lumen of the N-terminal TPRs, however, the data shown here indicate that those TPRs do not facilitate interactions that promote its binding or cleavage.

4.9.5 The conserved cysteines in HCF3R contribute collectively toward binding OGT

The HCF-1 proteolytic repeat contains a highly-conserved di-cysteine 'CX_4C' motif immediately preceding its glycosite (see Chapter 3.1). Cysteine residues with high degrees of conservation typically play important roles in interfacial recognition and often form disulfide bonds,[150] yet no role has been established for the conserved HCF-1 repeat cysteines. While there is no ladder of conserved cysteine residues in the TPRs, there are multiple lumen-facing cysteine residues (C158, C189, C257, C287, C313; Figure 4.24A) that could, theoretically, interface with the HCF3R cysteines, should they be proximal.

I chose to construct alanine point mutants of each of the seven cysteines in HCF3R: two cysteines from each repeat's 'CX_4C' motif, plus an additional cysteine in the spanning region between repeats 1 and 2. I also made an HCF3R-Δcys construct containing alanine mutations to all of the cysteines. I compared the relative binding affinity of all of these cysteine mutant HCF3Rs to that of wildtype HCF3R-AAA, and found that while all of the single cysteine mutants bound with low-nanomolar affinity, HCF3R-Δcys bound with an IC_{50} of 143 nM (Figure 4.24B). Thus, while each cysteine contributes negligibly to the strength of the OGT-HCF3R interaction, the cysteines as a whole clearly play some role, the exact nature of which remains unclear.

4.10 Re-analysis of cryo-EM data to extract meaning in light of biochemical findings

The findings presented above highlight the need to examine a core set of features within the OGT-HCF3R complex. First, a high-resolution density map of the monomeric catalytic domain must be solved in order to properly position the active site-engaging HCF-1 repeat. Second, the medial TPR domain must be resolved in order to identify structural features that are central to HCF3R binding and cleavage.

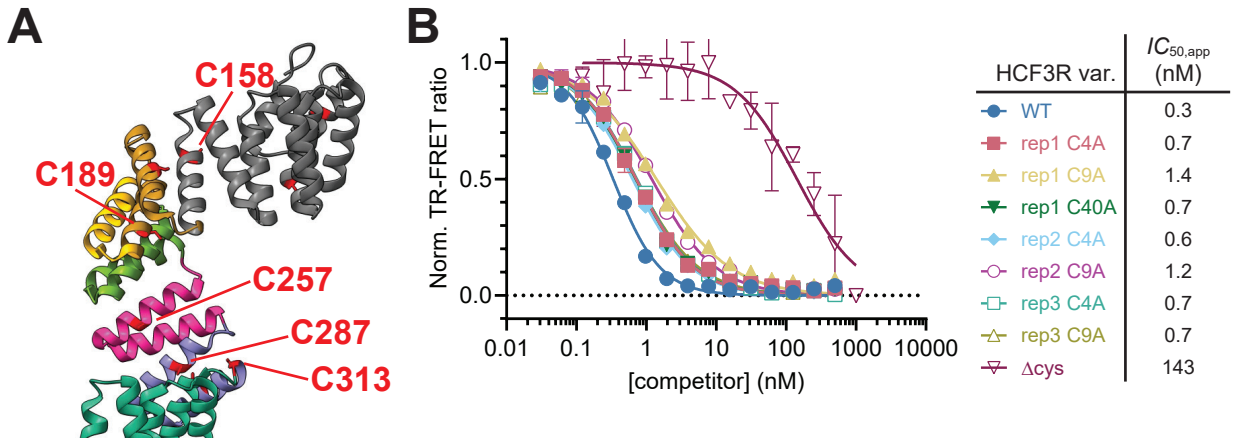


Figure 4.24: HCF3R cysteines contribute collectively towards binding. (A) OGT's N-terminal TPRs, with potential key cysteine residues highlighted. Figure produced in UCSF ChimeraX.[6–8] (B) Competition binding of HCF3R cysteine mutants to OGT-WT in the presence of UDP. Data are mean ± s.d. of triplicates, normalized to fractional binding and fitted to an '[inhibitor] vs response – four parameters' model in GraphPad Prism. $IC_{50,app}$ values are shown.

4.10.1 Local refinement of the medial TPRs reveals additional density

I first moved to work within CryoSPARC, using local refinement to improve the density of the medial TPRs and possibly uncover the density corresponding to bound peptide. I repeated steps done previously to generate a padded mask encompassing the region and performed multiple iterations of local refinement to optimize parameters. Ultimately, I was met with slight improvement of density in the region, especially at lower resolution contours (Figure 4.25A), while some density remained at higher resolution contours (Figure 4.25B). Unfortunately, I was still unable to model the bound peptide in this region's lumen due to poor density corresponding to sidechains.

4.10.2 Multibody refinement in RELION improved density of marginal regions

With my patience with CryoSPARC wearing thin, I attempted a long shot – porting my work into RELION and working with custom tools developed in that software. This was done by usage of the python3 package csparc2star,[208] which enabled direct export of a particle stack corresponding to a non-uniform 3D refinement into a format that RELION could read. Once in RELION, I was very interested in using two techniques: (1) focused classification with signal subtraction (FCwSS), a powerful method for classification of variable regions within a focused area of the structure;[193]

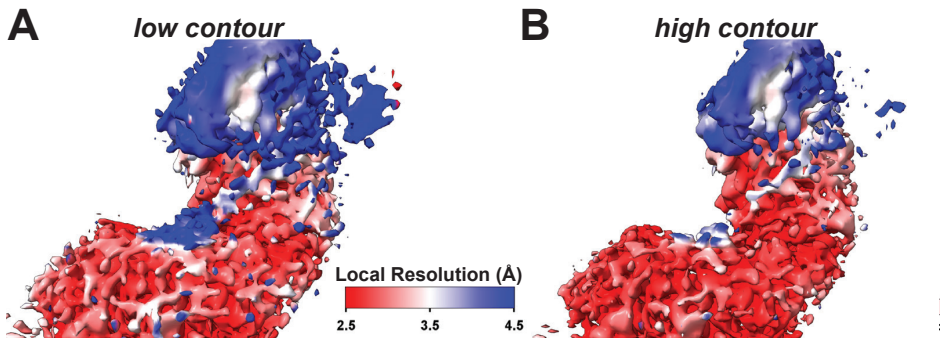


Figure 4.25: Local refinement improved density in the medial TPRs. Density maps from local refinement using a medial TPR mask shown at (A) low-resolution contours and (B) high-resolution contours indicate presence of potential peptide density. Figure produced in UCSF ChimeraX.[6–8]

and (2) multibody refinement, a newer methodology for separately and simultaneously resolving motion of different sub-areas of a complex.[209,210] Before I did any of this, I first ran a 3D refinement of my particle stack in RELION, confirming that I was able to work with my particles in the software (Figure 4.26A). This resulted in a density map resembling initial non-uniform refinements performed in CryoSPARC (*e.g.*, Figure 4.11): a good catalytic domain and very poor coverage of other features in the complex.

I then performed multibody refinement in RELION using four bodies – one encompassing each catalytic domain, and one encompassing each monomer's medial/N-terminal TPRs. Superimposition of all four bodies after refinement yielded an impressive improvement in density for the N-terminus and for the second monomer (Figure 4.26B), suggesting that it may be a useful tool moving forward.

I was also interested in performing FCwSS using both the input 3D refinement and the multibody refinement bodies, focusing on the region of interest that may be binding the peptide. However, my attempts in this were fruitless, and I instead ended up progressively winnowing down my particle stack without improving visible density. Future work may see a return to this strategy; however, I must improve my parameter tuning skills in order to make this possible.

4.11 Conclusion

This chapter presented a series of experiments aimed at understanding the physical nature of the OGT-HCF3R interaction, and what could be driving the picomolar affinity of an unstructured

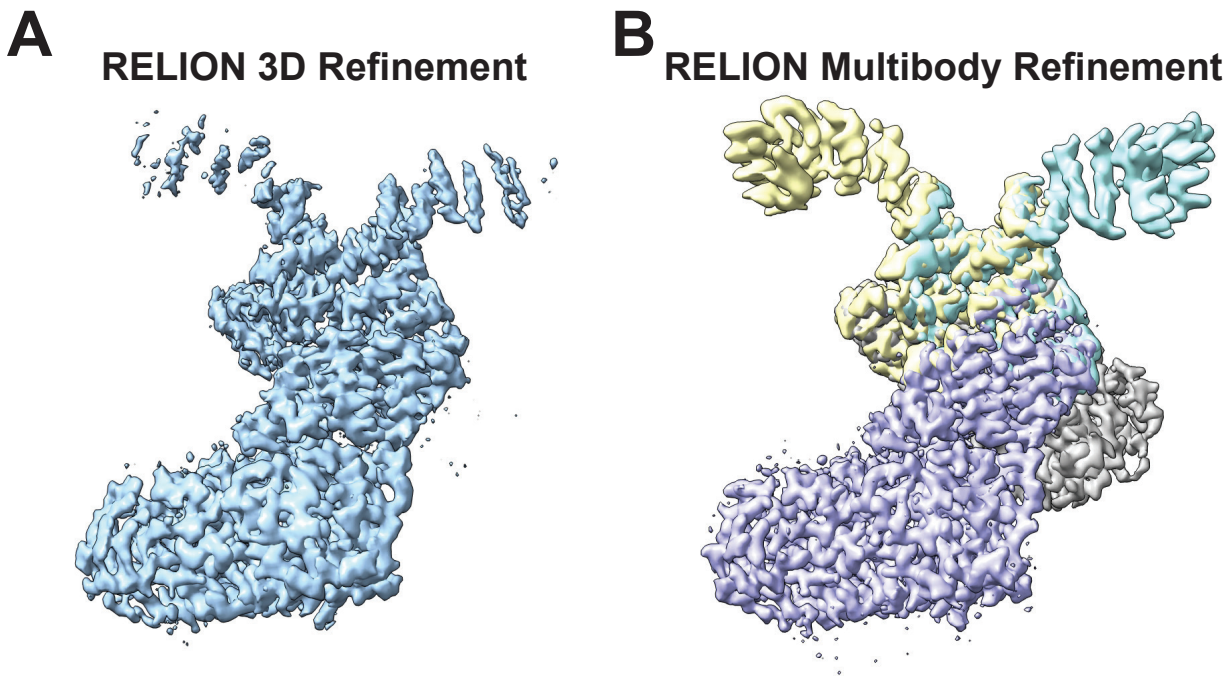


Figure 4.26: Multibody refinement improved density at distal regions of the complex. (A) 3D refinement of an input particle stack in RELION. (B) Result of 4-body refinement performed from the initial 3D refinement in RELION; all four bodies are colored separately. All density maps in both panels are viewed at the same contour level. Figure produced in UCSF ChimeraX.[6–8]

polypeptide for a flexible repeat protein. Using the binding assay characterized in Chapter 3 of this dissertation and single particle cryo-EM, performed with the assistance of Dr. Morgan Gilman and the Harvard Cryo-Electron Microscopy Center for Structural Biology, I identified regions of interest on both OGT and HCF3R that contribute to the interaction.

It is surprising that HCF3R's affinity for OGT is not governed at all by interactions in the N-terminal TPRs. The only published structure of a substrate bound in that region makes a series of contacts to lumenal residues,[57] and given the affinity we observed, we assumed HCF3R would exploit the same or similar contacts. However, this does explain data from Chapter 2 that showed OGT-8 and OGT-7 are still capable of binding HCF-1 in cells; the fact that those OGTs can still bind HCF3R tightly and catalyze its cleavage suggests those truncations are binding HCF-1 in a cleavage-oriented manner.

The high degree of conservation of HCF-1's proteolytic processing domain (PPD) is also interesting in light of the results demonstrated in this chapter. Namely, the dispensability of the third repeat for high-affinity interaction calls into question why that repeat is so highly conserved

across mammalian HCF-1 homologs. Future work will have to look at the cellular roles of the repeats, more specifically at the role of each fragment of the PPD post-cleavage.

My continuous struggles in resolving a high-resolution structure of the OGT-HCF3R complex are also not without note. Previous structural publications of OGT by cryo-EM did not note significant conformational heterogeneity, at least not to the degree I have noticed in my own work.[57,59] Fortunately, this sort of problem is one that is being actively pursued in the field of structural biology; future work will use the data presented here in conjunction with new tools and methods to achieve a high-resolution structure that adequately describes the contacts between OGT and HCF3R that drive their picomolar-affinity interaction.

4.12 Methods

General Vector Construction and Cloning

See Appendix A for information on the plasmids used in this chapter, as well as elsewhere in this dissertation. OGT heterologous expression plasmids were originally developed by Dr. John Janetzko. His_8-HCF3R expression plasmids were initially developed by Dr. John Janetzko, then later modified to include a C-terminal tri-tyrosine (YYY) tag by Dr. Cassie Joiner.

See Chapter 2 Methods for details on cloning protocols used throughout this chapter.

General staining methods

See Chapter 2 Methods for detailed procedures on SDS-PAGE electrophoresis and InstantBlue staining.

Expression and purification of polyhistidine-tagged OGT variants

See Chapter 2 Methods for detailed methodology.

Expression and purification of StrepTag-HaloTagged HCF1R and HCF3R variants

See Chapter 3 Methods for detailed methodology

Expression and purification of tag-free HCF3R variants (including HCF2R vars.)

See Chapter 2 and Chapter 3 Methods for detailed methodology on purification of polyhistidine-tagged (Chapter 2) and untagged (Chapter 3) HCF3R polypeptides.

Purification of the OGT-HCF3R-UDP complex for cryo-EM

OGT-WT and HCF3R-SAA were purified separately with N-terminal polyhistidine affinity tags, as described in Chapter 2 Methods, into HBS (20 mM HEPES, 150 mM NaCl, pH 7.4) supplemented with 20 mM $MgCl_2$, 1 mM DTT, and 0.5% (v/v) glycerol.

Purified OGT (1 mL, 9.2 μM) and purified HCF3R-SAA (1 mL, 127.5 μM) were combined and supplemented with 1 mM UDP. The mixture was incubated overnight at 4° with gentle rotation, then concentrated to <0.5 mL in a 3K MWCO Amicon concentrator. This concentrate was centrifuged at max speed for 10 minutes at 4° to pellet aggregated protein, then purified by size exclusion chromatography on an AKTA Pure 25 FPLC fitted with a Superdex 200 Increase 10/300 GL column into the same buffer, without UDP. Peak fractions were collected and assessed by SDS-PAGE and InstantBlue staining. Fractions containing OGT and HCF3R were combined, concentrated to 0.5 mg/mL (assuming 1:1 binding), and saved on ice for immediate vitrification.

Sample vitrification and grid preparation

3 μL of SEC-purified OGT-HCF3R-UDP at 0.5 mg/mL in 20 mM HEPES, pH 7.4, 150 mM NaCl, 20 mM $MgCl_2$, 1 mM DTT, and 0.5% (v/v) glycerol was applied to glow-discharged UltrAUfoil 300 mesh 1.2/1.3 grids, at 4° with 100% humidity. Grids were blotted for 7 seconds at a blot force of 5, after a 10 second wait time, using a Vitrobot Mark IV, then directly plunged into liquid nitrogen-cooled liquid ethane. Grids were clipped and loaded onto the microscope by core facility scientists at the Harvard Cryo-Electron Microscopy Center for Structural Biology.

Cryo-EM data collection

Cryo-EM data were collected on a 300 kV Titan Krios G3i microscope with a Gatan K3 electron detector and a GIF quantum filter in counted mode at the Harvard Cryo-Electron Microscopy Center for Structural Biology. A single video was collected in the center of nine holes per position using SerialEM version 4.1b.[211] Full parameters for data collection are found in Table 4.1.

Cryo-EM data processing in CryoSPARC

Pre-processing:

CryoSPARC data processing was conducted using CryoSPARC version 4, implemented via SBGrid Consortium.[182,212] 15,617 raw videos were patch motion corrected, dose-weighted, and patch CTF estimated using default parameters.

Initial processing:

Initially, particles were picked using the blob-picker and extracted with 2x2 binning. From these particles, a subset corresponding to OGT-containing complexes were identified through 2D classification and were used to generate *ab initio* 3D reconstructions of three volumes. All picked particles were heterogeneously refined into those 3D volumes, and particles corresponding to the best heterogeneous refinement class were separately refined by non-uniform refinement.

Re-processing using TOPAZ:

I later implemented TOPAZ to pick particles, and conducted a full data processing scheme, summarized here. 100 micrographs were selected at random for blob-picking, unbinned extraction (440 pixel box size), and 2D classification. The 2D classes resembling OGT were used to train a TOPAZ neural network to pick particles with rare poses.[199] The trained network was used to pick particles from all micrographs, which were combined and extracted with 4x4 binning (440 pixels to 110 pixels). Good particles were identified through 2D classification and pushed through 3D reconstruction and two rounds of heterogeneous refinement. Particles corresponding to the best class in the second round of heterogeneous refinement were re-extracted without binning and refined by non-uniform refinement without masking (dynamic mask start resolution of 1Å) using CTF parameter optimization per-group (tilt and anisotropic magnification). A mask was created using the output refinement map (1px dilation, 15px padding), and particles were corrected by reference-based motion correction against the initial patch CTF estimated micrographs using CryoSPARC's RBMC implementation. The RBMC-corrected particles were re-refined using the same non-uniform refinement parameters.

Local refinement:

The RBMC refinement density map was downloaded, blurred by Gaussian filtering using a standard deviation of 2, and segmented using the Segger implementation in UCSF ChimeraX.[213]

Segments corresponding to the catalytic domain and the N-terminal TPRs were grouped, re-sampled onto the original density map, and imported into CryoSPARC. Both volumes were converted into masks at thresholds determined by inspection of the maps in ChimeraX, padded significantly (CatD: 10px dilation/15px padding, NTerm: 18px dilation/15px padding), and used to locally refine the input map with custom parameters set (use pose/shift gaussian prior; recenter rotations/shifts each iteration; 5Å initial lowpass resolution).

3DFlex analysis:

3DFlex was performed using the implementation in CryoSPARC v4.

100,000 particles from the local refinement of OGT's catalytic domain were cropped to 256 pixels, then downsampled 2-fold to 128 pixels and prepared for 3DFlex. The initial RBMC-corrected non-uniform refinement map was lowpass filtered to 6Å, cropped to 256 pixels, and used to create a mask (8pix dilation, 3pix padding). This mask was segmented into four components (CatD-1, NTerm-1, NTerm-2, CatD-2) using the Segger implementation in UCSF ChimeraX,[213] and the segmented masks were dilated, padded, and used to create a tetrahedral mesh. Flex training was performed using the mesh and the prepared particles with 1 latent dimension, 0.0001λ rigidity, and 0.2 latent centering strength. A 41-frame volume series describing the flexible motion across latent space was generated using the training output and the 256 pixel-cropped input refinement.

Evenly-spaced volumes from the series and a focus mask encompassing the N-terminal TPR domain were used as inputs for hard 3D classification of all particles from the input refinement with custom parameters (0.2 initial O-EM learning rate, 2000 per-class O-EM batch size, RMS density change convergence check, 4Å target resolution). The output classes were reconstructed heterogeneously.

Model building:

While not directly discussed in this dissertation, model building was performed in WinCOOT v.0.9.8.7 using chains A and B from PDB deposition 4N3B as a basis for OGT and the HCF3R peptide, respectively.[46,214] Models were fitted into density maps using UCSF ChimeraX.

Cryo-EM data processing in RELION

Data conversion and import to RELION:

Aligned particles from non-uniform refinement following RBMC were converted into .star file format using the csparc2star python script in pyem.[208] The resultant particles.star file was imported into RELION v.5.

Initial modeling and generation of a high-resolution 3D structure:

The imported particles.star file was used to generate a 3D initial reference using standard parameters (T=4, 1 class, 440 pixel mask diameter), which was then refined by unsupervised 3D classification into one class (T=4, 1 class, 440 pixel mask diameter, 30Å initial lowpass filter, Blush regularization), and then refined by 3D refinement into a high-resolution density map (440 pixel mask diameter, 30Å initial lowpass filter, Blush regularization). A mask was created in RELION from the output map and used to sharpen the map with a B-factor of -50.

Multibody refinement:

The high-resolution refinement density map was downloaded and segmented into four components (1: CatD-1, 2: NTerm-1, 3: NTerm-2, 4: CatD-2) according to protocols described in above methods sections. The segments were imported back into RELION and converted to masks with generous padding. A .star file was created to detail relative rotation (1 rotates relative to 2, 2 rotates relative to 1, 3 rotates relative to 2, and 4 rotates relative to 3), standard deviations for angular searches (1: 10, 2: 20, 3: 30, 4: 30), and standard deviations for translational searches (1: 2, 2: 5, 3: 7, 4: 7). Multibody was then run using the high-resolution density map and the prepared .star file. Output density maps for each body were sharpened using their corresponding masks and B-factors of -50.

Competition binding of HCF-1 repeat polypeptides to OGT variants

See Chapter 3 Methods for detailed methodology.

Saturation binding of probe 2 to OGT variants

See Chapter 3 Methods for detailed methodology.

Determination of OGT dimer affinity in the presence of HCF3R and UDP

Aliquoted proteins were thawed on ice, then centrifuged for 10 minutes at 21130 rpm and

4°C to pellet aggregates. Protein concentrations were then re-calculated by A_{280} (and A_{340} or A_{495} for CoraFluor- or AF488-labeled OGT). Master mixes were prepared containing UDP (0.1 mM final), DTT (2.5 mM final), CoraFluor-1-tagged OGT (1 nM), and HCF3R (0, 1, 10, or 100 nM) in HBS FRET buffer (20 mM HEPES, pH 7.4, 20 mM $MgCl_2$, 150 mM NaCl, 0.5 mg/mL bovine serum albumin (BSA), 0.005% (v/v) Tween-20). AF488-tagged OGTs were diluted to the target concentration in HBS supplemented with 1 mM DTT and 0.3% (v/v) Tween-20, then dispensed in 1:2 dilution series across each column using a D300e digital dispenser (Tecan) fitted with a T8+ dispensehead (HP), with the final row having no AF488-OGT. Plates were covered in sealing tape (Perkin-Elmer), briefly centrifuged to eliminate bubbles (1500 g, RT, 1 minute), and incubated in the dark at room temperature for 4 hours. Plates were then moved to 4°C for overnight incubation, after which they were warmed to room temperature over a period of 4 hours prior to data collection.

TR-FRET measurements were acquired on a Perkin-Elmer EnVision plate reader using 340nm excitation with a 407nm/505nm advanced dual mirror and 520nm (25nm bandwidth, FITC) and 495nm (10nm bandwidth, Tb) filters, using a 50 μs delay and 300 μs integration. FRET ratio was taken as the ratio of 520nm/490nm emission on a per-well basis, then averaged across replicates. Data was fitted to a 'one site – total binding' model in GraphPad Prism 10. Due to usage of directly-labeled OGTs, data was not corrected or normalized to a model of fractional binding.

In vitro cleavage of HCF3R-EEE by OGT variants

See Chapter 2 Methods for detailed methodology.

Chapter 5

Summary and future directions

5.1 Summary of findings

While OGT's ability to bind and glycosylate myriad substrates has been well characterized in the literature, spanning decades of research,[35,90,152] the mechanisms by which it maintains interactions with those substrates have only been closely examined over the past 10 years. Additionally, the means by which a glycosyltransferase known for its promiscuity manages to bind a single substrate, HCF-1, with such high affinity that it can catalyze proteolysis, remained a mystery. Yes, the chemistry underlying the HCF-1 cleavage reaction had been described in full,[45,46,124] but the protein-protein interactions underpinning the whole process remained unknown. Moreover, the role these interactions play in OGT's essentiality had not been understood at all. In this dissertation, I offer my contributions towards understanding these mechanisms by which OGT not only engages its glycosylation substrates, but also binds and cleaves HCF-1, and maintains its pool of interactors. In doing so, I hope that this dissertation moves the field towards a better understanding of how OGT functions, and how those functions underpin its essential biological roles.

In Chapter 2, I examine a series of studies I performed assisting members of the Walker lab in characterizing the roles that various OGT features play on its cellular functions and essentiality. With Dr. Sarah Potter, I studied the role that the N-terminal TPRs play; we found that the N-terminal TPRs appear to be key drivers of *O*-GlcNAcylation and OGT-mediated scaffolding, and that removal of 6 or more TPRs is lethal to the cell.[108] With Dr. Potter and Dr. Victoria

Marando, I examined the role of OGT homodimerization, finding that while it is not essential for cell viability, its prevention significantly alters OGT's catalytic and noncatalytic activities.[108] And, with Dr. Cassandra Joiner, I conducted key experiments in assessing the role of the TPRs in glycosylation of three substrate proteins *in vitro*, showing that OGT's substrates can have unique and differential dependences on the TPR domain itself, and that global trends in glycosylation are not necessarily recapitulated on each and every substrate.[62]

In Chapter 3, I developed the first quantitative binding assay for OGT-substrate interactions based on a TR-FRET methodology in collaboration with Dr. N. Connor Payne in the Mazitschek lab, using a series of acceptor probes derived from small molecule inhibitors and HCF-1 repeat polypeptides.[109] Using this assay, I identified a polypeptide derived from three repeats of HCF-1's PPD, HCF3R, that binds OGT with picomolar affinity in the presence of the cofactor UDP. I demonstrated that this binding is dependent on cofactor occupancy in OGT's active site, and on key bidentate interactions between asparagine sidechains in OGT's TPR lumen and the HCF3R peptide backbone. Then, I showed that HCF3R functions as a potent inhibitor of OGT's catalytic activities in cell extracts and in MEFs under constitutive expression. Lastly, by tying HCF3R expression to an inducible promoter and fusing it to a degron protein, I developed a cellular inhibitor of OGT that can be switched on and off by addition of two small molecules, the first of its kind.

In Chapter 4, I continued my analysis of the OGT-HCF3R interaction using a combination of my binding assay and cryo-electron microscopy, which I conducted with the aid of Dr. Morgan Gilman from the Kruse lab and the Harvard Cryo-Electron Microscopy Center for Structural Biology. Using these tools, I identified that the medial region of the TPR domain is responsible for some key interactions between OGT and HCF3R that both drive its high-affinity binding and cleavage. In combination with data obtained from my work with Dr. Potter in Chapter 2, this may point to a role for the TPRs in driving HCF-1 cleavage.

The work presented in this dissertation provides valuable insights into the biophysical interactions at play between OGT and its substrates and further develops tools and methods that will have long-standing impacts on the field. Through this work, mechanisms underpinning the OGT-HCF3R interaction have not been fully elucidated, however, I will describe future work that may be able to discern these mechanisms below. Furthermore, the work in Chapter 2 largely

provided evidence for differences in OGT functions that may explain changes in cellular viability; below, I will lay out some future plans for digging into those altered functions.

5.2 Recommendations for technical improvements

Additional work to improve the inhibitory properties of HCF3R and move towards viability in animal models would benefit significantly from a technical overhaul of the delivery system. Currently, we are directly incorporating our inhibitor via lentiviral infection. For reasons that are both numerous and obvious, this would not be viable in a living organism other than as the first test of functionality. The most direct change we could make on this front would be to alter our delivery system, whether that be through usage of a cell-penetrating peptide,[180] fusion to some membrane-penetrating modality,[215,216] or usage of alternative delivery systems such as mRNA lipid nanoparticles to reach the target cells.[217] All of these technologies would separately benefit from reduction of the HCF3R polypeptide into the minimal inhibitory construct, an effort that we are actively seeking to achieve through work in the lab.

The other major point where technical improvements could yield clearer results is in the structural solution of the OGT-HCF3R complex, which has been thoroughly hampered by conformational heterogeneity. On the computational front, the simplest and most rigorously tested solution is also the most rigorous, itself: exhaustive usage of all available tools in both CryoSPARC and RELION, using empirically tested parameters, until the structure is fully resolved. However, this is likely not a 'smart' usage of time and various tools, both pre- and post-data collection, may be of use in solving this problem.

One strategy in the literature that has been employed to reduce protein flexibility in cryo-EM is to fuse the flexible regions other proteins to help them rigidify.[218,219] This has been used most often in structural studies of GPCRs by fusing proteins to the transmembrane helices, and has the side effect of adding size to the complex, also improving the ability of cryo-EM software to detect and refine the resultant complexes. One downside of this is that screening successful fusion requires many trips to the microscope, and one does not know if the fusion worked until data is fully worked up, which is a time-consuming process. However, if successful, it could fully resolve the flexibility problem that I have observed in the N-terminal TPR domain.

Otherwise, there are many computational tools that offer the solution to this problem, among them 3DFlex,[205] CryoDRGN,[220] and CryoSTAR.[221] All of these tools construct some sort of neural network, with which they analyze particle flexibility, motion, or heterogeneity otherwise. It may be that some combination of these tools provides structural answers, or that optimizing usage of existing non-machine learning tools in CryoSPARC and RELION could improve the complex resolution.

5.3 Discussion and future scientific directions

Over the course of my graduate studies, I developed a powerful toolkit for quantification of OGT-ligand and OGT-substrate interactions. I used this toolkit to identify that HCF3R, a polypeptide composed of three proteolytic repeats from HCF-1 and the contained spanning regions, binds OGT with picomolar affinity and functions as an inhibitor. However, this work only scratched the surface of what may be possible using this assay. While the OSMI series are incredibly potent *in vitro*, there is a need for new chemical matter in the space of OGT inhibitors due to their poor solubility and membrane permeability.[95] With a sensitive assay such as the one I developed using OSMI-4-FITC as a tracer, we could conduct new high-throughput screens with diverse chemical libraries to identify new moieties and fragments that could lead to new OGT inhibitors. Moreover, using peptide-based tracers that engage the TPR lumen, small molecules, peptides, or peptidomimetic compounds that bind the TPRs could be identified, which would represent a new means of engaging OGT. If such a compound were identified, one would be able to selectively inhibit glycosylation of just a subset of OGT's substrates, instead of all of them. Additionally, discovery of such inhibitors would enable further dissection of which of OGT's substrates make serine/threonine *O*-GlcNAcylation an essential function.

As mentioned in the previous section, another focus of future work will be modification of the HCF3R scaffold to improve its inhibitory properties. Furthermore, we are interested in whether a degron-tagged HCF3R can be used to degrade OGT in cells. The current iteration of the inhibitor, where the degron lies N-terminal to HCF3R, did not result in OGT degradation in HCT116 cells.[109] However, through minimization of the HCF3R construct (*i.e.*, removal of regions that are not necessary for inhibitory action) and shifting the location of the degron fusion,

we hope to be able to enable OGT co-degradation. This would allow for OGT inhibition and degradation in a wide variety of cell lines, without the need to directly modify OGT's genetic locus. Moreover, if alternative mechanisms that do not require genetic incorporation are used to deliver the HCF3R inhibitor, we would be met with a potent inhibitor and degrader of OGT that requires no genetic modifications whatsoever to the host cell, which would be a remarkably powerful tool to interrogate the biological functions of OGT and *O*-GlcNAc.

This work also highlights the need for complete, high-resolution information on the OGT-HCF3R interface. When I began my structural biology voyage a few years ago, I was hopeful that solving the structure of the OGT-HCF3R-UDP complex would be relatively simple – after all, HCF3R binds with picomolar affinity, so I reasoned some sort of TPR-rigidifying interactions were taking place. Sadly, as was made evident throughout Chapter 4 of this thesis, this was not the case. Nevertheless, I will continue working to characterize the interface through usage of structural and biochemical tools. Full structural solution of the interface would enable further sequence optimization of the peptide to drive additional OGT engagement, also a step that would improve its inhibitory properties.

With the information presented in this thesis, a hypothesis has emerged for how HCF3R engages OGT with picomolar affinity. One proteolytic repeat resides in the active site and the proximal TPRs, while a second repeat binds in the lumen of the medial TPRs; interactions between these two repeats form a locking hinge that anchors the first repeat in OGT's active site, enabling its cleavage should the glycosite residues be glutamates. This hypothesis explains the observed discrete densities corresponding to peptides in the proximal and medial TPRs, and also allows for the observed flexibility of the N-terminal TPRs, as the bound peptide does not necessarily rigidify the TPR domain itself, as it likely did in the case of the OGT-OGA structure solved in 2024.[57] One must also consider what the physiological role of this interaction may be. To do so, I offer the finding that HCF-1 glutamate glycosylation proceeds at a much slower rate than typical serine/threonine glycosylation; moreover, formation of the cyclized pyroglutamate intermediate also occurs on-enzyme.[45] Together, these steps would necessitate a longer residence time in the active site compared to serine/threonine glycosylation, thus a "locking" mechanism (that is released, much like a taut rope being cut in half, upon HCF-1 cleavage) is at least a somewhat plausible explanation for how that longer residence time is maintained.

Lastly, the work presented in Chapter 2 concerning OGT dimerization necessitates further work to uncover and characterize this mechanism in cells. While we worked with a "monomeric" version, we have no direct confirmation that OGT exists in its dimeric form in significant levels within the cell. Moreover, even if we suppose based on OGT's $K_{D,dimer}$ and estimates of its cellular concentration (39 and 100 nM, respectively) that a significant portion (>50%) of cellular OGT is dimerized, we do not know the nature of the distribution or function of dimeric or monomeric OGT pools. Our data on total protein levels and interactors of monomeric OGT suggest a specific role for it in immune response pathways and immune signaling, which while not entirely unheard of for OGT,[121,222] would still be worth characterizing. To study this, we first need a tool to visualize cellular pools of monomeric and dimeric OGT. I propose a simple FRET pair system, where donor- and acceptor-tagged OGTs are both expressed in a background where the endogenous OGT is tagged with a degron, facilitating its rapid degradation. By performing fluorescence microscopy and reading out direct emission from the donor and FRET emission from the acceptor after donor excitation, we could obtain cellular localization data for monomeric and dimeric OGT. This could also be repeated using known monomeric OGT variants to obtain negative control data, or under various stimuli to provoke immune signaling pathways. The data obtained could be cross-referenced with our proteomics data to help qualify some of the more observational points raised in Chapter 2.

The *O*-GlcNAc transferase field has made great strides since its inception in the 1980s. We are now closer to a full understanding of the mechanisms underpinning OGT's biological roles and its essentiality than we have ever been before. And yet, so many questions still remain. Further work is necessary to characterize specific mechanisms of substrate and interactor engagement that are crucial in unraveling OGT's essential biology. This thesis, and the work presented herein, improve our understanding of OGT's interactions with its substrates and interactors and demonstrates new tools that will impact studies in the future.

Appendix A

Protein constructs described in this dissertation

This dissertation contains a variety of protein constructs, used both for heterologous expression and purification and for cellular functional studies in mammalian cell lines. In this appendix, information concerning the amino acid sequences of these protein constructs can be found.

A.1 OGT protein constructs

A.1.1 OGT bacterial expression constructs

All of the following constructs were built from the pet24b-His_8-[HRV3C]-OGT-WT vector initially developed by Dr. John Janetzko and were used to express OGT variants in *E. coli*.[123] While not described in this thesis, OGT can also be expressed and purified using constructs containing C-terminal polyhistidine affinity tags. Molar extinction coefficients at 280 nm are given for quantification purposes.

Protein: His_8-[HRV3C]-OGT-WT
Initially cloned by: John Janetzko
Purpose: Expression of OGT-WT in *E. coli*
Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-mono (W198E/I201D)

Initially cloned by: Cassie Joiner

Purpose: Expression of OGT-mono in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 112080

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK

ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIELADHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-8

Initially cloned by: John Janetzko

Purpose: Expression of OGT-8 in *E. coli*

Molecular Weight (kDa): 101.2

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 100160

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSALGRLEEAKACYLKAIETQPNFAVAWS
NLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYINLGNVLKEARIFDRAVAAYLRALSLSPN
HAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANALKEKGSVAEAEDCYNT
ALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNLASVLQQQGKLQEA
LMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINPAFADAHSNLASIHKDS
GNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSIVADQLEKNRLPSV
HPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGRLRVGYVSSDFGN
HPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLSQIPCNGKAADRIH

QDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFMDYIITDQETSPAEVAEQ
YSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIV
KMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSISNGLATTQINNKA
ATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAV
GEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTP
MVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQRISSP
LFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVE

Protein: His_8-[HRV3C]-OGT-7*

Initially cloned by: John Janetzko

Purpose: Expression of OGT-7* in *E. coli*

Molecular Weight (kDa): 98.1

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 93170

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSGCVFNAQGEIWLAIHHFEKAVTLDPNF
LDAYINLGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIE
LQPHFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRL
YRKALEVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQD
VQGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCL
QIVCDWTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKI
NVLHKPPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDD
GTNFRVKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQ
AMWLGYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVI
DFKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIA
EAVIEMINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQ
LYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHV
RRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNR
QEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPD
HMIKPVE

Protein: His_8-[HRV3C]-OGT-7

Initially cloned by: John Janetzko

Purpose: Expression of OGT-7 in *E. coli*

Molecular Weight (kDa): 95.5

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 87670

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSDPNFLDAYINLGNVLKEARIFDRAVAA
YLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANALKEKGSV
AEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNLASVL
QQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINPAFADAH
SNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSIVAD
QLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGRLR
VGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLSQIP
CNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFMDYIITD
QETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDLKA
FLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSISNG
LATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVPNS
VLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTG
MDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRG
KVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVE

Protein: His_8-[HRV3C]-OGT-6

Initially cloned by: John Janetzko

Purpose: Expression of OGT-6 in *E. coli*

Molecular Weight (kDa): 91.7

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 84690

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSSPNHAVVHGNLACVYYEQGLIDLAIDT
YRRAIELQPHFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNI
EEAVRLYRKALEVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTL

KEMQDVQGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCN
LAHCLQIVCDWTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNL
CLDKINVLHKPPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYA
LSPDDGTNFRVKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALR
PAPIQAMWLGYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLK
KKAVIDFKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIP
MNTIAEAVIEMINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVY
CNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAP
KEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLE
LIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAA
GNKPDHMIKPVE

Protein: His_8-[HRV3C]-OGT-5

Initially cloned by: John Janetzko

Purpose: Expression of OGT-5 in *E. coli*

Molecular Weight (kDa): 87.9

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 80220

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSQPHFPDAYCNLANALKEKGSVAEAED
CYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNLASVLQQQGK
LQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINPAFADAHSNLASI
HKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSIVADQLEKN
RLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGRLRVGYVSS
DFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLSQIPCNGKA
ADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFMDYIITDQETSPAE
VAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDLKAFLDSLPD
VKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSISNGLATTQIN
NKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRF
PAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWA
GTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQR

ISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVE

Protein: His_8-[HRV3C]-OGT-4

Initially cloned by: John Janetzko

Purpose: Expression of OGT-4 in *E. coli*

Molecular Weight (kDa): 84.2

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 77240

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSCPTHADSLNNLANIKREQGNIEEAVRLY
RKALEVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDV
QGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQ
IVCDWTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKIN
VLHKPPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDG
TNFRVKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQA
MWLGYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVID
FKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAE
AVIEMINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQL
YKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVR
RGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNR
QEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPD
HMIKPVE

Protein: His_8-[HRV3C]-OGT-5N5A-1 (N322/356/390/424/458A)

Initially cloned by: Zebulon Levine[60]

Purpose: Expression of OGT-5N5A-1 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC

MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNALANIKREQGNIEEAVRLYRKALE
VFPEFAAAHSALASVLQQQGKLQEALMHYKEAIRISPTFADAYSAMGNTLKEMQDVQGALQ
CYTRAIQINPAFADAHSALASIHKDSGNIPEAIASYRTALKLKPDFPDAYCALAHCLQIVCDWT
DYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPP
YEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVK
VMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGY
PGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGH
IYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMIN
RGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPST
LQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLAD
VCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIA
VKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVE

Protein: His_8-[HRV3C]-OGT-5N5A-2 (N186/220/254/288/322A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-5N5A-2 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSALGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIA
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGALACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCALANALKEKGSVAEAEDCYNTALRLCPTHADSLNALANIKREQGNIEEAVRLYRKALE
VFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQ

CYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDW
TDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKP
PYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRV
KVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLG
YPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNG
HIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMI
NRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPS
TLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLA
DVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDI
AVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPV
E

Protein: His_8-[HRV3C]-OGT-5N5A-3 (N84/118/155/186/220A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-5N5A-3 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSALGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYIALAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGALLK
ALGRLEEAKACYLKAIETQPNFAVAWSALGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIA
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN

GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N322A

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of OGT-N322A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNALANIKREQGNIEEAVRLYRKALE
VFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQ
CYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDW
TDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKP
PYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRV
KVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLG
YPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNG
HIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMI
NRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPS
TLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLA
DVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDI
AVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPV

E

Protein: His_8-[HRV3C]-OGT-N356A

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of OGT-N356A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSALASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N390A

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of OGT-N390A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSAMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N424A

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of OGT-N424A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC

MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSALASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N458A

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of OGT-N458A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL

EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCALAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-D2A-1 (D420/454A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-D2A-1 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFAAAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPAAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL

GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-D2A-2 (D386/420A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-D2A-2 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFAAAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFAAAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE

DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-D2A-3 (D284/318A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-D2A-3 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
AAYCNLANALKEKGSVAEAEDCYNTALRLCPTHAASLNNLANIKREQGNIEEAVRLYRKALE
VFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQ
CYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDW
TDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKP
PYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRV
KVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLG
YPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNG
HIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMI
NRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPS
TLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLA
DVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDI
AVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPV
E

Protein: His_8-[HRV3C]-OGT-D2A-4 (D152/216A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-D2A-4 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSALGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLAAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-D2A-5 (D114/152A)

Initially cloned by: Cassie Joiner[62]

Purpose: Expression of OGT-D2A-5 in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIAGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSALGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-H19A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-H19A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAAREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP

DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-E80A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-E80A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAAAYSNLGNVYKERGQLQ
EAIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR

VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N84A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-N84A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSALGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ

LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-D114A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-D114A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIAGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N118A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-N118A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYIALAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-C148A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-C148A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYAVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-C189A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-C189A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGAVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIN

LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK
PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

Protein: His_8-[HRV3C]-OGT-N220A

Initially cloned by: Forrest Hammel

Purpose: Expression of OGT-N220A in *E. coli*

Molecular Weight (kDa): 118

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 117580

Protein Sequence:

MASMTGGQQMGRDPHHHHHHHHHLEVLFQGPGSADSTGLAELAHREYQAGDFEAAERHC
MQLWRQEPDNTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQE
AIEHYRHALRLKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLK
ALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYIA
LGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFP
DAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKAL
EVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGAL
QCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCD
WTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHK

PPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFR
VKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWL
GYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSN
GHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIE
MINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKID
PSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQ
LADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYE
DIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIK
PVE

A.1.2 OGT mammalian expression constructs

Protein: mKate2-2xFLAG-OGT-WT

Initially cloned by: Zebulon Levine[69]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-WT

Molecular Weight (kDa): 145

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFASSVGNVADSTGLAELAHREYQAGDFEAAERHCMQLWRQEPDN
TGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQEAIEHYRHALRL
KPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLKALGRLEEAKAC
YLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYINLGNVLKEARIF
DRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANAL
KEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHS
NLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINP
AFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKK

LVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKL
SDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFI
DLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFM
DYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLN
GIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITIN
GFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANIL
KRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLC
NGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLE
YLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

Protein: dTAG-OGT (FKBP12^{F36V}-2xHA-OGT-WT)

Initially cloned by: Zebulon Levine[69]

Purpose: Mammalian expression of (FKBP12^{F36V}-2xHA-tagged OGT-WT

Molecular Weight (kDa): 131

Protein Sequence:

MGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKVDSSRDRNKPFKFMLGKQEVIRGWE
EGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLEGGYPYDVPDYAGYPYD
VPDYAGIDRSTSLYKKAGFASSVGNVADSTGLAELAHREYQAGDFEAAERHCMQLWRQEPD
NTGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQEAIEHYRHALR
LKPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLKALGRLEEAKA
CYLKAIETQPNFAVAWSNLGCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYINLGNVLKEARI
FDRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANA
LKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAH
SNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINP
AFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKK
LVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKL
SDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFI
DLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFM
DYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLN
GIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITIN

GFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANIL
KRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLC
NGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLE
YLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

Protein: mKate2-2xFLAG-OGT-mono (W198E/I201D)

Initially cloned by: Sarah Potter[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-mono

Molecular Weight (kDa): 145

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFASSVGNVADSTGLAELAHREYQAGDFEAAERHCMQLWRQEPDN
TGVLLLLSSIHFQCRRLDRSAHFSTLAIKQNPLLAEAYSNLGNVYKERGQLQEAIEHYRHALRL
KPDFIDGYINLAAALVAAGDMEGAVQAYVSALQYNPDLYCVRSDLGNLLKALGRLEEAKAC
YLKAIETQPNFAVAWSNLGCVFNAQGEIELADHHFEKAVTLDPNFLDAYINLGNVLKEARIF
DRAVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANAL
KEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHS
NLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINP
AFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKK
LVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKL
SDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFI
DLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFM
DYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLN
GIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITIN
GFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANIL
KRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLC
NGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLE

YLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

Protein: mKate2-2xFLAG-OGT-8

Initially cloned by: Sarah Potter[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-8

Molecular Weight (kDa): 128

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFNLLKALGRLEEAKACYLKAIETQPNFAVAWSNLGCVFNAQGEIW
LAIHHFEKAVTLDPNFLDAYINLGNVLKEARIFDRAVAAYLRALSLSPNHAVVHGNLACVYY
EQGLIDLAIDTYRRAIELQPHFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNL
ANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFAD
AYSNMGNTLKEMQDVQGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKL
KPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGF
RKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMH
NPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYT
KGARNELFALRPAPIQAMWLGYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIG
DHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSS
NTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQ
YGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLP
QNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAA
SQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYL
QMWEHYAAGNKPDHMIKPVEVTESA

Protein: mKate2-2xFLAG-OGT-7*

Initially cloned by: Sarah Potter[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-7*

Molecular Weight (kDa): 125

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFCVFNAQGEIWLAIHHFEKAVTLDPNFLDAYINLGNVLKEARIFDR
AVAAYLRALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANALKE
KGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNL
ASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINPAFA
DAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSI
VADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDG
RLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLS
QIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFMDYII
TDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDL
KAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSIS
NGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVP
NSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHT
TGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKV
RGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

Protein: mKate2-2xFLAG-OGT-7

Initially cloned by: Cassie Joiner[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-7

Molecular Weight (kDa): 123

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN

LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFAVTLDPNFLDAYINLGNVLKEARIFDRAVAAYLRALSLSPNHAVV
HGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDAYCNLANALKEKGSVAEAEDCYNTALRLC
PTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNLASVLQQQGKLQEALMHY
KEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPE
AIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSIVADQLEKNRLPSVHPHH
SMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSH
LMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLSQIPCNGKAADRIHQDGIH
ILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFMDYIITDQETSPAEVAEQYSEKLA
YMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPD
GGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSISNGLATTQINNKAATGEEVP
RTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQ
YAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPG
ETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQY
TMELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

Protein: mKate2-2xFLAG-OGT-6

Initially cloned by: Cassie Joiner[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-6

Molecular Weight (kDa): 119

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFALSLSPNHAVVHGNLACVYYEQGLIDLAIDTYRRAIELQPHFPDA
YCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVF
PEFAAAHSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCY
TRAIQINPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTD
YDERMKKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPY

EHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKV
MAEANHFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYP
GTSGALFMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHI
YDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINR
GQIQITINGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTL
QMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLAD
VCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIA
VKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVE
VTESA

Protein: mKate2-2xFLAG-OGT-5

Initially cloned by: Cassie Joiner[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-5

Molecular Weight (kDa): 115

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFAIELQPHFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHAD
SLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAAHSNLASVLQQQGKLQEALMHYKEAIRIS
PTFADAYSNMGNTLKEMQDVQGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPEAIASYR
TALKLKPDFPDAYCNLAHCLQIVCDWTDYDERMKKLVSIVADQLEKNRLPSVHPHHSMLYP
LSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDLKLSDGRLRVGYVSSDFGNHPTSHLMQSI
PGMHNPDKFEVFCYALSPDDGTNFRVKVMAEANHFIDLSQIPCNGKAADRIHQDGIHILVN
MNGYTKGARNELFALRPAPIQAMWLGYPGTSGALFMDYIITDQETSPAEVAEQYSEKLAYMP
HTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIVLNGIDLKAFLDSLPDVKIVKMKCPDGG
DNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITINGFSISNGLATTQINNKAATGEEVPRTI
IVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYA
QNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETL

ASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTM
ELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

Protein: mKate2-2xFLAG-OGT-4

Initially cloned by: Cassie Joiner[108]

Purpose: Mammalian expression of mKate2-2xFLAG-tagged OGT-4

Molecular Weight (kDa): 111

Protein Sequence:

MVSELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKAVEGGPLPFAFDILATS
FMYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVN
FPSNGPVMQKKTLGWEASTETLYPADGGLEGRADMALKLVGGGHLICNLKTTYRSKKPAKN
LKMPGVYYVDRRLERIKEADKETYVEQHEVAVARYCDLPSKLGHRGGDYKDDDDKGDYKD
DDDKGIDRSTSLYKKAGFALRLCPTHADSLNNLANIKREQGNIEEAVRLYRKALEVFPEFAAA
HSNLASVLQQQGKLQEALMHYKEAIRISPTFADAYSNMGNTLKEMQDVQGALQCYTRAIQI
NPAFADAHSNLASIHKDSGNIPEAIASYRTALKLKPDFPDAYCNLAHCLQIVCDWTDYDERM
KKLVSIVADQLEKNRLPSVHPHHSMLYPLSHGFRKAIAERHGNLCLDKINVLHKPPYEHPKDL
KLSDGRLRVGYVSSDFGNHPTSHLMQSIPGMHNPDKFEVFCYALSPDDGTNFRVKVMAEAN
HFIDLSQIPCNGKAADRIHQDGIHILVNMNGYTKGARNELFALRPAPIQAMWLGYPGTSGAL
FMDYIITDQETSPAEVAEQYSEKLAYMPHTFFIGDHANMFPHLKKKAVIDFKSNGHIYDNRIV
LNGIDLKAFLDSLPDVKIVKMKCPDGGDNADSSNTALNMPVIPMNTIAEAVIEMINRGQIQITI
NGFSISNGLATTQINNKAATGEEVPRTIIVTTRSQYGLPEDAIVYCNFNQLYKIDPSTLQMWAN
ILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRIIFSPVAPKEEHVRRGQLADVCLDTPL
CNGHTTGMDVLWAGTPMVTMPGETLASRVAASQLTCLGCLELIAKNRQEYEDIAVKLGTDL
EYLKKVRGKVWKQRISSPLFNTKQYTMELERLYLQMWEHYAAGNKPDHMIKPVEVTESA

A.2 HCF-1 repeat protein constructs

A.2.1 HCF-1 repeat bacterial expression constructs

The following constructs for expression of derivatives of the HCF-1 proteolytic repeats in E. coli were developed from the pET47b-His_8-[HRV3C]-HCF3R-EEE plasmid initially cloned by Dr. John

Janetzko. Since many of these constructs were used after HRV 3C protease cleavage to remove the polyhistidine tag, I have annotated these by italicizing the cleaved portion and listing molecular weights and molar extinction coefficients pre- (not italicized) and post-cleavage (italicized).

Protein: His_8-[HRV3C]-HCF3R-EEE-YYY

Initially cloned by: Cassie Joiner

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCETHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNTATTATSNMAGQ
HGCSNPPCETHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY

Initially cloned by: Cassie Joiner

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-SAA-YYY

Initially cloned by: Cassie Joiner

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCSTHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-EAA-YYY

Initially cloned by: Cassie Joiner

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCETHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: StrepTag-[HRV3C]-HaloTag-HCF3R-AAA-YYY

Initially cloned by: Forrest Hammel

Purpose: Expression of HaloTagged HCF3R in *E. coli*

Molecular Weight (kDa): 49.5

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 69900

Protein Sequence:

MAWSHPQFEKSAALEVLFQGPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAAAS
LVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQHGCSNPPCATHETGTTNT
ATTAMSSVGANYYY

Protein: StrepTag-[HRV3C]-HaloTag-HCF3R-EEE-YYY

Initially cloned by: Forrest Hammel

Purpose: Expression of HaloTagged HCF3R in *E. coli*

Molecular Weight (kDa): 49.5

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 69900

Protein Sequence:

MAWSHPQFEKSAALEVLFQGPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPPGTVTLVCSNPPCETHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAAASL
VTSTVGQQNGSVVRVCSNPPCETHETGTTNTATTATSNMAGQHGCSNPPCETHETGTTNTAT
TAMSSVGANYYY

Protein: StrepTag-[HRV3C]-HaloTag-HCF1R-A

Initially cloned by: Forrest Hammel

Purpose: Expression of HaloTagged HCF1R in *E. coli*

Molecular Weight (kDa): 49.5

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 65430

Protein Sequence:

MAWSHPQFEKSAALEVLFQGPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPNGSVVRVCSNPPCATHETGTTNTATTATSN

Protein: His_8-[HRV3C]-HaloTag-HCF2R-1/2-AA ("HCF2R-1/2")

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of HaloTagged HCF2R-1/2-AA in *E. coli*

Molecular Weight (kDa): 45.8 (*43.6*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 59930 (*59930*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAAAS
LVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSN

Protein: His_8-[HRV3C]-HaloTag-HCF2R-2/3-AA-YYY ("HCF2R-2/3")

Initially cloned by: Forrest Hammel[109]

Purpose: Expression of HaloTagged HCF2R-2/3-AA in *E. coli*

Molecular Weight (kDa): 42.9 (*40.6*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 64400 (*64400*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPVRVCSNPPCATHETGTTNTATTATSNMAGQHGCSNPPCATHETGTTNTATTAM
SSVGANYYY

Protein: His_8-[HRV3C]-HaloTag-HCF2R-1/2Δ1-A ("HCF2R-1/2Δ1")

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF2R-1/2Δ1-A in *E. coli*

Molecular Weight (kDa): 42.8 (*40.5*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 59930 (*59930*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTN
TATTATSN

Protein: His_8-[HRV3C]-HaloTag-HCF2R-1/2Δ2-A ("HCF2R-1/2Δ2")

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF2R-1/2Δ2-A in *E. coli*

Molecular Weight (kDa): 43.2 (*41*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 59930 (*59930*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLH
GNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVL
VIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQ
NVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLST
LEISGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAAAS
LVTSTVGQQNGSV

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 1 C4A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVASNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 1 C9A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPAATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 1 C40A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVADRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 2 C4A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVASNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 2 C9A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPAATHETGTTNTATTATSNMAGQ
HGCSNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 3 C4A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGASNPPCATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, repeat 3 C9A

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGH
PQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQ
HGCSNPPAATHETGTTNTATTAMSSVGANYYY

Protein: His_8-[HRV3C]-HCF3R-AAA-YYY, ΔCys

Initially cloned by: Forrest Hammel

Purpose: Expression of HCF3R in *E. coli*

Molecular Weight (kDa): 16.1 (*13.8*)

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 5960 (*5960*)

Protein Sequence:

*MAHHHHHHHHSAALEVLFQ*GPGYQDPPGTVTLVASNPPAATHETGTTNTATTTVVANLGG
HPQPTQVQFVADRQEAAASLVTSTVGQQNGSVVRVASNPPAATHETGTTNTATTATSNMAG
QHGASNPPAATHETGTTNTATTAMSSVGANYYY

A.2.2 HCF-1 repeat mammalian expression constructs

Protein: mEGFP-2xFLAG ("GFP-2xFLAG")

Initially cloned by: Forrest Hammel[109]

Purpose: Mammalian expression of 2xFLAG-tagged monomeric EGFP (mEGFP)

Molecular Weight (kDa): 29.4

Protein Sequence:

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNT

PIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSGDYKDDDD
KGDYKDDDDKGSG

Protein: mEGFP-2xFLAG-HCF3R-AAA (“GFP-2xFLAG-HCF3R”)

Initially cloned by: Forrest Hammel[109]

Purpose: Mammalian expression of GFP-2xFLAG-tagged HCF3R-AAA

Molecular Weight (kDa): 41.9

Protein Sequence:

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSGDYKDDDD
KGDYKDDDDKGSGPGTVTLVCSNPPCATHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQ
EAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTNTATTATSNMAGQHGCSNPPCATHET
GTTNTATTAMSSVGAN

Protein: mEGFP-FKBP12^{F36V}-2xFLAG-HCF3R-AAA (“GFP-degron-2xFLAG-HCF3R”)

Initially cloned by: Victoria Marando[109]

Purpose: Mammalian expression of GFP-degron-2xFLAG-tagged HCF3R-AAA)

Molecular Weight (kDa): 53.8

Protein Sequence:

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSMGVQVETIS
PGDGRTFPKRGQTCVVHYTGMLEDGKKVDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSV
GQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLEGDYKDDDDKGDYKDDDDKGSGPG
TVTLVCSNPPCATHETGTTNTATTTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQN
GSVVRVCSNPPCATHETGTTNTATTATSNMAGQHGCSNPPCATHETGTTNTATTAMSSVGA

N

A.3 OGT substrate bacterial expression constructs

The following constructs for expression of OGT substrate point mutants in *E. coli* were developed from pET28a plasmids cloned by David Vocadlo's group at Simon Fraser University.

Protein: His_6-TAB1-WT

Initially cloned by: David Vocadlo lab[116]

Purpose: Expression of TAB1-WT in *E. coli*

Molecular Weight (kDa): 56.9

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 37360

Protein Sequence:

MGSSHHHHHHSSGLVPRGSHMMAAQRRSLLQSEQQPSWTDDLPLCHLSGVGSASNRSYSAD
GKGTESHPPEDSWLKFRSENNCFLYGVFNGYDGNRVTNFVAQRLSAELLLGQLNAEHAEAD
VRRVLLQAFDVVERSFLESIDDALAEKASLQSQLPEGVPQHQLPPQYQKILERLKTLEREISGGA
MAVVAVLLNNKLYVANVGTNRALLCKSTVDGLQVTQLNVDHTTENEDELFRLSQLGLDAG
KIKQVGIICGQESTRRIGDYKVKYGYTDIDLLSAAKSKPIIAEPEIHGAQPLDGVTGFLVLMSEGL
YKALEAAHGPGQANQEIAAMIDTEFAKQTSLDAVAQAVVDRVKRIHSDTFASGGERARFCPR
HEDMTLLVRNFGYPLGEMSQPTPSPAPAAGGRVYPVSVPYSSAQSTSKTSVTLSLVMPSQGQM
VNGAHSASTLDEATPTLTNQSPTLTLQSTNTHTQSSSSSSDGGLFRSRPAHSLPPGEDGRVEPYV
DFAEFYRLWSVDHGEQSVVTAP

Protein: His_6-TAB1-S395A

Initially cloned by: Forrest Hammel[62]

Purpose: Expression of TAB1-S395A in *E. coli*

Molecular Weight (kDa): 56.9

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 37360

Protein Sequence:

MGSSHHHHHHSSGLVPRGSHMMAAQRRSLLQSEQQPSWTDDLPLCHLSGVGSASNRSYSAD

GKGTESHPPEDSWLKFRSENNCFLYGVFNGYDGNRVTNFVAQRLSAELLLGQLNAEHAEAD
VRRVLLQAFDVVERSFLESIDDALAEKASLQSQLPEGVPQHQLPPQYQKILERLKTLEREISGGA
MAVVAVLLNNKLYVANVGTNRALLCKSTVDGLQVTQLNVDHTTENEDELFRLSQLGLDAG
KIKQVGIICGQESTRRIGDYKVKYGYTDIDLLSAAKSKPIIAEPEIHGAQPLDGVTGFLVLMSEGL
YKALEAAHGPGQANQEIAAMIDTEFAKQTSLDAVAQAVVDRVKRIHSDTFASGGERARFCPR
HEDMTLLVRNFGYPLGEMSQPTPSPAPAAGGRVYPVSVPYASAQSTSKTSVTLSLVMPSQGQ
MVNGAHSASTLDEATPTLTNQSPTLTLQSTNTHTQSSSSSSDGGLFRSRPAHSLPPGEDGRVEP
YVDFAEFYRLWSVDHGEQSVVTAP

Protein: His_6-CAMKIV-WT

Initially cloned by: David Vocadlo lab[116]

Purpose: Expression of CAMKIV-WT in *E. coli*

Molecular Weight (kDa): 54.2

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 48360

Protein Sequence:

MGSSHHHHHHSSGLVPRGSHMMLKVTVPSCSASSCSSVTASAAPGTASLVPDYWIDGSNRDA
LSDFFEVESELGRGATSIVYRCKQKGTQKPYALKVLKKTVDKKIVRTEIGVLLRLSHPNIIKLKEI
FETPTEISLVLELVTGGELFDRIVEKGYYSERDAADAVKQILEAVAYLHENGIVHRDLKPENLLY
ATPAPDAPLKIADFGLSKIVEHQVLMKTVCGTPGYCAPEILRGCAYGPEVDMWSVGIITYILLC
GFEPFYDERGDQFMFRRILNCEYYFISPWWDEVSLNAKDLVRKLIVLDPKKRLTTFQALQHPW
VTGKAANFVHMDTAQKKLQEFNARRKLKAAVKAVVASSRLGSASSSHGSIQESHKASRDPSP
IQDGNEDMKAIPEGEKIQGDGAQAAVKGAQAELMKVQALEKVKGADINAEEAPKMVPKAV
EDGIKVADLELEEGLAEEKLKTVEEAAAPREGQGSSAVGFEVPQQDVILPEY

Protein: His_6-CARM1-WT

Initially cloned by: David Vocadlo lab[116]

Purpose: Expression of CARM1-WT in *E. coli*

Molecular Weight (kDa): 68.1

Molar Extinction Coefficient, 280 nm ($M^{-1}cm^{-1}$): 61770

Protein Sequence:

MGSSHHHHHHSSGLVPRGSHMMAAAAAAVGPGAGGAGSAVPGGAGPCATVSVFPGARLL
TIGDANGEIQRHAEQQALRLEVRAGPDSAGIALYSHEDVCVFKCSVSRETECSRVGKQSFIITLG
CNSVLIQFATPNDFCSFYNILKTCRGHTLERSVFSERTEESSAVQYFQFYGYLSQQQNMMQDYV
RTGTYQRAILQNHTDFKDKIVLDVGCGSGILSFFAAQAGARKIYAVEASTMAQHAEVLVKSN
NLTDRIVVIPGKVEEVSLPEQVDIIISEPMGYMLFNERMLESYLHAKKYLKPSGNMFPTIGDVHL
APFTDEQLYMEQFTKANFWYQPSFHGVDLSALRGAAVDEYFRQPVVDTFDIRILMAKSVKYT
VNFLEAKEGDLHRIEIPFKFHMLHSGLVHGLAFWFDVAFIGSIMTVWLSTAPTEPLTHWYQVR
CLFQSPLFAKAGDTLSGTCLLIANKRQSYDISIVAQVDQTGSKSSNLLDLKNPFFRYTGTTPSPP
PGSHYTSPSENMWNTGSTYNLSSGMAVAGMPTAYDLSSVIASGSSVGHNNLIPLANTGIVNH
THSRMGSIMSTGIVQGSSGAQGSGGGSTSAHYAVNSQFTMGGPAISMASPMSIPTNTMHYGS

Chapter 6

References

References

(1) Editors *The Lancet* **1962**, *280*, 81–83.

(2) Baverstock, K. *Communicative & Integrative Biology* **2019**, *12*, 59–64.

(3) Crick, F. H. *Symposia of the Society for Experimental Biology* **1958**, *12*, 138–163.

(4) Crick, F. *Nature* **1970**, *227*, 561–563.

(5) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *Journal of Molecular Biology* **1987**, *194*, 531–544.

(6) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. *Protein Science: A Publication of the Protein Society* **2018**, *27*, 14–25.

(7) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. *Protein Science: A Publication of the Protein Society* **2021**, *30*, 70–82.

(8) Meng, E. C.; Goddard, T. D.; Pettersen, E. F.; Couch, G. S.; Pearson, Z. J.; Morris, J. H.; Ferrin, T. E. *Protein Science* **2023**, *32*, e4792.

(9) Wilkinson, M. E.; Charenton, C.; Nagai, K. *Annual Review of Biochemistry* **2020**, *89*, 359–388.

(10) Keenan, E. K.; Zachman, D. K.; Hirschey, M. D. *Molecular Cell* **2021**, *81*, 1868–1878.

(11) Lee, J. M.; Hammarén, H. M.; Savitski, M. M.; Baek, S. H. *Nature Communications* **2023**, *14*, 201.

(12) Ramazi, S.; Zahiri, J. *Database: The Journal of Biological Databases and Curation* **2021**, *2021*, baab012.

(13) Pawson, T.; Scott, J. D. *Trends in Biochemical Sciences* **2005**, *30*, 286–290.

(14) Murn, J.; Shi, Y. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 517–527.

(15) Yang, X.; Qian, K. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 452–465.

(16) Zachara, N. E.; Akimoto, Y.; Boyce, M.; Hart, G. W. In *Essentials of Glycobiology*, Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds., 4th; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2022.

(17) Levine, Z. G.; Walker, S. *Annual Review of Biochemistry* **2016**, *85*, 631–657.

(18) Li, W.; Ye, Y. *Cellular and Molecular Life Sciences* **2008**, *65*, 2397–2406.

(19) Rabut, G.; Peter, M. *EMBO Reports* **2008**, *9*, 969–976.

(20) Zhang, S.; Yu, Q.; Li, Z.; Zhao, Y.; Sun, Y. *Signal Transduction and Targeted Therapy* **2024**, *9*, 1–36.

(21) Rogers, L. D.; Overall, C. M. *Molecular & Cellular Proteomics* **2013**, *12*, 3532–3542.

(22) Cook, K. M.; Hogg, P. J. *Antioxidants & Redox Signaling* **2013**, *18*, 1987–2015.

(23) Gelens, L.; Saurin, A. T. *Developmental Cell* **2018**, *44*, 659–663.

(24) Torres, C. R.; Hart, G. W. *Journal of Biological Chemistry* **1984**, *259*, 3308–3317.

(25) Haltiwanger, R. S.; Holt, G. D.; Hart, G. W. *Journal of Biological Chemistry* **1990**, *265*, 2563–2568.

(26) Haltiwanger, R. S.; Blomberg, M. A.; Hart, G. W. *Journal of Biological Chemistry* **1992**, *267*, 9005–9013.

(27) Kreppel, L. K.; Blomberg, M. A.; Hart, G. W. *Journal of Biological Chemistry* **1997**, *272*, 9308–9315.

(28) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. *Nature Reviews Nephrology* **2019**, *15*, 346–366.

(29) Kumar, S.; Wang, Y.; Zhou, Y.; Dillard, L.; Li, F.-W.; Sciandra, C. A.; Sui, N.; Zentella, R.; Zahn, E.; Shabanowitz, J.; Hunt, D. F.; Borgnia, M. J.; Bartesaghi, A.; Sun, T.-p.; Zhou, P. *Nature Communications* **2023**, *14*, 1538.

(30) Alonso, J.; Schimpl, M.; van Aalten, D. M. F. *Journal of Biological Chemistry* **2014**, *289*, 34433–34439.

(31) Clarke, A. J.; HurtadoGuerrero, R.; Pathak, S.; Schüttelkopf, A. W.; Borodkin, V.; Shepherd, S. M.; Ibrahim, A. F. M.; van Aalten, D. M. F. *The EMBO Journal* **2008**, *27*, 2780–2788.

(32) Janetzko, J.; Walker, S. *Journal of Biological Chemistry* **2014**, *289*, 34424–34432.

(33) Hanover, J. A.; Yu, S.; Lubas, W. B.; Shin, S. H.; Ragano-Caracciola, M.; Kochran, J.; Love, D. C. *Archives of Biochemistry and Biophysics* **2003**, *409*, 287–297.

(34) Lazarus, B. D.; Love, D. C.; Hanover, J. A. *Glycobiology* **2006**, *16*, 415–421.

(35) Wulff-Fuentes, E.; Berendt, R. R.; Massman, L.; Danner, L.; Malard, F.; Vora, J.; Kahsay, R.; Olivier-Van Stichelen, S. *Scientific Data* **2021**, *8*, 25.

(36) Hu, J.; Xia, W.; Zeng, S.; Lim, Y.-J.; Tao, Y.; Sun, Y.; Zhao, L.; Wang, H.; Le, W.; Li, D.; Zhang, S.; Liu, C.; Li, Y.-M. *Nature Communications* **2024**, *15*, 2677.

(37) Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. *Annual Review of Biochemistry* **2011**, *80*, 825–858.

(38) King, D. T.; Serrano-Negrón, J. E.; Zhu, Y.; Moore, C. L.; Shoulders, M. D.; Foster, L. J.; Vocadlo, D. J. *Journal of the American Chemical Society* **2022**, *144*, 3833–3842.

(39) Ruan, H.-B.; Han, X.; Li, M.-D.; Singh, J. P.; Qian, K.; Azarhoush, S.; Zhao, L.; Bennett, A. M.; Samuel, V. T.; Wu, J.; Yates, J. R.; Yang, X. *Cell Metabolism* **2012**, *16*, 226–237.

(40) Paneque, A.; Fortus, H.; Zheng, J.; Werlen, G.; Jacinto, E. *Genes* **2023**, *14*, 933.

(41) Zachara, N. E.; Hart, G. W. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2004**, *1673*, 13–28.

(42) Katan, M.; Haigh, A.; Verrijzer, C. P.; van der Vliet, P. C.; O'Hare, P. *Nucleic Acids Research* **1990**, *18*, 6871–6880.

(43) Kristie, T. M.; Sharp, P. A. *Genes & Development* **1990**, *4*, 2383–2396.

(44) Zargar, Z. U.; Tyagi, S. *Transcription* **2012**, *3*, 187–192.

(45) Janetzko, J.; Trauger, S. A.; Lazarus, M. B.; Walker, S. *Nature Chemical Biology* **2016**, *12*, 899–901.

(46) Lazarus, M. B.; Jiang, J.; Kapuria, V.; Bhuiyan, T.; Janetzko, J.; Zandberg, W. F.; Vocadlo, D. J.; Herr, W.; Walker, S. *Science* **2013**, *342*, 1235–1239.

(47) Wilson, A. C.; LaMarco, K.; Peterson, M. G.; Herr, W. *Cell* **1993**, *74*, 115–125.

(48) Wilson, A. C.; Peterson, M. G.; Herr, W. *Genes & Development* **1995**, *9*, 2445–2458.

(49) Wilson, A. C.; Boutros, M.; Johnson, K. M.; Herr, W. *Molecular and Cellular Biology* **2000**, *20*, 6721–6730.

(50) Kristie, T. M.; Pomerantz, J. L.; Twomey, T. C.; Parent, S. A.; Sharp, P. A. *Journal of Biological Chemistry* **1995**, *270*, 4387–4394.

(51) Park, J.; Lammers, F.; Herr, W.; Song, J.-J. *Proceedings of the National Academy of Sciences* **2012**, *109*, 17430–17435.

(52) Julien, E.; Herr, W. *The EMBO Journal* **2003**, *22*, 2360–2369.

(53) Capotosti, F.; Guernier, S.; Lammers, F.; Waridel, P.; Cai, Y.; Jin, J.; Conaway, J. W.; Conaway, R. C.; Herr, W. *Cell* **2011**, *144*, 376–388.

(54) Lazarus, M. B.; Nam, Y.; Jiang, J.; Sliz, P.; Walker, S. *Nature* **2011**, *469*, 564–567.

(55) Blankenship, C. M.; Xie, J.; Benz, C.; Wang, A.; Ivarsson, Y.; Jiang, J. *Nature Chemical Biology* **2023**, *19*, 1423–1431.

(56) Alteen, M. G.; Meek, R. W.; Kolappan, S.; Busmann, J. A.; Cao, J.; O'Gara, Z.; Chou, Y.; Derda, R.; Davies, G. J.; Vocadlo, D. J. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2303690120.

(57) Lu, P.; Liu, Y.; He, M.; Cao, T.; Yang, M.; Qi, S.; Yu, H.; Gao, H. *Nature Communications* **2023**, *14*, 6952.

(58) Jínek, M.; Rehwinkel, J.; Lazarus, B. D.; Izaurralde, E.; Hanover, J. A.; Conti, E. *Nature Structural & Molecular Biology* **2004**, *11*, 1001–1007.

(59) Meek, R. W.; Blaza, J. N.; Busmann, J. A.; Alteen, M. G.; Vocadlo, D. J.; Davies, G. J. *Nature Communications* **2021**, *12*, 6508.

(60) Levine, Z. G.; Fan, C.; Melicher, M. S.; Orman, M.; Benjamin, T.; Walker, S. *Journal of the American Chemical Society* **2018**, *140*, 3510–3513.

(61) Joiner, C. M.; Levine, Z. G.; Aonbangkhen, C.; Woo, C. M.; Walker, S. *Journal of the American Chemical Society* **2019**, *141*, 12974–12978.

(62) Joiner, C. M.; Hammel, F. A.; Janetzko, J.; Walker, S. *Biochemistry* **2021**, *60*, 847–853.

(63) Ma, J.; Hou, C.; Li, Y.; Chen, S.; Wu, C. *International Journal of Molecular Sciences* **2021**, *22*, 9620.

(64) Allan, R. K.; Ratajczak, T. *Cell Stress & Chaperones* **2011**, *16*, 353–367.

(65) Hrit, J. et al. *eLife* **2018**, *7*, e34870.

(66) Ito, R.; Katsura, S.; Shimada, H.; Tsuchiya, H.; Hada, M.; Okumura, T.; Sugawara, A.; Yokoyama, A. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* **2014**, *19*, 52–65.

(67) Wang, L.; Li, G.; Zhou, Z.; Ge, C.; Chen, Q.; Liu, Y.; Zhang, N.; Zhang, K.; Niu, M.; Li, W.; Zhong, X.; Wu, S.; Zhang, J.; Liu, Y. *Oncogene* **2023**, *42*, 2329–2346.

(68) Karczewski, K. J. et al. *Nature* **2020**, *581*, 434–443.

(69) Levine, Z. G.; Potter, S. C.; Joiner, C. M.; Fei, G. Q.; Nabet, B.; Sonnett, M.; Zachara, N. E.; Gray, N. S.; Paulo, J. A.; Walker, S. *Proceedings of the National Academy of Sciences* **2021**, *118*.

(70) Pravata, V. M.; Muha, V.; Gundogdu, M.; Ferenbach, A. T.; Kakade, P. S.; Vandadi, V.; Wilmes, A. C.; Borodkin, V. S.; Joss, S.; Stavridis, M. P.; van Aalten, D. M. F. *Proceedings of the National Academy of Sciences* **2019**, *116*, 14961–14970.

(71) Gundogdu, M.; Llabrés, S.; Gorelik, A.; Ferenbach, A. T.; Zachariae, U.; van Aalten, D. M. F. *Cell Chemical Biology* **2018**, *25*, 513–518.e4.

(72) Mayfield, J. M.; Hitefield, N. L.; Czajewski, I.; Vanhye, L.; Holden, L.; Morava, E.; Aalten, D. M. F.; Wells, L. *Journal of Biological Chemistry* **2024**, *300*.

(73) Pravata, V. M.; Omelková, M.; Stavridis, M. P.; Desbiens, C. M.; Stephen, H. M.; Lefeber, D. J.; Gecz, J.; Gundogdu, M.; Õunap, K.; Joss, S.; Schwartz, C. E.; Wells, L.; van Aalten, D. M. F. *European Journal of Human Genetics* **2020**, *28*, 706–714.

(74) Omelková, M.; Fenger, C. D.; Murray, M.; Hammer, T. B.; Pravata, V. M.; Bartual, S. G.; Czajewski, I.; Bayat, A.; Ferenbach, A. T.; Stavridis, M. P.; van Aalten, D. M. F. *Disease Models & Mechanisms* **2023**, *16*, dmm049132.

(75) Tsherniak, A. et al. *Cell* **2017**, *170*, 564–576.e16.

(76) DepMap: The Cancer Dependency Map Project at Broad Institute, https://depmap.org/portal/.

(77) Shafi, R.; Iyer, S. P.; Ellies, L. G.; O'Donnell, N.; Marek, K. W.; Chui, D.; Hart, G. W.; Marth, J. D. *Proceedings of the National Academy of Sciences* **2000**, *97*, 5735–5739.

(78) O'Donnell, N.; Zachara, N. E.; Hart, G. W.; Marth, J. D. *Molecular and Cellular Biology* **2004**, *24*, 1680–1690.

(79) Hanover, J. A.; Forsythe, M. E.; Hennessey, P. T.; Brodigan, T. M.; Love, D. C.; Ashwell, G.; Krause, M. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 11266–11271.

(80) Love, D. C.; Ghosh, S.; Mondoux, M. A.; Fukushige, T.; Wang, P.; Wilson, M. A.; Iser, W. B.; Wolkow, C. A.; Krause, M. W.; Hanover, J. A. *Proceedings of the National Academy of Sciences* **2010**, *107*, 7413–7418.

(81) Vaidyanathan, K.; Wells, L. *Journal of Biological Chemistry* **2014**, *289*, 34466–34471.

(82) Bond, M. R.; Ghosh, S. K.; Wang, P.; Hanover, J. A. *PLOS ONE* **2014**, *9*, e113231.

(83) Simon, J. A.; Kingston, R. E. *Molecular Cell* **2013**, *49*, 808–824.

(84) Sinclair, D. A. R.; Syrzycka, M.; Macauley, M. S.; Rastgardani, T.; Komljenovic, I.; Vocadlo, D. J.; Brock, H. W.; Honda, B. M. *Proceedings of the National Academy of Sciences* **2009**, *106*, 13427–13432.

(85) Gambetta, M. C.; Oktaba, K.; Muller, J. *Science* **2009**, *325*, 93–6.

(86) Dassanayaka, S.; Brainard, R. E.; Watson, L. J.; Long, B. W.; Brittian, K. R.; DeMartino, A. M.; Aird, A. L.; Gumpert, A. M.; Audam, T. N.; Kilfoil, P. J.; Muthusamy, S.; Hamid, T.; Prabhu, S. D.; Jones, S. P. *Basic Research in Cardiology* **2017**, *112*, 23.

(87) Lee, B. E. et al. *Brain* **2020**, *143*, 3699–3716.

(88) Yang, X.; Su, K.; Roos, M. D.; Chang, Q.; Paterson, A. J.; Kudlow, J. E. *Proceedings of the National Academy of Sciences* **2001**, *98*, 6611–6616.

(89) Marshall, S.; Duong, T.; Wu, T.; Hering, M. A.; Yada, J.; Higgins, S.; Orbus, R. J.; Yan, Z.-H.; Rumberger, J. M. *Analytical Biochemistry* **2003**, *319*, 304–313.

(90) Kreppel, L. K.; Hart, G. W. *Journal of Biological Chemistry* **1999**, *274*, 32015–32022.

(91) Lubas, W. A.; Hanover, J. A. *Journal of Biological Chemistry* **2000**, *275*, 10983–10988.

(92) Gross, B. J.; Kraybill, B. C.; Walker, S. *Journal of the American Chemical Society* **2005**, *127*, 14588–14589.

(93) Gross, B. J.; Swoboda, J. G.; Walker, S. *Journal of the American Chemical Society* **2008**, *130*, 440–441.

(94) Ortiz-Meoz, R. F.; Jiang, J.; Lazarus, M. B.; Orman, M.; Janetzko, J.; Fan, C.; Duveau, D. Y.; Tan, Z.-W.; Thomas, C. J.; Walker, S. *ACS Chemical Biology* **2015**, *10*, 1392–1397.

(95) Martin, S. E. S.; Tan, Z.-W.; Itkonen, H. M.; Duveau, D. Y.; Paulo, J. A.; Janetzko, J.; Boutz, P. L.; Törk, L.; Moss, F. A.; Thomas, C. J.; Gygi, S. P.; Lazarus, M. B.; Walker, S. *Journal of the American Chemical Society* **2018**, *140*, 13542–13545.

(96) Kazemi, Z.; Chang, H.; Haserodt, S.; McKen, C.; Zachara, N. E. *Journal of Biological Chemistry* **2010**, *285*, 39096–39107.

(97) Umapathi, P.; Mesubi, O. O.; Banerjee, P. S.; Abrol, N.; Wang, Q.; Luczak, E. D.; Wu, Y.; Granger, J. M.; Wei, A.-C.; Reyes Gaido, O. E.; Florea, L.; Talbot, C. C.; Hart, G. W.; Zachara, N. E.; Anderson, M. E. *Circulation* **2021**, *143*, 1687–1703.

(98) Wang, H.-f.; Wang, Y.-x.; Zhou, Y.-p.; Wei, Y.-p.; Yan, Y.; Zhang, Z.-j.; Jing, Z.-c. *Acta Pharmacologica Sinica* **2023**, *44*, 8–18.

(99) Ma, J.; Hart, G. W. *Expert Review of Proteomics* **2013**, *10*, 365–380.

(100) Ma, Z.; Vocadlo, D. J.; Vosseller, K. *Journal of Biological Chemistry* **2013**, *288*, 15121–15130.

(101) Zhu, Q.; Zhou, H.; Wu, L.; Lai, Z.; Geng, D.; Yang, W.; Zhang, J.; Fan, Z.; Qin, W.; Wang, Y.; Zhou, R.; Yi, W. *Nature Chemical Biology* **2022**, *18*, 1087–1095.

(102) Itkonen, H. M.; Poulose, N.; Steele, R. E.; Martin, S. E. S.; Levine, Z. G.; Duveau, D. Y.; Carelli, R.; Singh, R.; Urbanucci, A.; Loda, M.; Thomas, C. J.; Mills, I. G.; Walker, S. *Molecular Cancer Research* **2020**, *18*, 1512–1521.

(103) Itkonen, H. M.; Urbanucci, A.; Martin, S. E.; Khan, A.; Mathelier, A.; Thiede, B.; Walker, S.; Mills, I. G. *Theranostics* **2019**, *9*, 2183–2197.

(104) Kamigaito, T.; Okaneya, T.; Kawakubo, M.; Shimojo, H.; Nishizawa, O.; Nakayama, J. *Prostate Cancer and Prostatic Diseases* **2014**, *17*, 18–22.

(105) Ma, Z.; Vosseller, K. *Journal of Biological Chemistry* **2014**, *289*, 34457–34465.

(106) Hart, G. W.; Huang, C.-W. *Brain* **2021**, *143*, 3515–3518.

(107) Zhu, Y.; Shan, X.; Yuzwa, S. A.; Vocadlo, D. J. *Journal of Biological Chemistry* **2014**, *289*, 34472–34481.

(108) Potter, S. C.; Gibbs, B. E.; Hammel, F. A.; Joiner, C. M.; Paulo, J. A.; Janetzko, J.; Levine, Z. G.; Fei, G. Q.; Haggarty, S. J.; Walker, S. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2401729121.

(109) Hammel, F. A.; Payne, N. C.; Marando, V. M.; Mazitschek, R.; Walker, S. *Journal of the American Chemical Society* **2024**, *146*, 26320–26330.

(110) Leavy, T. M.; Bertozzi, C. R. *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 3851.

(111) Malard, F.; Wulff-Fuentes, E.; Berendt, R. R.; Didier, G.; Olivier-Van Stichelen, S. *Database* **2021**, *2021*, baab039.

(112) Woo, C. M.; Lund, P. J.; Huang, A. C.; Davis, M. M.; Bertozzi, C. R.; Pitteri, S. J. *Molecular & Cellular Proteomics* **2018**, *17*, 764–775.

(113) Pathak, S.; Alonso, J.; Schimpl, M.; Rafie, K.; Blair, D. E.; Borodkin, V. S.; Albarbarawi, O.; van Aalten, D. M. F. *Nature Structural & Molecular Biology* **2015**, *22*, 744–750.

(114) Groves, J. A.; Lee, A.; Yildirir, G.; Zachara, N. E. *Cell Stress and Chaperones* **2013**, *18*, 535–558.

(115) Hein, M. Y.; Hubner, N. C.; Poser, I.; Cox, J.; Nagaraj, N.; Toyoda, Y.; Gak, I. A.; Weisswange, I.; Mansfeld, J.; Buchholz, F.; Hyman, A. A.; Mann, M. *Cell* **2015**, *163*, 712–723.

(116) Shen, D. L.; Gloster, T. M.; Yuzwa, S. A.; Vocadlo, D. J. *Journal of Biological Chemistry* **2012**, *287*, 15395–15408.

(117) Pekkurnaz, G.; Trinidad, J. C.; Wang, X.; Kong, D.; Schwarz, T. L. *Cell* **2014**, *158*, 54–68.

(118) Pozo, K. B. K.; Stephenson, F. A. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2011**, *1813*, 269–281.

(119) Martinez, M.; Renuse, S.; Kreimer, S.; O'Meally, R.; Natov, P.; Madugundu, A. K.; Nirujogi, R. S.; Tahir, R.; Cole, R.; Pandey, A.; Zachara, N. E. *Molecular & Cellular Proteomics* **2021**, *20*, 100069.

(120) Radzisheuskaya, A.; Shliaha, P. V.; Grinev, V. V.; Shlyueva, D.; Damhofer, H.; Koche, R.; Gorshkov, V.; Kovalchuk, S.; Zhan, Y.; Rodriguez, K. L.; Johnstone, A. L.; Keogh, M.-C.; Hendrickson, R. C.; Jensen, O. N.; Helin, K. *Molecular Cell* **2021**, S1097–2765(21)00098-8.

(121) Dong, H.; Liu, Z.; Wen, H. *Frontiers in Immunology* **2022**, *13*, 805018.

(122) Chang, Y.-H.; Weng, C.-L.; Lin, K.-I. *Journal of Biomedical Science* **2020**, *27*, 57.

(123) Merbl, Y.; Kirschner, M. W. *Proceedings of the National Academy of Sciences* **2009**, *106*, 2543–2548.

(124) Janetzko, J. C. OGT Catalyzes Three Distinct Post-translational Modifications via Glycosylation, Ph.D. Thesis, Boston, MA: Harvard University, 2017.

(125) Seo, H. G.; Kim, H. B.; Kang, M. J.; Ryum, J. H.; Yi, E. C.; Cho, J. W. *Scientific Reports* **2016**, *6*, 34614.

(126) Ramirez, D. H.; Yang, B.; D'Souza, A. K.; Shen, D.; Woo, C. M. *Analytical and Bioanalytical Chemistry* **2021**, *413*, 7385–7399.

(127) Decourcelle, A.; Loison, I.; Baldini, S.; Leprince, D.; Dehennaut, V. *Biochemical and Biophysical Research Communications* **2020**, *521*, 125–130.

(128) Yuan, H.; Mitchell, C. W.; Ferenbach, A. T.; Bonati, M. T.; Feresin, A.; Benke, P. J.; Tan, Q. K. G.; Aalten, D. M. F. *Stem Cell Reports* **2024**, *0*.

(129) Daou, S.; Mashtalir, N.; Hammond-Martel, I.; Pak, H.; Yu, H.; Sui, G.; Vogel, J. L.; Kristie, T. M.; Affar, E. B. *Proceedings of the National Academy of Sciences* **2011**, *108*, 2747–2752.

(130) Shu, X. E.; Mao, Y.; Jia, L.; Qian, S.-B. *Nature Chemical Biology* **2022**, *18*, 134–141.

(131) Kawahara, H.; Minami, R.; Yokota, N. *The Journal of Biochemistry* **2013**, *153*, 147–160.

(132) Lee, J.-G.; Ye, Y. *BioEssays* **2013**, *35*, 377–385.

(133) Krenciute, G.; Liu, S.; Yucer, N.; Shi, Y.; Ortiz, P.; Liu, Q.; Kim, B.-J.; Odejimi, A. O.; Leng, M.; Qin, J.; Wang, Y. *Journal of Biological Chemistry* **2013**, *288*, 20547–20557.

(134) Culver, J. The Role of Post-Translational Modifications in the Fate of Nascent Tail-Anchored Membrane Proteins, Ph.D. Thesis, New Haven, CT: Yale University, 2021.

(135) Payne, N. C.; Kalyakina, A. S.; Singh, K.; Tye, M. A.; Mazitschek, R. *Nature Chemical Biology* **2021**, *17*, 1168–1177.

(136) Potter, S. C. Impact of Structural Alterations on the Cellular Functions and Interactions of O-GlcNAc Transferase, Ph.D. Thesis, Boston, MA: Harvard University, 2023.

(137) Elias, J. E.; Gygi, S. P. *Nature Methods* **2007**, *4*, 207–214.

(138) Elias, J. E.; Gygi, S. P. *Methods in Molecular Biology* **2010**, *604*, 55–71.

(139) Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villén, J.; Haas, W.; Sowa, M. E.; Gygi, S. P. *Cell* **2010**, *143*, 1174–1189.

(140) DAVID Functional Annotation Bioinformatics Microarray Analysis, https://davidbioinformatics.nih.gov/.

(141) Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S.; Bork, P.; Jensen, L. J.; von Mering, C. *Nucleic Acids Research* **2023**, *51*, D638–D646.

(142) Lex, A.; Gehlenborg, N.; Strobelt, H.; Vuillemot, R.; Pfister, H. *IEEE Transactions on Visualization and Computer Graphics* **2014**, *20*, 1983–1992.

(143) Zhou, P.; Wang, Z.; Yuan, X.; Zhou, C.; Liu, L.; Wan, X.; Zhang, F.; Ding, X.; Wang, C.; Xiong, S.; Wang, Z.; Yuan, J.; Li, Q.; Zhang, Y. *Journal of Biological Chemistry* **2013**, *288*, 17532–17543.

(144) Emerson, F. J.; Chiu, C.; Lin, L. Y.; Riedel, C. G.; Zhu, M.; Lee, S. S. *Nature Communications* **2024**, *15*, 2320.

(145) Capotosti, F.; Hsieh, J. J.-D.; Herr, W. *Molecular and Cellular Biology* **2007**, *27*, 7063–7072.

(146) Gudkova, D.; Dergai, O.; Praz, V.; Herr, W. *Nucleic Acids Research* **2019**, *47*, 5792.

(147) Li, J.; Ebata, A.; Dong, Y.; Rizki, G.; Iwata, T.; Lee, S. S. *PLoS Biology* **2008**, *6*, e233.

(148) Kapuria, V.; Röhrig, U. F.; Waridel, P.; Lammers, F.; Borodkin, V. S.; van Aalten, D. M. F.; Zoete, V.; Herr, W. *Journal of Biological Chemistry* **2018**, *293*, 17754–17768.

(149) Janetzko, J.; Walker, S. *Journal of the American Chemical Society* **2017**, *139*, 3332–3335.

(150) Marino, S. M.; Gladyshev, V. N. *Journal of Molecular Biology* **2010**, *404*, 902.

(151) Chong, P. A.; Nosella, M. L.; Vanama, M.; Ruiz-Arduengo, R.; Forman-Kay, J. D. *Journal of Biological Chemistry* **2023**, *299*, 104629.

(152) Griffin, M. E.; Thompson, J. W.; Xiao, Y.; Sweredoski, M. J.; Aksenfeld, R. B.; Jensen, E. H.; Koldobskaya, Y.; Schacht, A. L.; Kim, T. D.; Choudhry, P.; Lomenick, B.; Garbis, S. D.; Moradian, A.; Hsieh-Wilson, L. C. *bioRxiv* **2023**, 2023.06.13.541482.

(153) PREDICTED: host cell factor 1 [Pseudopodoces humilis] - Protein - NCBI, `https://www.ncbi.nlm.nih.gov/protein/XP_014117713.1`.

(154) Gross, B. J.; Kraybill, B. C.; Walker, S. *Journal of the American Chemical Society* **2005**, *127*, 14588–14589.

(155) Liu, Y.; Nelson, Z. M.; Reda, A.; Fehl, C. *ACS chemical biology* **2022**, *17*, 2153–2164.

(156) Xu, B. et al. *Journal of Biological Chemistry* **2022**, *298*, 102115.

(157) Stephen, H. M.; Praissman, J. L.; Wells, L. *Journal of Proteome Research* **2021**, *20*, 1229–1242.

(158) Alteen, M. G.; Gros, C.; Meek, R. W.; Cardoso, D. A.; Busmann, J. A.; Sangouard, G.; Deen, M. C.; Tan, H.-Y.; Shen, D. L.; Russell, C. C.; Davies, G. J.; Robinson, P. J.; McCluskey, A.; Vocadlo, D. J. *Angewandte Chemie International Edition* **2020**, *59*, 9601–9609.

(159) Hu, C.-W.; Worth, M.; Fan, D.; Li, B.; Li, H.; Lu, L.; Zhong, X.; Lin, Z.; Wei, L.; Ge, Y.; Li, L.; Jiang, J. *Nature Chemical Biology* **2017**, *13*, 1267–1273.

(160) Bugg, T. D. H. In *Introduction to Enzyme and Coenzyme Chemistry*; John Wiley & Sons, Ltd: 2012, pp 26–49.

(161) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. *Journal of Biomolecular Screening* **1999**, *4*, 67–73.

(162) Los, G. V. et al. *ACS Chemical Biology* **2008**, *3*, 373–382.

(163) Cheng, Y.; Prusoff, W. H. *Biochemical Pharmacology* **1973**, *22*, 3099–3108.

(164) Wright, P. E.; Dyson, H. J. *Journal of Molecular Biology* **1999**, *293*, 321–331.

(165) Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E. *Pacific Symposium on Biocomputing* **1998**, 473–484.

(166) Huang, Y.; Liu, Z. *Chemistry: A European Journal* **2013**, *19*, 4462–4467.

(167) Lazar, T.; Tantos, A.; Tompa, P.; Schad, E. *Protein Science* **2022**, *31*, e4455.

(168) Bartlett, P. A.; Marlowe, C. K. *Biochemistry* **1983**, *22*, 4618–4624.

(169) Walsh, C. T.; Fisher, S. L.; Park, I. S.; Prahalad, M.; Wu, Z. *Chemistry & Biology* **1996**, *3*, 21–28.

(170) Bugg, T. D. H.; Wright, G. D.; Dutka-Malen, S.; Arthur, M.; Courvalin, P.; Walsh, C. T. *Biochemistry* **1991**, *30*, 10408–10415.

(171) Jencks, W. P. *Proceedings of the National Academy of Sciences* **1981**, *78*, 4046–4050.

(172) Sarkar, S.; Ballester, P.; Spektor, M.; Kataev, E. A. *Angewandte Chemie International Edition* **2023**, *62*, e202214705.

(173) Baron, U.; Bujard, H. In *Methods in Enzymology*, Thorner, J., Emr, S. D., Abelson, J. N., Eds.; Applications of Chimeric Genes and Hybrid Proteins - Part B: Cell Biology and Physiology, Vol. 327; Academic Press: 2000, pp 401–421.

(174) Nelson, M.; Hillen, W.; Greenwald, R. A., *Tetracyclines in Biology, Chemistry and Medicine*; Springer Science & Business Media: 2001.

(175) Gossen, M.; Bujard, H. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 5547–5551.

(176) Gossen, M.; Freundlieb, S.; Bender, G.; Müller, G.; Hillen, W.; Bujard, H. *Science* **1995**, *268*, 1766–1769.

(177) Das, A. T.; Tenenbaum, L.; Berkhout, B. *Current Gene Therapy* **2016**, *16*, 156–167.

(178) Nabet, B. et al. *Nature Chemical Biology* **2018**, *14*, 431–441.

(179) Nabet, B.; Ferguson, F. M.; Seong, B. K. A.; Kuljanin, M.; Leggett, A. L.; Mohardt, M. L.; Robichaud, A.; Conway, A. S.; Buckley, D. L.; Mancias, J. D.; Bradner, J. E.; Stegmaier, K.; Gray, N. S. *Nature Communications* **2020**, *11*, 4687.

(180) Bottens, R. A.; Yamada, T. *Cancers* **2022**, *14*, 5546.

(181) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. *Genome Research* **2004**, *14*, 1188–1190.

(182) Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P. C.; Timony, M. A.; Ottaviano, M.; Sliz, P. *eLife* **2013**, *2*.

(183) Shi, Y. *Cell* **2014**, *159*, 995–1014.

(184) Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W. *Expert Opinion on Drug Discovery* **2015**, *10*, 975–989.

(185) Smyth, M. S.; Martin, J. H. J. *Molecular Pathology* **2000**, *53*, 8–14.

(186) Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M. *Analytical Chemistry* **2021**, *93*, 1866–1879.

(187) Adrian, M.; Dubochet, J.; Lepault, J.; McDowall, A. W. *Nature* **1984**, *308*, 32–36.

(188) Milne, J. L. S.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E. H.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S. *The FEBS Journal* **2013**, *280*, 28–45.

(189) Wu, M.; Lander, G. C. *Current Opinion in Structural Biology* **2020**, *64*, 9–16.

(190) Lander, G. C.; Glaeser, R. M. *Biochemical Society Transactions* **2021**, *49*, 2287–2298.

(191) Zhang, K.; Wu, H.; Hoppe, N.; Manglik, A.; Cheng, Y. *Nature Communications* **2022**, *13*.

(192) Scheres, S. H. W. *Journal of Structural Biology* **2012**, *180*, 519–530.

(193) Bai, X.; Rajendra, E.; Yang, G.; Shi, Y.; Scheres, S. H. W. *eLife* **2015**, *4*, e11182.

(194) Jiang, W. How big should my particle box be? – The Jiang Lab, `https://jiang.bio.purdue.edu/how-big-should-my-particle-box-be/`, 2018.

(195) Punjani, A.; Zhang, H.; Fleet, D. J. *Nature Methods* **2020**, *17*, 1214–1221.

(196) Toader, B.; Sigworth, F. J.; Lederman, R. R. *Journal of Molecular Biology* **2023**, *435*, 168020.

(197) Tang, W. S.; Zhong, E. D.; Hanson, S. M.; Thiede, E. H.; Cossio, P. *Current Opinion in Structural Biology* **2023**, *81*, 102626.

(198) Case Study: End-to-end processing of an inactive GPCR (EMPIAR-10668) | CryoSPARC Guide, https://guide.cryosparc.com/processing-data/tutorials-and-case-studies/case-study-end-to-end-processing-of-an-inactive-gpcr-empiar-10668, 2024.

(199) Bepler, T.; Morin, A.; Rapp, M.; Brasch, J.; Shapiro, L.; Noble, A. J.; Berger, B. *Nature Methods* **2019**, *16*, 1153–1160.

(200) Tan, Y. Z.; Baldwin, P. R.; Davis, J. H.; Williamson, J. R.; Potter, C. S.; Carragher, B.; Lyumkis, D. *Nature Methods* **2017**, *14*, 793–796.

(201) Baldwin, P. R.; Lyumkis, D. *Progress in Biophysics and Molecular Biology* **2020**, *150*, 160–183.

(202) Baldwin, P. R.; Lyumkis, D. *Progress in Biophysics and Molecular Biology* **2021**, *160*, 53–65.

(203) Zivanov, J.; Nakane, T.; Scheres, S. H. W. *IUCrJ* **2019**, *6*, 5–17.

(204) Cianfrocco, M. A.; Kellogg, E. H. *Journal of Chemical Information and Modeling* **2020**, *60*, 2458–2469.

(205) Punjani, A.; Fleet, D. J. *Nature Methods* **2023**, *20*, 860–870.

(206) Jumper, J. et al. *Nature* **2021**, *596*, 583–589.

(207) Evans, R. et al. *bioRxiv* **2022**.

(208) Asarnow, D.; Palovcak, E.; Cheng, Y. asarnow/pyem: UCSF pyem v0.5, 2019.

(209) Nakane, T.; Kimanius, D.; Lindahl, E.; Scheres, S. H. *eLife* **2018**, *7*, e36861.

(210) Nakane, T.; Scheres, S. H. W. *Methods in Molecular Biology* **2021**, *2215*, 145–160.

(211) Mastronarde, D. N. *Journal of Structural Biology* **2005**, *152*, 36–51.

(212) Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A. *Nature Methods* **2017**, *14*, 290–296.

(213) Pintilie, G. D.; Zhang, J.; Goddard, T. D.; Chiu, W.; Gossard, D. C. *Journal of Structural Biology* **2010**, *170*, 427–438.

(214) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallographica* **2010**, *66*, 486–501.

(215) Zoltek, M.; Vázquez Maldonado, A. L.; Zhang, X.; Dadina, N.; Lesiak, L.; Schepartz, A. *ACS Central Science* **2024**, *10*, 860–870.

(216) Giudice, J.; Brauer, D. D.; Zoltek, M.; Vázquez Maldonado, A. L.; Kelly, M.; Schepartz, A. *bioRxiv* **2024**, 2024.04.05.588336.

(217) Lu, R.-M.; Hsu, H.-E.; Perez, S. J. L. P.; Kumari, M.; Chen, G.-H.; Hong, M.-H.; Lin, Y.-S.; Liu, C.-H.; Ko, S.-H.; Concio, C. A. P.; Su, Y.-J.; Chang, Y.-H.; Li, W.-S.; Wu, H.-C. *Journal of Biomedical Science* **2024**, *31*, 89.

(218) Zhang, K.; Wu, H.; Hoppe, N.; Manglik, A.; Cheng, Y. *Nature Communications* **2022**, *13*, 4366.

(219) Xu, J.; Chen, G.; Wang, H.; Cao, S.; Heng, J.; Deupi, X.; Du, Y.; Kobilka, B. K. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2414544121.

(220) Zhong, E. D.; Bepler, T.; Berger, B.; Davis, J. H. *Nature Methods* **2021**, *18*, 176–185.

(221) Li, Y.; Zhou, Y.; Yuan, J.; Ye, F.; Gu, Q. *Nature Methods* **2024**, *21*, 2318–2326.

(222) Ramakrishnan, P. *Journal of Biological Chemistry* **2024**, *300*.

heading=none