DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

OGT Catalyzes Three Distinct Post-Translational Modifications via Glycosylation

Citation
Janetzko, John Christopher. 2017. OGT Catalyzes Three Distinct Post-Translational Modifications via Glycosylation. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:41140981

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, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

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

Accessibility

OGT catalyzes three distinct post-translational modifications via glycosylation

A dissertation presented
by
John Christopher Janetzko
to
The Committee in Chemistry and Chemical Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Chemistry

Harvard University
Cambridge, Massachusetts

March 2017

© 2017 - John Christopher Janetzko

All rights reserved.

Dissertation Advisors

Suzanne Walker & Daniel Kahne

Author

John Christopher Janetzko

OGT catalyzes three distinct post-translational modifications via glycosylation

Abstract

It is estimated that the human genome encodes on the order of 20,000 genes, yet the number of proteins in the human proteome has been estimated to include more than one million proteins. One major source of diversity stems from protein post-translational modifications (PTMs). In metazoans, β-*O*-GlcNAcylation, the transfer of *N*-acetylglucosamine to serine and threonine (Ser/Thr) residues of numerous nuclear and cytoplasmic proteins, is a common PTM. Notably, all *O*-GlcNAc modifications are installed by a single essential protein, *O*-linked-β-*N*-acetylglucosamine transferase, OGT. Since its discovery over 30 years ago, scientists have been trying to understand the role of OGT in metazoan biology, with the vast majority of studies focused on the functional consequences of β-*O*-GlcNAcylation. This thesis elucidates the mechanisms of two other OGT-catalyzed modifications, amide backbone cleavage and isomerization of aspartate to isoaspartate, and lays the groundwork for addressing their physiological roles.

Host cell factor-1 (HCF-1) is a cell cycle regulatory protein that contains six 26-amino acid proteolytic repeats that must be cleaved for the protein to function properly. In 2011, OGT was implicated in HCF-1 cleavage, but it was unclear whether the process was autocatalytic and resulted from conformational changes related to Ser/Thr glycosylation or was directly catalyzed by OGT. Chapter two describes work demonstrating that OGT directly catalyzes HCF-1 cleavage using the same active site employed for Ser/Thr glycosylation. This study also showed that cleavage occurs between Cys9 and Glu10 of the proteolytic

repeat and resulted in a C-terminal product bearing a pyroglutamate residue.

Chapter three establishes the mechanism of HCF-1 cleavage using an OGT mutant that accumulates reaction intermediates. Cleavage proceeds via glutamate glycosylation followed by on-enzyme formation of an internal pyroglutamate, which hydrolyzes spontaneously after release from OGT.

Chapter four describes studies showing that aspartyl-containing model peptides can also be glycosylated by OGT, resulting in a peptide backbone rearrangement to produce isoaspartate through the intermediacy of a succinimide. Aspartate-to-isoaspartate isomerization in proteins occurs in cells, but was previously thought to be exclusively non-enzymatic. These studies suggest that enzymes capable of aspartate glycosylation, which include PARPs, may catalyze this transformation as well.

Contents

Abstract iii

Acknowledgments xvii

1 Introduction **1**

1.1 Post-translational modifications galore! 1

1.2 *O*-GlcNAc transferase and protein *O*-GlcNAcylation 3

1.3 Structure of *O*-GlcNAc transferase 8

1.4 Mechanism of Ser/Thr glycosylation by OGT 14

1.5 OGT is required for Host cell factor-1 proteolysis 20

1.6 Conclusions and focus of this dissertation 23

1.7 Resource: High-resolution mass spectrometry 23

2 HCF-1 is cleaved directly by OGT **30**

2.1 Introduction 31

2.2 Design of a repeat-only HCF-1 construct 34

2.3 Identifying requirements for HCF-1 cleavage 34

2.4 HCF-1 cleavage results in a pyroglutamate 37

2.5 Role of residues within the “CET” cleavage region 38

2.6 Structure of an HCF-1 repeat bound to OGT 42

2.7 Role of OGT active-site residues on cleavage and glycosylation 48
2.8 Structure-activity studies of the HCF-1/OGT interaction 51
2.9 Possible mechanisms for cleavage 57
2.10 Conclusions 62
2.11 Methods 64

3 Mechanism of HCF-1 cleavage **72**

3.1 Introduction 73
3.2 Development of an LC-MS-based HCF-1 cleavage assay 75
3.3 Detection of HCF-1 cleavage intermediates 79
3.4 Characterizing the anhydropeptide intermediate 86
3.5 Establishing the temporal sequence for intermediate formation 99
3.6 Characterizing the glycopeptide intermediate 100
3.7 Mechanism of HCF-1 cleavage by OGT 105
3.8 OGT (D554N) accelerates the glycosylation step of HCF-1 cleavage 109
3.9 The Nature of two anhydropeptide species 115
3.10 Conclusions 124
3.11 Methods 126

4 OGT catalyzes aspartate-to-isoaspartate isomerization via side chain glycosylation **135**

4.1 Introduction 136
4.2 A model substrate for aspartate glycosylation 137
4.3 Identifying and characterizing a succinimidyl peptide 140
4.4 Characterizing an isoaspartyl peptide 146
4.5 Conclusions 173
4.6 Methods 174

5 Conclusion **180**

A Supplementary Figures **185**

B Supplementary Tables **211**

C Glossary of Terms **222**

Bibliography **225**

List of Figures

1.1 Experimentally observed PTMs 2

1.2 OGT transfers GlcNAc from UDP-GlcNAc to Ser/Thr residues 3

1.3 The role of OGA and the HBP for protein *O*-GlcNAcylation 6

1.4 Schematic of human OGT domain architecture and X-ray structure of its TPRs 12

1.5 X-ray structure of the catalytic domain of human OGT 13

1.6 An extensive binding surface is formed by the nucleotide sugar 15

1.7 A generic GT-B Gtf mechanism 16

1.8 Active site of OGT 17

1.9 Reaction trajectory for Ser/Thr glycosylation by OGT 18

1.10 Two proposals for the operative proton transport mechanism in OGT . . . 20

1.11 Schematic of HCF-1 domain structure 21

1.12 Collision-induced MS/MS 25

1.13 Electron-capture dissociation MS/MS 28

2.1 Schematic of HCF-1, and HCF-1 constructs used for *in vitro* reactions . . . 33

2.2 Cleavage of HCF-3R requires UDP-GlcNAc 36

2.3 Cleavage of HCF-3R occurs between Cys9 and Glu10 and produces a pyroglutamate 38

2.4 The HCF-1 proteolytic repeat is highly conserved 39

2.5 Glutamate 10 is required for cleavage 40
2.6 Cysteine 9 is important, not essential for cleavage 41
2.7 Threonine 11 is not essential for cleavage 42
2.8 The threonine-rich region of HCF-1 binds in the TPRs of OGT 45
2.9 HCF-1 cleavage takes place in the glycosyltransferase active site of OGT . . 46
2.10 Glutamate 10 overlays with a glycosyl acceptor 47
2.11 OGT variants impaired in glycosylation but proficient at cleavage exist . . . 50
2.12 Cleavage ability of a repeat depends on its position within HCF-3R 52
2.13 Repeats shorter than HCF-3R are suitable substrates 53
2.14 TPR length in OGT is important for cleavage, but not glycosylation 55
2.15 A 32 amino acid HCF peptide is a suitable cleavage substrate 56
2.16 Possible cleavage mechanisms that proceed via a glycosylated glutamate . . 59
2.17 α_2-macroglobulin cleavage produces a pyroglutamate 61

3.1 Schematic of HCF-1 74
3.2 Proposed glycosyl glutamate linkage 74
3.3 Proposed ^{18}O transfer experiment 75
3.4 E10-32 undergoes cleavage 77
3.5 Plot of EICs for reaction of E10-32 with wild-type OGT 78
3.6 Two glycopeptides form during the cleavage reaction 79
3.7 OGT (D554N) is impaired for Ser/Thr glycosylation, but not HCF-1 cleavage 81
3.8 OGT (D554N) does not produce a glycopeptide byproduct 82
3.9 OGT (D554N) rapidly forms two intermediates 83
3.10 OGT (D554N) and wild-type have the same substrate requirements for HCF-1 cleavage 84

3.11 Wild-type OGT produces an anhydropeptide intermediate (intermediate 2) 85

3.12 Both peaks of Intermediate 2 have the same MS/MS fragmentation pattern 86

3.13 HCD MS/MS spectrum of Int-2 96

3.14 Possible Mechanisms, with or without cysteine participation 97

3.15 Plots of R9E10-32 cleavage time course with excess of OGT (D554N) 98

3.16 Temporal sequence of reaction intermediates 100

3.17 ECD MS/MS spectrum of Intermediate 1 101

3.18 The mechanism of HCF-1 cleavage proceeds via glycosylation on glutamate followed by decomposition to an internal pyroglutamate, which undergoes backbone hydrolysis 106

3.19 Plots for the reaction of Int-1 without OGT 107

3.20 Comparing the rate of substrate glycosylation for wild-type OGT and OGT (D554N) 110

3.21 Free-energy diagram for HCF-1 cleavage 111

3.22 Mutating Asp554 to asparagine does not affect the structure of the OGT active site 112

3.23 E10 of the HCF repeat is repelled by Asp554 114

3.24 Comparison of Int-2 formed on- and off-enzyme 116

3.25 Model dipeptides for imide vs. amide bond rotation 117

3.26 Computed structures for *s*-trans → *s*-cis interconversion of a model imide . 118

3.27 Computed structures for *s*-trans → *s*-cis interconversion of a model amide . 118

3.28 Model systems to probe epimerization of an internal pyroglutamate 119

3.29 Computed structures of Ala-Glu and Ala-pyroGlu model enolates 121

3.30 Computed structures of Cys-Glu and Cys-pyroGlu model enolates (thiolate form) 122
3.31 Computed structures of Cys-Glu and Cys-pyroGlu model enolates (thiol form) 123

4.1 Aspartate glycosylation, and related, substrates 137
4.2 A glycopeptide intermediate forms with D10-32, but not A10-32 139
4.3 Treatment of D10-32 with OGT leads to the formation of an anhydropeptide and a new species isobaric with starting material 141
4.4 Reaction of D10-32 requires both UDP-GlcNAc and OGT 142
4.5 Temporal sequence for species produced during reaction of D10-32 with OGT 143
4.6 Three possible mechanisms for anhydropeptide formation from a glycosyl aspartate 144
4.7 Cysteine 9 is not required for the formation of an anhydropeptide or an isobaric starting material species 145
4.8 Succinimide hydrolysis can produce isoaspartate or aspartate 146
4.9 Calculated isotope envelope for D10-32 (z = 4+) 147
4.10 Theoretical model for ^{18}O incorporation into D10-32 148
4.11 Observed ^{18}O incorporation into D10-32 149
4.12 ^{18}O incorporation requires active OGT 150
4.13 ^{18}O incorporation does not occur for E10-32 151
4.14 Fitting experimental ^{18}O incorporation to a theoretical model 152
4.15 HCD MS/MS reveals position of ^{18}O incorporation into D10-32 starting material isomer 162
4.16 CID MS/MS reveals position of ^{18}O incorporation into D10-32 starting material isomer 164

4.17 Rate of ^{18}O incorporation into D10-32 165
4.18 The S.M. isomer peak (D10-32) has the same LC retention time as an authentic isoD10-32 peptide 166
4.19 PIMT catalyzes isoaspartate methylation, producing a succinimide that hydrolyzes to a mixture of aspartate and isoaspartate 167
4.20 PIMT methylates isoaspartate, but not aspartate 168
4.21 The starting material isomer is an isoaspartate-containing peptide 169
4.22 isoD10-32 is a poor substrate for OGT 170
4.23 Comparison of glycosyl acceptor abilities 171

A.1 Role of the nucleotide sugar in E10-32 cleavage 186
A.2 Relative rates of cleavage for C9X peptides 187
A.3 Primary sequence of HCF proteolytic repeats in *Homo sapiens* 188
A.4 Time course for E10-32 cleavage by wild-type OGT 189
A.5 Time course for E10-32 cleavage by OGT (D554N), 100 minutes 191
A.6 Graphical representations of 100 minute time course for cleavage of E10-32 with stoichiometric OGT (D554N) 193
A.7 Time course for E10-32 cleavage by OGT (D554N), 3 minutes 194
A.8 Graphical representations of 3 minute time course for cleavage of E10-32 with excess OGT (D554N) 196
A.9 Time course for E10-32 cleavage by OGT (D554N), 30 minutes 198
A.10 Graphical representations of 30 minute time course for cleavage of E10-32 with excess OGT (D554N) 200
A.11 Graphical representations of 14 minute time course for cleavage of E10-32 with excess OGT (D554N) 201

A.12 Intermediate 2 is a dehydrated form of E10-32 (A) 202
A.13 Different HCF peptides produce the same intermediates with both wild-type OGT and OGT (D554N) 204
A.14 ECD MS/MS spectrum of Intermediate 1, with zooms 206
A.15 The GlcNAc moiety in Intermediate 1 is located at E11, not T12 208
A.16 Errors in MS/MS fragment ions for A, Intermediate 1 and Intermediate 2 . 209
A.17 Calibration curve for SID quantitation of N-term 210

List of Tables

3.1 List of y ions for Int-2 by HCD 87

3.2 List of singly- and triply-charged b ions for Int-2 by HCD 88

3.3 List of doubly-charged b ions for Int-2 by HCD 89

3.4 List of y ions for Int-2 by CID 90

3.5 List of b ions for Int-2 by CID 91

3.6 List of dehydrated b ions for Int-2 by CID 92

3.7 List of y ions for E10-32 by HCD 93

3.8 List of b ions for E10-32 by HCD 94

3.9 List of additional b ions for E10-32 by HCD 95

3.10 List of singly-charged c ions for Intermediate-1 by ECD 102

3.11 List of multiply-charged c ions for Int-1 by ECD 103

3.12 List of singly-charged z ions for Int-1 by ECD 104

3.13 List of doubly-charged z ions for Int-1 by ECD 105

3.14 Summary of calculated pK_as 120

4.1 List of y ions for isoD10-32 by HCD 153

4.2 List of singly-charged b ions for isoD10-32 by HCD 154

4.3 List of multiply-charged b ions for isoD10-32 by HCD 155

4.4 List of y ions for isoD10-32 by CID 156

4.5 List of b ions for isoD10-32 by CID 157

4.6 List of dehydrated b ions for isoD10-32 by CID 158

B.1 List of peptides employed in these studies 213

B.2 HCF-3R-XXX constructs 214

B.3 HCF-3R-EAA (C9X) constructs 215

B.4 HCF-3R-EAA (T11X) constructs 216

B.5 HCF-3R-XAA constructs 217

B.6 HCF-3R truncation constructs 218

B.7 List of OGT constructs employed in these studies 219

B.8 LC gradient method (q-Exactive) 220

B.9 LC gradient method (q-ToF) 220

B.10 Data collection and refinement statistics for hOGT4.5 (D554N):UDP:CKII glycopeptide ternary structure 221

C.1 List of abbreviated terms 224

Citations to previously published work

Chapter 1 has been adapted from the following publication: [108]; Chapters 2 and 3 include results that appeared in the following publications: [107, 141]; Chapter 4 appears in part in the following publication: [109]

> "The Making of a Sweet Modification: Structure and Function of *O*-GlcNAc Transferase" John Janetzko and Suzanne Walker. *The Journal of Biological Chemistry*. Volume 289, Number 50: 34424–34432. 2014.
>
> "HCF-1 is cleaved in the active site of *O*-GlcNAc transferase" Michael B. Lazarus, Jiaoyang Jiang, Vaibhav Kapuria, Tanja Bhuiyan, John Janetzko, Wesley F. Zandberg, David J. Vocadlo, Winship Herr and Suzanne Walker. *Science*. Volume 342, Number 6163: 1235–1239. 2013.
>
> "How the glycosyltransferase OGT catalyzes amide bond cleavage" John Janetzko, Sunia A. Trauger, Michael B. Lazarus and Suzanne Walker *Nature Chemical Biology*. Volume 12, Number 11: 892–893. 2016.
>
> "Aspartate Glycosylation Triggers Isomerization to Isoaspartate" John Janetzko and Suzanne Walker *Journal of the American Chemical Society*. Volume 139, Number 9: 3332–3335. 2017.

In addition to the above mentioned publications, I contributed to two other publications over the course of my doctorate which did not make it into this dissertation:[77, 191]

> "A Small Molecule That Inhibits OGT Activity in Cells" Rodrigo F. Ortiz-Meoz, Jiaoyang Jiang, Michael B. Lazarus, Marina Orman, John Janetzko, Chenguang Fan, Damien Y. Duveau, Zhi-Wei Tan, Craig J. Thomas and Suzanne Walker *ACS Chemical Biology*. Volume 10, Number 6: 1392–1397. 2015.
>
> "Development and Characterization of Potent Cyclic Acyldepsipeptide Analogues with Increased Antimicrobial Activity" Jordan D. Goodreid, John Janetzko, John P. Santa Maria Jr., Keith S. Wong, Elisa, Leung, Bryan T. Eger, Steve Bryson, Emil F. Pai, Scott D. Gray-Owen, Suzanne Walker, Walid A. Houry and Robert A. Batey *Journal of Medicinal Chemistry*. Volume 59, Number 2: 624–646. 2016.

Acknowledgments

"And that's when I know it's over.
As soon as you start thinking about the beginning,
it's the end." – Junot Díaz (*This Is How You Lose Her*)

Harvard is a place with many great scientists and mentors, and I have been privileged to work with two of them.

I will never forget my first meeting with Dan Kahne. That meeting took almost a month of planning; when we did finally meet, I told you I had never expressed a protein before and you simply smiled and said (paraphrase): "that's okay, you'll learn. It'll be good." Little did I know that these themes, a general disorganization, and a blind confidence in my abilities, would foreshadow the years to come. Dan, you took me in and gave me an opportunity to learn and grow as a scientist (and a place to work). You also introduced me to Suzanne, who would become my main scientific advisor. For these things I will always be grateful. So, to Dan, for our runs at the Harvard stadium, our mornings at *4a* coffee, and your endearing text messages.

To Suzanne Walker, who asked if she was boring me that first time we met because I yawned. For telling me about OGT, even though I was too naïve to know what I really wanted. For giving me space to learn, and to fail. For being patient with me. For teaching me to "focus". For being an excellent writer, and tolerating my mediocrity. For being one of the *best* scientists I know, and impressing me with her breadth and depth of knowledge. For meetings that last as long as the walk to her car, or all day. For showing me that family doesn't have to come at the expense of success. For letting me work out with Dan, as long as I didn't kill him.

So, to Suzanne, for making me the scientist I am, and for drinking scotch while writing grants in your office.

To my thesis committee, Stuart Schreiber and Tim Mitchison: Thank you for taking time out of your busy schedules for the GAC meetings, the email exchanges and the sporadic one-on-one meetings. Stuart, thank you for teaching me to keep the big picture in mind, and to always pursue the most important problems. Tim, thank you for your unfettered enthusiasm, and for approaching science with vigor and healthy skepticism.

To those who *actually* keep the Kahne and Walker labs running: Helen Corriero and Mike Quinn. I am so grateful to you both for helping me navigate university policies, schedule meetings, and ensuring that all of the things entailed in research, beyond running experiments, got done. Helen, you have on many occasions referred to yourself as my "lab mother", thank you for always looking out for me.

To members of the Kahne and Walker labs, past and present, thank you for giving me a home away from home, and making work seem like play. In particular: To Mike Lazarus and Jiaoyang Jiang, thank you for laying the foundation of what would become my Ph.D. work. To Mike in particular, for showing me the ropes when I first joined the lab, and for continuing to offer helpful input on my projects even after leaving the lab. For his patience, an for organized chaos. To Sara Martin, for invaluable lessons in humility, and for being one of my best friends in the lab. Poppyseed is lucky to have such a great mother. To Zebulon Levine, for being a fantastic sounding-board for scientific ideas, and for incoherent gels and the New England Patriots. To Rodrigo Ortiz-Meoz, for teaching me to run western blots, and tailgating Arcade Fire. To John Santa Maria Jr., for teaching me how to do bacterial MICs, and for all the volleyball. To Christian Gampe, who mentored me during my rotation project in the Kahne lab, for your encouragement, and for the struggles of protein folding.

To Goran Malojčić, for teaching me to solve protein crystal structures, and for having the best laugh. To Tristan Owens, David Westwood and Michael Mandler, for making trips to APS a little more enjoyable. To Joe Wzorek, for his incredible knowledge of chemistry, and his spirit on the volleyball court. To Matt Lebar, for our weekday gym routine, and for always bringing the party. To Fred Rubino, for the late-night conversations, and DIY western blotting reagents. To Mingyu Xue, for always knowing the best restaurants. To Katie Schaefer, for proving that great scientists can be well dressed. To Heidi Morris, for sharing my Canadian spirit. To McKay Wood ($V_{\max}$!), for help with the Walker lab q-ToF. To David Sherman, whose work ethic may only be rivaled by his sense of humor. To Janine May, who always had time to listen. To Dorothee Andres, for helping me through difficult times. To Alex George, for going out of the way to help a friend, and for his beautiful pictures. To Eileen Moison, for tolerating a locker room worth of gym clothes hanging around my (and therefore her) desk. To Benjamin Gross, for those forgotten tubes of His-Nup62. To the other members of #teamOGT: Fred Moss, Zhi Wei Tan, Harri Itkonen, Michael Melicher, Chenguang Fan and Marina Orman, for making all the long meetings a little easier to bear.

To those who I've had the opportunity to collaborate with over the years: David Vocadlo and Matt Alteen (SFU), for fruitful discussions and sharing peptides and nucleotide sugars with me. Piotr Sliz (HMS), for helping me with structure refinement. Winship Herr, Vaibhav Kapuria and Tanja Bhuiyan (UNIL), for our joint work on the HCF-1 cleavage project.

To my students, both at Harvard College and through the Canadian Chemistry Olympiad program, for always challenging me to be a better teacher.

To the Natural Sciences and Engineering Research Council of Canada (NSERC), for

generous fellowships throughout my scientific training thus far. In particular, thank you for sending Canadian dollars over the border, in the form of Postgraduate Scholarships.

To those who were instrumental in collecting macromolecular structure data. To the dedicated scientific staff of NE-CAT (APS Sector 24), in particular: Frank Murphy, Surajit Banerjee, Igor Kourinov, David Neau, Kay Perry, Jonathan Schuermann, and Narayanasami Sukumar. Also to Jay Nix, at ALS beamline 4.2.2. For all of their hard work, and for patiently helping me with data collection and analysis.

To those who helped me with/taught me about mass spectrometry: Sunia Trauger, Bogdan Budnik, Charles Vidoudez, John Neveu, and Christopher Thompson. In particular, to Sunia, much of this work would not have been possible without you. There are very few people I have met who are as helpful, accommodating and enthusiastic about their work as you are. Thank you for making so many days in the basement of Northwest labs a little brighter.

To those who inspired me to pursue research: Fred Fontaine, for Crabtree's *Organometallics*, and for encouraging me, even though I went to "the dark side" of chemistry. To Stanislaw Skonieczny, for taking me to Moscow in 2007, and eventually making an honest experimentalist out of me. To Andy Dicks, for his passion for teaching and for all the conferences, Olympiad camps, and late nights drinking rakı in Ankara. To Datong Song, for giving me my first chance in a research lab. To Farhad Nowrouzi, for treating me like a brother, his mentorship, and teaching me almost everything I know about organic chemistry. To Rob Batey, for believing in me as an undergraduate and for holding me to higher standards than I deserved.

To my friends, outside of the lab, you made working towards this Ph.D. possible: To Kevin Ryans, one of my closest friends; over the past ten years you have *always* been there

for me. Though we don't speak daily (or even weekly), I know you are never more than a phone call or text away. You have been a voice of reason in darker times, but also one of encouragement during my successes. To Hunter Snyder, for 4 am bike rides to CRI, the HBS sauna, and my terrible bowing of our double. I'm amazed every day how we became such good friends so quickly. To Tim Ramadhar, for cell phone holsters, delicious scotch, always making last call and for following me to Boston. To Dmitry Pichugin, for teaching me to climb, always knowing what to do when it comes to NMR, and for many unforgettable camping trips. To Jordan Goodreid, for our collaboration years after I left the Batey lab, and for running columns with your shirt off. To Anna Shneidman, for her love of walks and poor planning. To Mariah Hanley, for Russell House Tavern. To Juan Pablo Maianti, for ~~teaching~~ tolerating me in CHEM115, our trips to the gym, and for never forgetting to do planks. To Alan Nhieu, for reminding me what happens without pecs. To Adam Kafal, for his sense of adventure, and samba music. To Phil Miles, for always checking in. To Will Menegas, for only the best romantic advice. To Helen Yu, for her tremendous excitement for science.

To those who I've lived with over the years: Brian Lee and Jacob Sanders were my roommates for five years, spanning three addresses. It was through our discussions first year that I decided to join the Kahne and Walker labs, and for that I am incredibly grateful. To Brian, for his remarkable intellectual curiosity, and for clothes shopping on Amazon. To Jacob, one of the best "fake" organic chemists I know, for his characteristic walk, great conversations, and the occasional dirty dish. To the others who I have lived with: Hyungmok (Jim) Son, for teaching me to wax my skis; Stanislas Chaillou, for a healthy skepticism of Wikipedia; Michael Fryar, Mike Rouleau and Kyle Yu for all our shared meals.

To those I have rowed with: In rowing I found balance and structure. I learned to

push beyond my boundaries, and work harder than I previously thought possible. I do not believe I would be here today without rowing. To Dudley House rowing, where it all began. To Eliza Gettel, for organizing the Dudley crew, and (for some reason) using me as a model of erg technique. To everyone in the men's B2 boat (2015), and everyone in the winning men's A boat (2016), for more sunrises than I could ever dream of, and drinking champagne from silver trophies in Dudley House. To Weld boathouse, and Dan Boyne, thank you for opening your doors to me, for that beautiful white Filippi, and for racing in the Head of the Charles. In particular: To Joe Copeland, for those pesky bicycle accidents. To Glenn Beauchemin and Anna Harris, for teaching me to scull. To Community Rowing Inc. (CRI), for being a welcoming place filled with wonderful people, and for an armada of red Hudson boats. To the competitive men's team (at CRI), for giving me a place to call home six mornings a week and making me a half decent sweep rower. In particular: Emily Dwyer, for her coxing and (pro bono) coaching. To Ryan Andrews, for training with me in the pair, and not killing me. To Richard Martinez, for his love of tacos, and distaste for early mornings. To Sam Ames, for rowing in the double, and his love of bandannas. To Jonathan Bernblum, for his advice on love. To John Owens, for few words and a staggering resolve. To Nick Tsantes, for winning heats and skipping final races [for a Blink-182 concert].

To Judith Fox, Jon Bell and Hillary Bell for welcoming me into their family.

To my [soon-to-be] wife, Caitlin Bell: Thank you for giving me more than I deserve. For diligently reading this thesis; for "bro-jenga"; for keeping me out of jail when we go to see Taylor Swift; for our bright future together. I love you so much.

And finally, to my parents, Irene and Roy Janetzko: I am grateful for your unconditional love and support. The sacrifices you have made throughout your lives, and your hard work, gave me the chance to pursue my dreams. For that I will always be appreciative.

for my parents

Chapter 1

Introduction

> "It's a magical world, Hobbes, ol' buddy ...
> Let's go exploring!"
>
> – Calvin (Calvin and Hobbes, Bill Watterson)

A portion of this chapter originally appeared as [108]:

"The Making of a Sweet Modification: Structure and Function of *O*-GlcNAc Transferase" **John Janetzko** and Suzanne Walker. *The Journal of Biological Chemistry.* Volume 289, Number 50: 34424–34432. 2014.

1.1 Post-translational modifications galore!

The human proteome is vastly more complex than the human genome [252]: while the human genome encodes on the order of 20,000 genes [103], the human proteome has been estimated to include more than one million proteins [188]. A major source of diversity in going from genes to proteins is through protein post-translational modifications (PTMs). Two of the most widely known PTMs include phosphorylation (on serine, threonine and ty-

rosine) and glycosylation (both O-linked, occurring on serine and threonine, and N-linked, occurring on asparagine). It has been estimated that more than half of all proteins are glycosylated [126, 194]. Given the tremendous scale and the proteome, and the diversity of protein PTMs (Figure 1.1), it should not be surprising that until recently PTMs were generally characterized on a single-protein basis [53], and often with the exact site of modification remaining unknown [42].

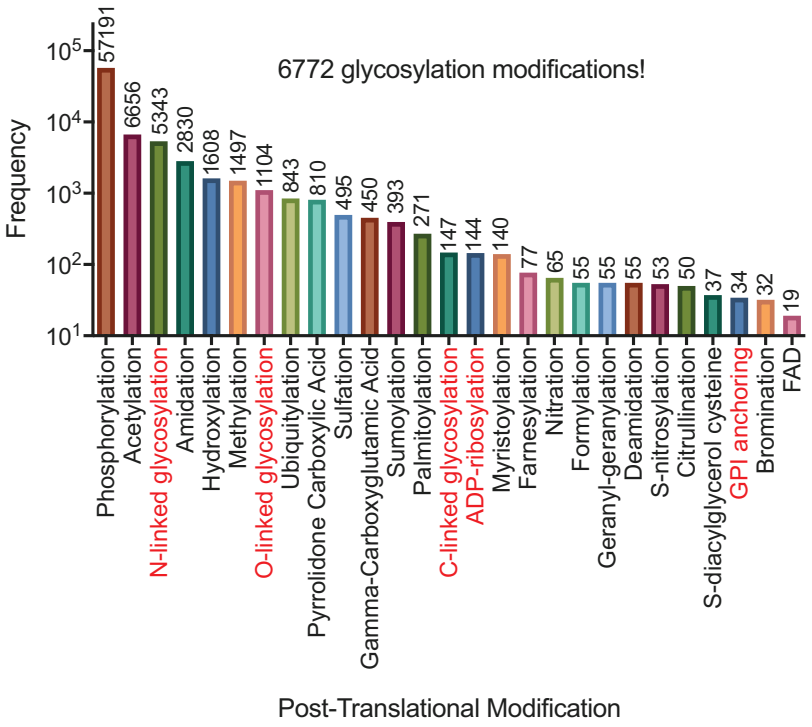


Figure 1.1: Experimentally observed PTMs. There were 87,308 experimentally observed PTMs for 530,264 proteins in the July 2011 release of the Swiss-Prot database. This figure was reproduced from reference: [126].

1.2 *O*-GlcNAc transferase and protein *O*-GlcNAcylation

Until the mid 1980s, many believed that glycoconjugates were exclusively derived from the secretory pathway [239]. Thirty years ago, when Torres and Hart made the unexpected discovery that many intracellular proteins are modified with *O*-GlcNAc [97, 230], this set in motion a change in paradigm for localization of glycoconjugates. Thirteen years later the enzyme responsible for this post-translational modification was finally cloned [130, 160]. Unlike the many glycosyltransferases (Gtfs) that reside in the endoplasmic reticulum (ER) or Golgi apparatus, *O*-linked-β-*N*-acetylglucosamine transferase (OGT) localizes to both the nucleus and cytoplasm of metazoan cells [108, 239]. OGT catalyzes transfer of a GlcNAc moiety from UDP-GlcNAc to specific serine and threonine residues in its many cellular targets (Figure 1.2)(henceforth termed *O*-GlcNAcylation). Remarkably, it is the sole enzyme responsible for this abundant PTM [86, 130, 160]. In mammalian cells, OGT is essential throughout life; however, the reason for this essentially is currently unknown [149].

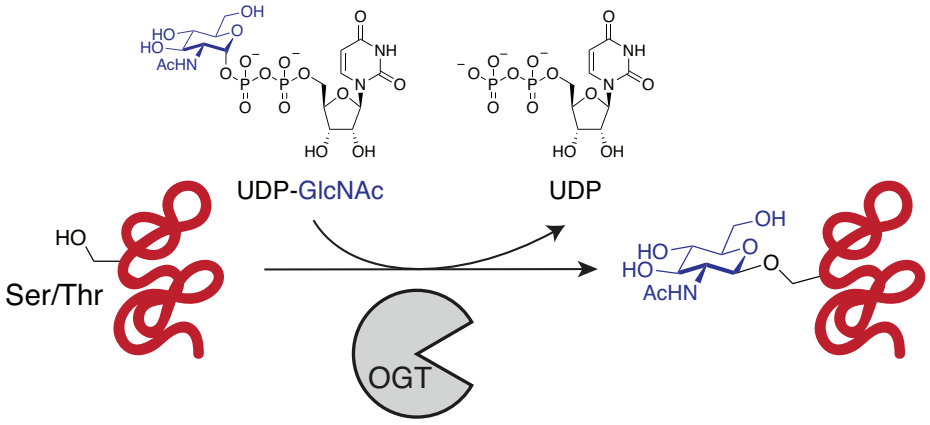


Figure 1.2: OGT transfers GlcNAc from UDP-GlcNAc to Ser/Thr residues.

The co-substrate for protein *O*-GlcNAcylation [87], UDP-GlcNAc bears intimate links to cellular metabolism [149, 216, 280]. UDP-GlcNAc is produced through the hexosamine

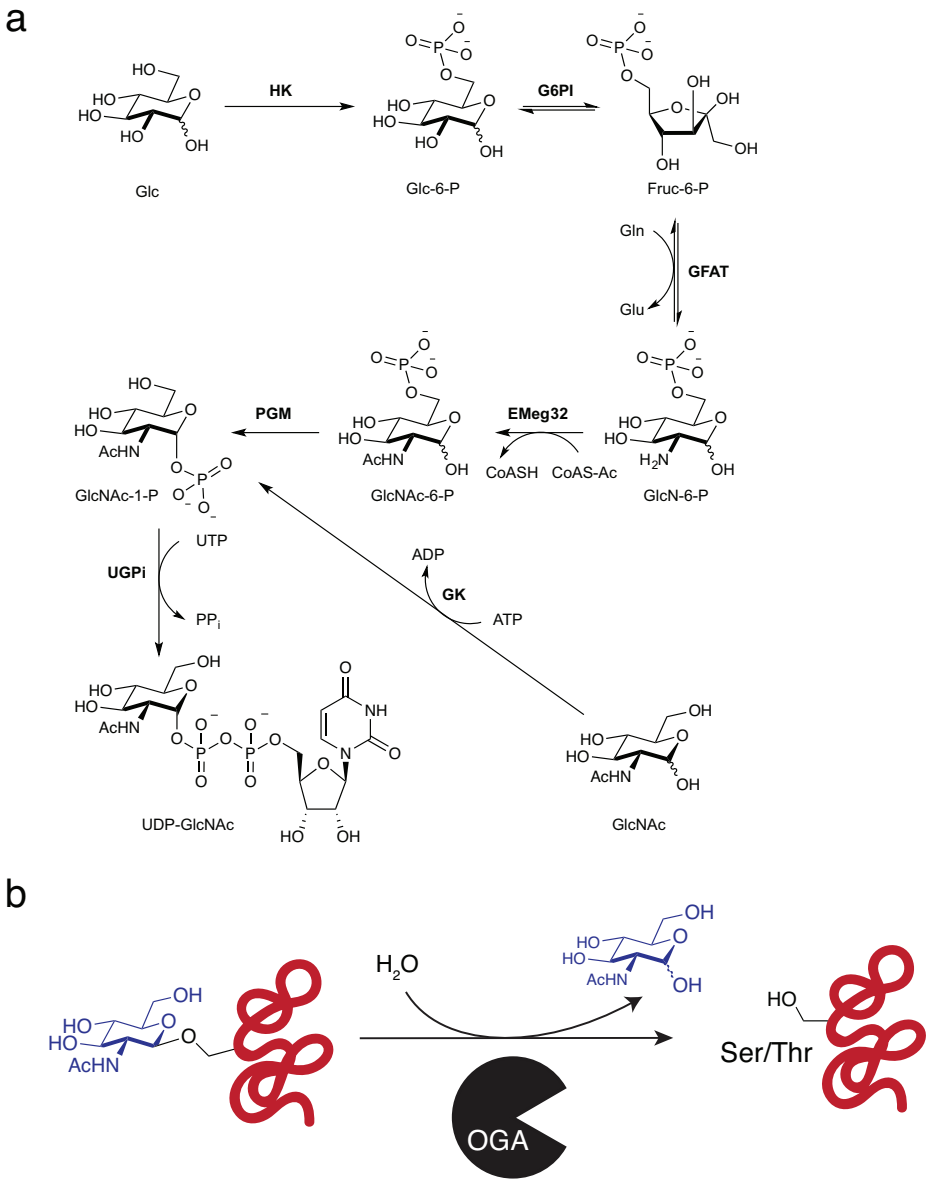
biosynthetic pathway (HBP, Figure 1.3a). About 2% of cellular glucose enters the HBP, via the rate-limiting enzyme glutamine-fructose amidotransferase (GFAT) [149, 232]. UDP-GlcNAc combines elements of glucose, glutamine, fatty acid metabolism and nucleotide biosynthesis, and the cellular concentration of UDP-GlcNAc has been shown to vary depending on the flux through these pathways [94, 168, 176, 249].

A glycosidase, *O*-GlcNAcase (OGA), capable of hydrolyzing the β-*O*-GlcNAc linkage, is also ubiquitous within the nucleus and cytoplasm of metazoan cells [4]. As a result, protein *O*-GlcNAcylation is a dynamic PTM, able to turnover multiple times over a protein's lifetime (Figure 1.3b) [26, 34, 93, 130, 206]. The dynamic nature of this modification in many ways makes it analogous to protein phosphorylation. However, unlike protein phosphorylation, for which there are 428 Ser/Thr kinases in the human genome [218], only OGT is responsible for all cellular *O*-GlcNAc [94].

Perhaps it is not surprising, given the large and diverse set of protein targets modified by OGT, that protein *O*-GlcNAcylation has been implicated in regulating gene expression [67, 90, 154, 192], cellular stress responses [119, 268, 270], metabolism [51, 207, 208, 266], and proteostasis [82, 88, 273, 278] in response to nutrient levels [20, 94, 95, 269], to name just a few [149]. Furthermore, given the myriad processes affected by OGT, there have been numerous associations between OGT and human disease. Perturbations in OGT expression level and/or global protein *O*-GlcNAcylation have been observed in cancer [164], diabetes [94, 237], Alzheimer's disease [267], and cardiovascular disease [166]. Whether OGT will prove to be a meaningful therapeutic target, however, remains to be seen [191].

This chapter will focus on structural and biochemical studies of OGT, including a discussion of its role in HCF-1 cleavage. Numerous comprehensive reviews of these topics have been published, including some of the earlier OGT work not discussed here. For additional reading, please see: [71, 93, 102, 170, 243].

Figure 1.3: *(On the following page)*The role of OGA and the HBP for protein *O*-GlcNAcylation. (a) The hexosamine biosynthetic pathway. Glucose (Glc) is converted to glucose-6-phosphate (Glc-6-P) by hexokinase (HK), which is isomerized to fructose-6-phosphate (Fruc-6-P) by glucose-6-phosphate isomerase (G6PI). Fruc-6-P is converted to glucosamine-6-phosphate (GlcN-6-P) by glutamine:fructose-6-phosphate amiotransferase (GFAT). GlcN-6-P is converted to *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P) by EMeg32, using acetyl coenzyme A (CoASAc); CoASH is free coenzyme A. GlcNAc-6-P is isomerized by phosphoGlcNAc mutase (PGM) to GlcNAc-1-P. Finally, UDP-GlcNAc pyrophosphorylase (UGPi) transfers uridine diphosphate from uridine triphosphate (UTP) to produce UDP-GlcNAc and inorganic pyrophosphate (PPi). The enzyme GlcNAc kinase (GK) is capable of phosphorylating GlcNAc with adenosine triphosphate (ATP), allowing for GlcNAc to be recycled [149]. (b) OGA is a hydrolase that removes β-*O*-GlcNAc linkages from proteins.


a
HK
G6PI
Glc
Glc-6-P
Fruc-6-P
Gln
Glu
GFAT
PGM
EMeg32
CoASH
CoAS-Ac
GlcNAc-1-P
GlcNAc-6-P
GlcN-6-P
UTP
UGPi
PPi
ADP
GK
ATP
UDP-GlcNAc
GlcNAc
b
H2O
OGA
Ser/Thr

1.3 Structure of *O*-GlcNAc transferase

The *OGT* gene, located on the X chromosome, encodes a protein with two major domains [91]. The N-terminal domain comprises a cluster of repeating motifs called tetratricopeptide repeats (TPRs), and the C-terminal domain bears the glycosyltransferase activity [130, 160] (Figure 1.4a). TPRs are degenerate 34-amino acid motifs that typically fold into paired anti-parallel α-helices [3, 137, 272]. Often found in tandem arrays, TPRs are known for their role in mediating protein-protein interactions [3, 18, 48]. Evidence implicated the TPRs of OGT in mediating interactions with protein substrates, as well as putative adapters and proteins shown to affect its enzymatic activity [31, 33, 104, 105, 131, 138, 172, 199, 205, 265, 208, 274]. There are three physiological isoforms of OGT, differing primarily in the number of N-terminal TPRs [91, 159, 187]. The longest isoform, with 13.5 TPRs, is referred to as nuclear and cytoplasmic OGT (ncOGT). A shorter form, with 9 TPRs and bearing a mitochondrial targeting sequence, is localized to the inner mitochondrial membrane (mOGT). The smallest of the three, sOGT, has only 2.5 TPRs and is perhaps the least studied. There have been conflicting reports as to whether mOGT has catalytic activity *in vivo* [159, 209, 231], and it has been speculated that mOGT and sOGT may somehow be involved in apoptosis [63, 220].

The first structure of a segment of OGT was reported in 2004. This structure comprised 11.5 TPRs from the N-terminus of the human protein and showed that the TPRs assemble into an elongated right-handed superhelix (Figure 1.4b) [111]. Several features of the TPR structure are noteworthy. In the asymmetric unit, two TPR molecules pack as a homodimer, with the interface centered on TPRs 6 and 7 [111]. Previous studies had suggested that OGT exists as an oligomer in solution and the TPRs were implicated in mediating its oligomerization [86, 131]. The TPR crystal structure suggested that OGT may

exist as a dimer, and consistent with this, mutating the residues at the TPR dimer interface afforded a lower order species. Because mutations that disrupted this interface did not substantially affect enzymatic activity *in vitro*, it remains unclear what role oligomerization serves for OGT function. The TPR structure also suggested a high degree of conformational flexibility, with one monomer bent by 40° between TPRs 9 and 10 [111]. While this deformation could be due to crystal-induced unfolding, the authors suggested the conformational flexibility might have implications for substrate recognition. Finally, phylogenetic comparisons showed strong conservation in solvent-exposed residues lining the interior of the TPRs, particularly several asparagine residues near the center of the cavity, and suggested an unexpected similarity to armadillo repeat-containing proteins, such as importin α and β-catenin [111]. These asparagine residues are highly conserved across OGT homologs, suggesting an important role in OGT function. The similarity to proteins known to recognize multiple related ligands is striking; for example: β-catenin, which binds to various cadherins [101], PEX5, which recognizes a peroxisomal targeting signal (PTS1) [68], karyopherin (importin) α, which recognizes a nuclear localization sequence [39], and the histocompatibility protein HLA-DR1, which recognizes pathogen-derived peptides [223], all employ asparagine-mediated peptide backbone contacts as a crucial component of their ligand binding. In these cases, there are two major reasons for this; backbone-binding interactions provide sequence-independent binding energy and predictably splay the side chains of the ligand, which in turn allows for sequence-selective recognition at defined sites along the polypeptide chain. In the context of OGT this might imply that a similar mode of substrate selection is employed for the engagement of adapter proteins or substrates.

Additionally, structures of a bacterial homologue of OGT from *Xanthomonas campestris*, xcOGT [36, 54, 171], and the catalytic domain of human OGT bearing 4.5 TPRs (hOGT4.5;

Figure 1.5) have been reported [142]. To date, structures of hOGT4.5 have been solved with UDP, UDP-GlcNAc, the hydrolysis-resistant analog UDP-5SGlcNAc [72], and several different peptide and glycopeptide substrates bound. Hence, binary as well as ternary substrate-like and product complexes are available [107, 110, 142, 140, 141, 198, 213]. A structure containing a bisubstrate analog has also been reported [21]. Taken together, these structures offer perhaps the most complete structural picture yet obtained for any Gtf. By combining the TPR structure with the catalytic domain structure, a model for human ncOGT, containing all 13.5 TPRs, can be created; in this model the TPRs form two complete superhelical turns and extend upwards from the active site.

The structures of OGT have revealed many important features. As expected, the catalytic domain of OGT possesses a GT-B fold, reminiscent of that found in MurG (CAZy family GT-28), GtfB (CAZy family GT-1), and many other Gtfs [23, 41, 85, 136, 158, 259]. GT-B Gtfs possess a catalytic domain comprising two lobes that each adopt a Rossmann-like fold [203]. OGT differs from other GT-B Gtfs in several key respects, however. The N-terminal catalytic domain bears two additional helices that are crucial components of the active site [142]. Moreover, it possesses an about 120 amino acid insertion (termed the intervening domain; Int-D) between its two catalytic lobes [142]. This insertion forms a domain that packs exclusively against the C-terminal catalytic domain. Although its structure is known (Figure 1.5), its function remains mysterious. The intervening domain does however contain a large number of surface-exposed basic residues, which might suggest a role in mediating interactions with anionic macromolecules or membrane interfaces [108, 142, 149]. It has been suggested that OGT is recruited to membranes through binding to phosphatidylinositol (3,4,5)-triphosphate (PIP3); [120, 264] however, mutational studies of the residues implicated in binding PIP3 are, at present, ambiguous [120, 142, 264], warranting further

study. The peptide substrate binds over the UDP-GlcNAc, and the peptide binding site spans a groove formed between the two catalytic lobes and the TPR domain (Figure 1.5). The extended conformation of the peptide substrate likely explains the preference for proline and β-branched amino acids near the site of glycosylation [142, 198]. There is an interaction between the α-phosphate of the nucleotide-sugar and the amide N-H of the acceptor serine that is proposed to play a role in peptide binding to OGT [142] and may further serve to orient the substrate for reaction (Figures 1.6 and 1.8) [140]. While several OGT residues contact the backbone of the peptide substrate, few make specific contacts to the peptide side chains, consistent with the lack of a clear consensus sequence [142, 155, 198]. Peptide binding is accompanied by widening of the cleft between the catalytic domain and the TPRs, and molecular dynamics (MD) simulations have suggested that the TPRs are able to pivot about a hinge region between TPRs 12 and 13, giving rise to open and closed states for the active site [142]. Such molecular motion would be crucial for accommodating larger substrates, such as loops in folded proteins. Moreover, the flexibility is consistent with that seen in the TPR crystal structure, and with the previous suggestion that the TPRs mediate substrate binding [111].

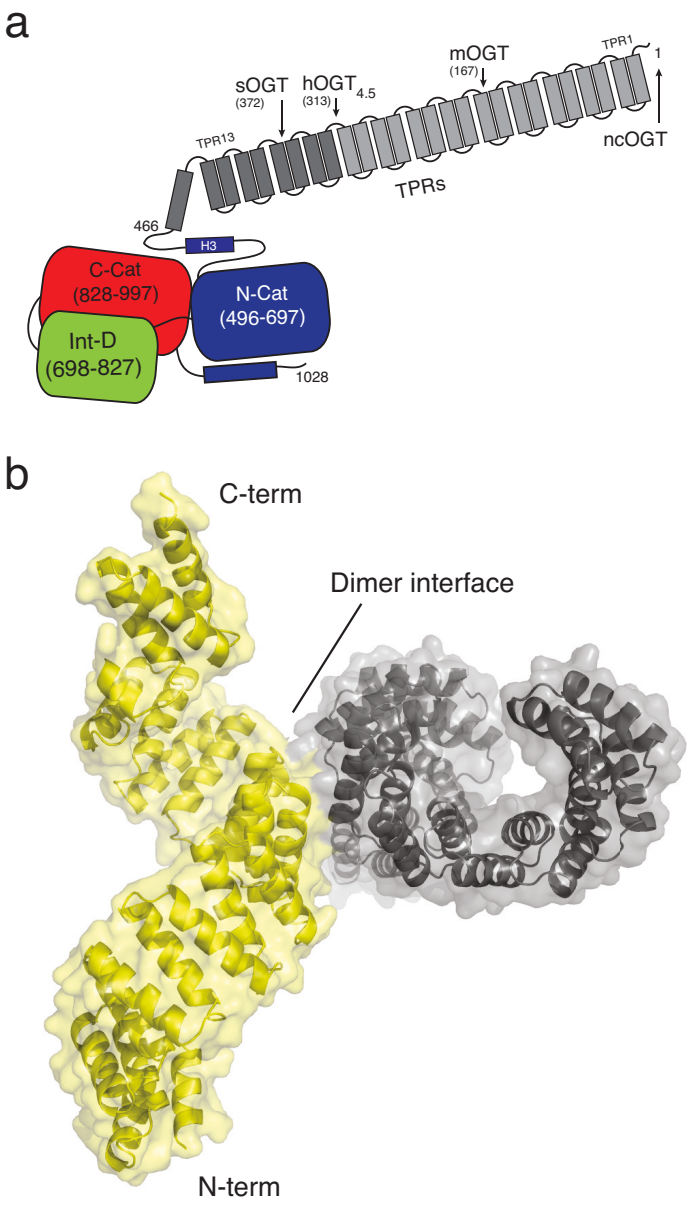


Figure 1.4: Schematic of human OGT domain architecture and X-ray structure of its TPRs. (a) A schematic representation of the domain architecture of OGT. (b) The structure of the TPR domain of OGT, spanning TPRs 1 through 11.5 (PDB: 1W3B). Chains are colored individually. The C- and N-termini are labeled, as is the dimer interface.

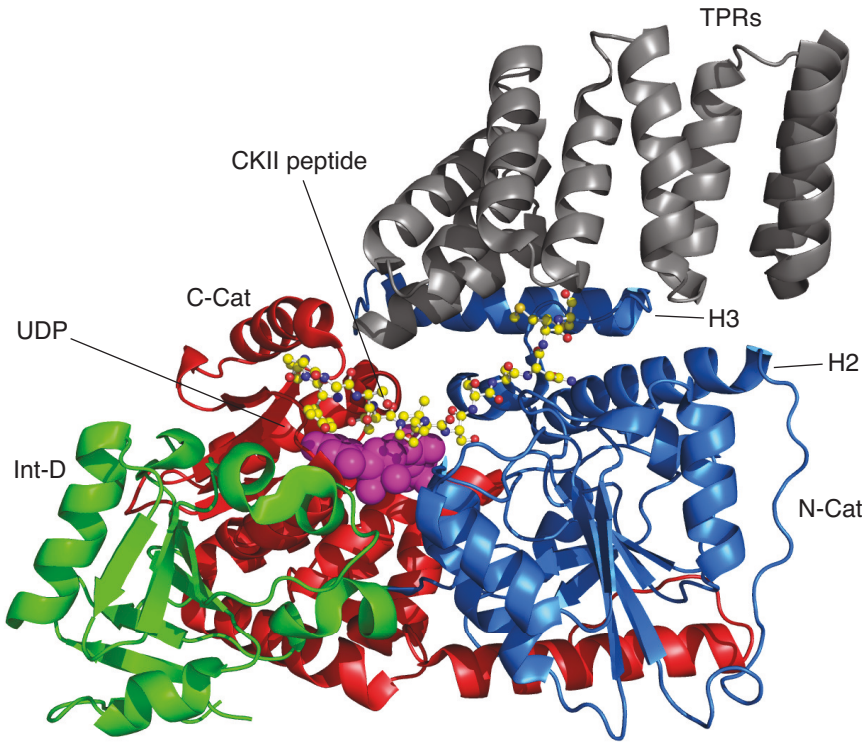


Figure 1.5: X-ray structure of the catalytic domain of human OGT. A cartoon representation of the catalytic domain of human OGT (PDB: 3PE4). Coloring matches the domain schematic shown in Figure 1.4a. The N-terminal catalytic domain is in blue; the C-terminal catalytic domain is in red; the intervening domain (Int-D) is in green; the TPR domain (TPRs 9–13.5) is in gray. In addition, Helices H2 and H3 are labeled, as are UDP (magenta spheres) and the peptide substrate (CKII) (yellow ball and sticks) derived from the well-studied OGT substrate, casein kinase II [142].

1.4 Mechanism of Ser/Thr glycosylation by OGT

OGT is an inverting Gtf that forms a β-glycosidic linkage between GlcNAc and the acceptor residue of the polypeptide substrate [149]. Although it was initially proposed that OGT follows a random bi-bi kinetic mechanism [131], structural information revealed that the polypeptide substrate bound above the UDP-sugar, which suggested an ordered mechanism in which UDP-GlcNAc binds first. Product inhibition studies also supported an ordered bi-bi mechanism [142]. Hence, UDP-GlcNAc binds first to OGT, followed by the polypeptide substrate, which makes extensive contacts with the sugar donor (Figure 1.6). Notably, kinetic studies have shown that the $K_{m,app}$ values for UDP-GlcNAc vary depending on the polypeptide substrate examined, ranging from $\approx 1\mu$M for Nup62 to $\approx 25\ \mu$M for CARM1, which underscores the importance of substrate-substrate interactions in the mechanism [131, 161, 217]. After the ternary substrate complex forms, the acceptor side chain attacks the anomeric carbon, with loss of UDP, to form a β-glycosidic linkage. Finally, the glycopeptide dissociates, followed by UDP.

In addition to binding the sugar donor and acceptor substrates in close proximity (in the case of OGT, UDP-GlcNAc and a polypeptide, respectively), it has long been presumed that Gtfs must do two things in order to catalyze glycosylation: (1) activate the leaving group of the glycosyl donor towards departure, and (2) remove the proton from the hydroxyl group on the glycosyl acceptor [136]. GT-A superfamily Gtfs use a metal ion to activate the leaving group, but OGT and other GT-B Gtfs are metal ion-independent enzymes and use a different strategy (Figure 1.7) [136].

Leaving group activation in OGT is achieved through several elements that stabilize the build-up of negative charge. First, the β-phosphate of UDP-GlcNAc is anchored by hydrogen bonds to the N-terminus of an α-helix; these include key interactions with His920,

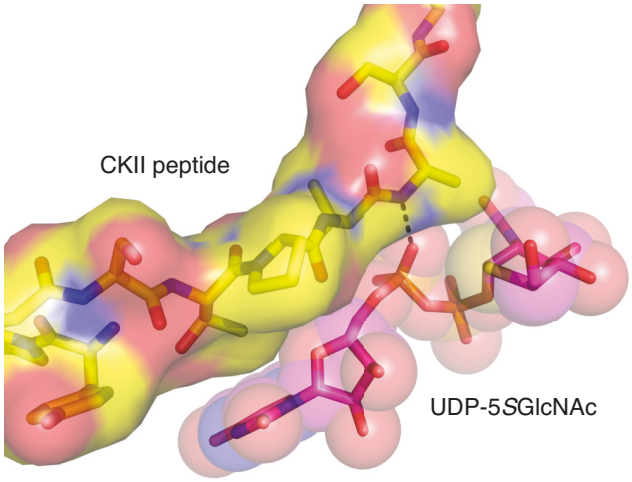


Figure 1.6: An extensive binding surface is formed by the nucleotide sugar. UDP-5*S*GlcNAc (shown as magenta spheres), a UDP-GlcNAc mimic resistant to hydrolysis [72], makes extensive interactions with a CKII peptide substrate (shown as yellow sticks and surface). A hydrogen bond between the backbone N–H of the acceptor peptide and the α-phosphate of the nucleotide sugar donor is shown as a black dotted line.

Thr921, and Thr922. Further stabilization from this helix is derived from the interactions between the negatively charged phosphate and the electric field generated by the net helix dipole [140]. Other GT-B superfamily members use a similar mode of activation [136] with some anchoring the α-phosphate, rather than the β-phosphate, of the leaving group [24, 100]. Second, the side chain of lysine 842 is positioned directly below the β-phosphate (Figure 1.8), and presumably further activates the leaving group towards departure. Biochemical data suggests that Lys842 plays an important role in OGTs enzymatic activity with mutation to methionine or alanine largely abolishing enzymatic activity [142, 140, 171, 213]. Since Lys842 is chemically reactive, as evidenced by its nucleophilicity towards a covalent OGT inhibitor [110], its pK_a may be suppressed due to its proximity to the aforementioned helix dipole; however, its protonation state remains to be established explicitly [110]. This

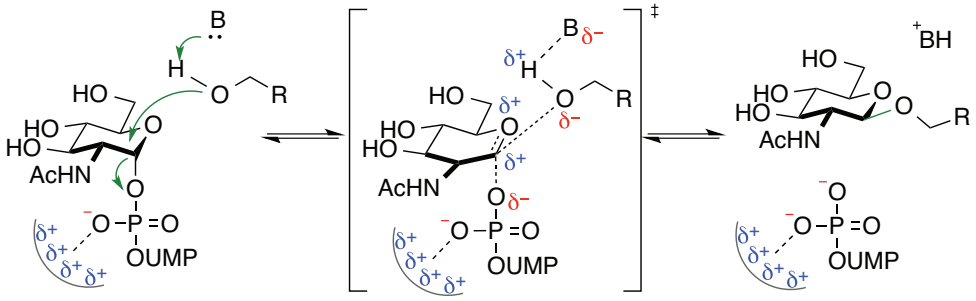


Figure 1.7: A generic GT-B Gtf mechanism. A Brønsted base (represented by "B"), and one or more groups to stabilize the accumulation of negative charge in the leaving group (represented by "δ^+") are thought to be required. A possible transition state, bearing oxocarbenium character, is shown.

combination of a helix dipole, carefully positioned hydrogen bonds from the N-terminus of the helix as well as a critical lysine contact to the β-phosphate of UDP-GlcNAc serves to activate the sugar donor towards nucleophilic attack. Nevertheless, glycosylation is favored over hydrolysis, raising the question of how selectivity for reaction with an acceptor hydroxyl instead of water is achieved, especially given that several crystallographically ordered water molecules are visible within the active site of all structures below 2.0 Å resolution (Figure 1.8). It seems likely that the aforementioned contact between the α-phosphate of UDP-GlcNAc and the amide N-H of the acceptor serine (Figure 1.8) is an important element for this selectivity. This contact may serve to bring the acceptor and donor into close proximity and position the acceptor hydroxyl in a more favorable position to form a bond with the anomeric carbon (labeled C1 in Figure 1.9) compared with any water molecule (Figure 1.8) [140]. It may also further activate the leaving group for departure.

When substrate and product ternary complexes are examined, modest changes in several positions can be seen, with the C1, C2 and O5 atoms of the GlcNAc displaying the

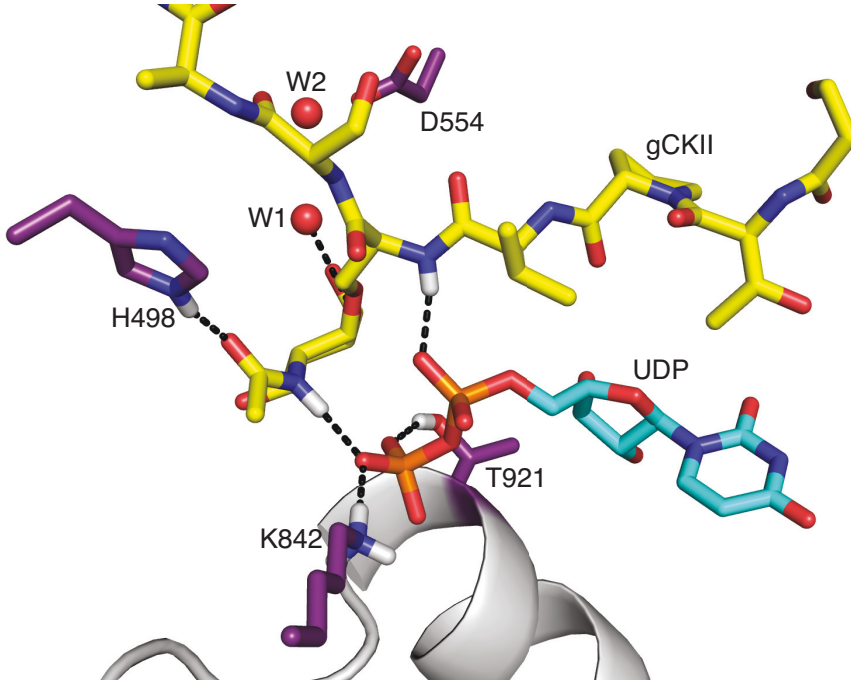


Figure 1.8: Active site of OGT. A view of the OGT active site in a ternary product complex of hOGT4.5, glycosylated CKII peptide (shown as yellow sticks) and UDP (shown as cyan sticks) (PDB: 4GYY). Key residues (side chains shown as purple sticks), two ordered water molecules, labeled W1 and W2 (shown as red spheres), and hydrogen bonds (black dotted lines) are shown.

greatest displacement (Figure 1.9). A net upward rotation of the sugar moves the anomeric carbon (C1) away from the β-phosphate of UDP and into bonding distance to the acceptor hydroxyl. The inferred reaction trajectory is consistent with an electrophilic migration-type mechanism [140]. Another notable change is the rotation of the C2-acetamide, which contacts His498 and rotates away from C1 along the reaction coordinate (Figure 1.9). The importance of the C2-acetamide was further demonstrated by the fact that OGT can transfer UDP-GalNAc, but not UDP-glucose or UDP-2-ketoGlc, a UDP-GlcNAc analog in which the C2 "N–H" is replaced by a "CH_2" [140]. Similar acetamide to β-phosphate interactions

have been observed for other glycosyltransferases [152, 234].

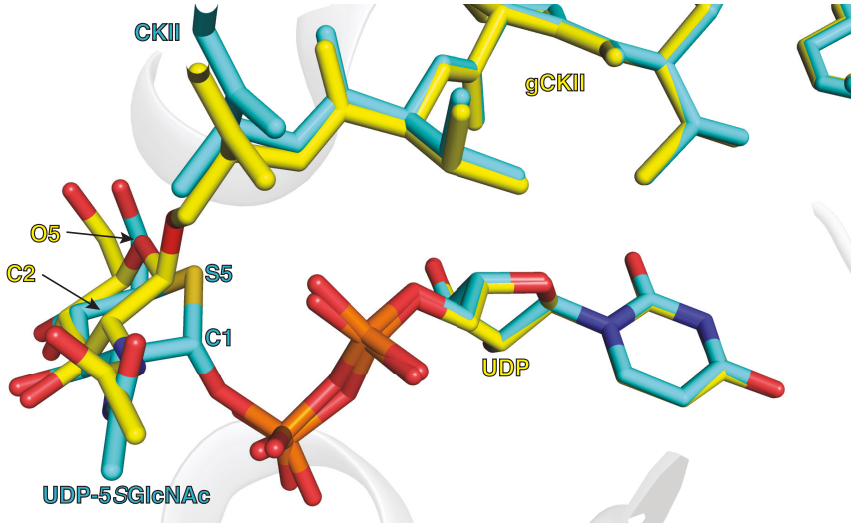


Figure 1.9: Reaction trajectory for Ser/Thr glycosylation by OGT. An overlay of: (1) A product ternary complex (PDB: 4GYW) of OGT, a glycosylated CKII peptide (gCKII), and UDP (both ligands shown in yellow sticks), and (2) A substrate ternary complex (PDB: 4GYY) of OGT, a CKII peptide and UDP-5*S*GlcNAc (both ligands shown in cyan sticks). Several positions on the NT-sugar(s) are labeled.

In addition to providing information about leaving group stabilization and reaction trajectory, the OGT ternary structures revealed a surprise: there is no OGT side chain that can serve as a catalytic base in the immediate vicinity of the reactants [140, 213, 257]. Two different side chains, His498 and His558, had previously been suggested as candidates for the catalytic base, but both were ruled out after the examination of ternary structures containing substrate analogs implicated them in other roles [140, 213]. Two alternatives for how proton transport is achieved have been suggested (Figure 1.10). In one, the α-phosphate of UDP is proposed to act as a base to deprotonate the acceptor hydroxyl (Figure 1.10, proposal I) [213]. Nuclear magnetic resonance studies showed a

pK_a of 6.4 for the pyrophosphoryl group of UDP-GlcNAc, making it a plausible base [106, 149, 213]. Further support stemmed from evidence suggesting that retaining Gtfs that proceed via an S_Ni-like mechanism [7, 59, 75, 145] may similarly employ a phosphate as the catalytic base [149]. In the other, a chain of water molecules leading to an aspartate side chain, Asp554, is proposed to facilitate transport of the proton liberated during the reaction from the active site (Figure 1.10, proposal II) [140]. In order to test the first hypothesis, diastereomeric phosphodithioate analogs of UDP-GlcNAc were prepared and it was found that while both had a similar binding affinity for OGT, only one of them functioned as a donor substrate [213]. While this finding suggests that the α-phosphate is involved in the catalytic mechanism, results involving substrate mimics must be examined cautiously. Replacing the acceptor serine with an aminoalanine may, depending on the protonation state of the amine, bring the amino group in closer proximity to the α-phosphate than it would be in the native substrate. Furthermore, oxygen to sulfur substitution on the phosphate would likely affect the previously mentioned hydrogen bond to the backbone of the acceptor amide (Figure 1.8). Given that the importance of this interaction has not been rigorously examined, its energetic contribution to binding should be quantified. The proposal that the proton is shuttled out of the active site by a Grotthuss-like mechanism [45, 80] invokes the active site ordered water molecules in the mechanism in lieu of implicating a catalytic base within OGT. In a transition state involving an oxocarbenium ion-like species, as computational studies have suggested for OGT [234], a general base may not be needed to deprotonate the acceptor hydroxyl for the reaction to occur. In this case the ordered water molecules leading to residue Asp554 may simply provide a pathway for proton extrusion from the active site. It has been reported that mutation of Asp554 to alanine[1] abolishes

[1]The primary sequence numbering of OGT used in reference [139] is based on mOGT. As a result, Asp554 is discussed as Asp438.

catalytic activity [139], but other studies have found that while mutation to asparagine impairs glycosylation, it does not abolish it [107, 213]. A computational study of these pathways suggested that the lowest energy path for catalysis employed the α-phosphate as the base; [135]. However, further study is required to resolve the issue of how the proton on the acceptor hydroxyl is ultimately removed from the active site.

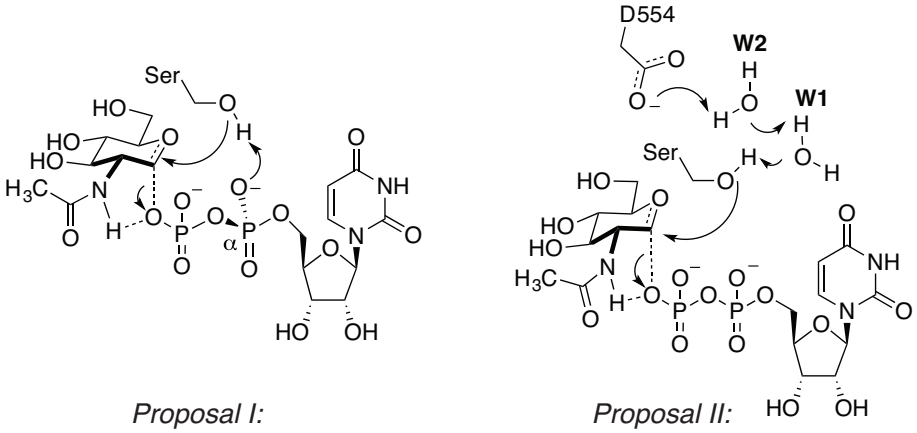


Figure 1.10: Two proposals for the operative proton transport mechanism in OGT. Proposal I involves the α-phosphate acting to deprotonate the glycosyl acceptor; Proposal II involves a Grotthuss-like mechanism wherein the proton is shuttled out of the active site through a chain of ordered water molecules.

1.5 OGT is required for Host cell factor-1 proteolysis

In 2011 OGT was implicated as the cellular factor responsible for the maturation of an essential transcriptional co-activator, host-cell factor 1 (HCF-1) [47, 27]. HCF-1 was first identified in the late 1980s as a component from HeLa cell lysate required for the stable formation of the *Herpes simplex* VP16-induced complex (VIC) and the expression of immediate-early genes during herpes viral infection [6, 70, 118, 133, 224, 244, 260, 262].

HCF-1 has since been implicated as a critical element in cell-cycle regulation, including both G_1-to-S and M-phase progression [116, 163, 165, 180, 236, 246, 261, 276]. Biochemically, the first report that described the purification of HCF-1 (then called C1), by Kristie and Sharpe [134], suggested its existence as a set of heterogeneous, but related, polypeptide fragments; later work revealed that these polypeptide fragments were derived from a large precursor protein (>200 kDa) by some proteolytic maturation process [254]. Though the protease responsible for this cleavage remained elusive, pulse-chase studies suggested that cleavage took place largely in the nucleus, but also, albeit to a lesser extent, in the cytoplasm [255].

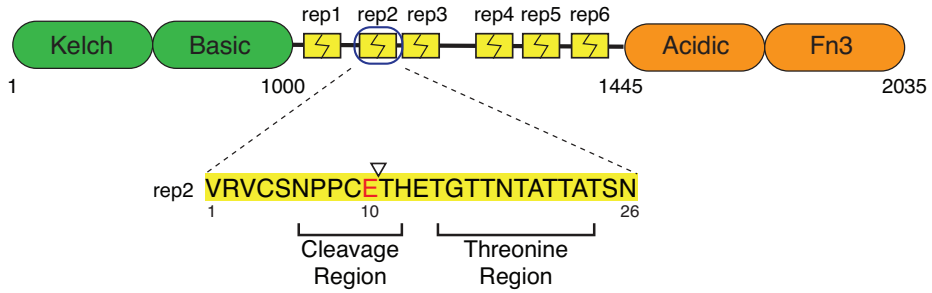


Figure 1.11: Schematic of HCF-1 domain structure. HCF-1 is translated as a 2035 amino acid polypeptide comprising: an N-terminal Kelch domain [83], followed by a region rich in basic amino acids (colored in green); the middle portion of the protein bears six 26-amino acid repeating sequences (colored in yellow); a C-terminal portion with an acid domain and a fibronectin type III (Fn3) repeat domain [253, 195] (colored in orange). The primary sequence of the second proteolytic repeat is shown, and the glutamic acid at the site of cleavage is shown in red. The cleavage site, identified by Edman degradation is denoted by the "▽" symbol, between Glu10 and Thr11.

Cleavage appeared to take place at one or more of six 26-amino acid repeating elements, which constituted an atypically large protease recognition motif (See Figure 1.11) [254, 255].

Based on N-terminal peptide sequencing, cleavage was reported to occur between Glu10 and Thr11 of a given repeat [132, 255]. One striking feature of this cleavage was the fact that even though it was shown to be required for proper HCF-1 function [115], the two halves of the protein appeared to remain non-covalently associated post-processing [195, 255, 253]. Expression of only the N-terminal domain of HCF-1 or of an HCF-1 variant incapable of being cleaved led to defects in cytokinesis and caused binucleation of cells [27, 115]. Binucleation was also observed when an HCF-1 variant lacking the proteolytic repeats, but bearing an ectopic protease site, was expressed in place of the endogenous copy [27]. Though several mechanisms for cleavage were considered [47, 27, 245], the finding that HCF-1 interacts stably with, and is modified by, OGT [27, 125, 175, 261] led to the suggestion that OGT may somehow promote cleavage. Of the two groups that implicated OGT in HCF-1 cleavage, the group of Winship Herr demonstrated that recombinant OGT and UDP-GlcNAc were necessary and sufficient for this process *in vitro* [27]. Alanine scanning of the repeat suggested that several of the threonine residues in the C-terminal region of the repeat were functionally important for cleavage. This led to a division of the proteolytic repeat into a "cleavage region" and a "threonine region" (as shown in Figure 1.11), with further experiments suggesting that OGT may interact with the HCF-1 repeat in a bipartite fashion [27]. Through a series of experiments, Capotosti *et al.* demonstrated that the glycosylation activity of OGT was required for HCF-1 cleavage. However, the questions of whether glycosylation by OGT could facilitate autoproteolysis, or whether Ser/Thr glycosylation of the N-terminal domain of HCF-1 was a requirement for cleavage remained unanswered [27].

1.6 Conclusions and focus of this dissertation

Though our mechanistic understanding of OGT has advanced tremendously since 1984, as has the available repertoire of chemical tools to study *O*-GlcNAcylation [258], the discovery that OGT is responsible for the cleavage of HCF-1 has led to an entirely new set of questions: How might a glycosyltransferase effect amide bond cleavage? Is a second OGT active site responsible for this new transformation? What is the mechanism of amide bond cleavage? And why would Nature connect OGT activity with the cleavage of HCF-1? (see also reference [89]) In the chapters to follow, I aim to answer these questions.

1.7 Resource: High-resolution mass spectrometry

High resolution mass spectrometry (MS) is a powerful analytic tool that exploits the intrinsic property of molecular mass to distinguish species [279]. However, a great deal of information can be garnered beyond simply the mass of a molecule. Using tandem mass spectrometry (MS^n), one can deduce structural characteristics of the molecules being examined. At its most fundamental level, MS operates by taking gaseous ions and analyzing them based on the ratio of their mass to their charge (m/z). Three different types of mass spectrometer were used in these studies: a time-of-flight (ToF) mass spectrometer [81], an Orbitrap mass spectrometer [99, 215], and a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer [167, 215]. The details of how gas-phase ions are produced (electrospray ionization (ESI) [62], in this case) and how each of these mass spectrometers manipulate ions are beyond the scope of this thesis and will not be discussed here. Rather, since aspects of the work described in Chapters 3 and 4 rely on tandem mass spectrometry experiments, a discussion of MS^n experiments for structure elucidation follows.

Today, peptide sequencing, as well as PTM determination, is routinely done by MS/MS [221]. The most common fragmentation methods, collision-induced dissociation (CID), and higher-energy collision-induced dissociation (HCD), effect fragmentation through vibrational excitation of the analyte ions [250, 279]. Excitation occurs via repeated collisions between the analyte ion and neutral gas molecules (Figure 1.12a). In the case of HCD, this process takes on the order of microseconds, allowing accumulated kinetic energy to be statistically redistributed across all degrees of freedom for the analyte ion [250]. As a result, the most labile linkage within this ion are subject to rupture first. In the case of a typical polypeptide, the weakest bond is the C–N (peptide) bond [250, 281], and scission results in the formation of b and y fragment ions (Figure 1.12b). However, if the peptide bears a PTM wherein a more labile linkage than an amide bond is present, e.g., phosphorylation or glycosylation, that linkage becomes preferentially subject to fragmentation. In these cases, loss of the PTM prior to the generation of b and y ions results in a loss of positional information, with respect to that PTM, making vibrational excitation ineffective for PTM analysis. Fortunately, alternative fragmentation approaches, involving the use of low energy electrons are able to retain sequence information for labile PTMs.

Electron-capture dissociation (ECD) and electron-transfer dissociation (ETD) are fragmentation methods that exploit ion-electron interactions to generate fragment species from the parent analyte ion [279, 281, 282]. The process of electron-capture-based dissociation involves the interaction of a multiply charged cation with an electron [281]. When an electron approaches a cation, there is a mutual Coulombic attraction; however, for this attraction to lead to capture the cation must dissipate the excess energy over a very short period of time (10^{-14} seconds) [281] (Figure 1.13a). This time-scale is too rapid for atomic motions to compensate and redistribute the energy across several degrees of freedom. For polyatomic

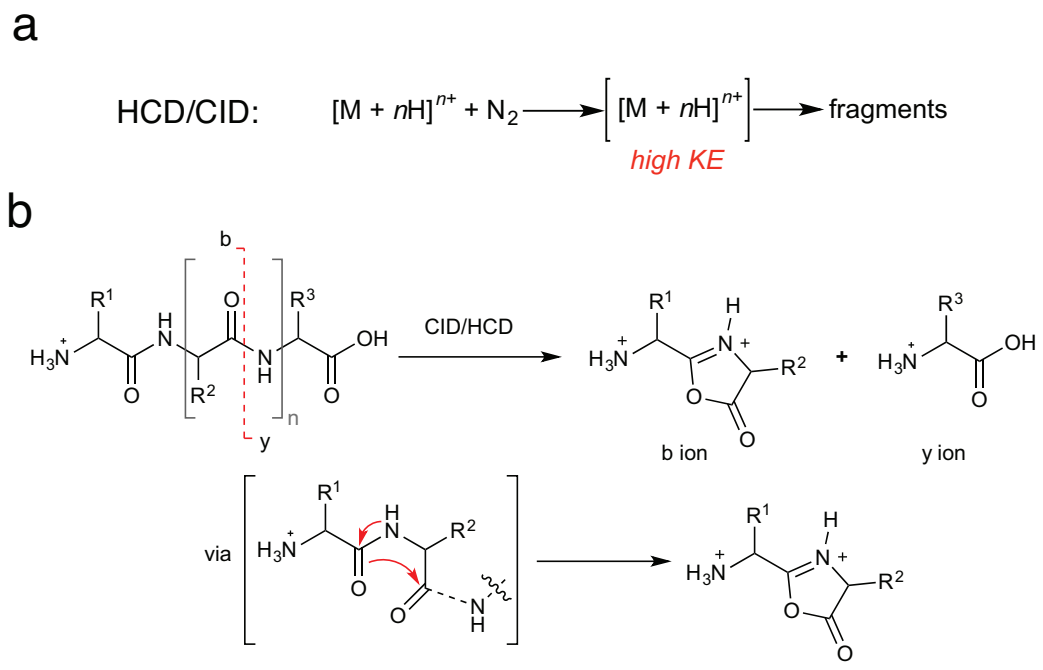


Figure 1.12: Collision-induced MS/MS. (a) A schematic showing the activation mode of CID/HCD. N_2 is used as a sample inert gas. (b) Representative structures for fragment ions produced by CID/HCD from a generic peptide (tripeptide when n=1). Formation of the oxazolone *b* ion structure is believed to occur concurrent with cleavage of the C–N bond (and departure of the *y* ion) [32].

cations a solution is a prompt dissociation event, which can allow for stabilization of the electron within time-scales on the order of 10^{-16}–10^{-15} s [281]. In this scenario, electron capture and bond cleavage appear to occur as a concerted process, known as dissociative recombination [281]. Further, bond cleavage under these conditions is mostly an electronic process, proceeding faster than atomic vibrations and rearrangements. Processes wherein bond dissociation occurs faster than energy redistribution are known as non-ergodic processes [281]; however, whether ECD is a non-ergodic process has been extensively debated [233, 283, 281].

There are several practical considerations for ECD/ETD worth discussing here: A consequence of electron capture is a net charge reduction of the parent ion which practically means that analyte peptides must be multiply charged in order to allow for detection. In ECD parent ions are bombarded directly with low energy electrons ($\approx$ 1–10 eV) under high vacuum (10^{-10} torr) [279]. As a result, a major limitation of ECD is that it is most commonly implemented on FT-ICR MS, whereas ETD is amenable to implementation on Orbitrap instruments, making it more readily accessible [279]. To circumvent the need for a direct electron source, ETD employs a "reagent" gas to facilitate the electron transfer process; the radical anion is produced separately using a chemical ionization source and allowed to react with analyte ions for a defined period of time [279]. As a result, the ion-ion interactions involved in ETD occur in an ion trap, which is held at higher pressure (by several orders of magnitude). This leads to a faster collisional cooling and reduces the internal energy of the post-electron attachment complex, reducing the likelihood of more energetic fragmentation pathways occurring. In this regard ECD imparts greater energy into an ion than ETD, often leading to additional peaks in an ECD spectrum, as compared to an ETD spectrum [279].

The result of ECD (or ETD) fragmentation is not the formation of b and y ions, but instead c' and $z^{\bullet}$ fragments (Figure 1.13b, "$'$" is pronounced "prime" and "$^{\bullet}$" is pronounced "dot") (Figure 1.13b). The prime symbol denotes the fragment to which the $H^{\bullet}$ is transfered [282]. In fact, an ECD spectrum is largely devoid of b and y ions. This has been attributed to several factors: First, peptide bond fragmentation proceeds by intermediates that require (relatively) long life times (up to microseconds), which are incompatible with the rate at which ECD is believed to occur ((10^{-12} s, or faster) [281]. As a result, only direct bond scission can occur in ECD. Second, the species formed in ECD are not even-electron species, as is the case for vibrational excitation (CID/HCD), but instead are hydrogen-abundant radical cations. Since the presence of the radical changes the proximal bond strengths, this has been cited as a potential explanation for the observed N–Cα bond cleavage [284, 281]. The preference for N–Cα bond cleavage has been rationalized through relative electron affinities [281]. The protonated side chains of lysine, arginine and histidine can be solvated by the amide backbone in the form of a species: $A\cdots H^{+}\cdots B$ [283]. An electron captured by such a species would form $^{\bullet}A–H + B$ if the electron affinity of $A^{+}H$ is greater than that of HB^{+} [283]. The electron affinity of a protonated nitrogen, such as an amide or amine is on the order of 4 eV, compared to 6 eV for a protonated carbonyl [283]; this also suggests that neutralization of the protonated carbonyl by electron capture would be an exothermic process. Though multiple mechanisms for bond scission have been described [279], a representative example, termed the "Cornell mechanism" is shown in Figure 1.13b.

Unlike C–N (peptide) fragmentation, where all amino acids are able to lead to fragmentation, tertiary amides are not susceptible to ECD fragmentation. Scission of the N–Cα bond for proline fails to produce two polypeptide fragments, so $c'/z^{\bullet}$ ion pairs are not expected at these positions [279, 281]. In addition to the even-electron c' and radical $z^{\bullet}$ ion

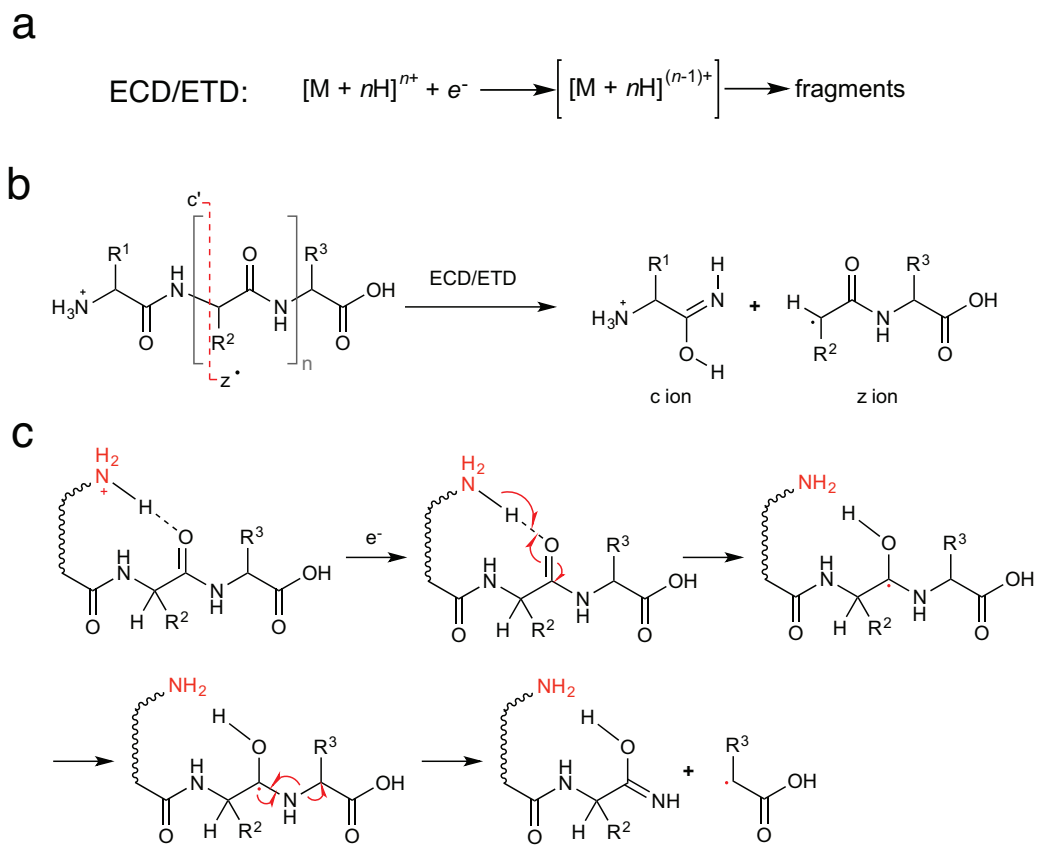


Figure 1.13: Electron-capture dissociation MS/MS. (a) A schematic showing the activation mode of ECD/ETD. (b) Representative structures for c' and $z^{\bullet}$ fragment ions produced by ECD/ETD from a generic peptide (tripeptide when n=1). (c) The "Cornell mechanism" for N–Cα bond cleavage in ECD/ETD [279]. The pendant protonated amine at the N-terminus represents a protonated residue located generically N-terminal to the site of bond scission. The product fragments shown represent the corresponding c' and $z^{\bullet}$ species.

pairs, it is possible to also observe radical $c^{\bullet}$ and even-electron z' ion pairs [210]. This has been suggested to occur via hydrogen atom transfer. After N–Cα bond cleavage of $[\mathrm{M} + n\mathrm{H}]^{n+}$, the $[c + z]^{(n-1)+\cdot}$ complex remains briefly held together by non-covalent interactions. If the lifetime of this complex is sufficiently long, the $z^{\bullet}$ radical can abstract a hydrogen atom from c' forming z' (where $z' = z^{\bullet} + \mathrm{H}$; a z+1 fragment ion) and $c^{\bullet}$ (where $c^{\bullet} = c' - \mathrm{H}$; a c–1 fragment ion) [210, 279].

Chapter 2

HCF-1 is cleaved directly by OGT

A portion of this chapter originally appeared as [141]:

"HCF-1 is cleaved in the active site of *O*-GlcNAc transferase" Michael B. Lazarus[‡], Jiaoyang Jiang[‡], Vaibhav Kapuria, Tanja Bhuiyan, **John Janetzko**, Wesley F. Zandberg, David J. Vocadlo, Winship Herr and Suzanne Walker. ([‡]equal contribution) *Science*. Volume 342, Number 6163: 1235–1239. 2013.

Contributions: The experiments presented in this chapter were conducted by John Janetzko, Michael B. Lazarus, and Jiaoyang Jiang. Michael B. Lazarus obtained crystallographic data, unless otherwise specified. Jiaoyang Jiang performed mass spectrometry experiments, unless otherwise specified. Michael B. Lazarus, Jiaoyang Jiang and I performed biochemical experiments, expressed and purified proteins and made various HCF-3R constructs.

Abstract

Host Cell Factor-1 (HCF-1), a transcriptional co-regulator of human cell-cycle progression, undergoes proteolytic maturation in which one or more of six repeated sequences is cleaved by the nutrient-responsive glycosyltransferase, β-*O*-linked-*N*-acetylglucosamine (*O*-

GlcNAc) transferase (OGT). Herein I present the findings that the tetratricopeptide-repeat (TPR) domain of OGT binds the C-terminal portion of an HCF-1 proteolytic repeat such that the cleavage region lies within the glycosyltransferase active site above UDP-GlcNAc. The conformation is similar to that of a glycosylation-competent peptide substrate. Cleavage occurs between cysteine and glutamate residues and results in a pyroglutamate product. Conversion of the cleavage site glutamate into serine converts an HCF-1 proteolytic repeat into a glycosylation substrate. Thus, protein glycosylation and HCF-1 cleavage occur in the same active site. The residues adjacent to the glutamate (residue 10) are not required for cleavage, and the important elements of OGT and HCF-1 required for cleavage are discussed.

2.1 Introduction

O-GlcNAc transferase (OGT) is a Ser/Thr glycosyltransferase that transfers a GlcNAc moiety from UDP-GlcNAc to hundreds of nuclear and cytoplasmic proteins, thus influencing their activity, localization, and overall function [90, 130, 251]. Because OGT activity is sensitive to UDP-GlcNAc concentrations, OGT is proposed to regulate cellular responses to nutrient status [51, 93, 266]. Human HCF-1 is a transcriptional co-regulator involved in regulating cell-cycle progression [260, 271]. In an unusual proteolytic maturation process [47, 132, 255], any of six centrally located 26 amino acid sequence repeats (Figure 2.1a) are cleaved by OGT in the presence of UDP-GlcNAc [27], providing a potential link between cell-cycle progression and nutrient levels. These repeats contain two essential regions for proteolysis: a C-terminal threonine-rich region, proposed to be an OGT-binding site, and the cleavage site, which contains a conserved Cys-Glu-Thr ("CET") sequence [132, 255]. OGT was previously shown to cleave a fragment of HCF-1, called HCF-1rep1 (Figure 2.1b),

which contains the first of the HCF proteolytic repeats plus a portion of the N-terminal HCF-1 sequence containing several *O*-GlcNAc sites [27].

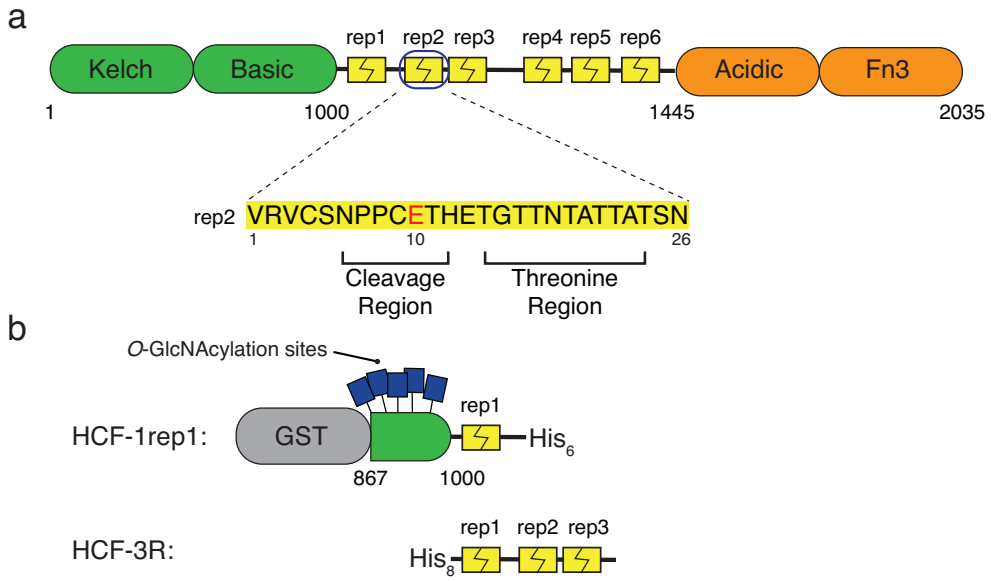


Figure 2.1: Schematic of HCF-1, and HCF-1 constructs used for *in vitro*. (a) HCF-1 is translated as a 2035 amino acid polypeptide comprising: an N-terminal Kelch domain [83], followed by a region rich in basic amino acids (colored in green); the middle portion of the protein bears six 26-amino acid repeating sequences (colored in yellow); a C-terminal portion with an acid domain and a fibronectin type III (Fn3) repeat domain [253, 195] (colored in orange). The primary sequence of the second proteolytic repeat is shown, and the glutamic acid (residue 10) at the site of cleavage is shown in red. (b) Two HCF-1 constructs, HCF-1rep1 and HCF-3R were employed in these studies. A Glutathione *S*-transferase (GST) tag was used at the N-terminus of the HCF-1rep1 construct; both constructs bear oligohistidine tags (His_6 for HCF-1rep1 and His_8 for HCF-3R).

2.2 Design of a repeat-only HCF-1 construct

In order to evaluate whether Ser/Thr glycosylation of HCF-1 by OGT was required for cleavage, Michael B. Lazarus and Jiaoyang Jiang designed a "repeat-only" HCF construct, comprising the first three proteolytic repeats of HCF-1 with an N-terminal His tag (Figure 2.1b), called HCF-3R. Notably, this construct did not undergo significant glycosylation under the conditions of our assays [141].

2.3 Identifying requirements for HCF-1 cleavage

Contribution: The experiments described in this section were conducted by Jiaoyang Jiang and Michael B. Lazarus, except for the nucleotide-sugar (NT-sugar) dependence experiments shown in Figure A.1.

In order to determine the requirements for cleavage, an HCF-3R construct where each repeat bears a glutamic acid at position 10 (HCF-3R-EEE) was treated with OGT under various conditions. No cleavage was observed when HCF-3R-EEE was incubated with wild-type OGT alone, or OGT and UDP. When treated with OGT and UDP-GlcNAc several products were observed by SDS-PAGE (Figure 2.2). These products were not observed when HCF-3R-EEE was incubated with OGT (K842A), a mutant incapable of catalyzing glycosylation (see mention in section 1.4) [142, 171], or when wild-type OGT was pre-treated with 1.5 equivalents of a covalent OGT inhibitor known to cross-link two active site residues, K842 and C917 [110]. Cleavage was also inhibited when UDP was added to reactions of wild-type OGT and UDP-GlcNAc; OGT is known to be potently product-inhibited by UDP [142], for which a K_{d} of 0.54 μM $\pm$ 0.01 μM for hOGT4.5 has been measured

by surface plasmon resonance (SPR) [213]. Consistent with these findings, the addition of calf alkaline phosphatase (CIP), which hydrolyzes UDP, enhanced cleavage [141]. Because UDP-GlcNAc, but not UDP, promoted cleavage, UDP-5*S*GlcNAc, a UDP-GlcNAc isostere was tested. UDP-5*S*GlcNAc has been shown to adopt the same conformation within the active site, yet is resistant to glycosylation and hydrolysis [72, 140]. Almost no peptide cleavage was observed (Figure 2.2), implying that the UDP-GlcNAc does not simply fulfill a structural function (see discussion in section 1.3 and Figure 1.6), but must react for HCF-3R cleavage to occur. Consistent with this proposal, UDP-5*S*GlcNAc was found to inhibit cleavage when added to reactions containing UDP-GlcNAc [141].

Further studies on the role of the nucleotide sugar, using a substrate described in Chapter 3, revealed that like for Ser/Thr glycosylation [140], the C2-acetamide is critical for HCF repeat cleavage (Figure A.1). This suggests that the C2-acetamide is required for activation of UDP-GlcNAc by OGT (see discussion in section 1.4).

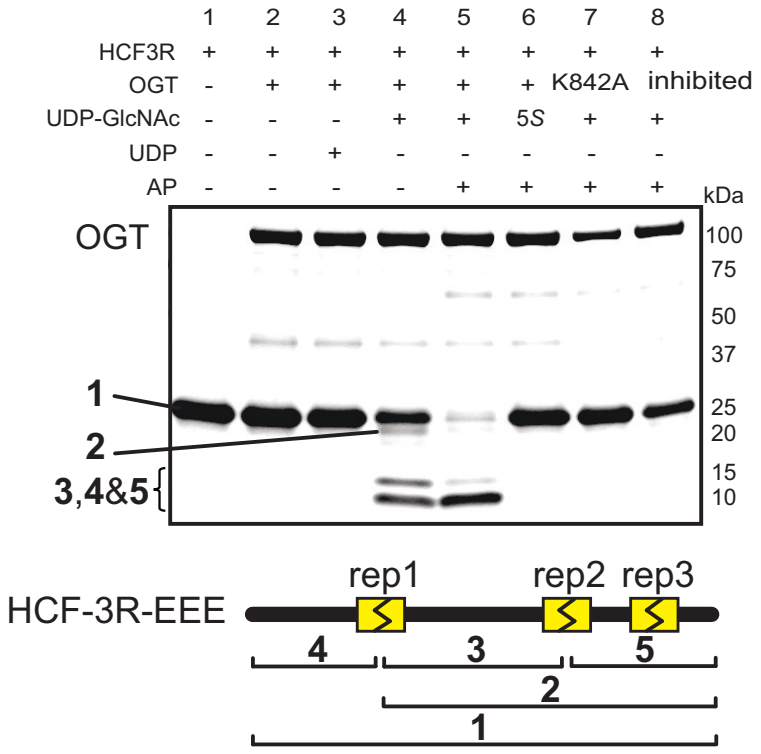


Figure 2.2: Cleavage of HCF-3R requires UDP-GlcNAc. HCF-3R-EEE (see Table B.2) was treated with wild-type OGT (lanes 2–6), with OGT (K842A) (lane 7), or with wild-type OGT that was pre-treated with a previously reported covalent inhibitor [110] (lane 8), in the presence of UDP (lane 3), UDP-GlcNAc (lanes 4, 5, 7, and 8), or UDP-5*S*GlcNAc (labeled "5*S*") [72] (lane 6). Calf alkaline phosphatase (AP) was added to reactions as indicated. Cleavage products were separated by SDS-PAGE and stained by Coomassie Blue.

2.4 HCF-1 cleavage results in a pyroglutamate

In order to obtain more precise information about the cleavage products they were analyzed using LC-MS. These experiments were conducted by Jiaoyang Jiang. Previous analysis of products isolated from cells concluded that cleavage occurs at the E–T peptide bond of the "CET" sequence [132, 255] but none of the four products observed had masses consistent with E–T cleavage. Rather, the masses observed were consistent with cleavage occurring at the preceding C–E bond, but only if all of the C-terminal fragments were 18 Da lower in mass than expected [141]. We hypothesized that the 18 Da difference resulted from generation of an N-terminal pyroglutamate. Using an HCF-3R construct bearing one functional and two inactivated (by E10A mutation) repeats, HCF-3R-EAA (Table B.2), Jiaoyang showed that cleavage of HCF-3R-EAA led to the production of one dehydrated C-terminal product [141]. Pyroglutamate aminopeptidase (PGP) [76] treatment reduced the mass of this product by 111 Da, corresponding to loss of pyroglutamate (Figure 2.3) [141]. The earlier proposed cleavage site was likely misidentified because N-terminal pyroglutamates would be unreactive under Edman sequencing conditions employed in those studies [56, 186], and cellular PGAPs may have processed some of the cleavage products [132, 141, 255].

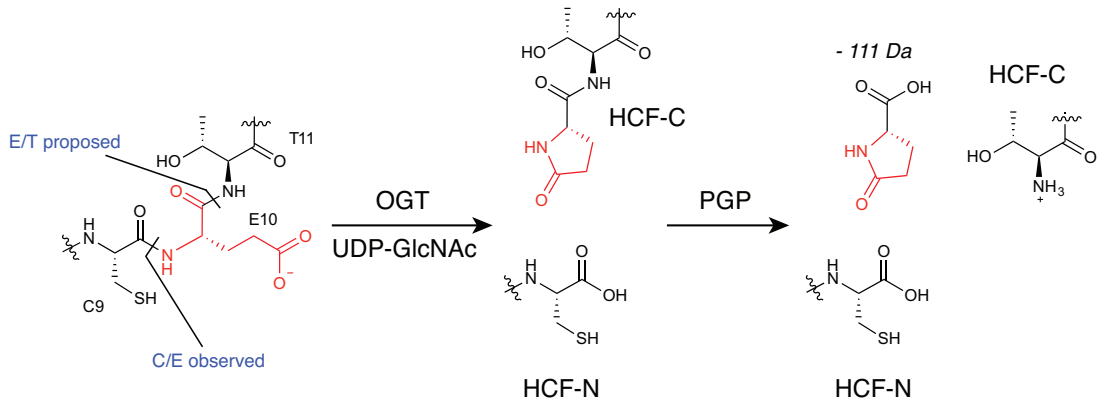
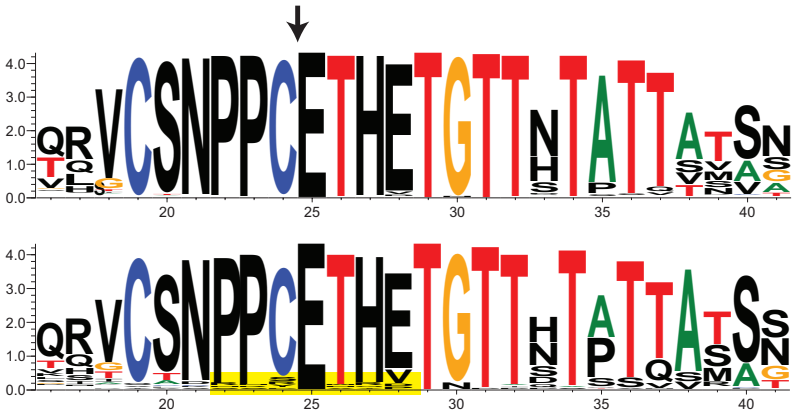


Figure 2.3: Cleavage of HCF-3R occurs between Cys9 and Glu10 and produces a pyroglutamate. Treating the products of HCF-1 cleavage by OGT with PGP affords a loss of 111 Da from the C-terminal fragment, supporting that the N-terminus of the HCF-C cleavage fragment bears a pyroglutamate.

2.5 Role of residues within the "CET" cleavage region

Having identified the correct cleavage site as the amide bond between C^9 and E^{10}, we investigated the effect of amino acid substitutions within the HCF repeat, using the HCF-3R-EAA as a parent construct. In order to make informed substitutions at positions 9 and 11, I examined all known sequences bearing resemblance to HCF-1 using Basic Local Alignment Search Tool (BLAST). Orthologous proteins to HCF-1 were identified in 54 organisms other than *Homo sapiens* (human). Organisms had between four and twelve repeats. There is very low sequence variability across repeats, both within a given organism and across taxa. I examined the most divergent repeats, which typically had multiple amino acid deviations from the cleavage consensus sequence, and based on these sequences and chemical intuition I generated a set of mutants at positions 9 and 11. Whether any of these divergent HCF ortholog repeats would be capable of undergoing cleavage was not known.



```
  X. laevis Repeat 6: ...VNVCSDPRSETYETGTINTPTTA...
  O. latipes Repeat 8: ...LNTYSNRPYETHPTGTTDTATQA...
  O. latipes Repeat 7: ...KSKCSNPPCENQVTGTTDTATQA...
O. divergens Repeat 3:  ...RPCSNPPRETHETGTTSTATTA...
 O. anatinus Repeat 3: ...QIVCSNPPCEVHETGTTNTATVA...
    M. zebra Repeat 3: ...NQVCSNPPCEMHETGTTNTATVA...
```

Figure 2.4: The HCF-1 proteolytic repeat is highly conserved. (top) A probability representation for repeats from 55 organisms shows little deviation from the consensus sequence. The (↓) arrow denotes the cleavage site. (middle) Using only the bottom 10% of sequences by identity shows the emergence of variability (highlighted in yellow). The sequences of some of the most divergent repeats are shown.

2.5.1 Role of Glutamate 10

It had previously been shown that mutation of E10 to alanine abolished cleavage [27]. I tested point mutants for their ability to undergo OGT-catalyzed cleavage. All substitutions: E10Q/D/N (Figure 2.5) and E10A/S (Figure 2.10) prevented cleavage. These mutants, except for E10N, have also been examined in the context of HCF-1rep1, with similar results being observed [141].

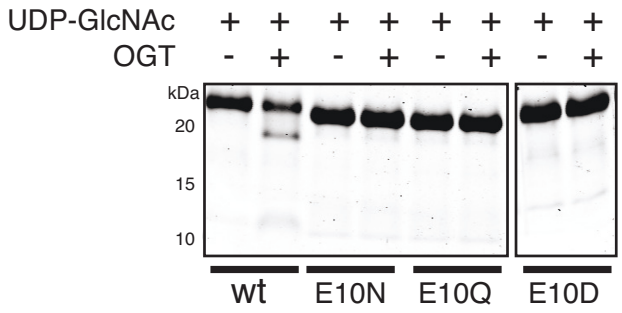


Figure 2.5: Glutamate 10 is required for cleavage. "wt" denotes HCF-3-EAA. Cleavage of HCF-3R-XAA by wild-type OGT. Cleavage reactions were separated by SDS-PAGE and visualized by Coomassie Blue.

2.5.2 Role of Cysteine 9

Previously, it had been shown that substitution of Cys9 with alanine greatly impaired cleavage [27]. I tested ten variants at position 9 (Figure 2.6) (see Table B.3 for sequences). Variants that substituted Cys9 for an acidic amino acid (D/E) showed no detectable cleavage. Substitution with small neutral amino acids (A/S/M) led to greatly reduced cleavage, but still a faintly detectable band. Methionine can be considered small if the side chain can exist in an extended conformation. Substitution with β-branched neutral amino acids (T/V) resulted in almost no cleavage. There is still possibly a band that corresponds to the cleaved product, but it is much weaker than in the case of (A/S). To our surprise substitution of Cys9 with a basic amino acid (K/R) led to robust cleavage. In addition, substitution with the aromatic amino acid tyrosine showed good cleavage, though less than K/R. These data demonstrate that while cysteine 9 is important for cleavage, it is not essential, and cleavage-pathways that may be cysteine-dependent (analogous to intein splicing) are not operative [5, 74]. While the precise role being played by this residue in the cleavage mech-

anism is at present unknown, one common feature of the residues well tolerated at this position is their ability to act as proton (or hydrogen bond) donors.

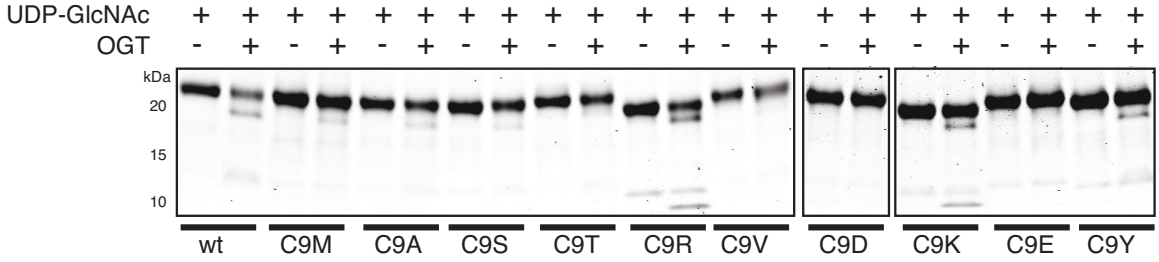


Figure 2.6: Cysteine 9 is important, not essential for cleavage. "wt" denotes HCF-3-EAA. Cleavage of HCF-3R-(C9X)EAA by wild-type OGT. Cleavage reactions were separated by SDS-PAGE and visualized by Coomassie Blue.

Subsequent studies, employing peptide-based substrates described in Chapter 3, allowed for a better determination of kinetic differences between variants at Cys9. These experiments showed that a Cys9 to Arg substitution improved cleavage proficiency relative to the wild-type (Cys9) allele. Further, serine at position 9 led to comparable cleavage to the wild-type sequence, while alanine led to a significant reduction in cleavage proficiency, though cleavage still occurred (Figure A.2).

2.5.3 Role of Threonine 11

I tested six variants at position 11 (see Table B.4 for sequences). All of the variants tested were cleavage-competent, albeit with reduced cleavage relative to the native sequence (Figure 2.7). These data suggest that T11 may not directly serve a functional role in cleavage, but rather may be optimally suited to facilitate a kink in the HCF peptide between the cleavage region and the threonine-rich region (structure is discussed in section 2.6).

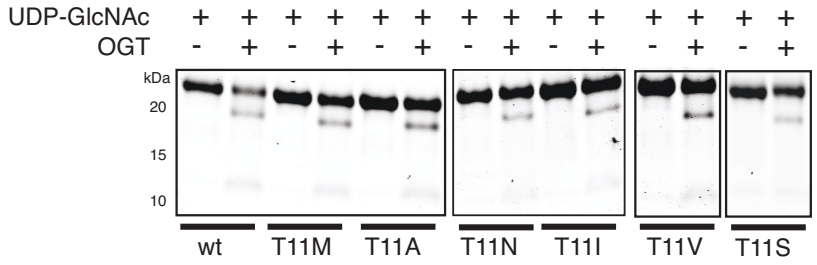


Figure 2.7: Threonine 11 is not essential for cleavage. "wt" denotes HCF-3-EAA. Cleavage of HCF-3R-(T11X)EAA by wild-type OGT. Cleavage reactions were separated by SDS-PAGE and visualized by Coomassie Blue.

2.6 Structure of an HCF-1 repeat bound to OGT

To gain additional insight into how OGT cleaves HCF-1, Michael B. Lazarus solved crystal structures of OGT bound to a 26 amino acid peptide (26mer, Table B.1). Using UDP-GlcNAc and the 26mer peptide, he was able to obtain well diffracting crystals, however the N-terminal portion of the peptide (spanning residues 1 through 12) could not be resolved. We reasoned this may be due to reaction *in crystallo*. The use of a shortened 16 amino acid peptide comprising only the threonine-rich region of the HCF repeat (HCF-C, Table B.1) afforded a higher-resolution (1.8 Å) ternary complex of hOGT4.5:UDP:HCF-C (PDB: 4N39). The use of a longer peptide, afforded lower resolution data and did not show density for additional residues. Using an unreactive HCF-1 peptide (an E10A variant) he was able to obtain a 1.9 Å structure (PDB: 4N3A), but again density was only observed for the threonine-rich portion of the repeat (Figure 2.8) [141]. These two structures show the threonine-rich peptide bound in an extended conformation within the lumen of the TPR domain [111, 142, 161] of OGT (Figure 2.8a) [141]. Five conserved asparagine residues within the TPR domain form a series of interactions, four being bidentate, with the amides of alternating residues along the peptide backbone. A lysine side chain also

contacts the peptide backbone. Four aspartates and a glutamine of OGT form hydrogen bonds to the threonine side chains of the HCF-1 repeat (Figure 2.8b) [141]. The binding mode in the crystal structures is consistent with mutational data demonstrating the importance of the OGT asparagines and the conserved C-terminal threonines for both cleavage and binding [27, 141].

Because intact UDP-GlcNAc is required for OGT-catalyzed cleavage of the HCF-3R substrates, we thought the UDP-5*S*GlcNAc analogue might stabilize density for a full repeat. Indeed, Michael B. Lazarus obtained a 2.17 Å structure of OGT with UDP-5*S*GlcNAc and a 26 amino acid HCF peptide corresponding to HCF rep2 (see Figure 2.1a), but with an E10Q substitution (PDB: 4N3B). In this structure the C-terminal threonine-rich region binds to the TPR domain as described above, and the N-terminal cleavage region is now visible (Figure 2.9) and forms an extensive binding interface with UDP-5*S*GlcNAc. He also obtained a ternary complex with all wild-type substrates (PDB: 4N3C): UDP-GlcNAc:26mer:hOGT4.5, confirming the binding mode of the E10Q peptide. Perhaps unsurprisingly the density for this ternary complex was poorer than for the mutant ternary complex with UDP-5*S*GlcNAc and an E10Q peptide. Both of the complexes (with E10 and UDP-GlcNAc and Q10 with UDP-5*S*GlcNAc) show poor density for the side chain of residue 10 beyond the β-carbon.

Remarkably, the cleavage region binds in a mode almost identical to that of a glycosylation-competent peptide substrate [140], and residue 10 aligns perfectly with the glycosyl acceptor amino acid (Figure 2.10a). The structures suggested an E10S mutation would be glycosylated at residue 10. Indeed, HCF-3R-SAA is not cleaved, and unlike HCF-3R-EAA, the E10S analog (HCF-3R-SAA) was glycosylated effectively (Figure 2.10b). Thus, the identity of the amino acid at position 10 of an HCF-1 proteolytic repeat glutamate or serine can

dictate whether OGT cleaves or glycosylates the substrate. The experiment shown was conducted by John Janetzko.

2.6.1 Attempts to trap a reactive intermediate structurally

Having identified variants of the HCF-1 cleavage repeat that, while impaired, still underwent cleavage (see section 2.5.2), I obtained co-complexes of hOGT4.5 with either UDP-5*S*GlcNAc or UDP-GlcNAc and C9R-26, C9Y-26, C9S-26, or $C9NH_2$-26 (amino alanine substituted for cysteine) (see Table B.1 for sequences). While C9R, C9S and $C9NH_2$ yielded various ternary complexes, the C9Y peptide was poorly soluble and was not suitable for crystallographic trials. Crystals were frozen at various time points in an attempt to trap a reactive intermediate. Unfortunately, in all cases where UDP-GlcNAc was used, robust difference density was only seen for the C-terminal region of the HCF peptide, corresponding to residues E13 through T24. When UDP-5*S*GlcNAc was used, the full length peptide could be resolved, for both C9R-26 and C9S-26. I was able to refine these structures to 2.22 Å and 3.34 Å, respectively. In both cases the HCF peptide bound analogously to the wild-type 26mer peptide. The side chain of Glu10 was disordered beyond the β-carbon in both cases, as was the case for the other two full-length HCF repeat ternary complexes [141]. In addition, for the C9R-26mer peptide, the arginine side chain was also disordered beyond the β-carbon.

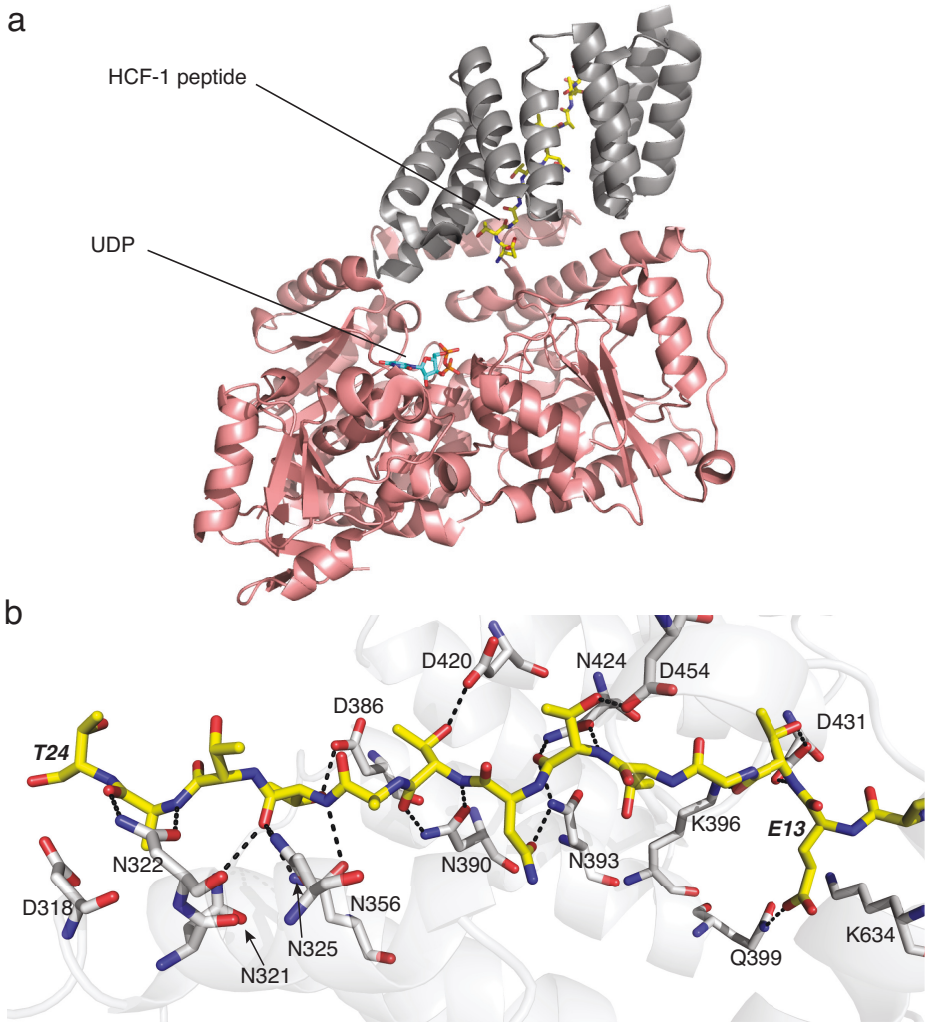


Figure 2.8: The threonine-rich region of HCF-1 binds in the TPRs of OGT. (a) Overall structure of hOGT4.5:UDP:HCF-C ternary complex (PDB: 4N39). The OGT catalytic domain (colored in salmon) and the TPR domain (colored in gray), along with the HCF-C peptide (in yellow sticks) and UDP (in cyan sticks) are shown. (b) A close-up of the OGT–HCF-C interactions. The peptide is shown in yellow sticks, specific OGT side chains making contacts to the peptide are shown as sticks and labeled. Hydrogen bonds are shown as black dotted lines.

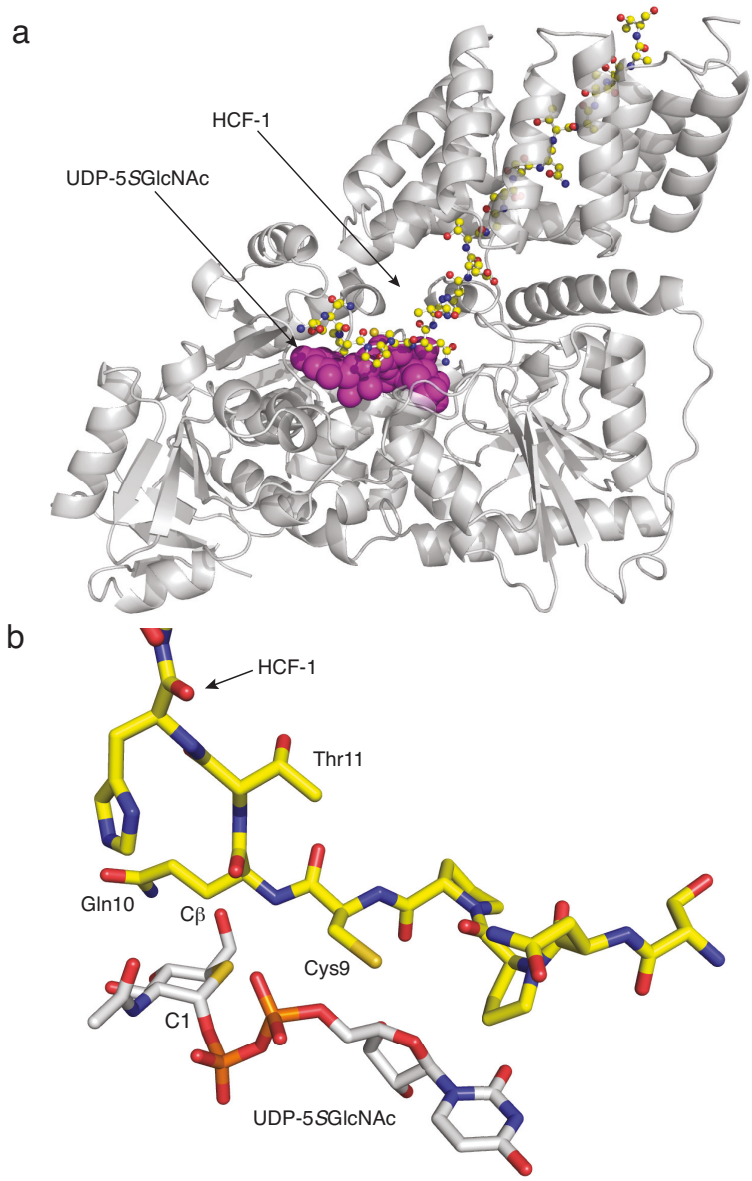


Figure 2.9: HCF-1 cleavage takes place in the glycosyltransferase active site of OGT. (a) Overall structure of hOGT4.5:UDP-5*S*GlcNAc:Q10-26 (PDB: 4N3B). UDP-5*S*GlcNAc is shown as purple spheres. Q10-26 is shown as yellow balls and sticks. (b) A close-up of two substrate analogs shown in the OGT active site. Density is visible for the entire cleavage region, and the "C-E10Q-T" region is annotated. The anomeric carbon (C1) of the nucleotide sugar is labeled.

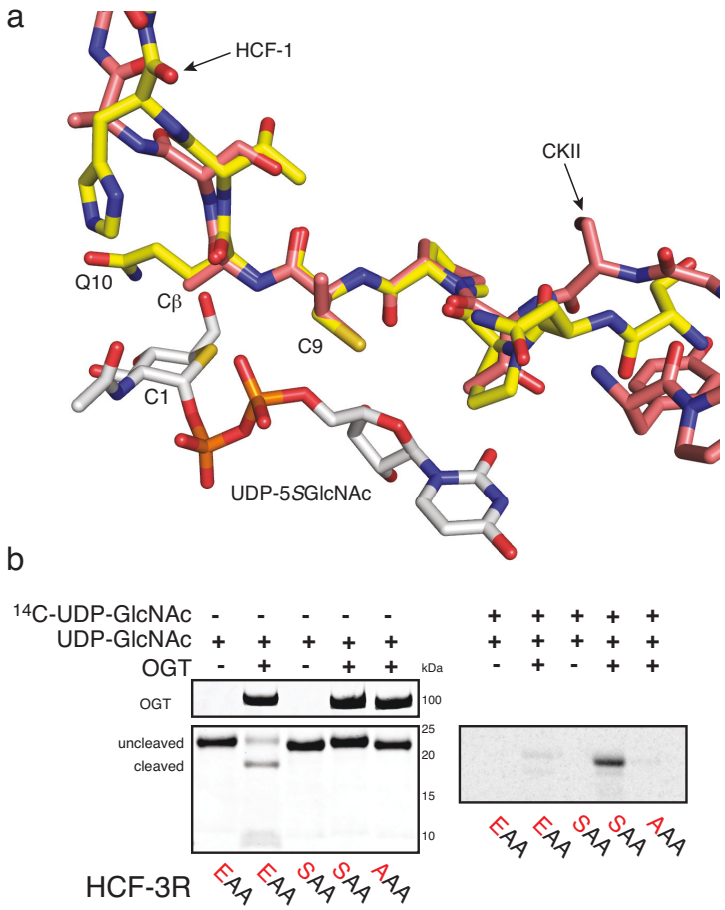


Figure 2.10: Glutamate 10 overlays with a glycosyl acceptor. (a) Overlay of substrate analogs from the ternary complexes hOGT4.5:UDP-5*S*GlcNAc:Q10-26 (PDB: 4N3B) and hOGT4.5:UDP-5*S*GlcNAc:CKII-A (PDB: 4GYY) [140]. UDP-5*S*GlcNAc is only shown from one structure as they overlay almost perfectly. The HCF repeat peptide is shown as yellow sticks, whereas the CKII peptide is shown as salmon sticks. (b) Mutating Glu10 to serine within the first repeat of HCF-3R (HCF-3R-SAA) converts a cleavage substrate into a glycosylation substrate. HCF-3R-SAA is not cleaved by OGT. (left) Cleavage products are separated by SDS-PAGE and visualized by Coomassie Blue. (right) Glycosylation of HCF substrates was carried out with radiolabeled UDP-GlcNAc; reaction mixtures were separated by SDS-PAGE and visualized by autoradiography.

2.7 Role of OGT active-site residues on cleavage and glycosylation

Building on the work of others, I sought to compare a set of OGT mutants for their ability to catalyze Ser/Thr glycosylation and HCF-1 cleavage. I wanted to simultaneously examine whether there were differences between two types of glycosylation substrate: one that we knew bound within the TPR domain of OGT (HCF-3R-SAA), and the well-studied, heavily glycosylated, protein nucleoporin62 (Nup62) [98, 160, 162], which may or may not bind within the TPRs. Four OGT active site residues implicated in Ser/Thr glycosylation were evaluated: K842, which is involved in binding and activation of UDP-GlcNAc for glycosyl-transfer; H498, which contacts the C2-*N*-acetyl group of UDP-GlcNAc; H558, which contacts a backbone carbonyl of glycosylation substrates, and D554 which had been implicated in transporting the proton of the acceptor hydroxyl out of the active site (Figure 2.11) [36, 108, 142, 140, 171, 213, 257]. Previous work showed that OGT constructs wherein K842 was mutated were impaired for glycosylation of: HCF-1rep1 [141] (ncOGT K842M), Tab1 protein [213] (hOGT4.5 K842M), and the Nup62 protein [171] (ncOGT K842A). I confirmed that OGT (K842M) is also impaired for Nup62 glycosylation, and unable to effect the cleavage of an HCF-3R construct (Figure 2.11). Its importance in cleavage suggests that UDP-GlcNAc activation is a key component of the cleavage mechanism [141]. It is noteworthy that all of these assays are an end-point measurement, wherein enzymatic deficiencies may be marginalized if the time point taken is sufficiently long. Substitution of H498 or H558 with alanine decreased S/T glycosylation, as evidenced by both a weaker signal and a mass-shift for the glycosylated Nup62 protein, which was consistent with previous data that showed an impairment of these mutants for glycosylation of

a CKII-derived peptide [142]. In contrast, these mutations appeared to have little effect on cleavage of HCF-3R-EAA after 6 hours (Figure 2.11). The use of HCF-1rep1 as a substrate showed a similar trend with impairment of S/T glycosylation, but minimal impact on cleavage [141]. Substitution of D554 with alanine or asparagine showed an impairment for Ser/Thr glycosylation much like H498A and H558A, when Nup62 was used, but a more dramatic impairment was observed when HCF-3R-SAA was used (Figure 2.11). Previous data showed that OGT (D554A) was impaired for glycosylation of a CKII-derived peptide (Jiaoyang Jiang, unpublished) and for S/T glycosylation of HCF-1rep1 [117]. We reasoned that if D554 serves a role in proton transport out of the active site, this differential reactivity (for cleavage vs. S/T glycosylation) observed for D554(A/N) may be a consequence of the reactive glutamate side chain existing as a carboxylate, thereby making a "catalytic base" within OGT dispensable.

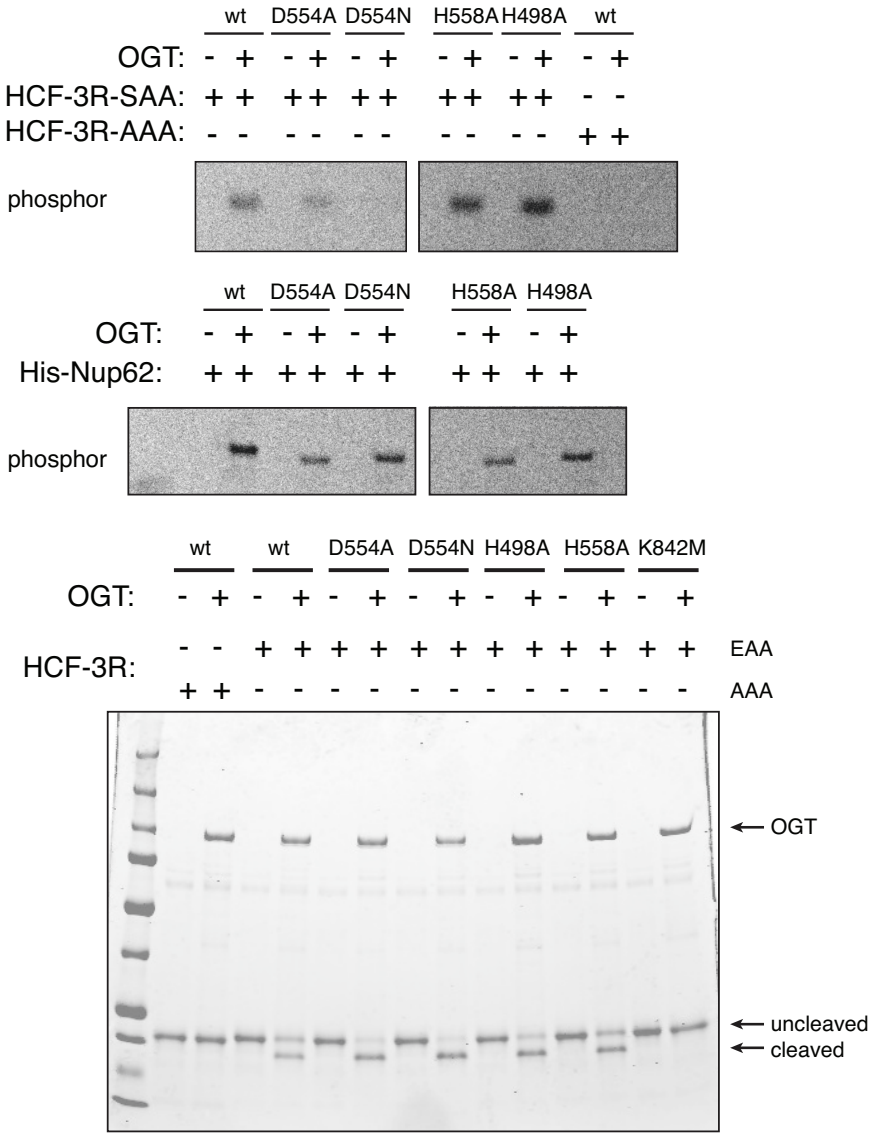


Figure 2.11: OGT variants impaired in glycosylation but proficient at cleavage exist. Glycosylation of HCF-3R-SAA (top) and nucleoporin62 (middle) by various OGT point mutants. Glycosylation reactions were separated by SDS-PAGE and visualized by autoradiography. Cleavage of HCF-3R-EAA (bottom) by various OGT point mutants. Cleavage reactions were separated by SDS-PAGE and visualized by Coomassie Blue.

2.8 Structure-activity studies of the HCF-1/OGT interaction

The HCF substrates employed up this point have been polypeptides on the order of 150 amino acids (HCF-3R), or small proteins ≈425 amino acids (HCF-1rep1) [141]. Even though we had structural information about the HCF-1/OGT interaction from X-ray crystallography, a 26 amino acid peptide (26mer, see Table B.1) was not a cleavage substrate in solution. I set out to determine both the minimal OGT construct capable of effecting HCF-1 cleavage, and the minimal HCF-1 construct that would serve as a cleavage substrate for *in vitro* experiments.

I tested whether the position of the reactive repeat (the one bearing glutamate at position 10, vs. alanine) mattered (Figure 2.12). HCF-3R-EEE showed the greatest fraction of cleavage products, relative to uncleaved polypeptide. HCF-3R-EAA was a better substrate than HCF-3R-AEA, and HCF-3R-AAE was entirely unreactive under these reaction conditions. These findings suggested that something C-terminal of the cleavage site was required for robust cleavage. Notably, when Jiaoyang Jiang had previously examined the cleavage products of HCF-3R-EEE by LC-MS, she noted that fragment "5", as shown in Figure 2.2, was not observed. Given that the primary sequences for each of the human repeats (Figure A.3) are almost identical, taken together these data suggest that recognition of the HCF repeat extended beyond the 26 amino acid consensus sequence.

I prepared a series of constructs (Table B.6) where the order of repeats in HCF-3R was permuted (Figure 2.13). The region labeled "spacer" corresponds to the 36 amino acids between the the 26th residue of repeat 1 and the first residue of repeat 2. Attempts to express constructs consisting of only a single repeat were unsuccessful. By comparing two

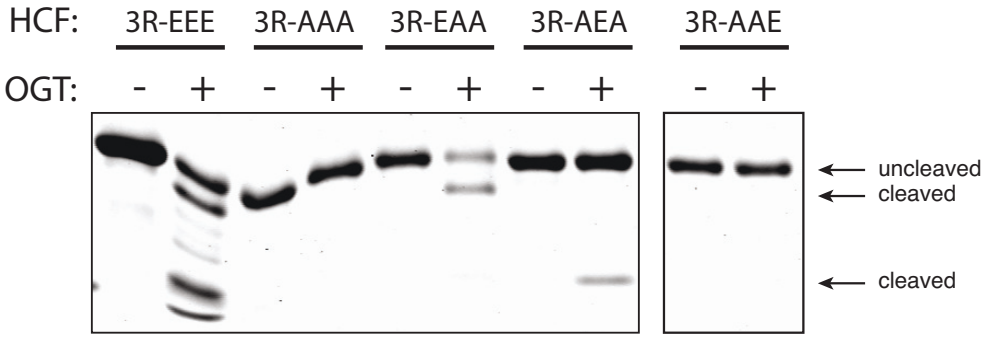


Figure 2.12: Cleavage ability of a repeat depends on its position within HCF-3R. Cleavage of various single-site HCF-3R constructs with wild-type OGT. Cleavage reactions were separated by SDS-PAGE and visualized by Coomassie Blue.

constructs that each bear only a single repeat (C15 vs. C18), it becomes evident that the position of the repeat sequence within the polypeptide was extremely important. Further, it appeared that in C19 cleavage had occurred at *both* repeats, suggesting that the lack of reactivity seen for HCF-3R-AAE was not due to human rep3 being non-functional. In order to confirm this, I analyzed the the reaction products of C19 cleavage by LC-MS. Fragments corresponding to cleavage at both repeats were observed; however, fragments that would correspond to cleavage *only* at the C-terminal repeat (rep3) were not observed. These data provided further evidence that 26 amino acids were not sufficient to define the cleavage site, but also, given the reactivity of C15, that the the minimal sequence required was not more than 36 amino acids longer than a single repeat.

I prepared a series of OGT constructs wherein the TPRs were sequentially truncated (Table B.7). With HCF-3R-EAA as a cleavage substrate, there was a dramatic effect of OGT TPR length on cleavage proficiency (Figure 2.14). Cleavage was not detectable for constructs bearing fewer than 6.5 TPRs, and very weak for OGT7.5 (Δ6 TPR). The activity of an OGT construct bearing 9 TPRs was similar to that of full-length OGT. The

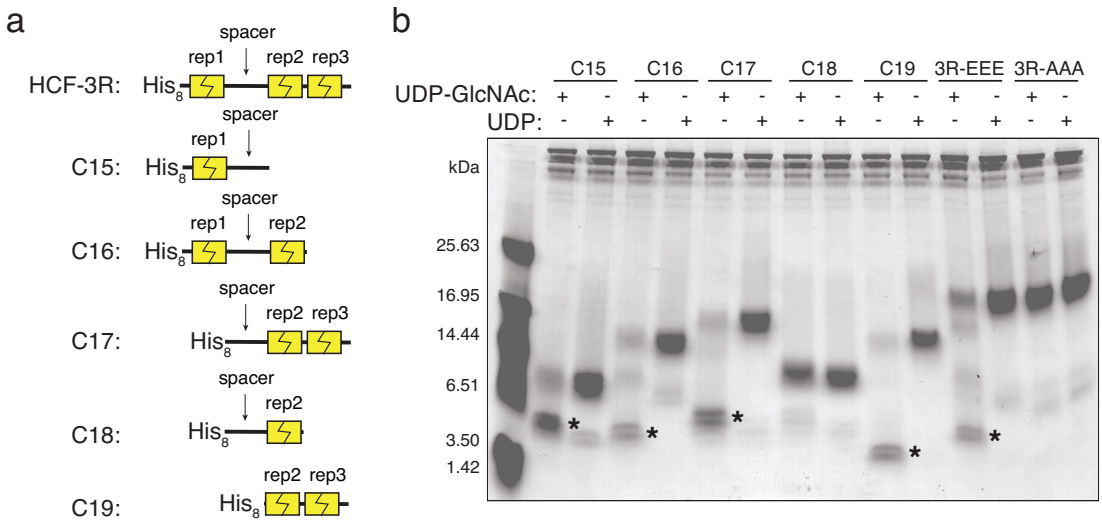


Figure 2.13: Repeats shorter than HCF-3R are suitable substrates. Cleavage of various HCF-3R truncation constructs with wild-type OGT. Cleavage reactions were separated by SDS-PAGE and visualized by Coomassie Blue. *denotes cleavage product(s).

same trend was not observed for glycosylation of the related peptide, HCF-3R-SAA. This suggested that cleavage and glycosylation have different dependencies on the TPRs. During the course of these studies, similar findings on dependence of TPR length for cleavage was published [117].

2.8.1 A peptide substrate for HCF-1 cleavage

Given that the cleavage activity of OGT bearing 9 TPRs was comparable to that of full-length OGT, we reasoned that the minimal number of contacts for an HCF substrate to be cleavable were confined within this region. In a composite structural model of full-length OGT (PDBs 1W3B+4N3B), the distance from the most C-terminal residue of the HCF-1 peptide in PDB 4N3B (Thr24) to the end of TPR 9 of OGT was less than 30 Å. Given that the approximate distance between two side chains in a linear dipeptide is ≈7 Å we

reasoned that appending several amino acids may be sufficient to convert the 26mer into a viable cleavage substrate. I obtained two peptides, one with 5 amino acids appended to the C-terminus of the 26mer, and one with 10 amino acids appended, and incubated them with wild-type OGT. Indeed, the peptides with C-terminally appended residues were better substrates for cleavage by OGT (Figure 2.15), with the 32 amino acid peptide undergoing ≈60% cleavage in four hours at 37 °C.

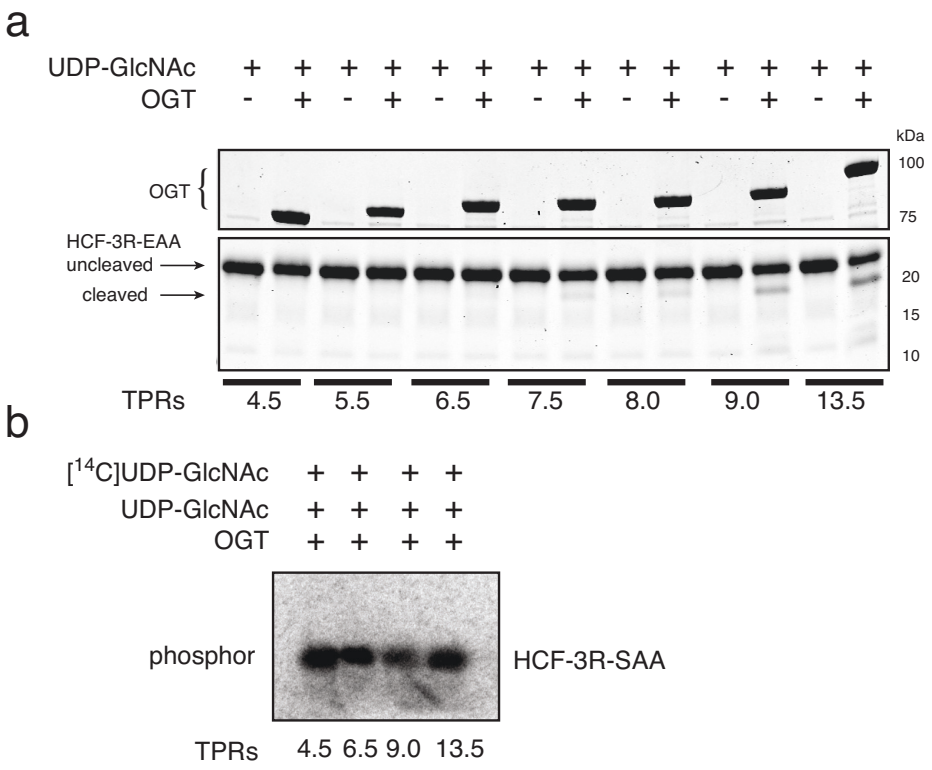


Figure 2.14: TPR length in OGT is important for cleavage, but not glycosylation. Cleavage of HCF-3R-EAA by various OGT constructs. The stated number of TPRs are counted from the catalytic domain of OGT (from C-terminus), e.g. 4.5 corresponds to the hOGT4.5 construct used for structural studies. Reactions were separated by SDS-PAGE; cleavage reactions were visualized by Coomassie Blue and glycosylation reactions visualized by autoradiography.

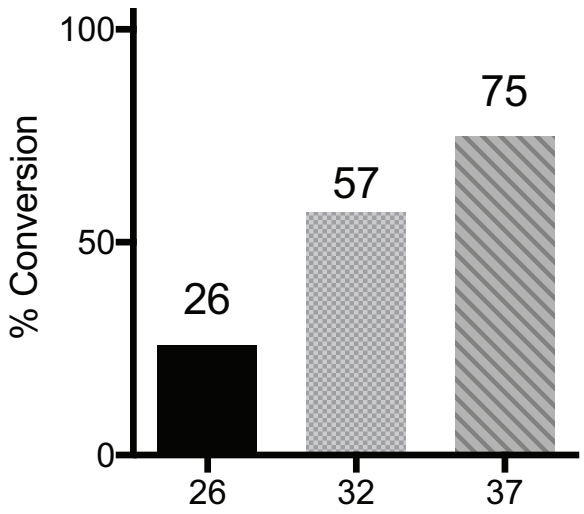


Figure 2.15: A 32 amino acid HCF peptide is a suitable cleavage substrate. Using LC-MS, conversion was quantified by comparing the fraction of starting material remaining after a four hour incubation under standard conditions. "26", "32", and "37" denote the peptides: 26mer, 32mer and 37mer, respectively (see Table B.1).

2.9 Possible mechanisms for cleavage

Regarding the mechanism of cleavage, previous hypotheses suggested that OGT contains a dedicated protease active site or acts as a co-protease to template HCF-1 autocatalysis [27]. Instead, OGT promotes cleavage of the HCF repeat using the same catalytic region as for glycosylation. The C-terminal threonine-rich region of the HCF repeat binds within the lumen of the TPR domain of OGT, stabilized by an extensive network of hydrogen bonds. When bound, the cleavage region resides within the active site of OGT and binds above UDP-GlcNAc in the same conformation that a glycosylation substrate would, with the glutamate side chain positioned near the anomeric carbon of the sugar. Since a pyroglutamate product is formed, and spontaneous cyclization of N-terminal glutamates is kinetically very slow, on the order of 0.001 day^{-1} [156, 235], the glutamate side chain is likely activated by formation of an ester species as part of the cleavage mechanism. We speculate that the glutamate side chain traps a transient oxocarbenium ion formed within the active site, producing a glutamyl ester that can undergo intramolecular attack (Figure 2.16, initial step), leading ultimately to formation of the N-terminal pyroglutamate. Three possible mechanisms for cleavage proceeding from a glutamyl ester are suggested (Figure 2.16). Though Cys9 was shown not to be required for cleavage (see section 2.5.2), this does not preclude mechanisms 2 and 3 from being operative when Cys9 *is* present. Based on the results of the C9X mutagenesis experiments, it seems plausible that Cys9 plays a role in transferring a proton to the GlcNAc leaving group, or interacting with the glycosyl ester to facilitate attack by the amide nitrogen.

Although direct physical evidence for a glutamyl-sugar intermediate has yet to be obtained, glycosylation of the glutamate side chain prior to cleavage is consistent with the structural data, the strict requirement for glutamate at the cleavage site, the formation

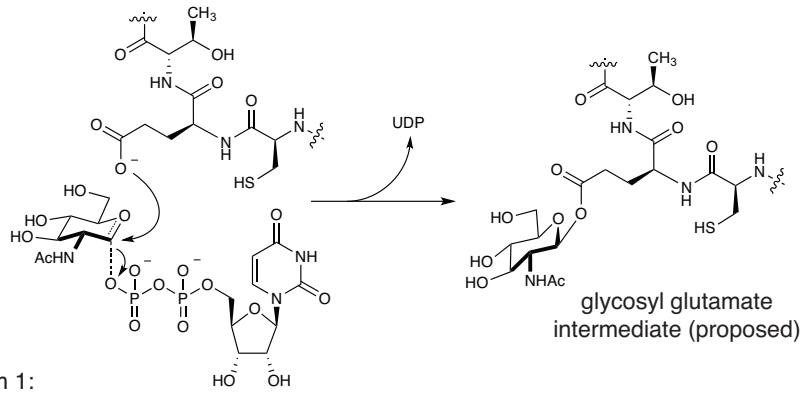
of a pyroglutamate product, and the observation that cleavage of HCF-1 repeats requires UDP-GlcNAc and depends on an OGT residue, K842, which is essential for catalyzing glycosylation [141].

2.9.1 α_2-macroglobulin as an example of internal pyroglutamate hydrolysis

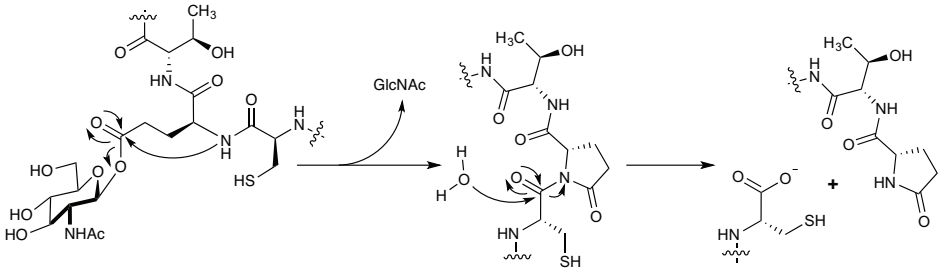
Though enzymatic glycosylation on glutamate has not, to our knowledge, been reported (see also discussion in section 2.9), there is precedent for the spontaneous hydrolysis of internal pyroglutamates that form by other mechanisms (Figure 2.17) [58, 122, 123, 124]. In this system, a model of the α_2-macroglobulin metastable binding site, the residues Cys-Gly-Ala-Glu exist as an equilibrium mixture of a internal pyroglutamate and macrocyclic thioester. Kinetic analysis showed that the preferred pathway is hydrolysis of the internal pyroglutamate, to produce two fragments, rather than thioester hydrolysis to regenerate the free cysteine and glutamic acid residues.

Figure 2.16: *(On the following page)* Possible cleavage mechanisms that proceed via a glycosylated glutamate. We speculate that cleavage occurs after formation of a glutamyl-sugar intermediate. After this initial step there are several ways this species could proceed to cleavage. We outline three possible mechanisms here. Mechanisms 2 and 3 implicate Cys9 through the formation of thioester intermediates. Other biological mechanisms with peptide backbone cleavage involve cysteines or pyroglutamate [5, 58, 74, 122, 123, 124]. However, the fact that the Cys9 is not essential for cleavage of HCF-1 (see section 2.5.2) supports the first mechanism, although other mechanisms are possible. The formation of a β-thiolactone species has been observed previously from activated cysteine-like esters [43].

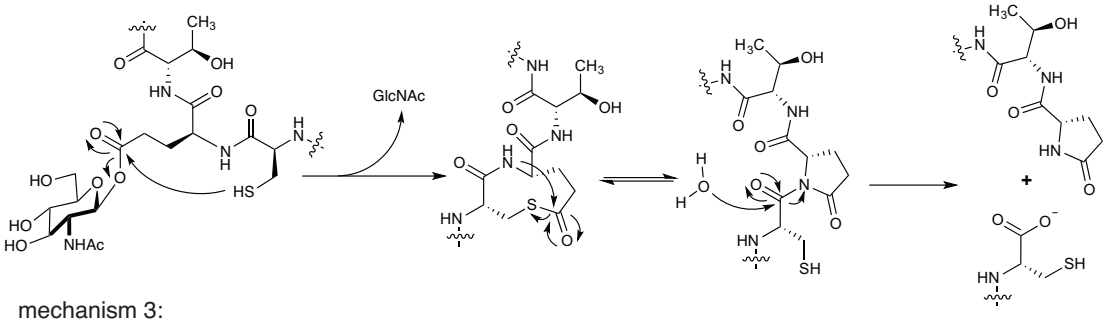
Initial step:



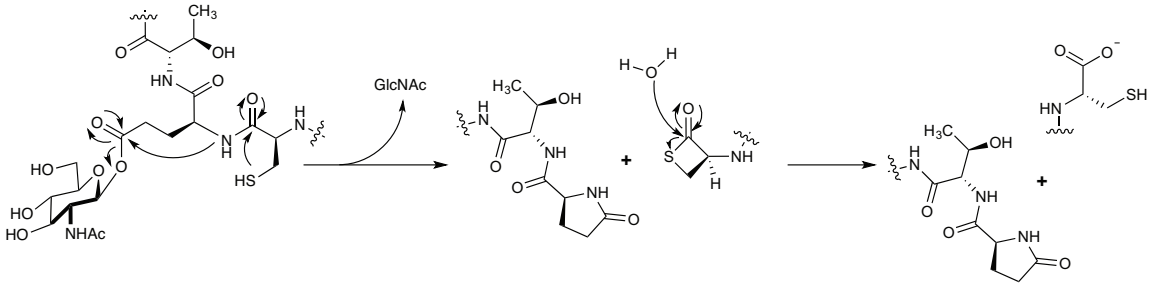
mechanism 1:



mechanism 2:



mechanism 3:



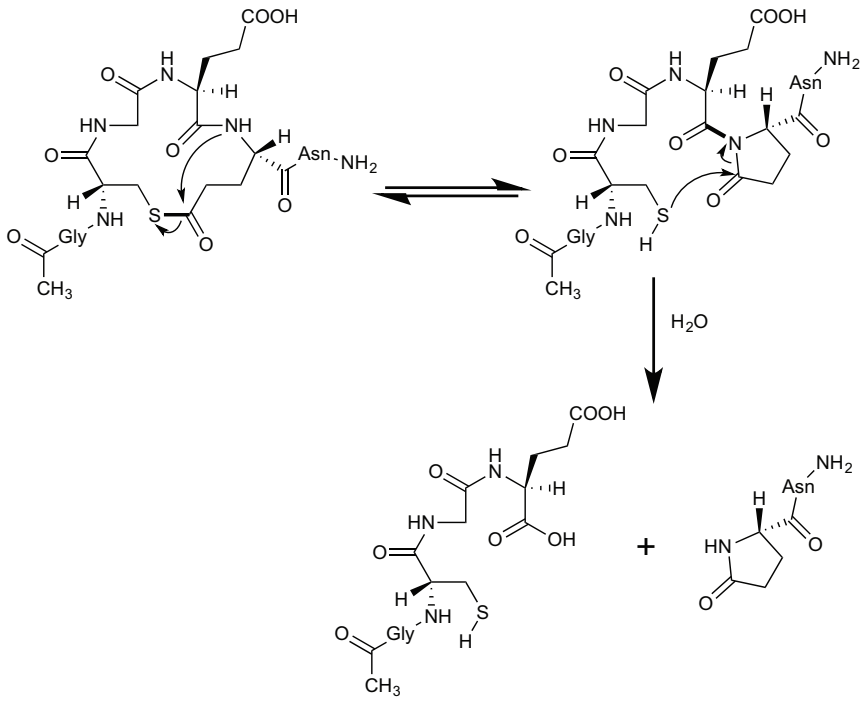


Figure 2.17: α_2-macroglobulin cleavage produces a pyroglutamate. The mechanism by which the metastable binding site in α_2-macroglobulin undergoes spontaneous fragmentation [58, 122, 123, 124].

2.10 Conclusions

The structural studies described here, involving the HCF-1 proteolytic repeat bound to OGT, provided the first information about how the TPRs of OGT recognize at least some of its substrates [108]. The structures show an elaborate network of hydrogen bonds that envelop the threonine-rich C-terminus of the HCF-1 repeat within the TPRs of OGT. Within this region, HCF-1 is bound in an extended conformation with the highly conserved ladder of OGT asparagine residues engaging the peptide, mostly through interactions with the amide backbone. By engaging the peptide backbone, this asparagine network positions the side chains, which in this case are largely threonines, to make additional hydrogen bonds to polar residues, mostly aspartates, that line the interior of the TPR cavity; this likely provides sequence selectivity as large or hydrophobic residues would be disfavored. These structures confirmed that there is indeed an analogy between OGTs mode of substrate binding and that of some other proteins that recognize multiple related ligands [111]. However, it remains to be seen whether other substrates, adapters or enhancers interact with OGT analogously [108].

Overall, these studies have provided information about an unprecedented mechanism of proteolysis in which OGT uses UDP-GlcNAc as a co-substrate to effect the cleavage of HCF-1. Furthermore, these studies revealed a remarkable similarity between how a cleavage substrate and how a glycosylation substrate interact with OGT. Through the use of site-directed mutagenesis, we identified key residues in both OGT and the HCF repeat. Using a series of HCF truncation constructs, we were ultimately able to identify a short peptidic substrate for future experiments. We suggested that cleavage occurs via the intermediacy of an unusual glutamyl GlcNAc ester. We also showed that two very different post-translational protein modifications proteolysis and addition of a sugar residue

(Ser/Thr glycosylation) can occur in the same active site with the outcome determined by the identity of a single amino acid in the polypeptide substrate.

Acknowledgments

I am grateful for support from the National Science and Engineering Research Council (NSERC) of Canada as a PGS-M and PGS-D3 fellowship recipient. This work was funding by a National Institutes of Health grant (R01 GM094263) to Suzanne Walker. Though many of their experiments were not discussed here, Vaibhav Kapuria and Tanja Bhuiyan in the lab of Winship Herr (Université de Lausanne, Switzerland) contributed a great deal to this work, and to our understanding of HCF-1 cleavage. Wesley F. Zandberg and David J. Vocadlo (Simon Fraser University, Vancouver, BC, Canada) provided UDP-5SGlcNAc and D.J.V offered useful comments on portions of this work. Amanda McFedries and Prof. Alan Saghatelian (now at the Salk institute) provided access to and assistance with a q-ToF mass spectrometer in the Saghatelian lab, used for the initial peptide cleavage experiments.

X-ray diffraction data obtained by Michael B. Lazarus were collected at the National Synchrotron Light Source at Brookhaven National Laboratory (Beamline X25). Coordinates and structure factors were deposited under accession codes 4N39, 4N3A, 4N3B, and 4N3C in the PDB. X-ray diffraction data obtained by John Janetzko were collected at the Advanced Photon Source at Argonne National Laboratory (Beamline 24 ID, NE-CAT). Coordinates have not yet been deposited in the PDB.

2.11 Methods

2.11.1 Material Sources

DNA for HCF-3R-EEE, -AAA, -SAA, -EAA, and -AEA were obtained from Jiaoyang Jiang or Michael B. Lazarus. DNA for hOGT4.5 (Δ9 OGT), Δ8 OGT, Δ7 OGT and OGT (A624T) were obtained from Michael B. Lazarus. His-Nup62 was obtained from Benjamin J. Gross who expressed and purified it according to previously published methods [162]. UDP-5*S*GlcNAc was obtained from Wesley F. Zandberg and David J. Vocadlo, who prepared it as previously described [72].

2.11.2 Reagents and General Methods

^{14}C-UDP-GlcNAc was obtained from American Radiochemicals (ARC). UDP-GlcNAc was obtained from Sigma as a lyopholized solid, and reconstituted in deionized water (to ≈25 mM). THP [Tris(3-hydroxypropyl)phosphine] was obtained from Novagen as a 0.5 M solution in water. Calf alkaline phosphatase (CIP) was obtained from New England Biolabs. Peptides were obtained from Biomatik (Cambridge, ON, Canada); their purity was evaluated by HPLC and mass spectrometry, and they were used as received. Lyophilized peptides were dissolved in water to ≈5 mM and added to reactions according to the final concentration desired. Peptide concentrations, for peptides with an N-terminal Tyr, were determined by measuring A_{274} (λ_{max}, tyrosine) in water, assuming an extinction coefficient of 1400 $M^{-1} \cdot cm^{-1}$. A complete list of peptides used in this study is provided in Table B.1.

2.11.3 Cloning and gene construction

Wild-type OGT, based on *Homo sapiens* (human) nuclear and cytoplasmic OGT (ncOGT), was subcloned from OGT (A624T) into a pET 24b vector. In instances where

only "OGT" is stated, wild-type enzyme was employed. This construct bears an N-terminal T7 tag followed by octa-His and an HRV3C protease site upstream of the gene. For expression, OGT plasmids were transformed into *E. coli* BL21(DE3) ΔArnA (made by Michael B. Lazarus); HCF plasmids were transformed into *E. coli* BL21(DE3) for expression. All Constructs were passaged in NovaBlue cells. Cell stocks were stored as a glycerol stock (12.5% glycerol, final) at -80 °C. Point mutants, and truncations were prepared by Quickchange mutagenesis from the respective parental constructs.

2.11.4 Production of HCF-3R and truncation constructs

HCF-3R constructs (Tables B.2, B.3, B.4, B.5) and HCF truncation constructs (Table B.6) were purified from *E. coli* BL21(DE3) as follows: LB media, supplemented with 50 μg/mL kanamycin, was warmed to 37 °C and inoculated with a 1:100 dilution of an overnight culture. Growth was monitored at OD_{600}, and once the OD reached ≈1.1 the temperature was reduced to 16 °C over 30 minutes. Overexpression was initiated by addition of IPTG to a final concentration of 0.2 mM and allowed to proceed for 16 h at this temperature. After this time, cells were harvested by pelleting at 5,000×g for 20 min (4 °C). Cell pellets were resuspended in 1×TBS (150 mM NaCl and 20 mM Tris) pH 7.4, supplemented with: 0.1 mg/mL lysozyme, 0.1 mg/mL DNAse I and 1 mM PMSF. Cells were lysed using a cell disruptor. Lysates were supplemented with imidazole to a final concentration of 40 mM and clarified by centrifugation at 15,000×g for 20 minutes (4 °C). Clarified lysates were flowed over Ni-NTA agarose resin that had been equilibrated with 1×TBS supplemented with 40 mM imidazole and at pH 8. The flow-through was collected, and re-applied to the resin. After applying the lysate the second time, the resin was washed with 10 column volumes of 1×TBS supplemented with 50 mM imidazole, pH 8. Lastly, the protein was eluted with 4

column volumes of 1×TBS supplemented with 250 mM imidazole, pH 8. The eluate was dialyzed (twice) against 1×TBS pH 8.0 using 3.5 kDa MWCO cassettes to remove imidazole. After dialysis, samples were collected and supplemented with THP to a final concentration of 1 mM. Constructs were evaluated for purity by SDS-PAGE and typically do not require additional purification. Concentrations of HCF-3R constructs were determined by DCTM (detergent compatible) protein assay (Bio-Rad). HCF-3R constructs were flash frozen and stored at -80 °C; once thawed, they were kept for up to one week at 4 °C.

2.11.5 Production of OGT

hOGT4.5, used for crystallography, was prepared as previously reported [142]. ncOGT constructs: wild-type OGT, OGT (D554N), OGT (K842M), OGT (D554A), OGT (H498A), and OGT (H558A) (sequences outlined in Table B.7) were purified from *E. coli* BL21(DE3) ΔArnA as follows:*E. coli* BL21(DE3) ΔArnA were grown in LB media, supplemented with 50 μg/mL kanamycin overnight. To LB pre-warmed to 37 °C, was added a 1:100 inoculum from the overnight culture. Growth was monitored at OD_{600}, and once the OD reached ≈1.1 the temperature was reduced to 16 °C over 30 minutes. Overexpression was initiated by addition of IPTG to a final concentration of 0.2 mM and allowed to proceed for 16 h at this temperature. After this time, cells were harvested by pelleting at 5,000×g for 20 min (4 °C). Cell pellets were resuspended in 1×TBS (150 mM NaCl and 20 mM Tris) pH 7.4, supplemented with: 0.1 mg/mL lysozyme, 0.1 mg/mL DNAse I and 1 mM PMSF. Cells were lysed using a cell disruptor. Lysates were supplemented with imidazole to a final concentration of 40 mM and clarified by centrifugation at 15,000×g for 20 minutes (4 °C). Clarified lysates were applied to Ni-NTA agarose resin that had been equilibrated with 1×TBS supplemented with 40 mM imidazole and at pH 7.4. Once applied to the

resin, the supernatant was incubated with end-over-end rotation for ≈60 minutes at 4 °C, after which the flow through was collected. The resin was washed with 10 column volumes of 1×TBS supplemented with 50 mM imidazole, pH 7.4. Lastly, the protein was eluted with 4 column volumes of 1×TBS supplemented with 250 mM imidazole, pH 7.4. The eluate was supplemented with THP to a final concentration of 1 mM. For ncOGT constructs, N-terminal tags were not cleaved. Instead, protein was directly purified by size-exclusion chromatography (SEC) (Superdex 200, GE Healthcare) with an isocratic elution using 1×TBS as a mobile phase. Fractions were pooled, supplemented with THP to a final concentration of 1 mM and evaluated for purity by SDS-PAGE. Generally, no further purification is required. OGT used in the experiments described below was quantified by NanoDrop (A_{280}) using an extinction coefficient, ϵ_{280} = 118,955 $M^{-1} \cdot cm^{-1}$, and assumed to have a molecular weight of 118,047 Da. OGT can be flash-frozen and stored at -80 °C. Once thawed, OGT activity does not seem to appreciably change over a week at 4 °C. Generally, OGT from the same batch (either thawed freshly, or thawed several days prior) should always be used for experiments that will be directly compared.

2.11.6 *In vitro* HCF-3R cleavage for SDS-PAGE analysis

In a 20 μL reaction volume was mixed: 1 mM UDP-GlcNAc, 1 mM THP, 100 μM HCF-3R construct, and 0.1 U CIP in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT, where added, was added to a final concentration of 2.67 μM, and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After 6 hours, reactions were quenched by the addition of an equal volume (20 μL) of 2×Laemmli loading buffer (with β-mercaptoethanol). Samples were boiled for 10 minutes, and pelleted at 15,000×g for 10 minutes. 5 μL of sample was loaded per lane

onto a 4-20% Tris-Gly gradient gel and run for 45 minutes at 200 V. The gel was stained with Coomassie Brilliant Blue overnight, and destained the following day until bands were clearly visible and a suitable background was obtained.

2.11.7 *In vitro* glycosylation for SDS-PAGE (autoradiography) analysis

In a 20 μL reaction volume was mixed: 1 mM UDP-GlcNAc (250 μM hot; 750 μM cold), 1 mM THP, 100 μM HCF-3R construct (or Nup62), and 0.1 U CIP in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT, where added, was added to a final concentration of 2.67 μM, and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After 5 hours, reactions were quenched by addition of an equal volume (10 μL) of 2×Laemmli loading buffer (with β-mercaptoethanol). Samples were boiled for 10 minutes, and pelleted at 15,000×g. 7 μL of sample was loaded per lane onto a 4-20% Tris-Gly gradient gel and run for 5 minutes at 90 V, followed by 65 minutes at 150 V. The gel was dried, using a gel drier and placed in an exposure cassette. A storage phosphor was placed on top and the phosphor and exposed for 24–36 hours before being imaged with a Typhoon imager.

2.11.8 *In vitro* HCF truncation cleavage for SDS-PAGE analysis

In a 20 μL reaction volume was mixed: 1 mM UDP-GlcNAc, 1 mM THP, 100 μM HCF-3R construct, and 0.1 U CIP in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT, where added, was added to a final concentration of 2.67 μM, and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After 5 hours, reactions were quenched by the addition of an equal volume (20 μL) of 2×Tricine loading buffer (with β-mercaptoethanol). Samples were boiled for 5

minutes, and pelleted at 15,000×g for 10 minutes. 5 μL of sample was loaded per lane onto a pre-cast 16.5% Tris-Tricine-SDS gel (Bio-Rad) and run for 105 minutes at 100 V. The gel was fixed, stained and destained according to manufacturer guidelines.

2.11.9 General LC-MS Methods

General measurements of peptides were performed on a q-Exactive Plus FT-mass spectrometer (Thermo Scientific^{TM}) equipped with an Ultimate 3000 uHPLC. Reaction mixture components were separated using a 33 minute reverse phase HPLC method (Table B.8); the last half of the LC method was empirically optimized to wash the column and re-equilibrate between runs to maintain column performance. with a Kinetex® C18 (2.6 μm particle size, 100 Å pore size, 2.1 mm ID, 150 mm length) at a constant flow rate of 0.200 mL/minute and column temperature of 35 °C. 5 μL of sample was injected, unless otherwise specified. Detection on the q-Exactive Plus mass spectrometer was performed between 300-2000 *m/z*, using an acquisition target of 3×10^6, maximum IT of 200 ms at a resolution of 70,000 for MS and 35,000 for MS/MS data.

2.11.10 *In vitro* HCF-1 cleavage for LC-MS analysis

Typical cleavage reactions were conducted as follows: 20 μL reaction volume with components having the following final concentrations: 1 mM UDP-GlcNAc, 1 mM THP, 12 μM HCF-1 peptide, 0.1 U CIP and 12 μM OGT in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT was added last and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After the desired reaction time, reactions were quenched by addition of four volumes (80 μL) of ice-cold methanol supplemented with 0.1% formic acid. Reactions were then centrifuged at

16,100×g for 10 minutes to remove any precipitated protein (4 °C), and transferred to fresh tubes. Samples were concentrated to dryness using a speedvac (with T = 30 °C), for ≈60 minutes. Samples were re-suspended in 100 μL 0.1% formic acid, mixed by vortexing, and centrifuged for 10 minutes at 15,000×g (4 °C) to remove any insoluble material. Samples were then transferred to LC vials, and maintained at 4 °C until they were analyzed.

2.11.11 Crystallization of ternary complexes

All complexes were prepared by incubating hOGT4.5 at 7 mg/ml with 1 mM nucleotide-sugar and 3 mM peptide. All crystals were grown with the hanging drop method by combining 2 μl of protein complex with 1 μl of reservoir solution. Crystals of the ternary complex containing OGT and either UDP-GlcNAc or UDP-5SGlcNAc and a 26 amino acid HCF peptide were grown with reservoir solution consisting of 1.4 M sodium malonate pH 7.0 and 0.1 M Bis-Tris Propane pH 7.0. Crystals were grown at room temperature and cryoprotected in solutions consisting of the reservoir solution plus 28% xylitol and flash frozen in liquid nitrogen.

2.11.12 Data Collection, Structure Determination, and Refinement

Data for HCF mutants were collected at APS beamline 24-ID at Argonne National Laboratory. Data was processed using iMosflm [13] and scaled with Scala or Aimless in the CCP4 crystallography suite [256]. The PHENIX software package was used for molecular replacement and refinement [1]. The previously determined OGT-HCF ternary complex (PDB: 4N3B) was used as a search model for molecular replacement, using the program Phaser. Models were subsequently refined by rigid body refinement and multiple rounds of simulated annealing, minimization, atomic displacement parameter (ADP or B-factor)

refinement and TLS refinement (determined using the TLSMD server [193]), with interspersed manual adjustments using Coot [57]. Geometric restraints for UDP-GlcNAc and UDP-5*S*GlcNAc were generating using PHENIX Elbow, and these restraints were used throughout refinement. All structural figures were made with Pymol [214] except for the electrostatic surface figures, which were created with CCP4mg [178]. Crystallographic software accessed through SBGrid [183].

2.11.13 HCF sequence bioinformatics

HCF-1 repeat-like sequences were identified by BLAST using a query of "CSNPPCETHETGTTNTATTT" using the program Geneious. The database search was performed on the non-redundant (nr) sequence database using the "blastp" search program. A maximum of 1,000 hits were retrieved. Hits were scored according to the BLOSUM80 matrix with a gap cost of 10 1 (Open Extend), a word size of 2 and a maximum E-value of 1×10^{-5}. Sequence representations were prepared using Geneious or WebLogo [44].

Chapter 3

Mechanism of HCF-1 cleavage

> "In God we trust; all others must bring data."
>
> – W. Edwards Deming

A portion of this chapter originally appeared as [107]:
"How the glycosyltransferase OGT catalyzes amide bond cleavage" **John Janetzko**, Sunia A. Trauger, Michael B. Lazarus and Suzanne Walker *Nature Chemical Biology.* Volume 12, Number 11: 892–893. 2016.

Contributions: Except for the following items, I performed all experiments described in this chapter. Michael B. Lazarus expressed, purified, crystallized and solved the structure of the hOGT4.5 (D554N):UDP:glycoCKII ternary complex. Jacob N. Sanders ran computations with my input. Sunia A. Trauger helped me collect and analyze some of the mass spectrometry data.

Abstract

The essential human enzyme OGT, known for modulating the functions of nuclear and cytoplasmic proteins through Ser/Thr glycosylation, was unexpectedly implicated in the

proteolytic maturation of the cell cycle regulator HCF-1. In this chapter I will describe my work that showed that HCF-1 cleavage occurs via glycosylation of a glutamate side chain followed by on-enzyme formation of an internal pyroglutamate, which undergoes spontaneous backbone hydrolysis. I will discuss the state of our mechanistic understanding, as well as remaining questions.

3.1 Introduction

Host cell factor-1 is an essential chromatin-associated protein important in cell cycle control [196, 236]. It contains six centrally located 26 amino acid proteolytic repeats, each consisting of a threonine-rich region and a region that undergoes cleavage (Figure 3.1) [27, 254, 255]. Proper cell cycle regulation requires HCF-1 cleavage [115, 165]. The protease responsible for cleaving HCF-1 went unidentified for many years until studies implicated O-GlcNAc transferase (OGT) in the process [27, 47]. OGT is an essential nucleocytoplasmic glycosyltransferase best known for its role as a nutrient sensor [20, 92, 108]. The discovery that OGT is involved in HCF-1 processing raised questions about how a glycosyltransferase might promote amide bond hydrolysis.

OGT cleavage of HCF-1 gives a C-terminal product containing an N-terminal pyroglutamate (Figure 3.1) [141], implying that the cleavage mechanism involves an intermediate in which the glutamate side chain is activated, e.g., through formation of an ester. As UDP-GlcNAc is required for cleavage [27, 141], but a non-hydrolyzable analog inhibits cleavage [141], we proposed that OGT glycosylates the glutamate side chain (Figure 3.2). A mechanism involving glutamate glycosylation would serve to link HCF-1 cleavage to nutrient status through the abundance of UDP-GlcNAc. Other enzymes that use cellular metabolites as co-substrates in amide bond cleavage reactions have been described. For

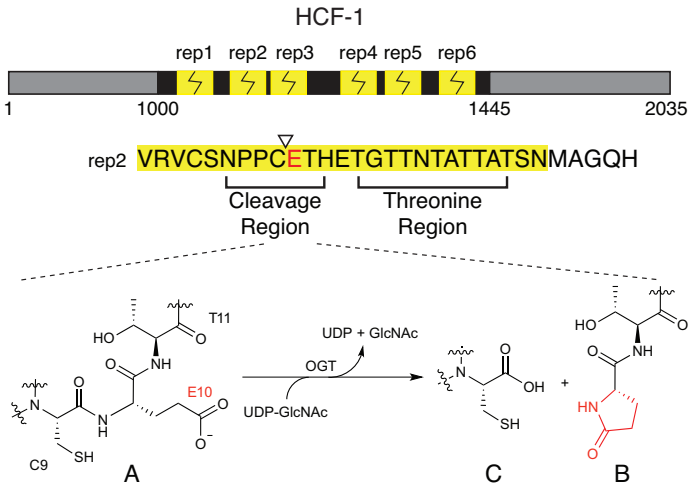


Figure 3.1: HCF-1 contains six similar proteolytic repeats that are cleaved between residues C9 and E10, producing a C-terminal pyroglutamate-containing product, C, that was proposed to form via the intermediacy of a glycosyl ester [141].

example, sirtuins use NAD^+ to activate the acetyl-lysine amide bond, thus linking cellular nutrition to protein deacetylation [15, 61, 240]. To the best of our knowledge, prior to this work, no NT-sugar glycosyltransferases had been reported to catalyze glycosylation on glutamate, and we did not have direct evidence for this intermediate or for other intermediates that would illuminate the cleavage mechanism.

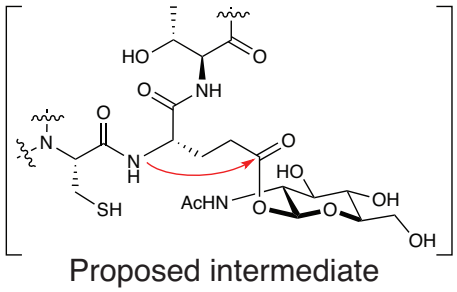


Figure 3.2: Proposed glycosyl glutamate linkage.

3.2 Development of an LC-MS-based HCF-1 cleavage assay

In order to identify reaction intermediates, I required an assay with higher resolution than our previous gel-based assay; this would be imperative for the detection of low-abundance molecular species, or those expected to be of (relatively) similar molecular-weight to the starting peptide, and therefore likely indistinguishable by SDS-PAGE. In section 2.8.1 I showed that a 32 amino acid peptide could serve as a suitable substrate to study HCF-1 cleavage by OGT.

At the outset of this work, there were two main ideas for how to demonstrate the intermediacy of a glycosyl glutamate linkage. The first, and most compelling, would involve direct observation of such a species. However, given the proposed instability of such an intermediate [129], a second approach, involving transfer of an ^{18}O label from the carboxylate side chain of E10 to the anomeric carbon of GlcNAc would allow us to infer that a covalent bond had formed between them during the cleavage process (Figure 3.3). Given

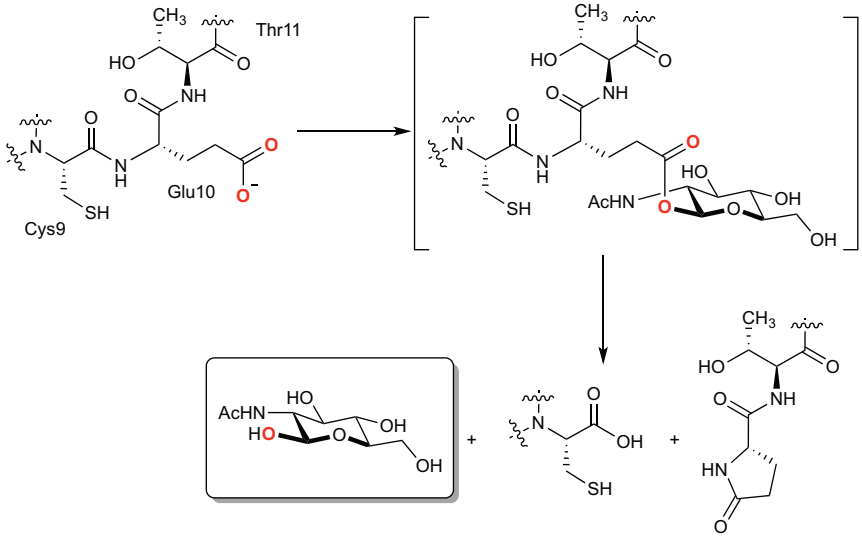
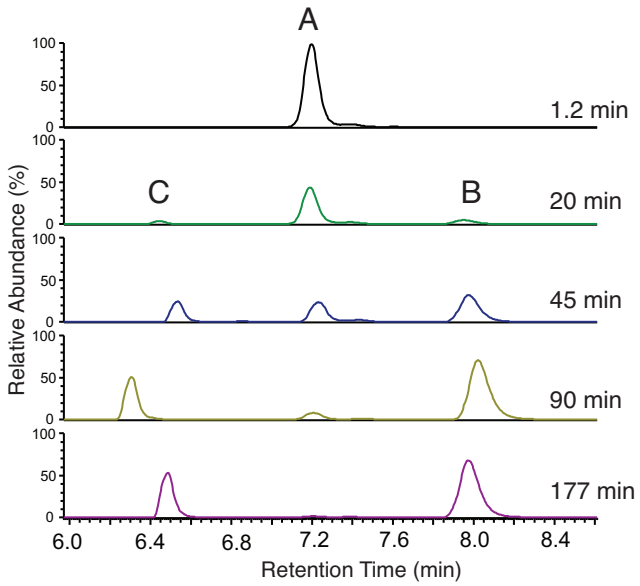


Figure 3.3: Proposed ^{18}O transfer experiment. ^{18}O labeled atoms are shown in red.

my intentions of conducting ^{18}O experiments, I designed an HCF-1 substrate, E10-32, (See Table B.1) wherein the only carboxylate was that of E10. E13 (of the native rep2 sequence) was substituted with Gln, and the C-terminus was amidated. Further, an N-terminal Tyr was added to allow for absorbance-based standardization of peptide concentration.

I monitored cleavage E10-32 as a function of time by high-resolution Fourier-transform mass spectrometry (FTMS) (discussed in section 1.7) and found that under these conditions the reaction was complete within about three hours (Figure 3.4, see also Figure A.4). Integrating individual extracted ion chromatograms (EICs) shows the disappearance of starting material (fit to a first-order decay, Equation 3.1), and increases in B and C, which appear to reach a maximal signal around 90 minutes before plateauing (Figure 3.5).

$$Y = Ae^{-kX} + b \tag{3.1}$$



peptide	*m/z*
A: YV[1]RVCSNPPCETHQTG[15] ... [31]	845.1319 (z = 4+)
B: YV[1]RVCSNPPC[9]	569.2626 (z = 2+)
C: pE[10]THQTG[15] ... [31]	747.6698 (z = 3+)

Figure 3.4: E10-32 undergoes cleavage. A time course for cleavage of E10-32 by wild-type OGT is shown; traces are EICs for the exact masses of species A, B, and C (within ± 5 ppm). The peptide sequence flanking the cleavage site and the *m/z* of each of the three species are shown under the traces. "pE" denotes the pyroglutamate residue. The 32 amino acid cleavage substrate was designed with an N-terminal Y to enable concentration determination, but the sequence shown is numbered starting from the second amino acid to be consistent with the numbering used to describe HCF-1 repeats. The peptide sequence continues after G^{15} as in the repeat in Figure 3.1; the complete sequence of the cleavage substrate is provided in Table B.1.

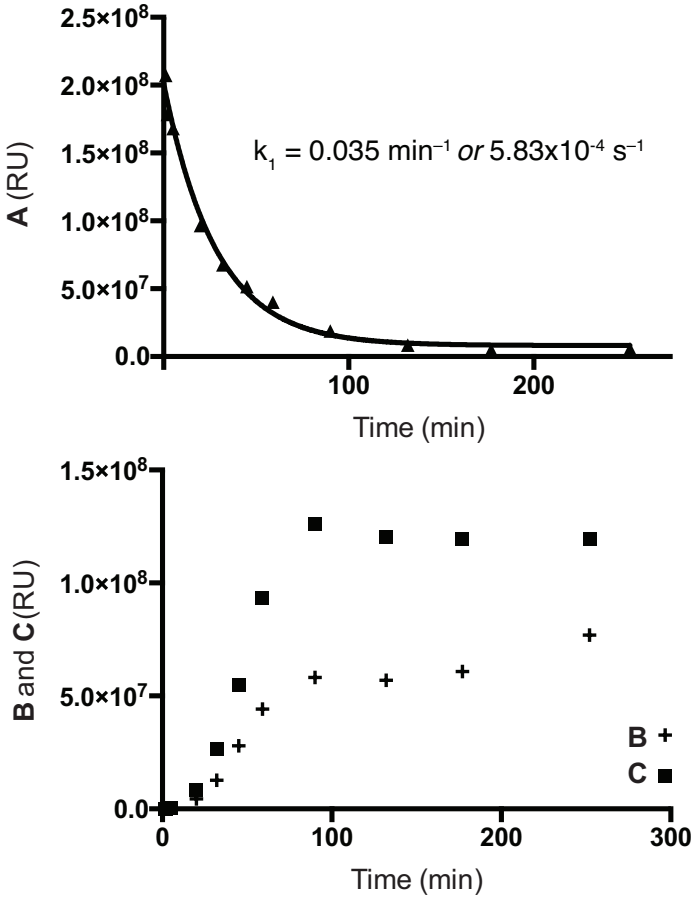


Figure 3.5: Plot of EICs for reaction of E10-32 with wild-type OGT. Integration of individual EICs (for data shown in Figure A.4), corrected against EIC signal from a blank run.

3.3 Detection of HCF-1 cleavage intermediates

Careful examination of the LC-MS traces revealed two low abundance peaks having a m/z of 895.9017 (4+), consistent with a glycopeptide [A+GlcNAc] (Figure 3.6). One of these peaks formed earlier in the reaction and decayed over time, suggesting it was a glycopeptide intermediate; the other formed more gradually, but increased throughout the cleavage reaction, suggesting it was a glycopeptide byproduct (BP) that prevented backbone cleavage. Glycosylation on T11, the threonine adjacent to the glutamate, or on one of several threonines in the C-terminal region of the peptide that binds to the tetratricopeptide (TPR) domain of OGT, could prevent cleavage by hindering glycosylation at E10.

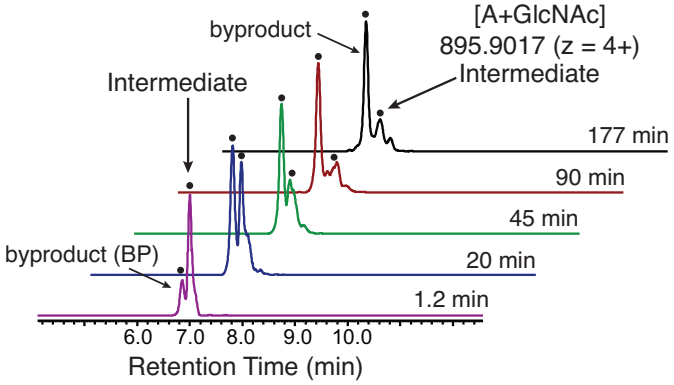


Figure 3.6: Two glycopeptides form during the cleavage reaction. LC traces showing EICs for m/z 895.9017 ± 5 ppm, corresponding to [A+GlcNAc]. The y-axis scaling is such that the tallest peak is 100% (axis not shown). Two peaks having the same exact mass were observed, one of which is a reaction intermediate while the other is a byproduct glycopeptide.

With the goal of characterizing the glycopeptide intermediate in mind, I sought to eliminate the byproduct glycopeptide that obscured it. Thinking that T11, being adjacent to the reactive glutamate, was a likely off-target glycosylation site I obtained a peptide variant

in which T11 was replaced with valine, E10T11V-32. I had previously shown that HCF-3R-EAA (T11V) (see Table B.4 for sequence) was able to undergo cleavage (see section 2.5.3). A glycopeptide byproduct still formed with this peptide, so we wondered whether it would instead be possible to minimize glycopeptide byproduct formation by using an OGT mutant impaired for Ser/Thr glycosylation, but capable of catalyzing cleavage. I had previously examined a set of OGT active site mutants, and identified one with this behavior (see section 2.7).

We reasoned that a residue that serves the role of a general base for the glycosylation reaction would be dispensable in a scenario where the glycosyl accepter was already deprotonated (e.g., for a carboxylate). Our lab, and others, found that aspartate 554 plays an important role in Ser/Thr glycosylation [140, 213], and we believed that it was acting (either directly or indirectly) to remove the proton liberated from the acceptor hydroxyl from the active site [108]. I re-tested the ability of a OGT (D554N) single mutant to catalyze Ser/Thr glycosylation and to effect peptide backbone cleavage using two nearly identical polypeptide substrates, HCF-3R-SAA and HCF-3R-EAA, respectively (see Tables B.5 and B.2 for sequences, respectively). In one polypeptide, the glutamate in the proteolytic repeat was replaced with serine. Compared to wild-type OGT, OGT (D554N) was dramatically impaired in its ability to glycosylate the serine-containing polypeptide; however, both enzymes were proficient at cleaving the glutamate-containing polypeptide (See Figure 3.7; see previous experiments, Figure 2.11). Over the course of my work, the group of Winship Herr published two papers that made use of OGT (D554H/H558D), a double mutant that seems to show similar behavior to OGT (D554N), albeit with a further reduced ability to catalyze Ser/Thr glycosylation [16, 117].

Next, I tested the ability of OGT (D554N) to cleave the E10-32 peptide; the reac-

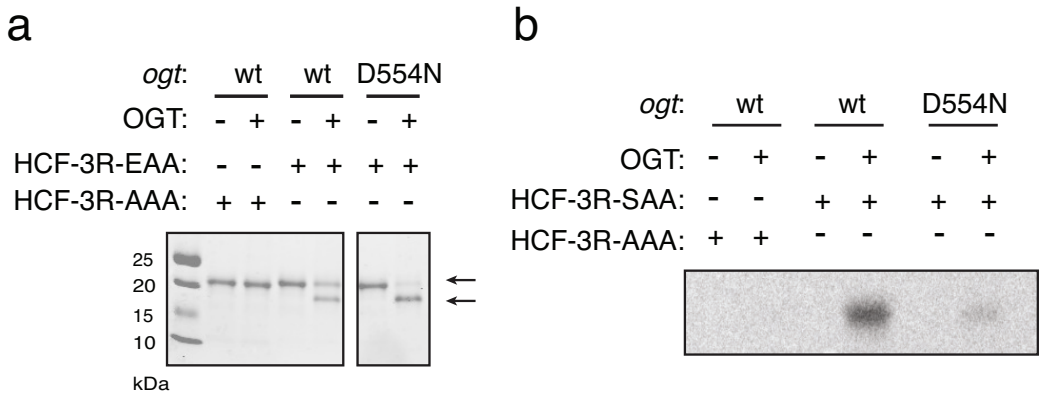


Figure 3.7: OGT (D554N) is impaired for Ser/Thr glycosylation, but not HCF-1 cleavage. (a) SDS-PAGE analysis detected with Coomassie showed that cleavage of a HCF-3R-EAA construct occurs with both wild-type OGT and the OGT (D554N) mutant. (b) Autoradiograph detecting transfer of ^{14}C-GlcNAc to the HCF-3R-SAA polypeptide substrate for wild-type OGT and the D554N mutant. The HCF-3R-AAA peptide was used as a negative control.

tion was complete within 100 minutes, and we did not detect any glycopeptide byproduct (Figure 3.8). I conducted several time-course experiments using stoichiometric, and excess quantities of OGT (D554N) (Figures A.5, A.6, A.7, A.8, A.9, A.10, and A.11).

I found that the starting material was almost entirely consumed within the first three minutes of the reaction (compare Figures A.4 and 3.9), and two intermediates had accumulated to high levels. One of these had the same retention time and exact mass as the glycopeptide intermediate identified in the wild-type OGT cleavage reaction (Figure 3.6). The other had an m/z of 840.8784 (4+), corresponding to loss of water from the starting material (Figure 3.9).

The dramatic increase in the abundance of the glycopeptide intermediate, together with elimination of the byproduct glycopeptide species, suggested that OGT (D554N) would be

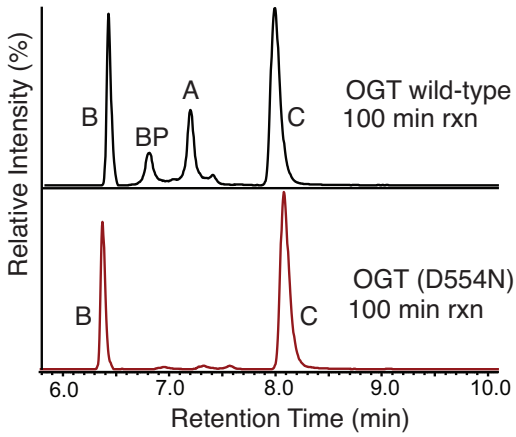


Figure 3.8: OGT (D554N) does not produce a glycopeptide byproduct. EICs for cleavage reactions of E10-32 with wild-type OGT (top) and OGT (D554N) (bottom) showing A, B, C and [A+GlcNAc].

a useful tool to study the mechanism of HCF-1 cleavage. In order to be confident that wild-type OGT and OGT (D554N) were effecting cleavage by the same mechanism, I tested whether the two enzymes had the same cleavage requirements, and produced the same products. Indeed, both enzymes produced products B and C (with the same LC retention times), and required UDP-GlcNAc and the glutamate at position 10 in the HCF-1 sequence for cleavage to occur (Figure 3.10). As such, I used OGT (D554N) for the production and study of intermediates 1 and 2.

We observed the same anhydropeptide species (intermediate 2) in cleavage reactions with wild-type enzyme (Figure 3.11), but it accumulated to lower levels (compare Figure A.4 and Figure A.5).

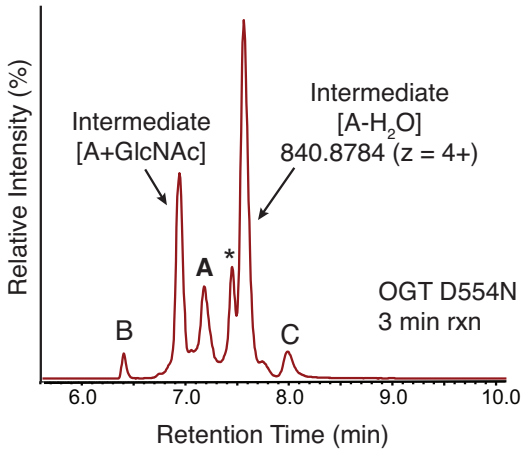


Figure 3.9: OGT (D554N) rapidly forms two intermediates. An LC trace showing a 3 minute time point for cleavage of E10-32 using OGT (D554N). Peaks for A, B, C, the [A+GlcNAc] intermediate, and another intermediate, [A-H_2O], are shown. The "*" denotes a second peak having the same m/z as the [A-H_2O] species. A complete time course is shown in Figure A.5.

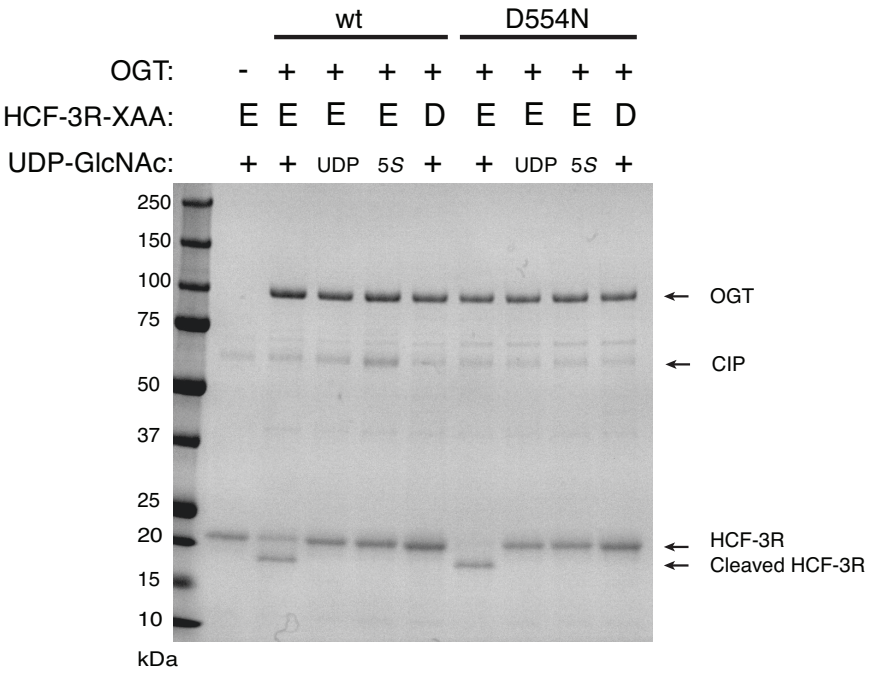


Figure 3.10: OGT (D554N) and wild-type have the same substrate requirements for HCF-1 cleavage. SDS-PAGE analysis detected with Coomassie showing substrate requirements for cleavage of HCF-3R constructs (see Table B.1). The "X" in HCF-3R-XAA, is the identity of residue 10 in the first repeat, either Glu (E10) or Asp (D10); UDP and *5S* denote the use of UDP and UDP-*5S*GlcNAc, respectively, rather than UDP-GlcNAc in the reaction. These data show that cleavage occurs when UDP-GlcNAc, but not UDP or UDP-*5S*GlcNAc [140], is used and that the repeat must contain glutamic acid, rather than aspartate, at position 10.

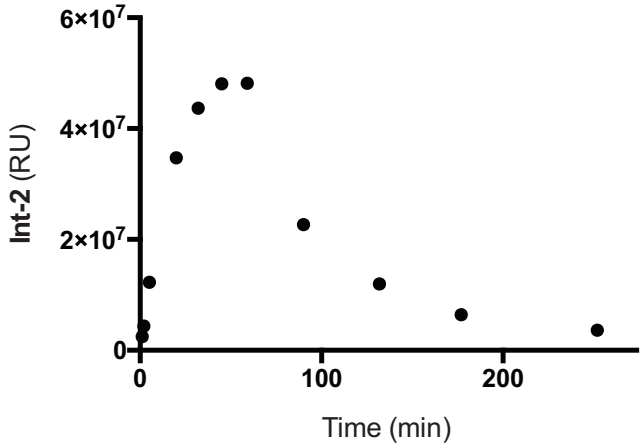


Figure 3.11: Wild-type OGT produces an anhydropeptide intermediate (intermediate 2). Integration of the EIC for Int-2 (intermediate 2) (m/z 840.8784; 4+) (data shown in Figure A.4). Integration for Int-2 included both peaks. Values were corrected based on the signal of the EIC from a blank run.

3.4 Characterizing the anhydropeptide intermediate

Since the anhydropeptide ([M-H_2O]) intermediate appeared as two peaks, I compared their MS/MS fragmentation patterns and, except for minor differences in fragment ion intensities, they were identical (Figure 3.12).

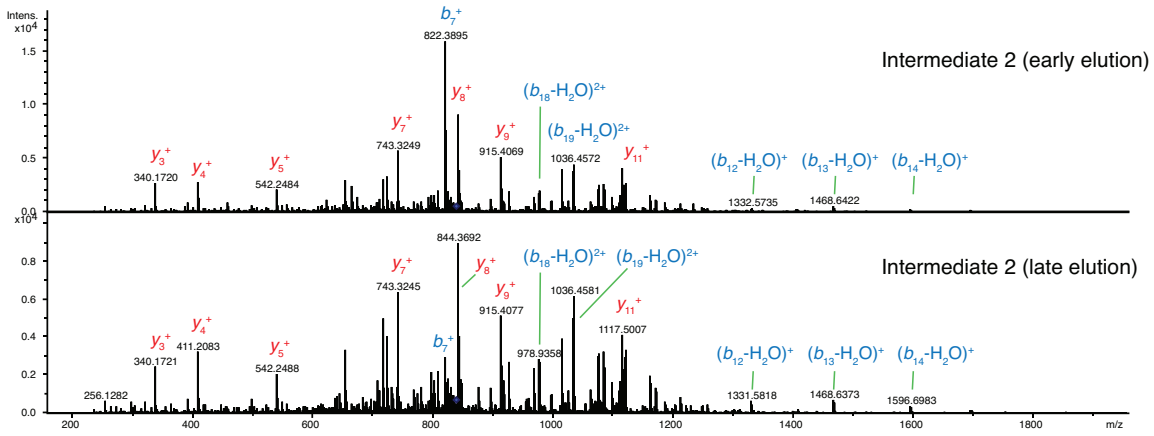


Figure 3.12: Both peaks of Intermediate 2 have the same MS/MS fragmentation pattern. Comparison of the CID MS/MS spectra for the two peaks present in the Int-2 (intermediate 2) (m/z 840.8784; 4+) EIC. Both by CID (shown) and HCD, the fragmentation patterns are essentially indistinguishable other than slight differences in relative abundance for some fragment ions.

The fragmentation data localized the dehydration to the C^9E^{10} region of the peptide, rather than the C-terminal threonine-rich region, which is also prone to dehydration during the CID/HCD fragmentation process [10] (see Figures 3.13 and A.12, and Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9). These LC-MS/MS data for the anhydropeptide species were consistent with formation of an internal pyroglutamate (as indicated by the red arrow in Figure 3.2), but we were not sure why there were two peaks.

To rule out the possibility that one of the anhydropeptide species was a thiolactone

Residue	C-term.	y^+ obs. (m/z)†	$(y\text{-}H_2O)^+$ obs. (m/z)†	y^{2+} obs. (m/z)†	$(y\text{-}H_2O)^{2+}$ obs. (m/z)†
H(amidated)	1	155.0926	–	–	–
Q	2	283.1510	–	–	–
G	3	340.1724	–	–	–
A	4	411.2095	–	–	–
M	5	542.2503	–	–	–
N	6	656.2925	–	–	–
S	7	743.3261	–	–	–
T	8	844.3714	–	422.6895	413.6846
A	9	915.4088	897.3971	–	449.2028
T	10	1016.4529	998.4398	–	499.7266
T	11	1117.5004	1099.4962	–	550.2511

Table 3.1: List of y ions for Int-2 by HCD. † Errors are shown in Figure A.16.

Residue	N-term.	b^+ obs. (m/z)†	$(b\text{-}H_2O)^{3+}$ obs. (m/z)†
Y	1	–	–
V	2	263.1386	–
R	3	419.2395	–
V	4	518.3076	–
C	5	–	–
S	6	–	–
N	7	822.3903	–
P	8	–	–
P	9	–	–
N	19	–	–
T	20	–	724.6581
A	21	–	–

Table 3.2: List of singly- and triply-charged b ions for Int-2 by HCD. † Errors are shown in Figure A.16. The double horizontal line marks a break in sequence numbering.

Residue	N-term.	$(b\text{-}H_2O)^{2+}$ obs. (m/z)†	$(b\text{-}2H_2O)^{2+}$ obs. (m/z)†
C	10	–	–
E	11	615.7719	–
T	12	666.2944	–
H	13	734.8252	–
Q	14	798.8547	–
T	15	849.3768	840.3713
G	16	877.8874	868.8799
T	17	928.4115	919.4068
T	18	978.9337	969.9290
N	19	1035.9566	1026.9483
T	20	1086.4812	1077.4728
A	21	1121.9928	–

Table 3.3: List of doubly-charged b ions for Int-2 by HCD. † Errors are shown in Figure A.16.

Residue	C-term.	y^+ obs. (m/z)†	y^{2+} obs. (m/z)†	$(y\text{-}H_2O)^{2+}$ obs. (m/z)†
H(amidated)	1	–	–	–
Q	2	–	–	–
G	3	340.1720	–	–
A	4	411.2090	–	–
M	5	542.2484	–	–
N	6	656.2907	–	–
S	7	743.3247	–	–
T	8	844.3689	422.6908	–
A	9	915.4070	458.2063	449.2025
T	10	1016.4545	508.7373	499.7283
T	11	1117.5017	–	550.2511
A	12	1188.5368	–	–
T	13	1289.5869	–	–

Table 3.4: List of y ions for Int-2 by CID. † Errors are shown in Figure A.16.

Residue	N-term.	b^+ obs. (m/z)†	b^{2+} obs. (m/z)†
C	5	–	–
S	6	708.3482	–
N	7	822.3897	411.7006
P	8	–	460.2247
P	9	–	–
C	10	–	560.2535

Table 3.5: List of b ions for Int-2 by CID. † Errors are shown in Figure A.16.

resulting from cysteine participation (Figure 3.14), which was previously suggested as a mechanistic possibility (Figure 2.16 (mechanism 2), and [141]), I replaced Cys9 in the E10-32 peptide with Arg, an amino acid found in a proteolytic repeat from the *Odobenus rosmarus divergens* (walrus) HCF-1 ortholog (see section 2.5.2). R9E10-32 was a good cleavage substrate (Figure 3.15); we again observed two peaks corresponding to an anhydropeptide intermediate in reactions with both wild-type OGT and OGT (D554N) (Figure A.13). We also observed two anhydropeptide peaks with E10V11-32 (Figure A.13), S9E10-32, and A9E10-32 (data not shown). Taken together, these data exclude the possibility that the second species is the result of cysteine participation. An expanded discussion on the nature of the two anhydropeptide species in provided in section 3.9 on page 115.

Residue	N-term.	$(b\text{-}H_2O)^+$ obs. (m/z)†	$(b\text{-}H_2O)^{2+}$ obs. (m/z)†
T	12	1331.5869	–
H	13	1468.6412	734.8235
Q	14	1596.6996	798.8551
T	15	1697.7511	849.3767
G	16	–	877.8901
T	17	–	928.4123
T	18	–	978.9359
N	19	–	1035.9563
T	20	–	1086.4824
A	21	–	1121.9988
T	22	–	1172.53
T	23	–	1223.05
A	24	–	1258.5627

Table 3.6: List of dehydrated b ions for Int-2 by CID. † Errors are shown in Figure A.16.

Residue	C-term.	y^+ obs. (m/z)†	y^{2+} obs. (m/z)†	$(y\text{-}H_2O)^+$ obs. (m/z)†	$(y\text{-}H_2O)^{2+}$ obs. (m/z)†
H(amidated)	1	155.0928	–	–	–
Q	2	–	–	–	–
G	3	340.1729	–	–	–
A	4	411.2099	–	–	–
M	5	542.2510	–	–	–
N	6	656.2933	–	–	–
S	7	743.3254	–	–	363.1610
T	8	844.3726	422.6898	826.3614	413.6851
A	9	915.4108	458.2095	897.3971	449.2036
T	10	1016.4541	–	998.4438	499.7275
T	11	1117.5040	–	1099.4923	550.2521
A	12	–	–	–	–

Table 3.7: List of y ions for E10-32 by HCD † Errors are shown in Figure A.16.

Residue	N-term.	b^+ obs. (m/z)†	b^{2+} obs. (m/z)†
V	2	263.1392	–
R	3	419.2400	–
V	4	518.3087	–
C	5	621.3179	–
S	6	708.3492	–
N	7	822.3923	411.7000 (402.6949)‡
P	8	–	460.2264 (451.2238)‡
P	9	–	508.7544
C	10	–	560.2573
E	11	–	624.7779
T	12	–	675.3015
H	13	–	743.8321 (734.8258)‡
Q	14	–	807.8615 (798.8545)‡
T	15	–	858.3858 (849.3779)‡
G	16	–	886.8935 (877.8891)‡
T	17	–	937.4193 (928.4130)‡
T	18	–	987.9402 (978.9350)‡
N	19	–	1044.9629 (1035.9553)‡
T	20	–	1095.4865 (1086.4810)‡

Table 3.8: List of b ions for E10-32 by HCD. † Errors are shown in Figure A.16. ‡ Denotes $(b\text{-}H_2O)^{2+}$ ions also observed.

Residue	N-term.	b^{2+} obs. (m/z)†	$(b\text{-}2H_2O)^{2+}$ obs. (m/z)†	b^{3+} obs. (m/z)†
T	18	–	969.9297	–
N	19	–	–	–
T	20	–	–	730.6610 (724.6583)‡
A	21	1131.0024	–	–

Table 3.9: List of additional b ions for E10-32 by HCD. † Errors are shown in Figure A.16. ‡ Denotes $(b\text{-}H_2O)^{3+}$ ions.

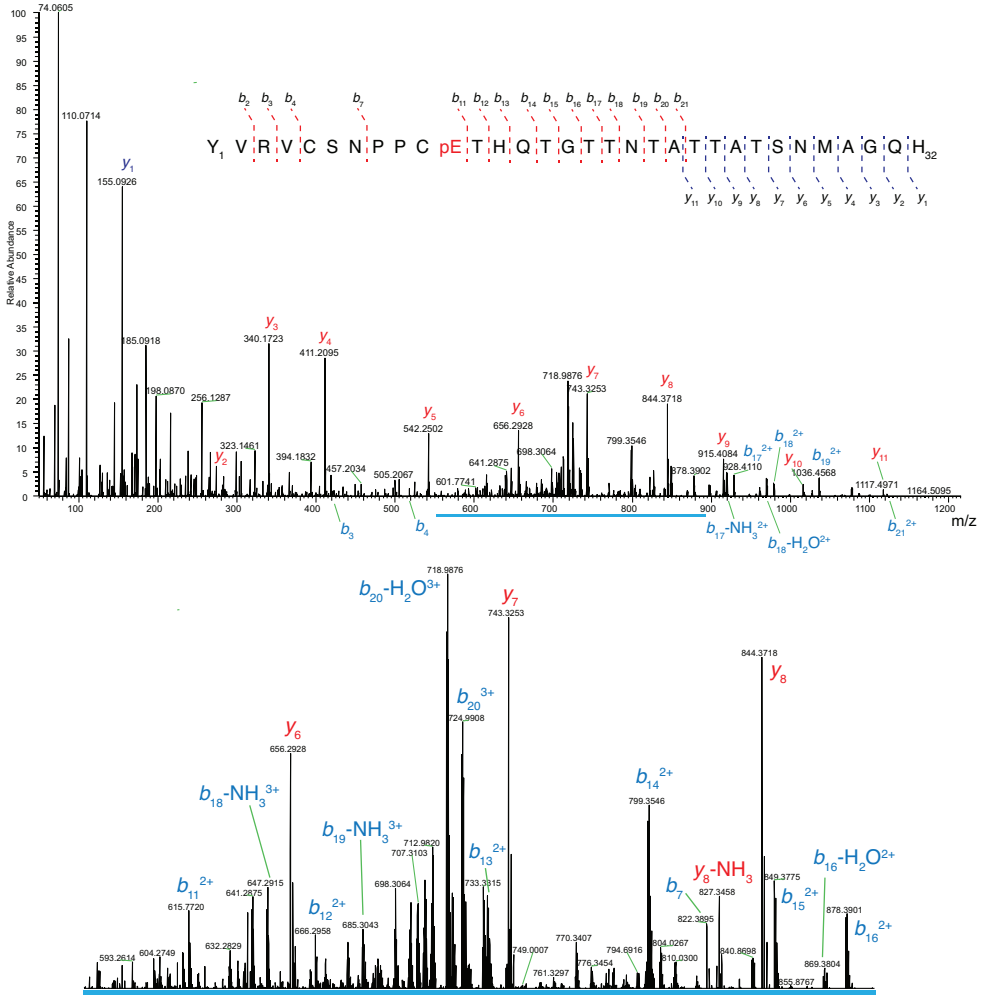


Figure 3.13: HCD MS/MS spectrum of Int-2. *b* and *y* ions obtained by HCD fragmentation are shown for major peaks. Not all peaks are labeled for clarity. Observed ions for Int-2, both by CID and HCD, are provided in Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6. Some ions derived from ammonia or water loss are also not annotated. The top spectrum is the full MS/MS spectrum, and the bottom spectrum is a zoom of the region from about 550–900 *m/z*, as indicated by the blue bar. Based on the collective fragmentation data, we mapped the site of dehydration to the C^9E^{10} region of the peptide (HCF repeat numbering).

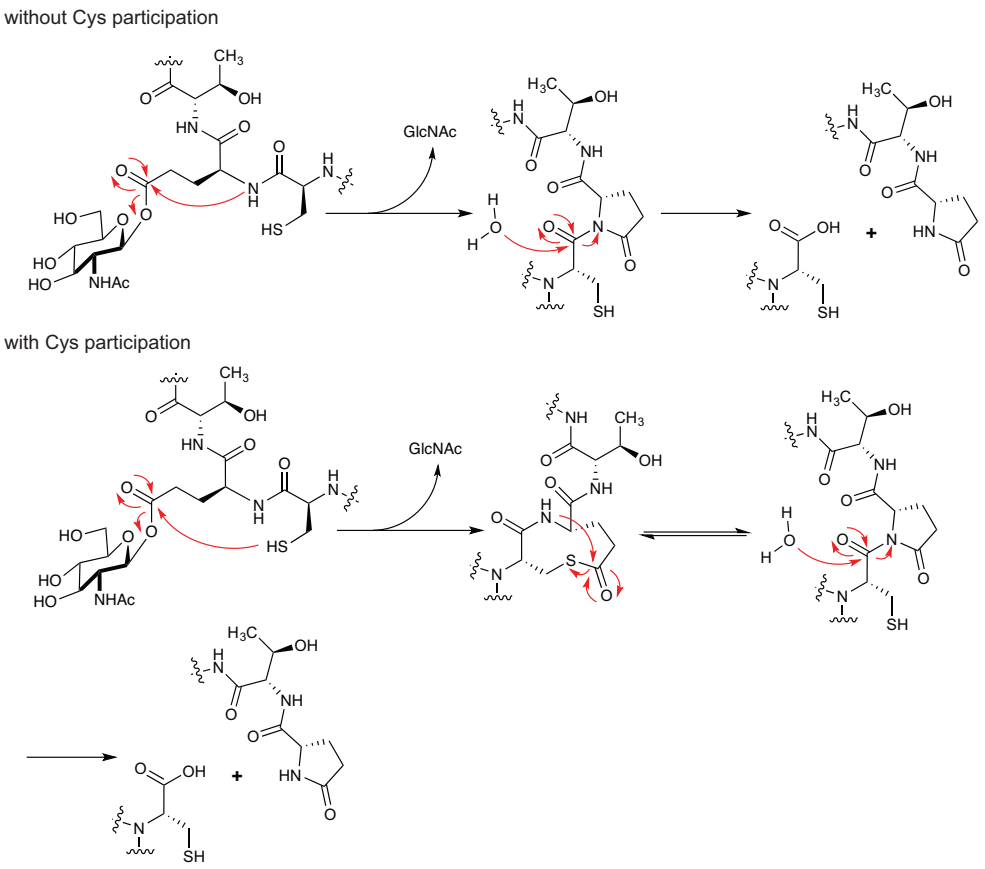


Figure 3.14: Possible Mechanisms, with or without cysteine participation. After initial glycosylation of the glutamate side chain (see Figure 2.16), we reason that formation of the requisite internal pyroglutamate can occur either in a Cys-independent or a Cys-dependent manner. In the first case, the activated glutamate side chain is trapped by the C–E amide nitrogen to form an internal pyroglutamate directly. We also consider the case where the ester is trapped by the thiol of Cys to form a macrocyclic thiolactone. Such a species would potentially be in equilibrium with an internal pyroglutamate [58, 122, 123, 124]. In both cases hydrolysis of the internal pyroglutamate leads to product formation.

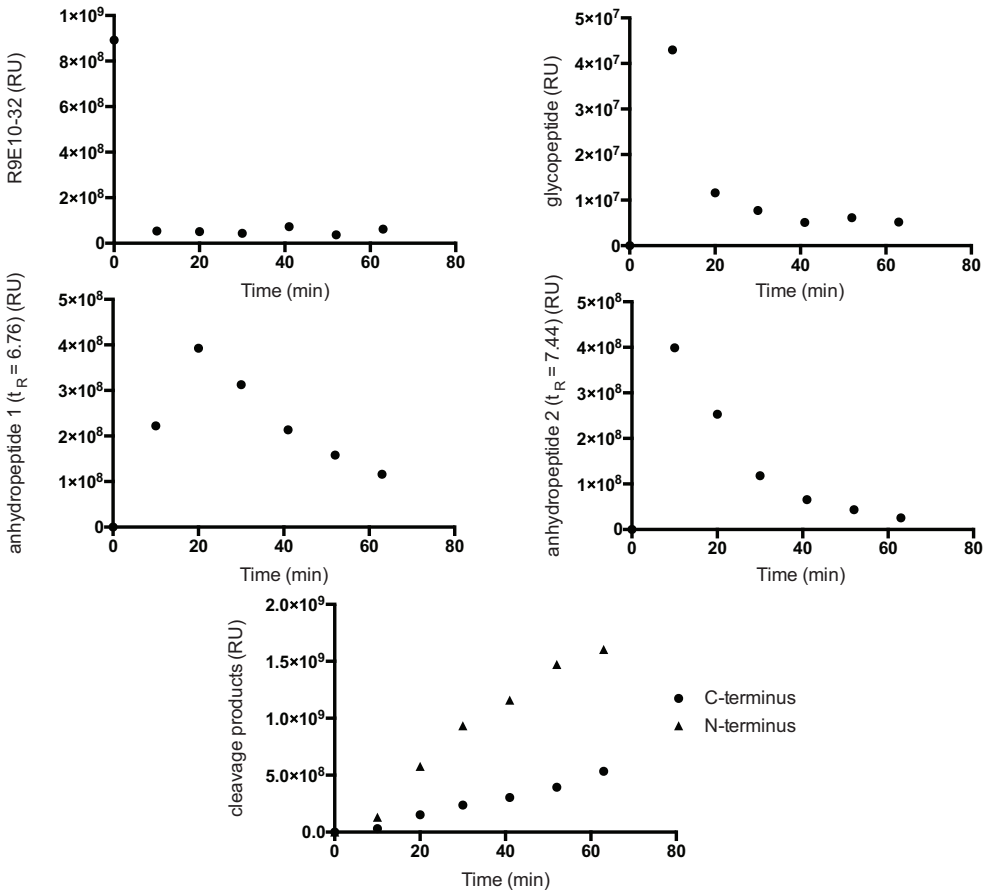


Figure 3.15: Plots of R9E10-32 cleavage time course with excess of OGT (D554N). Integration of individual EICs provided quantitation of response units for each species at each time. Starting material, [R9E10-32] (both charge states,m/z 858.4048 (4+) and 686.9253 (5+), were integrated), decays analogously to E10-32. A glycopeptide (m/z 727.5412 (5+)) forms rapidly, and subsequently decays as two anhydropeptide species (m/z 683.3232 (5+)) form. Both of these species decay as the reaction progresses and products form (N-terminus, m/z 397.5415 (3+) and C-terminus, m/z 747.6698 (3+); this is the same C-terminal product as E10-32). All values used are corrected based on the signal of an EIC from a blank run.

3.5 Establishing the temporal sequence for intermediate formation

Having conducted several time courses under conditions where OGT was in excess of the peptide substrate (single-turnover conditions, STO), I was able to unequivocally establish the temporal sequence of events for HCF-1 cleavage. Within 60 seconds the starting material had converted to the glycopeptide intermediate. As this species decayed, one anhydropeptide species appeared (Figure 3.16; see also Figures A.7 and A.8), and the second grew in as the reaction progressed. Both species were consumed by the end of the reaction, indicating that both are on-pathway to cleavage products. Hence, the reaction involves initial formation of a glycopeptide intermediate (Intermediate 1) followed by conversion to an internal pyroglutamate (Intermediate 2) and then backbone cleavage. I measured a rate constant of 0.08 s^{-1} for the consumption of A (starting material) by OGT (D554N), which is approximately two orders of magnitude faster than that of wild-type OGT (compare Figures 3.5 and A.8). Hence, replacement of aspartate for asparagine at position 554 dramatically accelerated formation of the glycopeptide intermediate. Overall cleavage rates for wild-type OGT and OGT (D554N) differed only by a factor of two to three, however, evidently because hydrolysis of the internal pyroglutamate is rate determining.

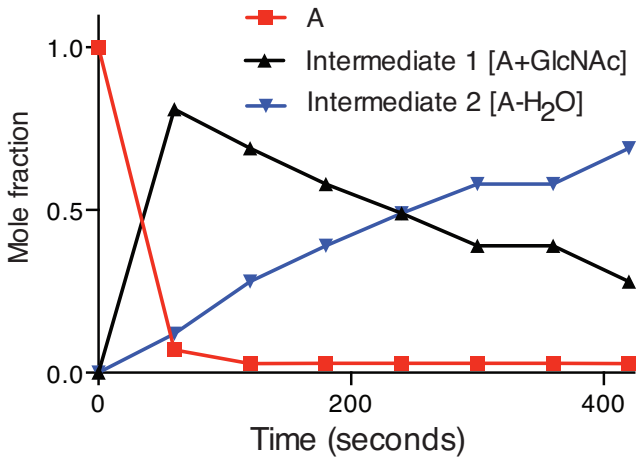


Figure 3.16: Temporal sequence of reaction intermediates. The temporal sequence of events was established from a time course of cleavage using a large excess of OGT (D554N). Mole fractions of A (starting material), [A+GlcNAc] and [A-H_2O], were plotted as a function of time. The data were truncated at 420 seconds, such that products represent a negligible mole fraction. Mole fractions were determined based on the maximal signal for each species, normalized over all species present at that time point.

3.6 Characterizing the glycopeptide intermediate

Because the use of OGT (D554N) results in a substantial accumulation of the glycopeptide intermediate at early time points, I was able to obtain sufficient material for MS/MS analysis by quenching OGT (D554N) cleavage reactions after about one minute. The intermediate proved stable enough to allow for LC purification, and fragmentation by electron-capture dissociation (ECD) allowed unambiguous assignment of the GlcNAc modification to glutamate 10 (Figures 3.17, A.14, and A.15 and Tables 3.10, 3.11, 3.12, and 3.13).

While the stability of the glycosylated intermediate was unexpected, aspartate- and

glutamate-linked glycosyl esters produced by protein ADP-ribosylation are, at least in some cases, also sufficiently stable to allow for isolation and characterization [96, 227].

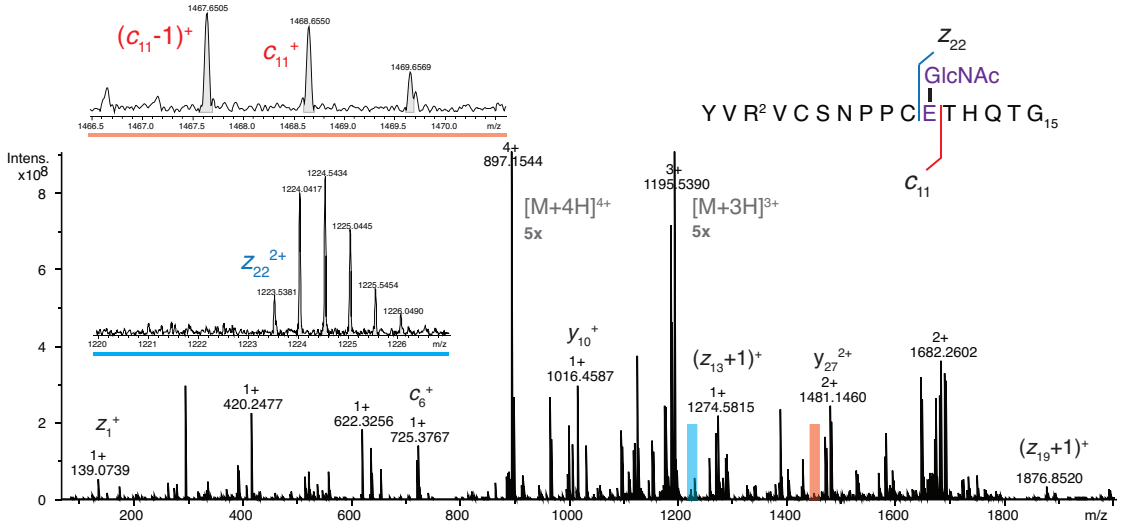


Figure 3.17: ECD MS/MS spectrum of Intermediate 1. Intermediate 1 is a glutamate-linked glycopeptide. Insets show zooms of ions $c_{11}{}^{+}$ and $z_{22}{}^{2+}$ that flank the glycosylated residue. Additional spectra are presented in Figures A.14 and A.15.

Residue	N-term.	c^+ obs. (m/z)†	$(c\text{-}1)^+$ obs. (m/z)†
Y	1	–	180.0869
V	2	–	279.1579
R	3	436.2665	435.2585
V	4	535.3350	534.3274
C	5	638.3452	637.3369
S	6	725.3767	724.3687
N	7		
P	8		
P	9	1033.5258	1032.5172
C	10	1136.5349	1135.5292
E*	11	1468.6550	1467.6505
T	12	1569.7027	1568.6995
H	13	1706.7614	–
Q	14	–	–
T	15	–	–
G	16	–	–

Table 3.10: List of singly-charged c ions for Int-1 by ECD. * Denotes a glycosylated glutamic acid residue (E* = E+GlcNAc). † Errors are shown in Figure A.16. Grayed out cells denote ions not observed in ECD fragmentation.

Residue	N-term.	c^{2+} obs. (m/z)†	$(c\text{-}1)^{2+}$ obs. (m/z)†
H	13	–	853.3816
Q	14	–	917.4110
T	15	–	967.9349
G	16	–	Occluded
T	17	–	1046.9699
T	18	–	1097.4938
N	19	–	1154.5155
T	20	1205.5424	–
A	21	–	1240.5735
T	22	–	1291.0829
T	23	–	1341.6088
A	24	–	1377.1273
T	25	1427.6520	–
S	26	1471.6694	–
N	27	1528.6907	–
M	28	1594.2122	–
A	29	1629.7262	–
G	30	1658.2428	–
Q	31	1722.2699	–

Table 3.11: List of multiply-charged c ions for Int-1 by ECD. † Errors are shown in Figure A.16.

Residue	C-term.	z^+ obs. (m/z)†	$(z+1)^+$ obs. (m/z)†
Q	19	–	1876.8520
T	18	1747.7835	1748.7915
G	17	1646.7325	1647.7439
T	16	–	–
T	15	–	1489.6792
N	14	1387.6171	1388.6247
T	13	1273.5738	1274.5815
A	12	–	–
T	11	–	–
T	10	1000.4400	1001.4460
A	9	899.3884	900.4018
T	8	828.3543	829.3626
S	7	727.3067	728.3149
N	6	640.2746	–
M	5	526.2315	527.2339
A	4	395.1909	396.1989
G	3	324.1538	325.1616
Q	2	267.1324	268.1402
H(amidated)	1	139.0739	140.0817

Table 3.12: List of singly-charged z ions for Int-1 by ECD. † Errors are shown in Figure A.16.

Residue	N-term.	z^{2+} obs. (m/z)†	$(z+1)^{2+}$ obs. (m/z)†
N	26	1429.1156	–
P	25		
P	24		
C	23	–	–
E*	22	–	1224.0417
T	21	–	1057.9808
H	20	–	1007.4568

Table 3.13: List of doubly-charged z ions for Int-1 by ECD. * Denotes a glycosylated glutamic acid residue (E* = E+GlcNAc). † Errors are shown in Figure A.16. Grayed out cells denote ions not observed in ECD fragmentation.

3.7 Mechanism of HCF-1 cleavage by OGT

These studies show that OGT effects HCF-1 amide bond cleavage by the mechanism in Figure 3.18. Glycosylation on glutamate is followed by conversion to an internal pyroglutamate, which is susceptible to spontaneous hydrolysis [58, 122, 123, 124, 185]. Because an internal pyroglutamate can form non-enzymatically if a glutamate side chain is activated [185], we wondered whether Intermediate 1 converts to Intermediate 2 on or off OGT.

In order to test this, I isolated the glycopeptide intermediate (Intermediate 1) and re-subjected it to the reaction conditions, but without OGT. In order to differentiate newly formed products from those that may have formed during the isolation process, I used ^{18}O-water for reactions conducted with isolated glycopeptide. When ^{18}O-water is used as the reaction medium, the incorporation of a single ^{18}O atom into the C-terminal Cys residue

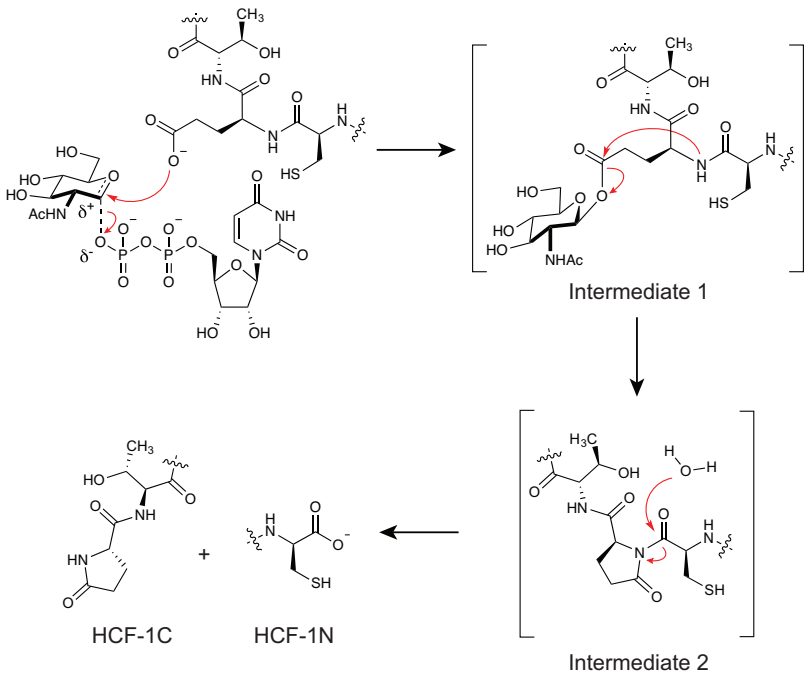
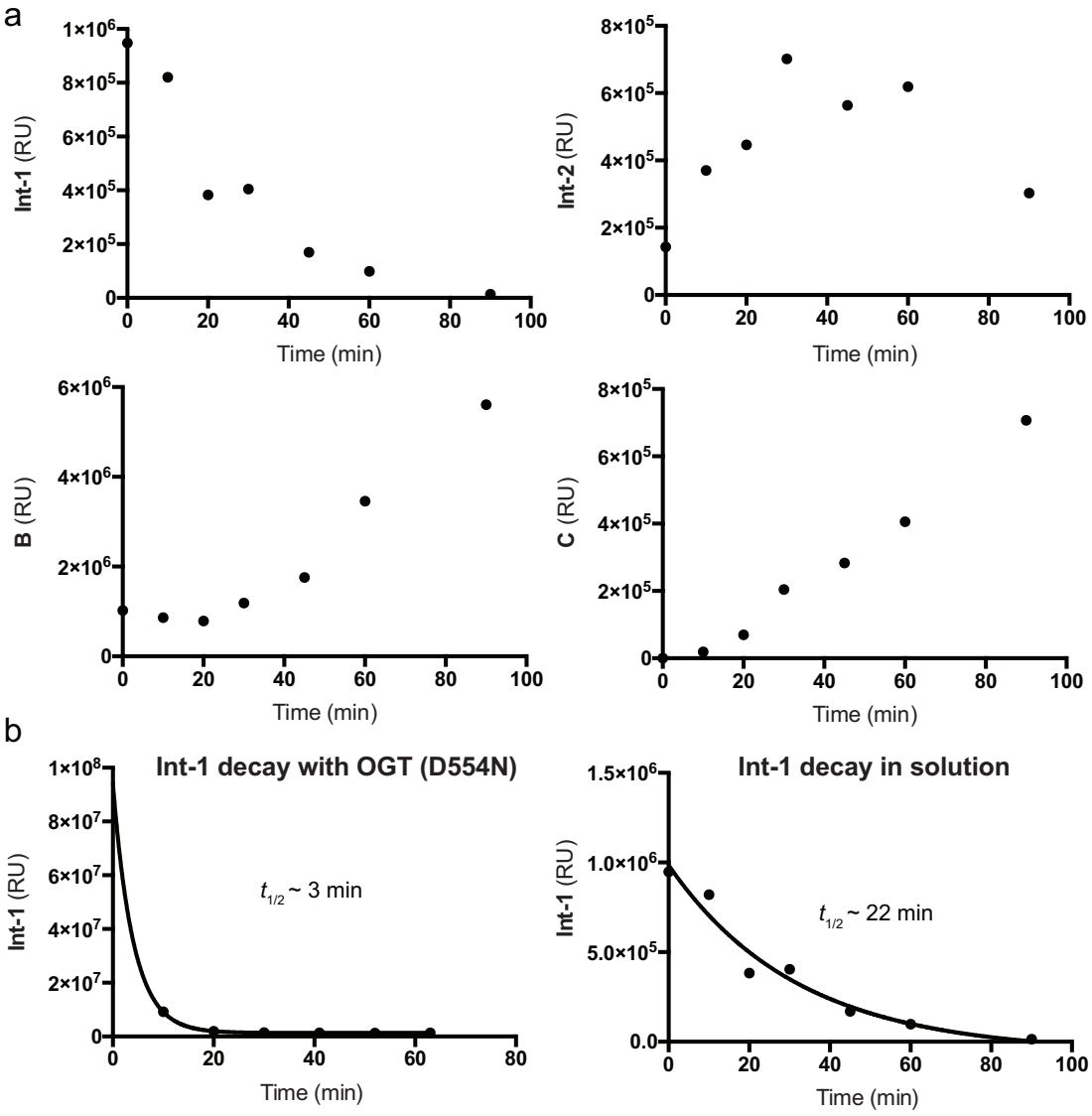


Figure 3.18: The mechanism of HCF-1 cleavage proceeds via glycosylation on glutamate followed by decomposition to an internal pyroglutamate, which undergoes backbone hydrolysis. Chemical mechanism of HCF-1 cleavage by OGT. Both glycosylation on glutamate and conversion to the internal pyroglutamate (Intermediate 2) occur on the enzyme (see text and Figure 3.19), but hydrolysis likely occurs after dissociation.

of the N-terminal cleavage product is observed. I observed formation of an anhydropeptide followed by hydrolytic cleavage over 90 minutes. Because the rate of conversion of glycopeptide to internal pyroglutamate was an order of magnitude slower than observed in the presence of OGT (Figure 3.19), we conclude that both the first and second steps shown in Figure 3.18 occur while the peptide is bound to OGT. Hydrolysis likely occurs after dissociation from the enzyme.

Figure 3.19: *(On the following page)* Plots for the reaction of Int-1 without OGT. Data is based on a time course where Int-1 is incubated in reaction buffer without OGT. (a) Integration of individual EIC traces provides quantitation, in response units, for each species (A, Int-1, Int-2, B and C) at each time. The m/z for each species is the same as in previous captions. The amount of starting material (A) (not shown) does not change over the course of the incubation; this material co-purified with the glycopeptide intermediate. There is a distinct lag in product formation, as seen in other time courses, shown in Figures A.6, A.8, and A.10. As seen previously, Int-2 forms and then decays, concomitant with the formation of products. (b) The disappearance of Int-1 was fit to a first-order decay (Equation 3.1), both in the presence (left) and absence (right) of OGT. Data for Int-1 conversion, in the presence of enzyme, is taken from the 100 minute time course shown in Figure A.5 (see also Figure A.6). The curve is fit from the time point where maximum Int-1 has accumulated.


a
Int-1 (RU)
Int-2 (RU)
B (RU)
C (RU)
Time (min)
b
Int-1 decay with OGT (D554N)
$t_{1/2}$ ~ 3 min
Int-1 decay in solution
$t_{1/2}$ ~ 22 min

3.8 OGT (D554N) accelerates the glycosylation step of HCF-1 cleavage

Though OGT (D554N) accelerates the formation of the two cleavage reaction intermediates, the apparent overall reaction rate was similar for wild-type OGT and OGT (D554N). Comparing the rates of E10-32 consumption by wild-type OGT and OGT (D554N), suggested that OGT (D554N) catalyzes glycosylation on glutamate much faster than the wild-type enzyme (Figure 3.20). This suggests that OGT (D554N) may provide a lower-energy pathway to the glycopeptide intermediate (Int-1)(Figure 3.21). Assuming that the free-energy of TS2 and the internal pyroglutamate (Int-2) are relatively unchanged for OGT (D554N), we would expect that accelerating the first step (glycosylation) would lead an accumulation of intermediates 1 and 2, for a process where the initial barrier for the first step is high and the rate-determining step is the conversion of intermediate 2 to products. This does not account for the possibility that dissociation of intermediate 2 from OGT may also be rate determining.

It might seem curious that a mutant impaired with respect to Ser/Thr glycosylation (see Figure 3.7) would be improved with respect to glutamate glycosylation. Aspartate 554, however, is proposed to serve as a base to shuttle the proton released from the acceptor hydroxyl out of the active site following glycosylation [108, 140] and this function would dispensable for glycosylation of a carboxylate (see also section 2.7). While this explanation resolves a seeming paradox, it does not address *how* OGT (D554N) can catalyze glutamate glycosylation faster than the wild-type enzyme.

Michael B. Lazarus had previously solved a 2.05 Å ternary structure of hOGT4.5 (D554N) bound to UDP and a CKII glycopeptide (Figure 3.22 and Table B.10). This

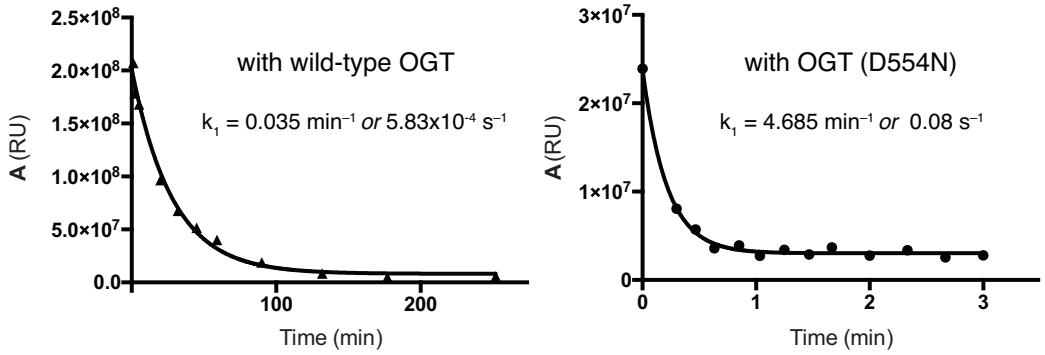


Figure 3.20: Comparing the rate of substrate glycosylation for wild-type OGT and OGT (D554N). Data for the rate of starting material consumption by wild-type OGT and OGT (D554N) are reproduced from Figures 3.5 and A.7, respectively.

structure showed that the D554N mutant protein is virtually identical to the wild-type enzyme (0.36 Å root mean squared deviation (RMSD)) [140], implying that the difference in activity is not due to a conformational change in the protein. Instead, we reasoned that glutamate glycosylation may be disfavored by a repulsive interaction with the D554 side chain in wild-type OGT. In order for glutamate 10 to attack the anomeric carbon of UDP-GlcNAc, it must approach within 4.5 Å of the D554 carboxylate (Figure 3.23), and this unfavorable interaction is eliminated in the D554N variant.

A repulsive interaction involving the reactive glutamate of the HCF proteolytic repeat and the OGT co-substrate UDP-GlcNAc was previously suggested [16]. Whether Asp554 disfavors glycosylation directly through electrostatic repulsion, or indirectly remains to be precisely determined.

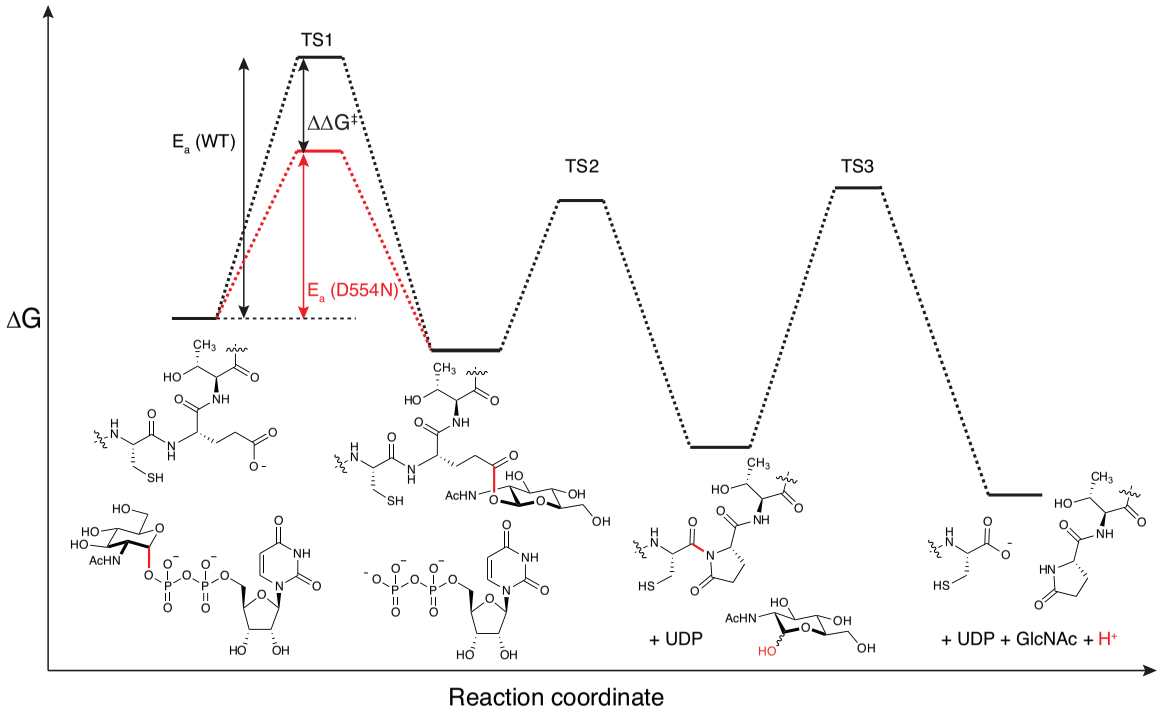
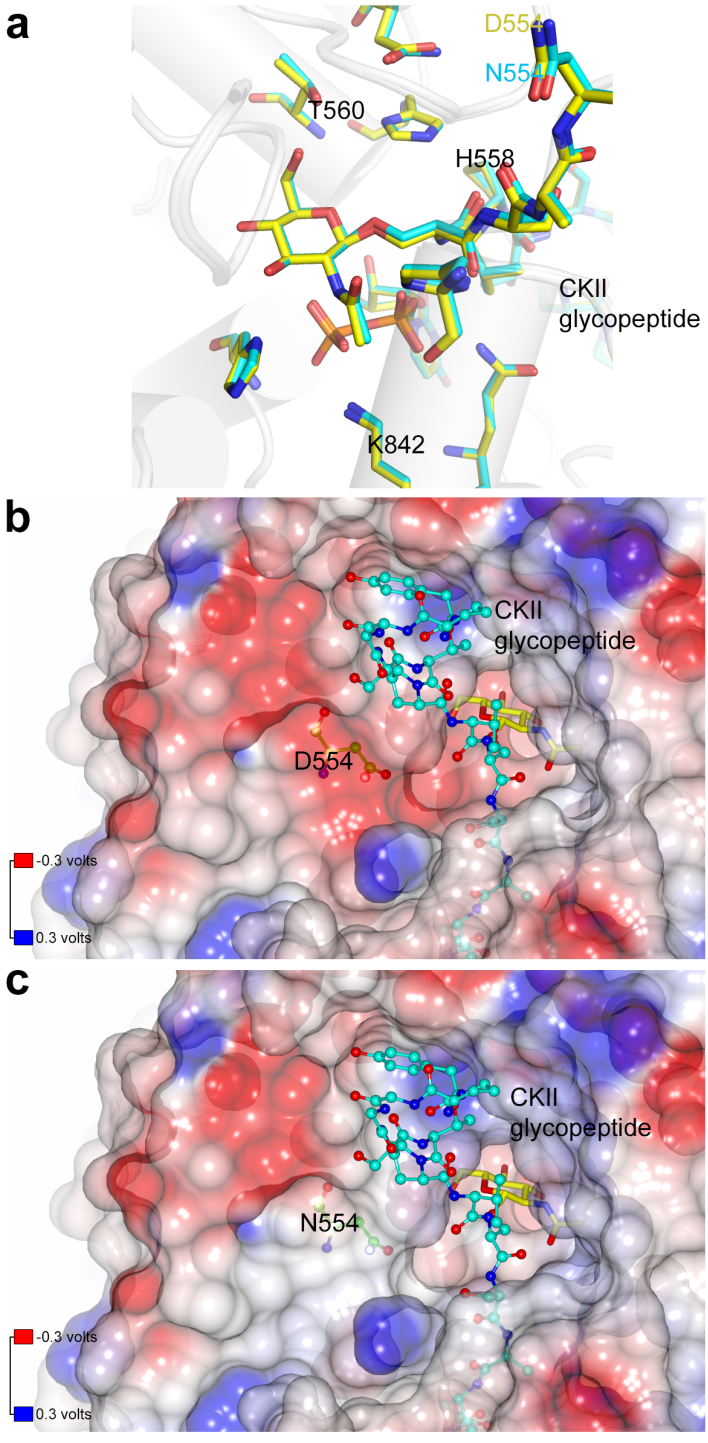


Figure 3.21: Free-energy diagram for HCF-1 cleavage. A pictorial representation of free energy vs. reaction coordinate for the steps of HCF-1 cleavage. The vertical axis (ΔG) is not to scale, and is intended to aid in qualitative discussions made in the text.

Figure 3.22: *(On the following page)* Mutating Asp554 to asparagine does not affect the structure of the OGT active site. (a) Overlay of the UDP:hOGT4.5 (D554N):CKII glycopeptide ternary product complex, with the wild-type ternary product complex (PDB: 4GYW). The glycopeptide and important side chains are shown in stick representation. The wild-type structure is shown in yellow and the D554N structure is shown in cyan. The structures are nearly identical with an RMSD of 0.36 Å. (b) Electrostatic surface rendering of the wild-type UDP:hOGT4.5:glycoCKII ternary product complex crystal structure. The peptide is shown in cyan and the GlcNAc is shown in yellow. The D554 side chain is shown in ball and stick representation, highlighting the negative charge (red) around that residue. Electrostatic scale is shown in the lower left corner. (c) Electrostatic surface rendering of the UDP:hOGT4.5 (D554N):glycoCKII ternary product complex crystal structure. The peptide is shown in cyan and the GlcNAc in yellow as in (b). The Asn554 side chain is shown and the neutral charge (white) in the vicinity of this side chain is indicated by the white color in the surface rendering.


a
D554
N554
T560
H558
CKII
glycopeptide
K842
b
CKII
glycopeptide
D554
-0.3 volts
0.3 volts
c
CKII
glycopeptide
N554
-0.3 volts
0.3 volts

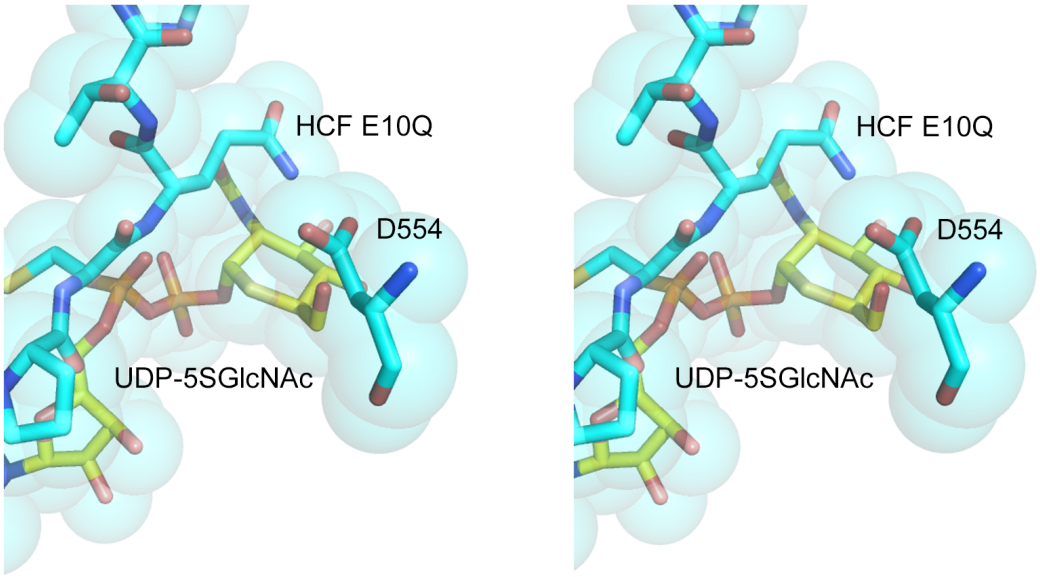


Figure 3.23: E10 of the HCF repeat is repelled by Asp554. Stereoimage of a view from a crystal structure of OGT with a bound nucleotide-sugar analog and an analog of an HCF-1 proteolytic repeat in which E10 is replaced with glutamine (PDB: 4N3B). The peptide and D554 residue are shown in cyan and the UDP-5SGlcNAc is shown in yellow. The side chains of D554 (from OGT) and the E10Q side chain (from the HCF-1 repeat) are 6.4 Å away in the crystal structure (as shown). In order for E10 to reach the anomeric carbon of the sugar, the side chain must rotate towards it, which would place it within 4.5 Å of the D554 carboxylate. The resulting repulsive interaction is unfavorable and we propose that the D554N mutant glycosylates E10 more rapidly because this electrostatic repulsion is relieved.

3.9 The Nature of two anhydropeptide species

For all peptide substrates examined, EIC traces of OGT-formed anhydropeptide showed two distinct peaks. Their relative abundances varied with time, with one forming earlier in the reaction than the other (see Figure A.7). However, both peaks are consumed by the end of the reaction, suggesting that they are both on-pathway intermediates. As mentioned previously in this chapter (see section 3.4), MS/MS spectra (both CID and HCD) collected for both peaks showed no unique fragment ions, though several fragments differed slightly in their relative abundance (Figure 3.12). This suggested that the two species were likely stereoisomers.

When Int-1 was isolated and allowed to convert to products (via Int-2), only a *single* anhydropeptide species was observed (corresponding to the anhydropeptide formed second in the presence of enzyme) (Figure 3.24). Based on this, we speculate that the anhydropeptide species formed second in the enzymatic reaction is the result of a Int-2a $\rightarrow$ Int-2b process (where a and b denote the species formed first and second, respectively), which occurs after dissociation from OGT. If this process were sufficiently fast, we would not expect to observe Int-2a in the absence of enzyme. We reasoned that there were two possibilities for the difference between Int-2a and Int-2b: (1) they were rotational isomers about the internal pyroglutamate bond, analogous to that seen for some proline-containing peptides [179]; (2) they were diastereomers [73]. Attempts to isolate each of these species individually (using an approach analogous to the isolation of Int-1), to test whether they were able to interconvert, or whether only a single one led directly to products, were fraught with difficulty, and sufficient quantities of material have yet been unobtainable.

Together with Jacob Sanders, we investigated the feasibility of each possibility computationally. For a model dipeptide (Figure 3.25), he computed the *s*-cis and *s*-trans con-

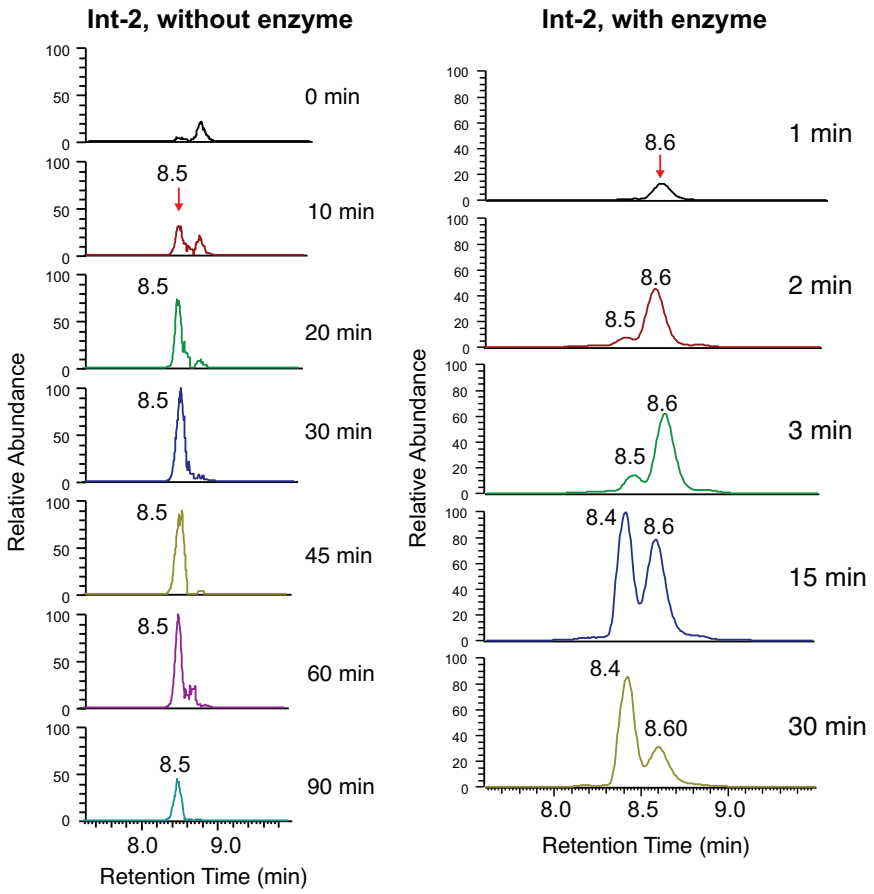


Figure 3.24: Comparison of Int-2 formed on- and off-enzyme. (Left) EIC traces showing Int-2 produced as a result of spontaneous Int-1 decay in the absence of enzyme (see Figure 3.19). A single Int-2 peak is observed in this case. A small peak for Int-2 was observed in the time-zero fraction. Based on examining the preceding blank, this is likely carry-over from previous runs isolating Int-1. The vertical axis is scaled to a fixed intensity across all samples since only Int-2 is shown in this case. (Right) EIC traces showing Int-2 produced in the presence of enzyme (taken from Figures A.7 (1, 2 and 3 minute time points) and A.9 (15 and 30 minute time points). These traces are similarly scaled vertically to a fixed intensity since only Int-2 is shown.

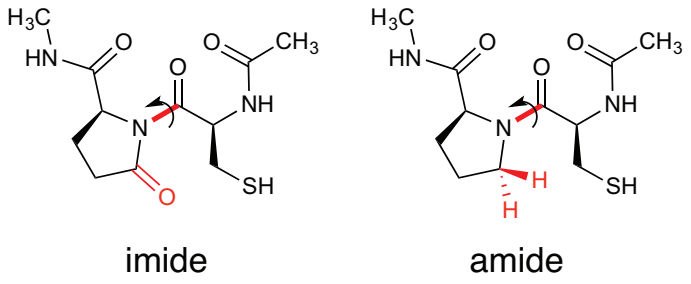


Figure 3.25: Model dipeptides for imide vs. amide bond rotation. Rotation about the bolded bond (colored in red) is examined.

formers, as well as the transition state for rotation about the imide bond. These values were then compared to those obtained for the corresponding amide. We used the Eyring-Polanyi equation (Equation 3.2, where k_B is Boltzmann's constant, R is the gas constant, and h is Planck's constant) to relate rate and the computed free-energy values. The barrier to rotation for the amide was 22.5 kcal/mol (at 298 K), consistent with values previously reported [121, 211]. In the case of the imide, the rotational barrier was 12.8 kcal/mol (at 298 K), which is more in-line with those seen for esters. These free-energy differences correspond to rates of *s*-cis $\rightarrow$ *s*-trans interconversion of $\approx 2.0 \times 10^{-4}$ s^{-1} and $\approx 2.6 \times 10^{4}$ s^{-1}, for the amide and imide models, respectively. This rate of exchange for the imide would be inconsistent with observing two peaks by LC for Int-2.

$$k = \frac{k_B T}{h} e^{\frac{-\Delta G^\circ}{RT}} \tag{3.2}$$

At near-neutral pH (cleavage reaction conditions), peptide epimerization is extremely slow. Thus, in order for the two Int-2 species to be diastereomers, a proton in the Int-2a would have to be epimerization-prone. There is precedent for succinimide-containing peptides to be epimerization prone [69], and it has been previously demonstrated that the Cα proton

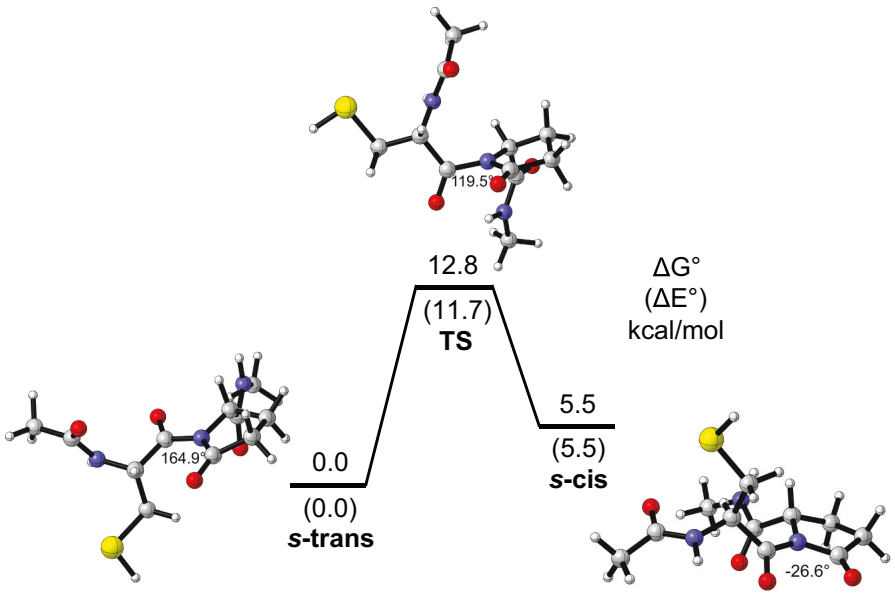


Figure 3.26: Computed structures for *s*-trans → *s*-cis interconversion of a model imide. Free-energy diagram not to scale.

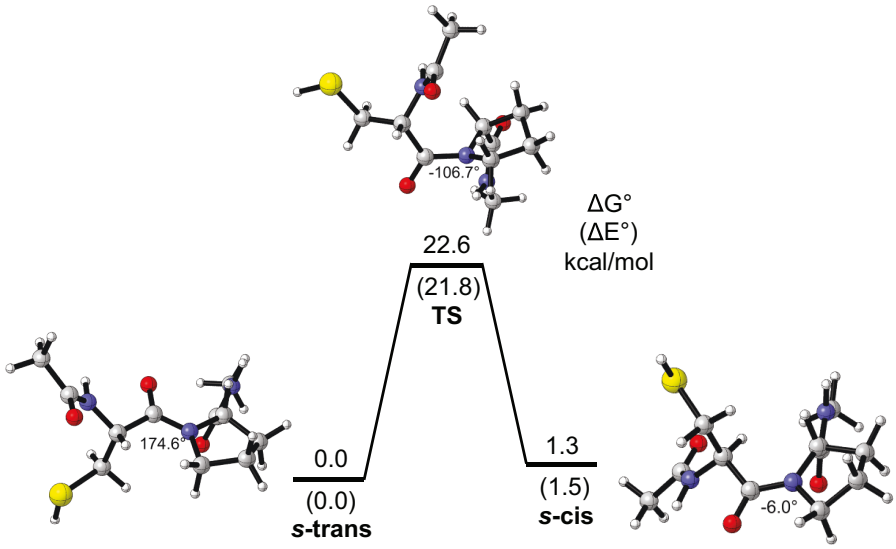


Figure 3.27: Computed structures for *s*-trans → *s*-cis interconversion of a model amide. Free-energy diagram not to scale.

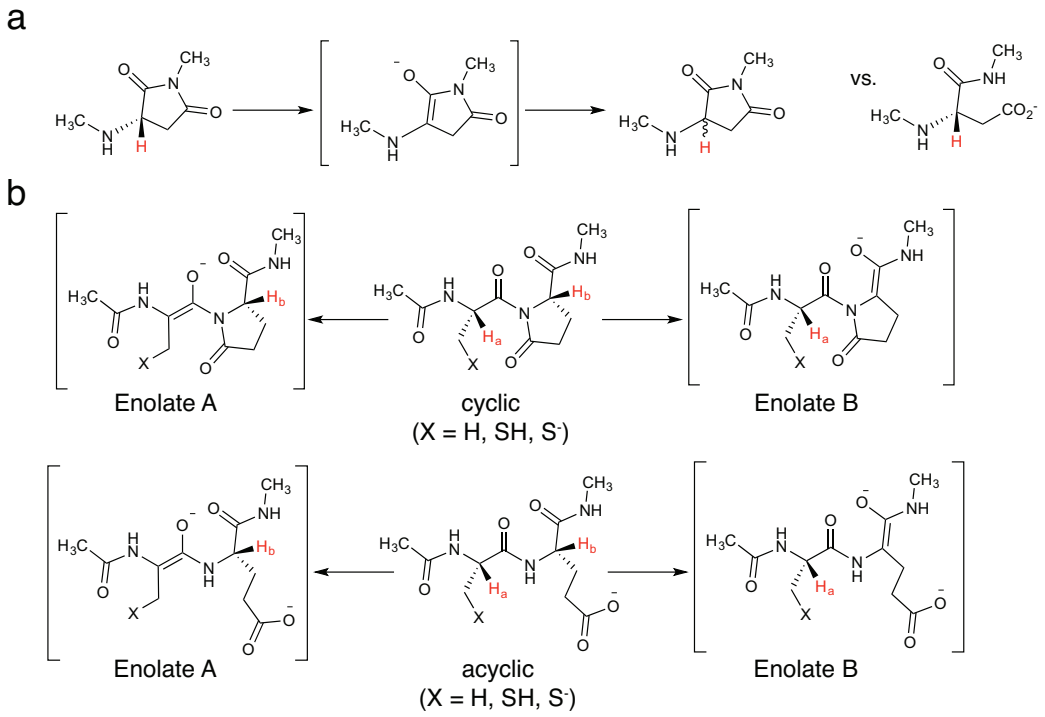


Figure 3.28: Model systems to probe epimerization of an internal pyroglutamate.

epimerizes on the order of 10^5 times faster than the corresponding proton in an acyclic structure (Figure 3.28a, red proton) [202]. We calculated the energy of enolates formed by abstraction of either H_a or H_a for a series of model internal pyroglutamate-containing dipeptides (cyclic) (Figure 3.28b), and the corresponding glutamate-containing peptides (acyclic). Since we observed two Int-2 species regardless of the identity of the residue at position 9, we computed both cysteine and alanine dipeptides.

There was little difference between the calculated pK_a values for proton leading to "Enolate B" in the context of glutamate- or pyroglutamate-containing model dipeptides (Figures 3.29, 3.30 and 3.31). Strikingly, the pK_a for the proton leading to "Enolate A" was on the order of 10 pK_a units *lower* for peptides with a pyroglutamate residue (cyclic) (Figures 3.29 and 3.30). Results are summarized in Table 3.14. "Enolate A" could not

	pK_a	
Structure	Glu	pyroGlu
Cysteine thiol		
enolate A	N.D.‡	13.8
enolate B	26.5	27.1
Cysteine thiolate		
enolate A	29.5	21.1
enolate B	28.0	28.0
Alanine		
enolate A	26.2	15.7
enolate B	26.9	26.5

Table 3.14: Summary of calculated pK_as. pK_a values for model dipeptides shown in Figures 3.29, 3.30 and 3.31. ‡ E1cb elimination observed.

be computed for the cysteine thiol-derived peptide (X = SH) as this species underwent an E1cb-type elimination of a hydrogen sulfide anion. These calculations suggest that Int-2 may be prone to epimerization at the Cα–H of Cys9.

If Int-2b is indeed the diastereomer of Int-2a, conducting the cleavage reaction in D_2O should lead to deuterium incorporation at the Cα position of Cys9. After amide bond hydrolysis occurs this position will no longer be exchangeable, making isotope incorporation stable. Experiments to test this hypothesis are underway.

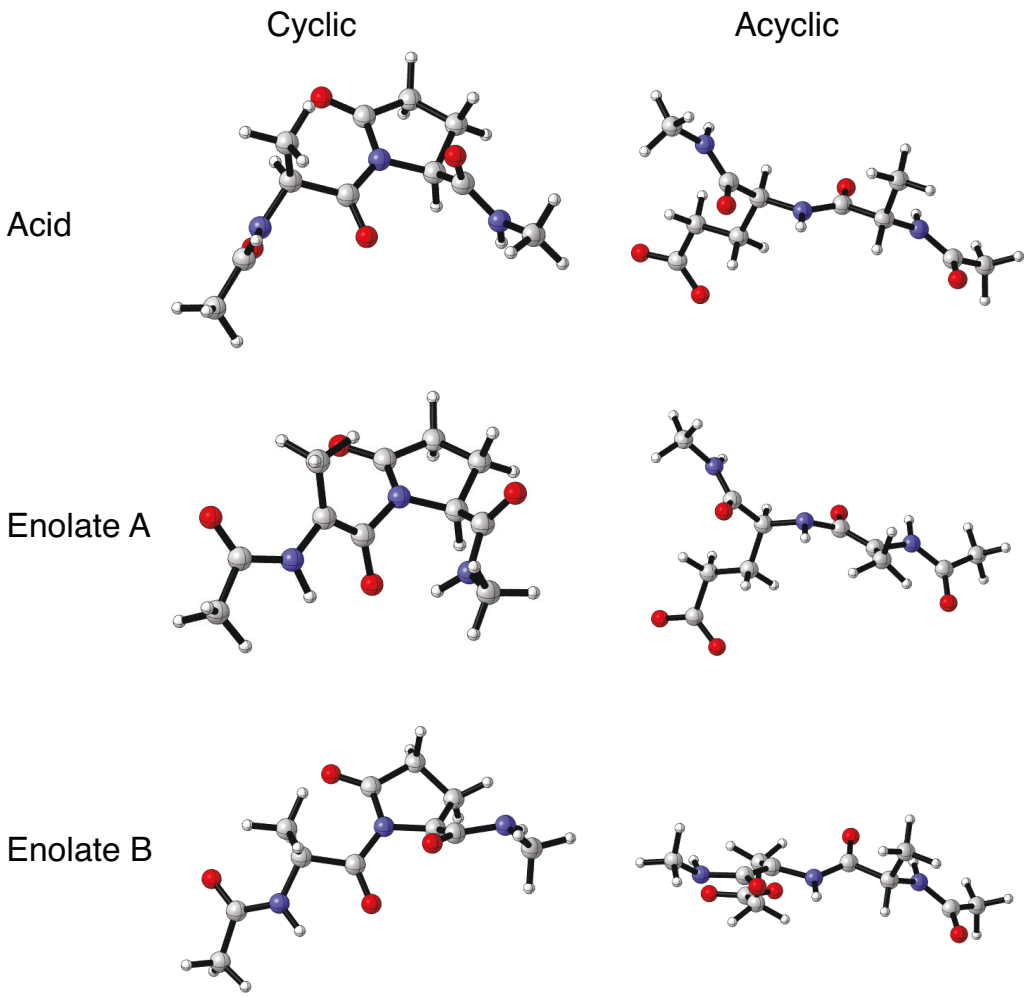


Figure 3.29: Computed structures of Ala-Glu and Ala-pyroGlu model enolates.

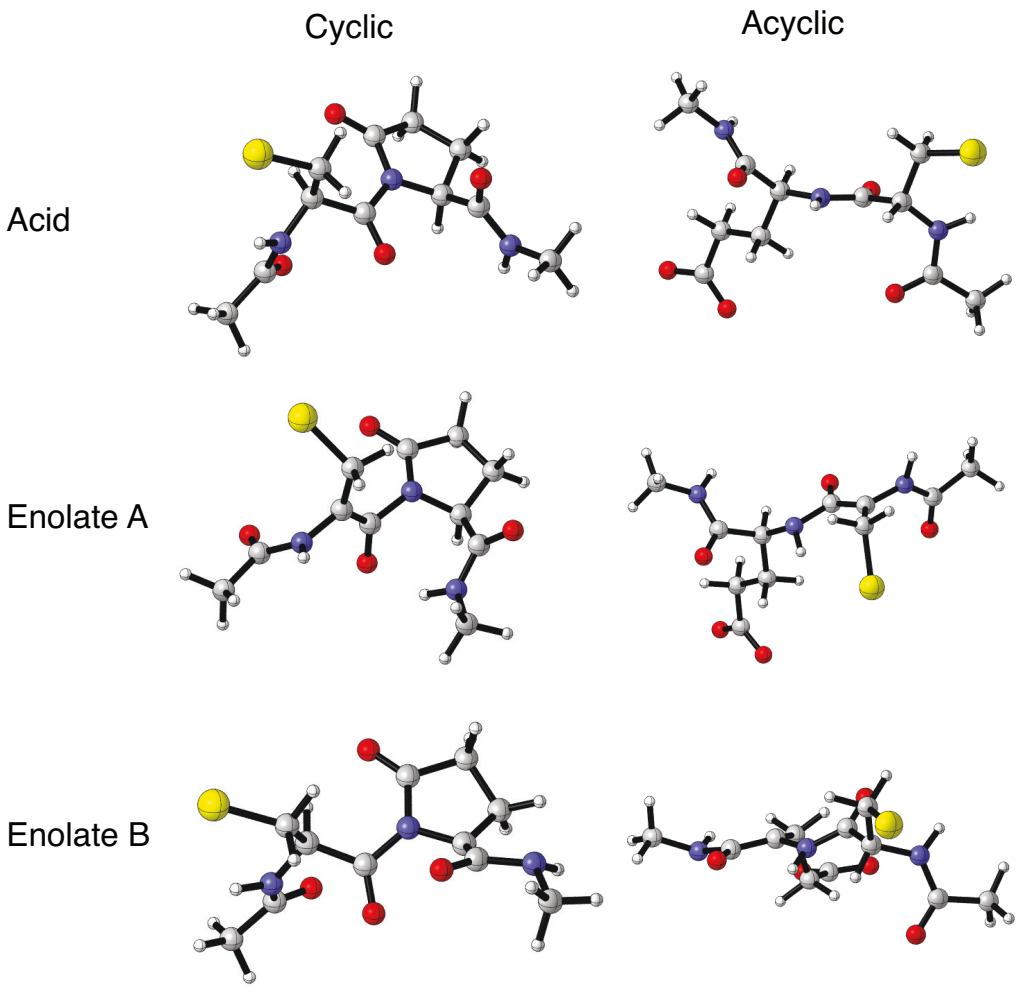


Figure 3.30: Computed structures of Cys-Glu and Cys-pyroGlu model enolates (thiolate form).

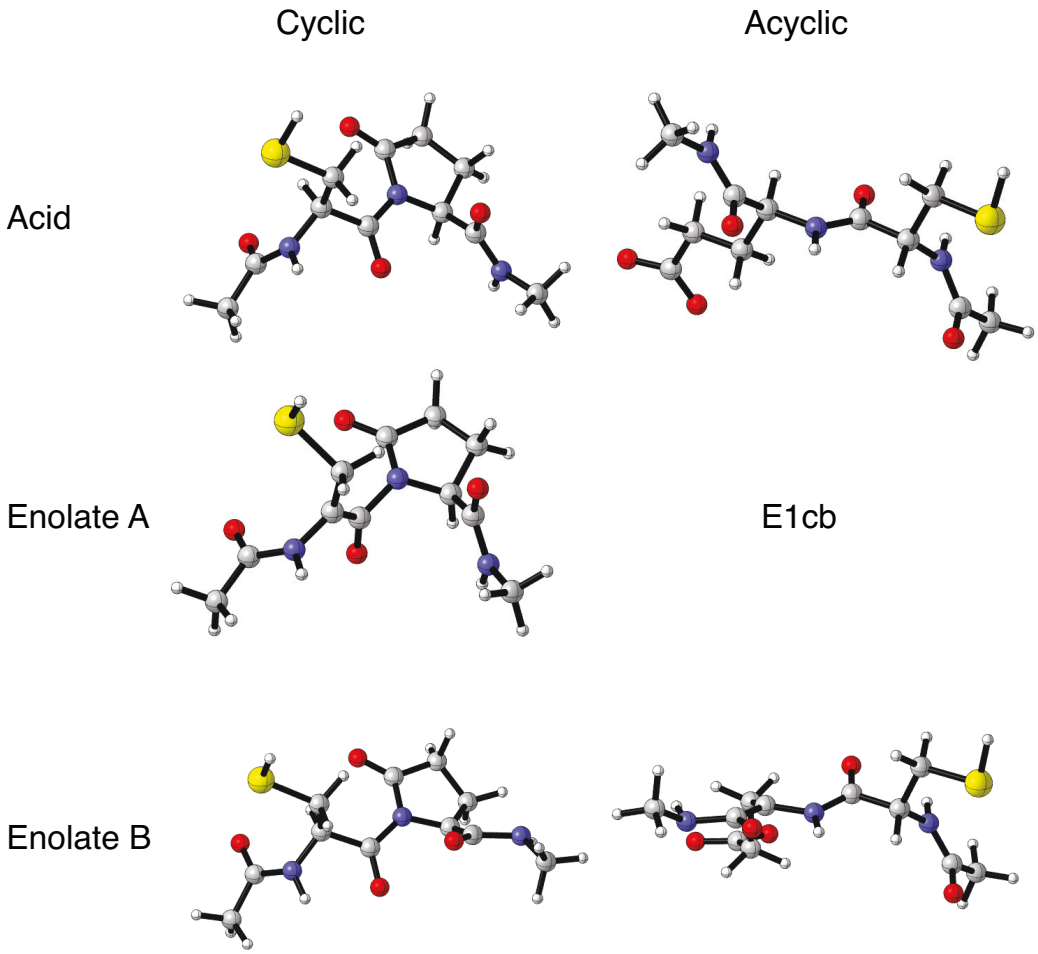


Figure 3.31: Computed structures of Cys-Glu and Cys-pyroGlu model enolates (thiol form). "E1cb" denotes that the structure underwent an E1cb elimination, precluding the calculation of this species.

3.10 Conclusions

Though these studies provide a mechanistic picture for how OGT effects HCF-1 cleavage, several mechanistic details still require further study. Though mutagenesis data has provided insight into the requirements, or lack thereof, for the residues adjacent to the glutamic acid, their precise roles are still unknown. Furthermore, there is an absence of information regarding the proton-transfer dynamics occurring during the cleavage reaction: what is the source of the proton that is transfered to the GlcNAc upon formation of the pyroglutamate? How is the proton removed from the amide N–H that attacks the glycosyl ester? The mechanistic picture provided from this work currently implicates no residues in OGT for catalyzing any part of HCF-1 cleavage beyond the initial glutamate glycosylation. Regarding the internal pyroglutamate species, whether the two species are indeed diastereomers, or whether the fact that there are two species has any biological relevance has yet to be shown explicitly.

Overall, the work reported here illuminates how the glycosyltransferase OGT catalyzes amide bond cleavage. Although the rates of cleavage of the model peptide studied here may appear slow, they are consistent with the slow rates estimated from pulse–chase data for cleavage of HCF-1 in cells [255]. The relevant timescale for HCF-1 cleavage is the length of the mammalian cell cycle, and the rates of peptide cleavage measured here are fully compatible with that timescale. To the best of our knowledge, HCF-1 is the first example of a protein that undergoes a physiologically relevant backbone cleavage reaction triggered by glutamate glycosylation, but a glycosidase was recently reported to undergo *in vitro* cleavage and a similar mechanism was proposed [129]. Therefore glutamate glycosylation may be a general way to effect amide bond hydrolysis in proteins. OGT may have other cleavage substrates, and because 1-2% of genomic space throughout all kingdoms of life

is occupied by glycosyltransferases [239, 49], there may exist other enzymes that catalyze glycosylation-mediated amide bond cleavage.

Acknowledgments

I am grateful for support from the National Science and Engineering Research Council (NSERC) of Canada as a PGS-M and PGS-D3 fellowship recipient. This work was funding by a National Institutes of Health grant (R01 GM094263) to Suzanne Walker. I thank Bruker Daltonics (Billerica, MA), specifically Christopher Thompson, who graciously provided access to and assistance with a SolariX XR 7T q-Q-FT-ICR mass spectrometer. This was crucial for obtaining structural information about the glycopeptide intermediate. I would like to thank David J.Vocadlo (Simon Fraser University, Vancouver, BC, Canada) for providing UDP-5*S*GlcNAc. I also thank JoAnne Stubbe (MIT), Daniel Kahne, and Walker lab members Zebulon G. Levine and Sara E. S. Martin for critical reading of the *Nat. Chem. Biol.* manuscript. B. McKay Wood provided assistance with the q-ToF mass spectrometer in the Walker lab.

X-ray diffraction data were collected at the National Synchrotron Light Source at Brookhaven National Laboratory (Beamline X25). Coordinates and structure factors were deposited under accession code 5HGV in the PDB.

3.11 Methods

3.11.1 Reagents and General Methods

^{14}C-UDP-GlcNAc was obtained from American Radiochemicals (ARC). UDP-GlcNAc was obtained from Promega as a 100 mM solution in water. THP was obtained from Novagen as a 0.5 M solution in water. Calf alkaline phosphatase (CIP) was obtained from New England Biolabs. Peptides were obtained from Biomatik (Cambridge, ON, Canada); their purity was evaluated by HPLC and mass spectrometry, and they were used as received. Lyophilized peptides were dissolved in water to ≈5 mM and added to reactions according to the final concentration desired. Peptide concentrations were determined by measuring A_{274} (λ_{max}, tyrosine) in water, assuming an extinction coefficient of 1400 $M^{-1} \cdot cm^{-1}$. A complete list of peptides used in this study is provided in Table B.1. The general cleavage peptide, E10-32, is derived from human HCF-1 repeat 2, but contains Gln13 rather than Glu13. This E13Q point mutant shows no appreciable difference in reaction rate from a peptide bearing the native sequence, and was used so that a single system could be employed for both reaction monitoring and isotope labeling studies. Numbering starts from the second residue in the HCF-1 peptides (valine) to be consistent with HCF-1 repeat numbering in the literature.

3.11.2 Production of HCF-3R constructs

HCF-3R constructs (3R-EAA, 3R-SAA, 3R-DAA and 3R-AAA; see Table B.5) were expressed and purified as outlined in section 2.11.4.

3.11.3 Production of OGT

hOGT4.5 (D554N), used for crystallography, was prepared by Michael B. Lazarus, as previously reported [142]. ncOGT constructs: wild-type OGT, OGT (D554N), OGT (K842M) (sequences outlined in Table B.7) were expressed and purified as outlined in section 2.11.5.

3.11.4 *In vitro* HCF-3R cleavage for SDS-PAGE analysis

In a 20 μL reaction volume was mixed: 1 mM UDP-GlcNAc, 1 mM THP, 100 μM HCF-3R construct, and 0.1 U CIP in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. Either OGT (to a final concentration of 2.5 μM) or water were added, and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After 6 hours, reactions were quenched by the addition of an equal volume (20 μL) of 2×Laemmli loading buffer (with β-mercaptoethanol). Samples were boiled for 10 minutes, and pelleted at 15,000×g for 10 minutes. 5 μL of sample was loaded per lane onto a 4-20% Tris-Gly gradient gel and run for 45 minutes at 200 V. The gel was stained with Coomassie Brilliant Blue overnight, and destained the following day until bands were clearly visible and a suitable background was obtained.

3.11.5 *In vitro* HCF-3R glycosylation for SDS-PAGE (autoradiography) analysis

In a 10 μL reaction volume was mixed: 1 mM UDP-GlcNAc (100 μM hot; 900 μM cold), 1 mM THP, 55 μM HCF-3R construct, and 0.1 U CIP in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. Either OGT (to a final concentration of 2.5 μM) or water were added, and reactions were mixed by pipetting then incubated at 37 °C

in a thermomixer to ensure uniform heating. After 5 hours, reactions were quenched by addition of an equal volume (10 μL) of 2×Laemmli loading buffer (with β-mercaptoethanol). Samples were boiled for 10 minutes, and pelleted at 15,000×g. 7 μL of sample was loaded per lane onto a 4-20% Tris-Gly gradient gel and run for 5 minutes at 90 V, followed by 65 minutes at 150 V. The gel was dried, using a gel drier and placed in an exposure cassette. A storage phosphor was placed on top and the phosphor and exposed for about 40 hours before being imaged with a Typhoon imager.

3.11.6 *In vitro* HCF-1 cleavage for LC-MS analysis

Typical cleavage reactions were conducted as follows: 20 μL reaction volume with components having the following final concentrations: 1 mM UDP-GlcNAc, 1 mM THP, 12 μM HCF-1 peptide, 0.1 U CIP and 12 μM OGT in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT was added last and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After the desired reaction time, reactions were quenched by addition of four volumes (80 μL) of ice-cold methanol supplemented with 0.1% formic acid. Reactions were then centrifuged at 16,100×g for 10 minutes to remove any precipitated protein (4 °C), and transferred to fresh tubes. Samples were concentrated to dryness using a speedvac (with T = 30 °C), for about 60 minutes. Samples were re-suspended in 100 μL 0.1% formic acid, mixed by vortexing, and centrifuged for 10 minutes at 15,000×g (4 °C) to remove any insoluble material. Samples were then transferred to LC vials, and maintained at 4 °C until they were analyzed.

3.11.7 Enzyme Stoichiometry for LC-MS experiments

Typical reactions, as described in 3.11.6, were conducted with a 1:1 peptide:OGT stoichiometry. Where noted, single turnover (STO) experiments were conducted with 3μM peptide and 27.5 μM OGT to ensure that all peptide was bound to OGT.

3.11.8 General LC-MS Methods

General measurements of peptides were performed on a q-Exactive Plus FT-mass spectrometer (Thermo Scientific^{TM}) equipped with an Ultimate 3000 uHPLC. Reaction mixture components were separated using a 33 minute reverse phase HPLC method (Table B.8); the last half of the LC method was empirically optimized to wash the column and re-equilibrate between runs to maintain column performance. with a Kinetex® C18 (2.6 μm particle size, 100 Å pore size, 2.1 mm ID, 150 mm length) at a constant flow rate of 0.200 mL/minute and column temperature of 35 °C. 5 μL of sample was injected, unless otherwise specified. Detection on the q-Exactive Plus mass spectrometer was performed between 300-2000 *m/z*, using an acquisition target of 3×10^6, maximum IT of 200 ms at a resolution of 70,000 for MS and 35,000 for MS/MS data. For MS/MS analysis, dynamic exclusion was turned off and a global inclusion list was used to maximize S/N for the targeted analysis. A collision energy of 30 eV and an isolation window of 6 *m/z* was used for higher-energy collision-induced dissociation (HCD).

For some experiments additional MS/MS information was obtained using a q-ToF (CID). These measurements were conducted on a Bruker Impact HD q-ToF mass spectrometer coupled to an Agilent 1290 uHPLC. A collision energy of 35 eV and an isolation window of 6 *m/z* was used for collision-induced dissociation (CID). The source and ToF parameters were optimized for the transmission of higher molecular weight ions. A counter

current drying gas flow rate of 12 L/min., and a nebulizing gas pressure of 45 bar were used. Funnel 1 was set to 150 V, and funnel 2 was set to 36 V, with a hexapole RF of 300 V_{p-p}. A calibration segment was present at the beginning and end of the run for *m/z* calibration with sodium formate cluster using the Bruker HPC algorithm for the curve fitting.

Unless otherwise stated, extracted ion chromatograms displayed are for the indicated *m/z* ±5 ppm tolerance.

3.11.9 Purification of glycopeptide intermediate by LC

In order to obtain a suitable quantity of glycopeptide intermediate, four parallel reactions were run as follows: 20 μL reaction volume with components having the following final concentrations: 1 mM UDP-GlcNAc, 1 mM THP, 12 μM E10-32, 0.1 U CIP and 12 μM OGT (D554N) in a buffer of 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT was added last and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After 40 seconds, reactions were quenched by addition of four volumes (80 μL) of ice-cold methanol supplemented with 0.1% formic acid. Reactions were then centrifuged at 16,100×g for 10 minutes to remove any precipitated protein (4 °C), and transferred to fresh tubes. Samples were concentrated to dryness using a speedvac (T = 30 °C) for about 30 minutes. Samples were combined by first re-suspending one sample in 150 μL 0.1% formic acid, vortexing, and using the solution to re-suspend the next sample, etc. The final sample was centrifuged for 10 minutes at 15,000×g (4 °C) to remove any insoluble material. The sample was then transferred to an LC vial and maintained at 4 °C. In order to simultaneously detect and collect the glycopeptide, I employed an in-line T between the column and the mass spectrometer inlet. To compensate for this change, the flow rate was adjusted to 0.2 mL/min, and the instrument tuning was adjusted

for maximum sensitivity under these conditions. For each run, 10 μL of sample was injected; each sample collected was concentrated at 30 °C by speedvac and when completely dry these were kept on ice until all had been collected. Collected fractions were combined when possible and dried by the same method until the entire initial sample had been purified. This sample was stored dry at -80 °C until analysis was performed. An aliquot of one collected fraction was re-injected (1:5 dilution) to evaluate purity and stability, and showed glycopeptide to be the predominant peak in the sample, with some starting material (which happens to elute at a close retention time) as a contaminant. Attempts to exclude starting material from the sample resulted in a low recovery of glycopeptide intermediate, likely due to the low volumes being collected. I cannot exclude the possibility that some hydrolysis of this glutamyl ester contributes to the starting material observed in the purified material.

3.11.10 Reaction of purified glycopeptide intermediate

Purified glycopeptide intermediate, obtained as described in 3.11.9, was resuspended in neutral water. Equal volumes of glycopeptide were distributed into each of seven reaction tubes. In a 20 μL final reaction volume, glycopeptide was added to buffer containing 1 mM THP, 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. At specified time points, the reactions were quenched by the addition of four volumes (80 μL) of ice-cold methanol supplemented with 0.1% formic acid. Samples were treated as described for standard LC-MS experiments, except dried samples were reconstituted in 20 μL rather than 100 μL.

3.11.11 Electron-capture dissociation (ECD) MS/MS experiments

A Bruker 7.0 T Solarix Q-q-FT-ICR mass spectrometer was used for electron capture dissociation (ECD) measurements (and ETD measurements). A dry sample of the glycopep-

tide intermediate was re-suspended in 40 μL of 50% acetonitrile/water supplemented with 0.1% formic acid. The sample was introduced at a rate of 20 μL/h using a syringe pump and a Captive Spray^{TM} nanospray source. A capillary voltage of 1000 V was used to get a stable nanospray signal. The quadruply charged species, $[M+4H]^{4+}$, for the glycopeptide was isolated using the quadrupole and subjected to ECD in the ICR cell using an ECD voltage of 1.0 V, and a time of 0.100 s. A two-megabyte acquisition file size was used to acquire 100 scans that were summed to obtain a single ECD mass spectrum. The data was internally calibrated using single point lock-mass correction using the predicted monoistopic m/z of 896.9017 for the precursor ion after an external multi-point linear calibration with arginine clusters. Data were manually inspected for the presence or absence of predicted ECD fragment ions (as determined using the UCSF MS-Product server) within a 5 ppm mass accuracy window.

3.11.12 CID and HCD MS/MS experiments

Structural information about the anhydropeptide species were obtained by MS/MS using both q-Exactive (HCD) and q-ToF (CID) instruments. For this species, and given the instrument tuning, these instruments provided complementary information, with the q-ToF datasets including better data for higher molecular weight fragment ions, particularly those with higher charge states. In order to unequivocally interpret the MS/MS data obtained for the anhydropeptide intermediate we compared it to the MS/MS spectrum of the starting material, E10-32, under identical conditions. In the case of both peptides serine/threonine dehydration in the C-terminal region of the peptide was observed as a result of CID fragmentation; only the anhydropeptide species that forms in the reaction showed dehydration in the Cys9–Glu10 region. Data were then manually inspected for

predicted fragment ions (as determined using the UCSF MS-Product server) within a 5 ppm mass accuracy window.

3.11.13 Crystallization of hOGT4.5 (D554N)

Ternary complexes were prepared by incubating hOGT4.5 (D554N) at 7 mg/mL with 1 mM UDP-GlcNAc and 3 mM CKII3K peptide. Crystals of the UDP-hOGT4.5 (D554N)-CKII3K glycopeptide ternary product complex were obtained by seeding the hOGT4.5 (D554N) protein with wild-type crystals, using the seed bead kit (Hampton) after mixing 2 μL of complex with 1 μL of reservoir consisting reservoir containing 1.6 M Li_2SO_4 and 0.1 M Bis Tris Propane, pH 7.0, as previously reported [140]. Crystals were cryoprotected in solutions consisting of the reservoir solution, supplemented with 28% xylitol and flash frozen in liquid nitrogen.

3.11.14 Data Collection, Structure Determination, and Refinement

Data were collected at NSLS beamline X25 at Brookhaven National Laboratory. Data was processed using iMosflm [13] and scaled with Scala or Aimless in the CCP4 crystallography suite [256]. The PHENIX software package was used for molecular replacement and refinement [1]. The previously determined OGT-CKII glycopeptide ternary product complex (PDB: 4GYW) was used as a search model for molecular replacement, using the program Phaser. Models were subsequently refined by rigid body refinement and multiple rounds of simulated annealing, minimization, atomic displacement parameter (ADP or B-factor) refinement and TLS refinement (determined using the TLSMD server [193]), with interspersed manual adjustments using Coot [57]. Geometric restraints for UDP-GlcNAc were generating using PHENIX Elbow, and these restraints were used throughout refine-

ment. All structural figures were made with Pymol [214] except for the electrostatic surface figures, which were created with CCP4mg [178]. Crystallographic software accessed through SBGrid [183].

3.11.15 Computational methods

The dipeptide Cys9-Gln10 was extracted from the small peptide present in PDB 4N3B, and was capped with an acetyl group at the N-terminus and an N-methyl group at the C-terminus based on the positions of the two residues flanking the dipeptide in the crystal structure. Gln10 was mutated *in silico* to glutamate to generate a model of the acyclic dipeptide, and the side chain of Gln10 was cyclized *in silico* onto its backbone amide nitrogen to generate a model of the cyclic dipeptide. For both cyclic and acyclic dipeptides, the structure along with two conjugate base structures deprotonated at the alpha positions of Cys9 and Glu10 were optimized in Gaussian '09 [64] at the B3LYP/6-31G+(d) [14, 144] level of theory using an IEFPCM [181, 182, 197] model for the water solvent; a frequency calculation was used to confirm the presence of local minima and calculate free energies at 298 K. All calculations were repeated with Cys9 present as a thiolate and mutated to alanine. To estimate pK_a values from free energy differences between each acid and conjugate base, the related molecule N,N'-dimethylacetamide was used as a standard, with its computed free energy difference calibrated to the known experimental pKa value of 29.4 in water [204]. Structures were rendered in CYLview [146].

The dipeptide model for amide/imide bond rotation was generated as described above. ΔG° values were calculated at 298 K.

Chapter 4

OGT catalyzes aspartate-to-isoaspartate isomerization via side chain glycosylation

> "But that's just how it is. You don't see it until you do, and then you wonder why you didn't see it all along."
>
> – Roger Howe (Yale Mathematics)

Apart from minor modifications, a large portion of this chapter originally appeared as [109]: "Aspartate Glycosylation Triggers Isomerization to Isoaspartate" John Janetzko and Suzanne Walker *Journal of the American Chemical Society.* Volume 139, Number 9: 3332–3335. 2017.

Contributions: I performed all experiments described in this chapter.

Abstract

OGT is an essential human enzyme that glycosylates numerous nuclear and cytoplasmic proteins on serine and threonine. It also cleaves Host cell factor-1 (HCF-1) by a mechanism in which the first step involves glycosylation on glutamate. Replacing glutamate with aspartate in an HCF-1 proteolytic repeat was shown to prevent peptide backbone cleav-

age, but whether aspartate glycosylation occurred was not examined. In this chapter I will describe the finding that OGT glycosylates aspartate much faster than it glycosylates glutamate in an otherwise identical model peptide substrate; moreover, once formed, the glycosyl aspartate reacts further to form a succinimide intermediate that hydrolyzes to produce the corresponding isoaspartyl peptide. Aspartate-to-isoaspartate isomerization in proteins occurs in cells, but was previously thought to occur exclusively via non-enzymatic means. These findings suggest that there may also be an enzyme-catalyzed component. In addition to OGT, enzymes that may catalyze aspartate-to-isoaspartate isomerization would include PARPs, enzymes known to ribosylate aspartate residues in the process of poly(ADP-ribosyl)ation.

4.1 Introduction

As discussed in earlier chapters, OGT is an essential mammalian enzyme known to catalyze two distinct post-translational modifications (PTMs) using the same active site [108, 149]. One is the transfer of *N*-acetylglucosamine to Ser/Thr residues of nuclear and cytoplasmic proteins [20, 92], and the other is cleavage of Host cell factor-1 (HCF-1) [27], a transcriptional co-regulator found in several chromatin-associated complexes [196, 236]. Ser/Thr *O*-GlcNAcylation affects localization, stability, and enzymatic activity of a wide variety of proteins and may serve to link cell state to glucose levels [20, 92]; HCF-1 cleavage is required for cell cycle progression [27, 115, 165]. HCF-1 cleavage occurs via glycosylation on glutamate followed by on-enzyme formation of an internal pyroglutamate that undergoes spontaneous hydrolysis [107]. Glycosylation is a common feature of both OGT-catalyzed PTMs. Nevertheless, the discovery that OGT can glycosylate glutamate was surprising because there are major steric and electronic differences between Ser/Thr and Glu side

chains. Given OGTs capacity to glycosylate glutamate, we wondered whether it could also glycosylate aspartate.

4.2 A model substrate for aspartate glycosylation

Glycosylation on aspartate was assessed using a variant of the peptide previously employed to investigate the mechanism of HCF-1 cleavage by OGT [107]. The cleavage peptide, E10-32 (Discussed in Chapter 3; Figure 4.1 and Table B.1), contains a threonine-rich region that binds to the tetratricopeptide repeat domain of OGT and a cleavage region consisting of "CET", with the site of cleavage being the amide bond between Cys9 and Glu10 (HCF-1 repeat numbering) [141]. It was demonstrated previously that the residues adjacent to the glutamate can be varied without abolishing cleavage (see Chapters 2 and 3); however, the glutamate itself is indispensable [27, 107, 117, 141].

I replaced glutamate with aspartate (D10-32) and incubated this peptide with UDP-GlcNAc and wild-type OGT for varying amounts of time before analyzing the reaction mixture by high-resolution LC-MS (henceforth, unless otherwise specified, "OGT" implies the use of wild-type enzyme). A glycopeptide [M+GlcNAc, 892.3978 (z=4+)] formed within a minute of initiating the reaction and continued to accumulate for approximately fifteen minutes before decreasing. When aspartate was replaced with alanine (A10-32), this gly-

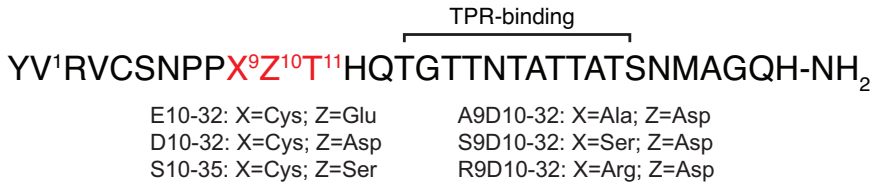


Figure 4.1: Aspartate glycosylation, and related, substrates. A complete list of peptides used in these studies is provided in Table B.1.

copeptide was not observed (Figure 4.2). We therefore concluded that glycosylation had occurred on the aspartate side chain.

It is worth noting, as mentioned briefly in the caption to Figure 4.2, that the signal observed in the [M+GlcNAc] EIC increases linearly over time for all peptides tested. Furthermore, the signal becomes diffuse after fifteen minutes due to numerous species that appear to be only partly resolved by LC, combined with the low absolute abundance of the glycopeptide species.

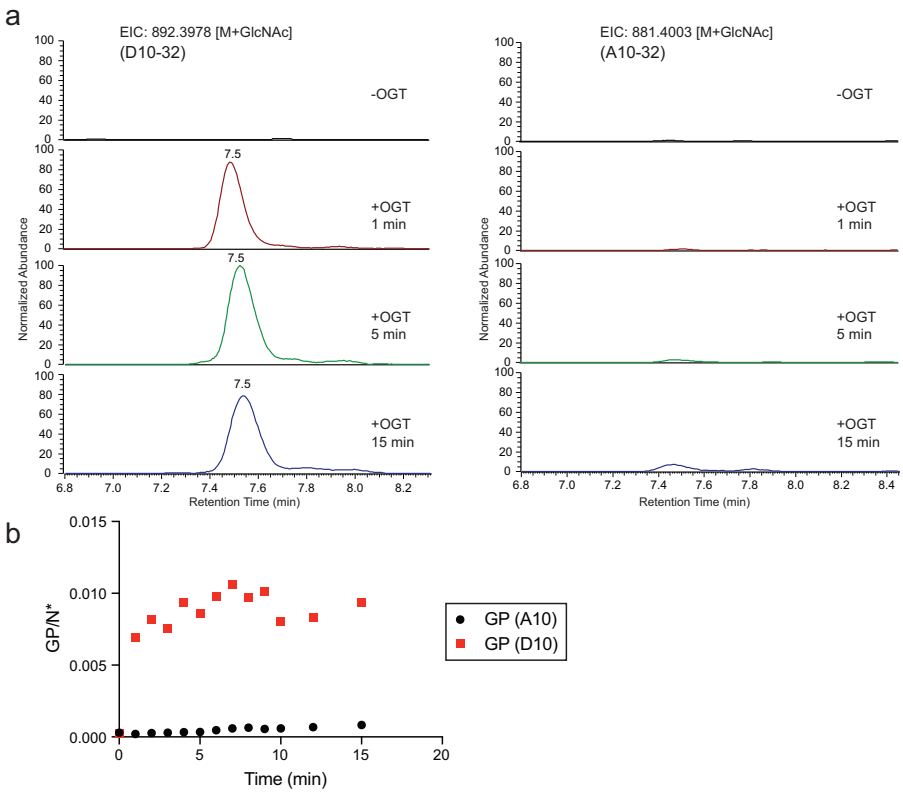


Figure 4.2: A glycopeptide intermediate forms with D10-32, but not A10-32. (a) EIC traces for the expected [M+GlcNAc] with D10-32 (left) and A10-32 (right); Vertical axes are normalized to 2.83×10^5 counts. It is noteworthy that with all peptides used a gradual accumulation of several glycopeptides is observed, which we attribute to glycosylation within the T-rich region. (b) Graphical representation of short time courses performed (with representative EIC traces shown in part a) with A10-32 and D10-32. D10-32 shows rapid formation of a glycopeptide (GP) that appears to remain at a steady-state level over 15 minutes. Quantities are expressed as a ratio relative to an internal standard, N^*.

4.3 Identifying and characterizing a succinimidyl peptide

The decrease in glycopeptide abundance over time suggested that this species was being converted to other species. Further analysis of the reaction mixtures revealed the appearance of an anhydropeptide [M-H_2O, m/z 837.1253 (z=4+)] as well as a peptide with the same exact mass (isobaric) as the starting material but with a different retention time by LC (henceforth S.M. isomer)(Figure 4.3). The formation of both species was dependent on the presence of both OGT and UDP-GlcNAc (Figure 4.4).

A time course, conducted under substoichiometric conditions, showed that glycopeptide appeared before anhydropeptide, which in turn appeared before the starting material isomer (Figure 4.5). From these observations we concluded that the anhydropeptide had formed from the glycosyl aspartate, and we considered three pathways by which a glycosyl aspartate could progress to an anhydropeptide (Figure 4.6).

Pathway "c" seemed unlikely as it required the formation of a strained β-lactam ring (see Figure 4.6 caption). To test whether the anhydropeptide was a cyclic thioester (pathway "a"), I replaced Cys9 with alanine (A9D10-32), serine (S9D10-32) and arginine (R9D10-32). These substitutions were previously shown to be compatible with glutamate glycosylation and cleavage [107, 141]. For each peptide we observed both an anhydropeptide and a peak isobaric with the corresponding starting material (Figure 4.7). Therefore, the anhydropeptide observed in reactions with D10-32 was not a cyclic thioester. The other pathway (pathway "b") leads to a succinimide. Succinimides have precedent in biological systems, where they are known to form spontaneously from certain aspartic acid (by dehydration) or asparagine residues (by deamidation) [9, 69, 79, 127, 173, 177, 222, 225].

Succinimides are known to be hydrolytically unstable under physiological conditions ($t_{\frac{1}{2}} = \approx 138$ min at pH 7.4 and 37 °C)[69, 177]; hydrolysis can occur at either of the two

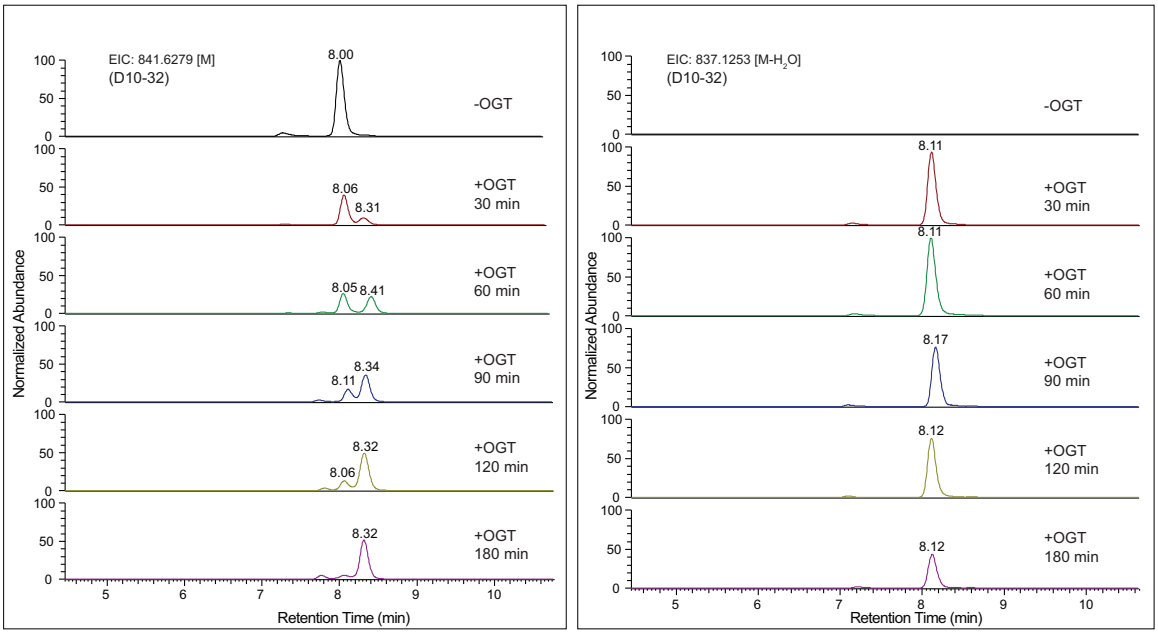


Figure 4.3: Treatment of D10-32 with OGT leads to the formation of an anhydropeptide and a new species isobaric with starting material. (left) EIC traces for D10-32 show consumption of starting material over time, and the formation of a new species isobaric with starting material, but having a different retention time (S.M. isomer). Vertical axes are normalized to 1.33×10^7 counts. (right) EIC traces show the appearance of an anhydropeptide species. This species disappears over time. Vertical axes are normalized to 1.31×10^6 counts.

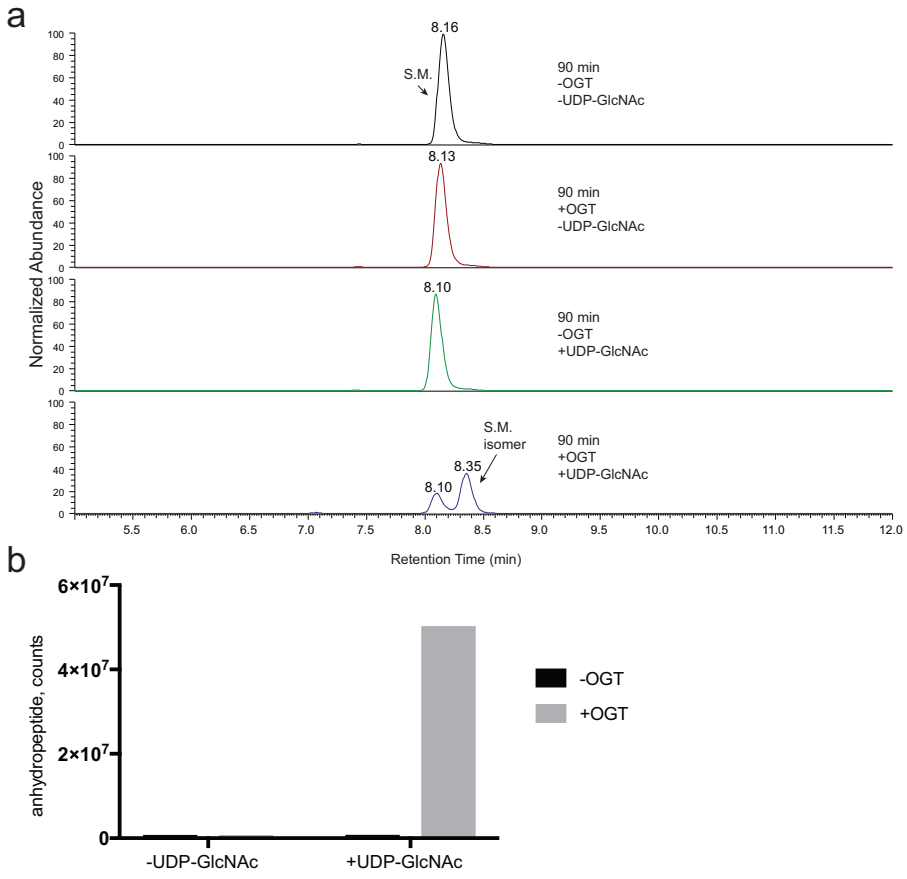


Figure 4.4: Reaction of D10-32 requires both UDP-GlcNAc and OGT (a) EIC traces for D10-32 starting material (841.6279 m/z). Vertical axes are scaled to 4×10^7 counts. Reactions were conducted under standard isomerization conditions, except UDP-GlcNAc was omitted from the top two reactions. (b) Anhydropeptide formation (measured by integration of EIC for 837.1253 m/z) is observed only for reactions where both UDP-GlcNAc and OGT are included.

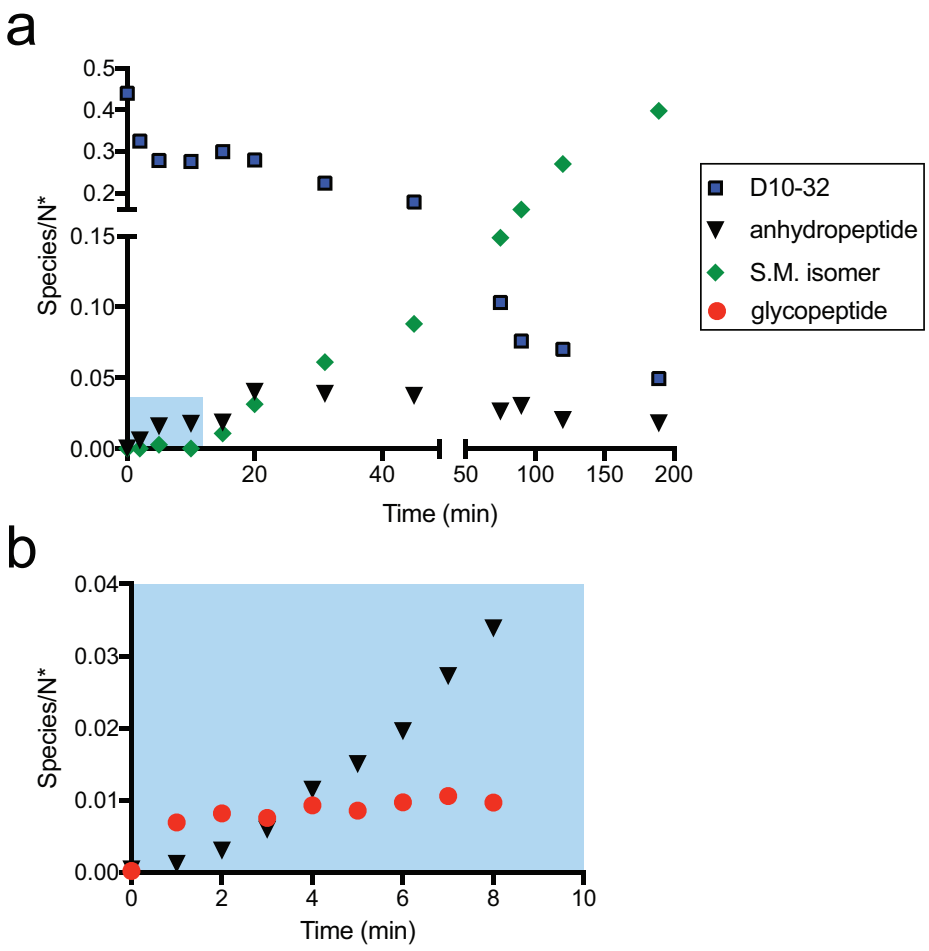


Figure 4.5: Temporal sequence for species produced during reaction of D10-32 with OGT. Under standard reaction conditions with wild-type OGT, D10-32 consumption is concomitant with the formation of glycopeptide, anhydropeptide and an isobaric species (S.M. isomer), in that order. (a) A graphical representation of a ratio of integrated ion areas for D10-32/S.M. isomer (EIC 841.6279 m/z) and anhydropeptide (EIC 837.1253 m/z) relative to the internal standard N^* (EIC 574.2668 m/z). (b) A zoom of data collected from zero to eight minutes of reaction of D10-32 with OGT, showing only integrated ion areas for glycopeptide (EIC 892.3978 m/z) and anhydropeptide (EIC 837.1253 m/z) relative to the internal standard N^* (EIC 574.2668 m/z).

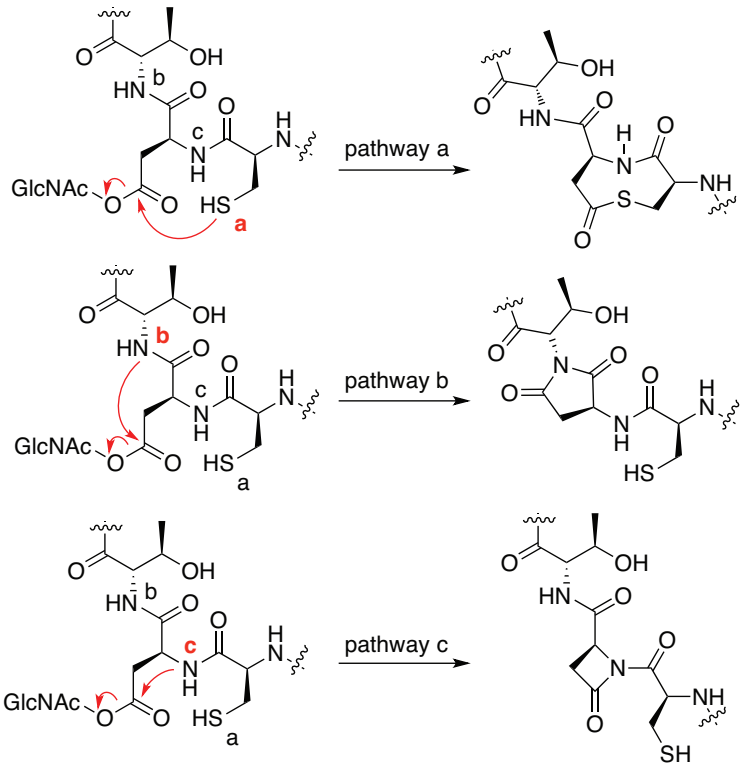


Figure 4.6: Three possible mechanisms for anhydropeptide formation from a glycosyl aspartate. Pathway "a" involves trapping by the cysteine thiol, N-terminal to the glycosyl ester; pathway "b" involves attack of the C-terminal amide nitrogen, while pathway "c" involves attack of the N-terminal amide nitrogen. The result of pathway "b" is a succinimide, compared to pathway "c", which affords an *N*-acyl-β-lactam. The product of pathway "c" is likely hydrolytically unstable [241], with the potential to be attacked by water at either the endocyclic carbonyl to regenerate an aspartic acid residue, or at the exocyclic carbonyl to affect peptide backbone scission, the latter of which is not observed when D10-32 is treated with OGT/UDP-GlcNAc.

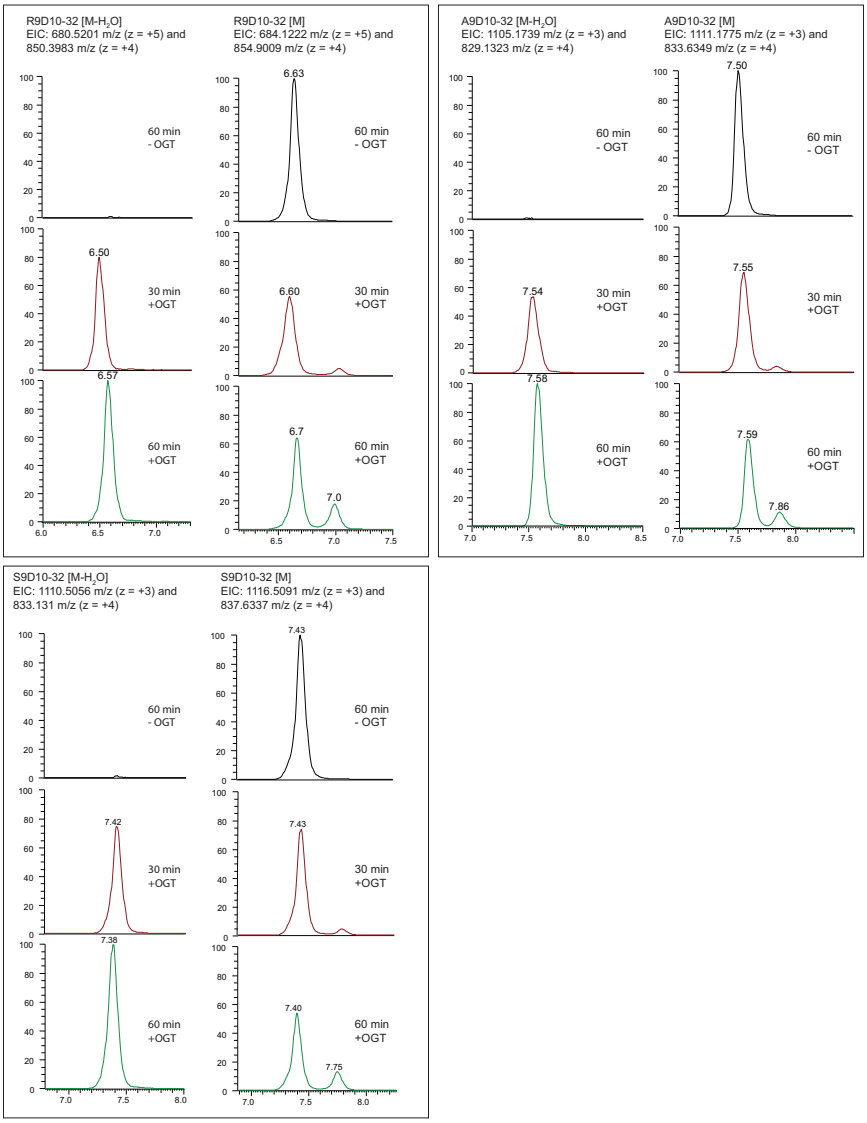


Figure 4.7: Cysteine 9 is not required for the formation of an anhydropeptide or an isobaric starting material species. EIC traces corresponding to [M-H_2O] and [M] for C9X (X = A, S, R) variants are shown. Each peptide shows the formation of an anhydropeptide and a peak isobaric with the corresponding starting material. Vertical axes are normalized to the maximum number of counts observed for each given species. Horizontal axes are retention time in minutes. Reactions were conducted under standard isomerization conditions.

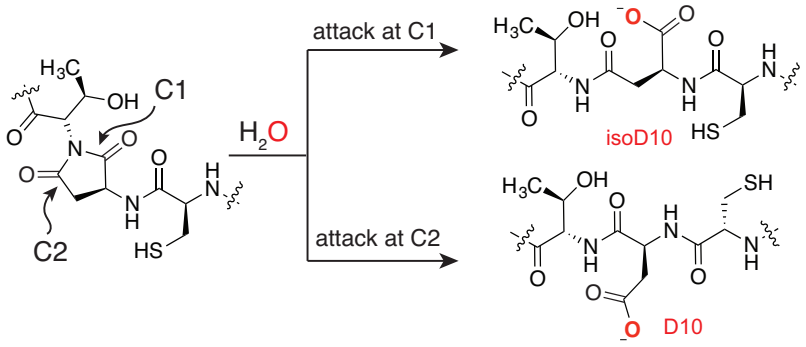


Figure 4.8: Succinimide hydrolysis can produce isoaspartate or aspartate. Succinimide hydrolysis can occur at either C1 or C2 to produce isoaspartate and aspartate, respectively. The oxygen atom incorporated from bulk water is shown in red.

endocyclic carbonyl groups to produce either an aspartic acid or isoaspartic acid species (Figure 4.8), however the regiochemistry of this attack favors the formation of isoaspartate by between 3:1 to 5:1 [84, 128].

4.4 Characterizing an isoaspartyl peptide

4.4.1 ^{18}O incorporation supports succinimide and isoaspartate formation

As isoaspartate and aspartate are isomeric, hydrolysis of a succinimide would explain the isobaric peaks observed in reactions of the D10-containing peptides with OGT. This would also be consistent with the fact that the isobaric species formed last in these reactions (Figure 4.5). Incorporation of ^{18}O from bulk water is commonly employed to demonstrate the presence of a succinimide within a polypeptide [79, 153, 228].

Incorporation of ^{18}O can be monitored through changes in the isotope distribution for a given charge state of the peptide [114]. For each ^{18}O atom to replace an ^{16}O atom there is a $+\frac{2}{z}$ shift in the isotope envelope, where z is the charge state of the peptide,

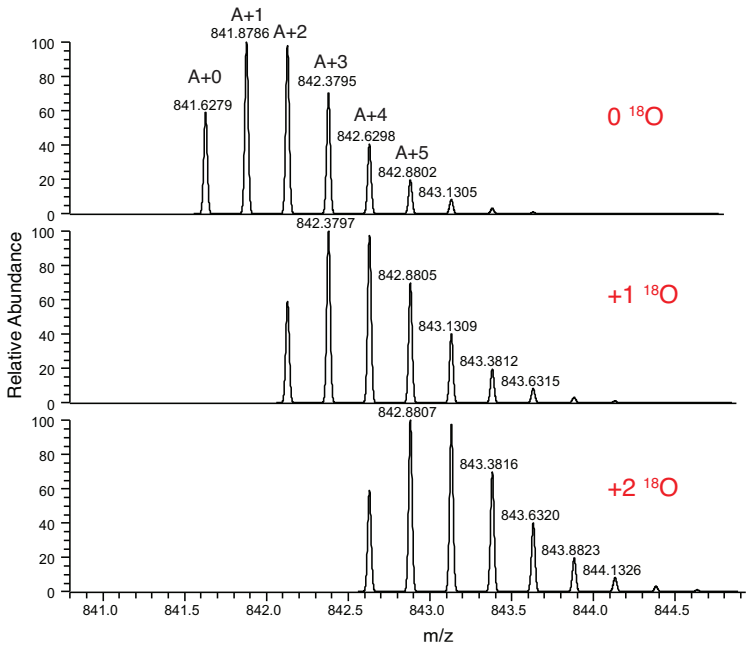
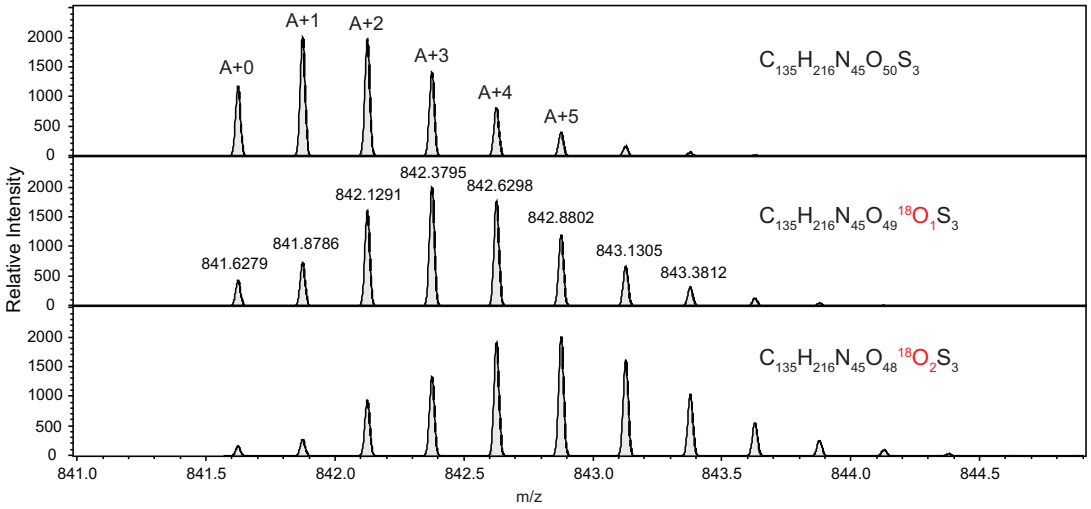


Figure 4.9: Calculated isotope envelopes for D10-32. Theoretical model for sequential ^{18}O incorporation into D10-32. Incorporation is modeled at 100 atom% abundance, and shown for the quadruply charged ion (z=4+).

as shown in Figure 4.9. I reasoned that I could accurately quantify any changes in the isotope distribution by measuring the ratio of each isotopomer peak (A+n) relative to the monoisotopic peak (A+0) over time. Since the experimental resolution for the quadruply charged state of the D10-32 peptide (m/z 841.62794, monoisotopic), was 38,506 (see section 4.6.7), this meant that the maximum Δm resolvable under these conditions would be 0.02186 Da (Equation 4.1, where R is the resolution (at FWHM), Δm is the difference (m/z) between two peaks and M is the m/z for the observed peak). Since this resolution in insufficient to resolve the mass difference between ^{16}O+2$\times^{13}$C and ^{18}O+2$\times^{12}$C (Δm = 0.002466 Da), we expect to observe accidental coincidence in the mass spectrum, rather than peak-splitting.

$$R = \frac{M}{\Delta m} \tag{4.1}$$



isotopomer	m/z	% of Total Zero ^{18}O	One ^{18}O	Two ^{18}O
A+0	841.62794	13.89	4.59	1.52
A+1	841.87863	24.32	8.04	2.66
A+2	842.12906	24.44	17.4	8.84
A+3	842.37965	17.92	22.24	12.75
A+4	842.63013	10.56	19.86	18.23
A+5	842.88055	5.26	13.72	19.44
A+6	843.13093	2.28	7.8	15.88
A+7	843.38128	0.88	3.79	10.42
A+8	843.63162	0.31	1.62	5.74

(A+n)/(A+0)	Zero ^{18}O	One ^{18}O	Two ^{18}O
A+ 0	1.00	1.00	1.00
A+ 1	1.75	1.75	1.75
A+ 2	1.76	3.79	5.82
A+ 3	1.29	4.85	8.39
A+ 4	0.76	4.33	11.99
A+ 5	0.38	2.99	12.79
A+ 6	0.16	1.70	10.45
A+ 7	0.06	0.83	6.86
A+ 8	0.02	0.35	3.78

Figure 4.10: Theoretical model for ^{18}O incorporation into D10-32. A model for ^{18}O incorporation (at 67 atom% abundance, see section 4.6.7) was constructed to obtain expected $\frac{A+n}{A+0}$ ratios.

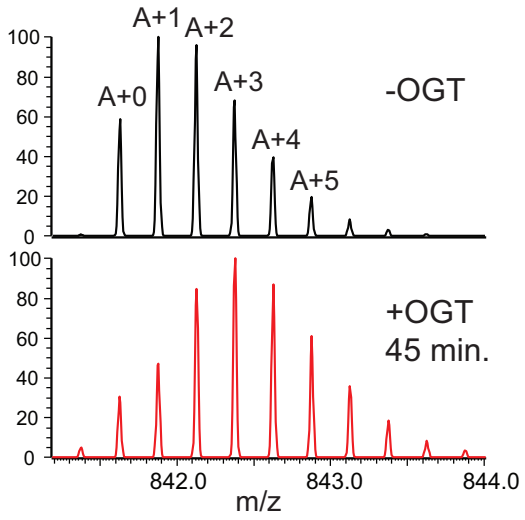


Figure 4.11: Observed ^{18}O incorporation into D10-32. A comparison of the isotope envelope for the quadruply charged state of D10-32 in ^{18}O-water, without OGT (top), or after 45 minutes with wild-type OGT (bottom).

Under the experimental conditions employed, the maximum isotope enrichment of ^{18}O would be 67 atom% (see section 4.6.6). As such, I constructed a model isotopic distribution where zero, one or two molar equivalents of ^{18}O had been incorporated into the D10-32 peptide (at this abundance), and determined the expected $\frac{A+n}{A+0}$ ratios (Figure 4.10). I then conducted several time course experiments using ^{18}O-water. With wild-type OGT, under single-turnover conditions, a clear shift in the isotope envelope was observed (Figure 4.11). ^{18}O incorporation is time-dependent, and required catalytically active OGT, as evidenced by the lack of incorporation when OGT (K842M) was employed (Figure 4.12). OGT (K842M) is compromised in it's ability to activate UDP-GlcNAc and does not allow for glycosyltransfer [213]. Unsurprisingly, the use of wild-type OGT and ^{16}O-water also showed no shift in the D10-32 isotope envelope. When E10-32 was reacted with wild-type OGT in the presence of ^{18}O-water, no shift in the isotope envelope was observed (Figure 4.13). This

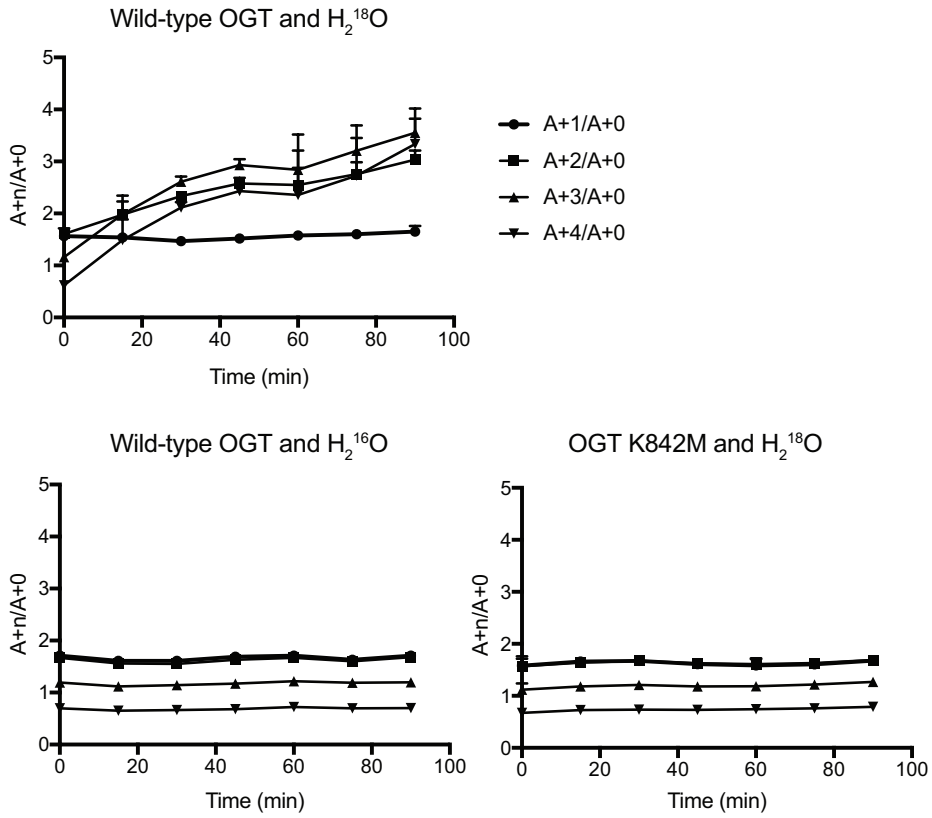


Figure 4.12: ^{18}O incorporation requires active OGT. Plots of $\frac{A+n}{A+0}$ ratios over time for D10-32 treated with wild-type OGT and ^{16}O water (bottom left), with OGT (K842M) and ^{18}O water (bottom right), and with wild-type OGT and ^{18}O water (top left). Experiments were run in duplicate and all points are mean ± s.e.m.

finding suggested that while both D10-32 and E10-32 form glycosyl ester intermediates, the glutamyl ester does not revert to starting material (via direct hydrolysis) to any appreciable extent. Presumably the same is true for the aspartyl GlcNAc ester, which lends further support to the notion that hydrolysis occurs from the succinimide species, rather than the glycosyl ester. I compared the observed $\frac{A+n}{A+0}$ ratios to theoretical values for ^{18}O incorporation at various levels of isotope enrichment. Under my experimental conditions, the abundance of ^{18}O is 67%, which means that for each molar equivalent of water to be

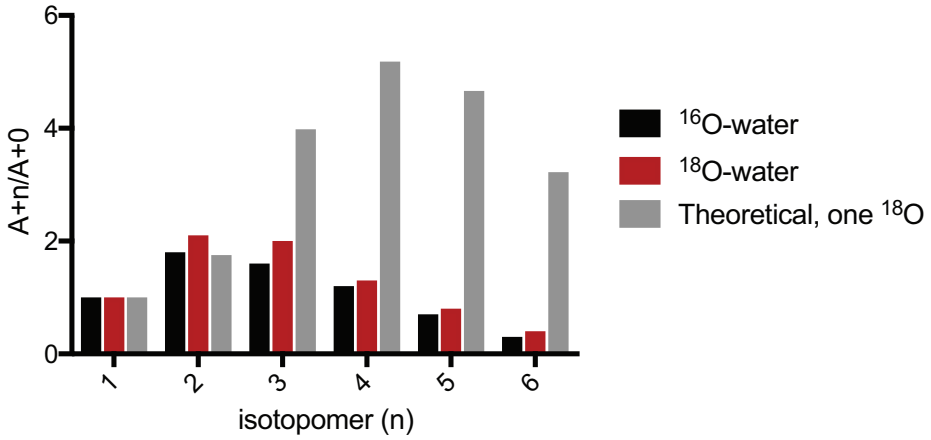


Figure 4.13: ^{18}O incorporation does not occur for E10-32. Reaction of E10-32 with wild-type OGT in either ^{16}O or ^{18}O water shows no ^{18}O incorporation into the starting material peak. Reactions were conducted once, and run under single turnover conditions for 90 minutes. Theoretical incorporation ratios are for 67 atom% ^{18}O.

incorporated/exchanged, 67% would be ^{18}O and 33% would be ^{16}O. Indeed, $\frac{A+n}{A+0}$ ratios for the incorporation of one ^{18}O atom fit well to experimental data (Figure 4.14) suggesting that under these reaction conditions a single ^{18}O was incorporated into the D10-32 peptide.

I mapped the site of ^{18}O incorporation using MS/MS (Figures 4.15 and 4.16, Tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6). Compared to fragment ions for reactions carried out in ^{16}O water, no shift was observed for b ions up to b_9 or y ions up to y_{11}, which was as far as we could detect; however, a systematic shift of +2 Da was observed for each ion from b_{12} through b_{20}. These data showed that a single ^{18}O atom had been incorporated within the tripeptide sequence spanning residues 9 to 11 (Figure 4.15). Taken together with other data, we concluded that ^{18}O had been incorporated into the side chain of Asp10 via hydrolysis of a succinimide produced from a glycosyl aspartate.

Knowing that a single ^{18}O atom was being incorporated from bulk solution, we reasoned

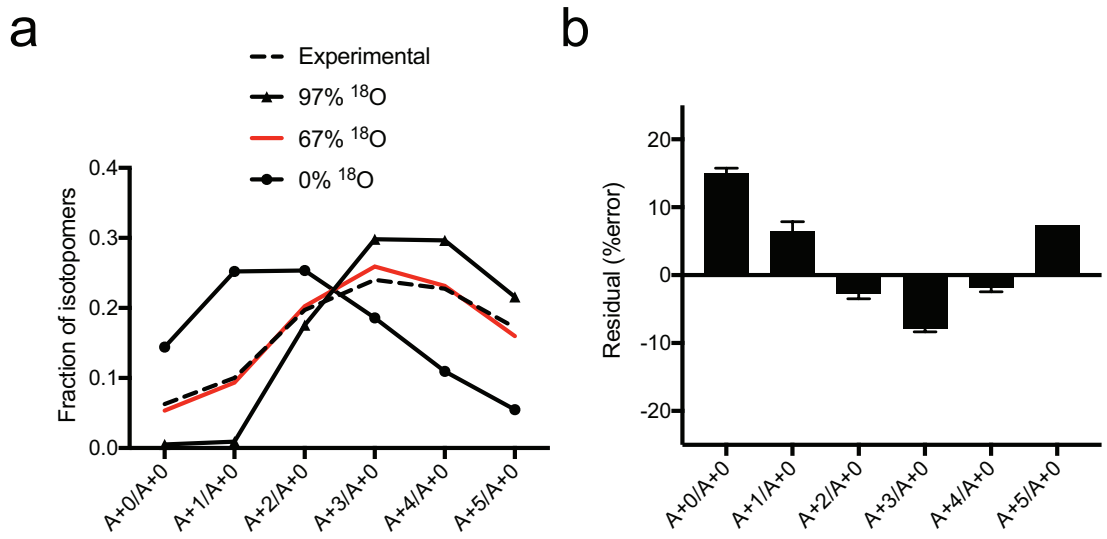


Figure 4.14: Fitting experimental ^{18}O incorporation to a theoretical model. (a) Modeled isotope patterns are shown for D10-32 with 0% ^{18}O, 67% ^{18}O and 97% ^{18}O, along side the experimentally observed pattern for D10-32 after treatment with wild-type OGT and ^{18}O-water for 90 minutes. The y-axis represents the fraction that a given isotopomer comprises relative to the sum of the first six isotopomers (A+0 through A+5). (b) Residuals, determined as the difference between the experimental isotopomer abundance at 90 minutes and the calculated abundance expected for 67 atom% incorporation of a single ^{18}O atom. Residuals are expressed as a % error, which was computed by dividing the obtained residual by the experimental value and expressing it as a percentage.

Residue	C-term.	y^+ obs. (m/z)	Error (ppm)
H(amidated)	1	155.0926	0.64
Q	2	283.1510	1.06
G	3	340.1723	1.47
A	4	411.2096	0.73
M	5	542.2501	0.55
N	6	656.2922	1.68
S	7	743.3243	1.35
T	8	844.3715	1.78
A	9	915.4083	1.97
T	10	1016.4527	5.02
T	11	1117.5010	4.03

Table 4.1: List of y ions for isoD10-32 by HCD. isoD10-32 (S.M. isomer formed by reaction of D10-32 with OGT/UDP-GlcNAc).

Residue	N-term.	b^+ obs. (m/z)	Error (ppm)
Y	1	–	–
V	2	263.1389	0.38
R	3	419.2399	0.48
V	4	518.3092	1.35
C	5	621.3170	1.13
S	6	708.3501	0.42
N	7	822.3917	1.22
P	8	Ambiguous	–
P	9	Ambiguous	–
C	10	Interference with $y_{10}+$	–
D	11	Ambiguous	–

Table 4.2: List of singly-charged b ions for isoD10-32 by HCD. isoD10-32 (S.M. isomer formed by reaction of D10-32 with OGT/UDP-GlcNAc).

Residue	N-term.	b^{2+} obs. (m/z)	Error (ppm)†	Mass shift (Da)‡
N	7	411.6997	0.73	–
P	8	460.2260	0.87	–
P	9	508.7533	1.18	–
C	10	Ambiguous	–	–
D	11	Ambiguous	–	–
T	12	669.2836	1479.89	1.98
H	13	737.8244	1357.58	2.00
Q	14	801.8534	1248.67	2.00
T	15	852.3778	1175.27	2.00
G	16	880.8879	1136.39	2.00
T	17	931.4098	1072.64	2.00
T	18	981.9353	1019.13	2.00
N	19	1038.9545	960.93	1.99
Residue	N-term.	b^{3+} obs. (m/z)	Error (ppm)†	Mass shift (Da)‡
T	20	726.6568	918.88	2.00

Table 4.3: List of multiply-charged b ions for isoD10-32 by HCD. isoD10-32 (S.M. isomer formed by reaction of D10-32 with OGT/UDP-GlcNAc). †Error is measured as the difference in m/z relative to predicted b ions, expressed in ppm. ‡Mass shift measured as the difference in the monoisotopic peak within an isotope envelope between isoD10-32 peptide produced in ^{16}O-water and ^{18}O-water

Residue	C-term.	y^+ obs. (m/z)	Error (ppm)	y^{2+} obs. (m/z)	Error (ppm)
H(amidated)	1	–	–	–	–
Q	2	–	–	–	–
G	3	340.1735	2.06	–	–
A	4	411.2108	2.19	–	–
M	5	542.2505	0.18	–	–
N	6	656.2941	1.22	–	–
S	7	743.3257	2.58	–	–
T	8	844.3737	2.47	–	–
A	9	915.4116	4.41	458.2085	0.44
T	10	1016.4561	4.13	508.7343	3.54
T	11	1117.5062	1.29	559.2561	0.54

Table 4.4: List of y ions for isoD10-32 by CID. isoD10-32 (S.M. isomer formed by reaction of D10-32 with OGT/UDP-GlcNAc).

Residue	N-term.	b^{2+} obs. (m/z)	Error (ppm)†	Mass shift (Da)‡	b^{3+} obs. (m/z)	Error (ppm)†	Mass shift (Da)‡
N	19	1037.9571	0.00	2.00	–	–	–
T	20	1088.4815	0.55	2.00	725.9908	1.52	2.00
A	21	1124.0000	0.44	2.00	749.6701	1.73	2.00
T	22	1174.5367	11.41	2.00	783.3510	0.38	2.00
T	23	1225.0385	7.10	2.00	Ambiguous	–	–
A	24	1260.5637	1.59	2.00	Ambiguous	–	–
T	25	–	–	–	874.3946	1.03	2.00
S	26	–	–	–	Ambiguous	–	–
N	27	–	–	–	941.4193	1.27	2.00
M	28	–	–	–	985.1014	0.81	2.00
A	29	–	–	–	1008.7813	1.59	2.00

Table 4.5: List of b ions for isoD10-32 by CID. isoD10-32 (S.M. isomer formed by reaction of D10-32 with OGT/UDP-GlcNAc). †Error is measured as the difference in m/z relative to predicted b ions, expressed in ppm. ‡Mass shift measured as the difference in the monoisotopic peak within an isotope envelope between isoD10-32 peptide produced in ^{16}O-water and ^{18}O-water.

Residue	N-term.	b^{3+}–H_2O obs. (m/z)	Error (ppm)†	Mass shift (Da)‡
T	20	719.9888	3.61	2.00
A	21	743.6646	0.81	2.00
T	22	777.3516	4.89	2.00
T	23	811.0291	1.60	2.00
A	24	834.7090	0.48	2.00
T	25	868.3880	4.61	2.00
S	26	897.4034	0.89	2.00
N	27	935.4134	3.74	2.00

Table 4.6: List of dehydrated b ions for isoD10-32 by CID. isoD10-32 (S.M. isomer formed by reaction of D10-32 with OGT/UDP-GlcNAc). †Error is measured as the difference in m/z relative to predicted b ions, expressed in ppm. ‡Mass shift measured as the difference in the monoisotopic peak within an isotope envelope between isoD10-32 peptide produced in ^{16}O-water and ^{18}O-water.

that the first-order rate constant for succinimide hydrolysis could be obtained from the rate of ^{18}O incorporation. I divided the observed $\frac{A+n}{A+0}$ ratio at each time point by the theoretical $\frac{A+n}{A+0}$ expected for the addition of one molar equivalent of ^{18}O in order to obtain a fractional conversion. I then plotted conversions measured from $\frac{A+3}{A+0}$ and $\frac{A+4}{A+0}$ as a function of time, and fit the conversions to a linear regression. This data yielded an average first-order rate constant of 0.0039 min^{-1} ($t_{\frac{1}{2}} \approx 177$ min) (see Figure 4.17). These values are comparable to rates previously measured for the spontaneous hydrolysis of succinimides in solution under similar conditions [69, 177]. From these data it is unclear whether hydrolysis of the succinimide occurs while bound to OGT or not; however, the rate of hydrolysis is consistent with it being a non-enzymatic process.

4.4.2 LC-MS/MS and PIMT provide further evidence

To confirm that the isobaric peak observed in these reactions was an isoaspartyl peptide, as would be expected from hydrolysis of a succinimide intermediate, I obtained an authentic peptide standard, isoD10-32 (Table B.1). This peptide had the same retention time as the isobaric peptide peak observed in the OGT reactions with D10-32 (Figure 4.18).

Isoaspartate and aspartate cannot be readily distinguished in standard mass spectrometry experiments since they have the same *m/z* [55]. Several other methods have had success distinguishing Asp and isoAsp [29], including: Edman degradation [248], site-specific ^{3}H-incorporation [52, 225], carboxypeptidase Y cleavage [112], enzymatic methylation [37, 66, 78], and nuclear magnetic resonance (NMR) spectroscopy [30]. In addition, several tandem mass spectrometry-based methods have shown the ability to distinguish isoaspartate from aspartate within a polypeptide [50, 147, 150, 189]. These methods employ ETD/ECD-based fragmentation to detect a unique ion pair, c'+57 and $z^{\bullet}$–57, which

is produced when an isoaspartate is present. I obtained ETD spectra for both D10-32 and isoD10-32 and qualitatively evaluated these spectra for differences. Unfortunately there were no fragments unique to one species, and no notable differences in fragment ion intensity.

Seeking an alternative route to characterize the starting material isomer species, I turned to protein isoaspartyl methyl transferase (PIMT), an enzyme that repairs isoaspartate-containing proteins by selectively transferring a methyl group from SAM to the carboxylic acid of isoaspartate. Methylation promotes succinimide formation, and even though succinimide hydrolysis favors the re-formation of an isoaspartate, repeated cycles of methylation and succinimide hydrolysis eventually convert isoaspartate back to aspartate (Figure 4.19) [66, 78]. In the past, PIMT has been used to demonstrate the presence of an isoaspartate by quantifying the SAM-derived byproducts produced in the reaction, either methanol or *S*-Adenosyl-L-homocysteine (AdoHcy; SAH) [28]. More recently, the use of PIMT in conjunction with ^{18}O labeling has been used to identify isoaspartate residues within several polypeptides [153]. I treated D10-32 and isoD10-32 with PIMT and SAM, and only isoD10-32 reacted (Figure 4.20). Next, I treated a sample of D10-32, which had previously been incubated with wild-type OGT and UDP-GlcNAc to produce the putative isoaspartyl peptide, with PIMT/SAM; the peptide was consumed rapidly concomitant with a substantial increase in succinimide abundance (Figure 4.21). Therefore, we concluded that OGT catalyzes isomerization of D10-32 to isoD10-32.

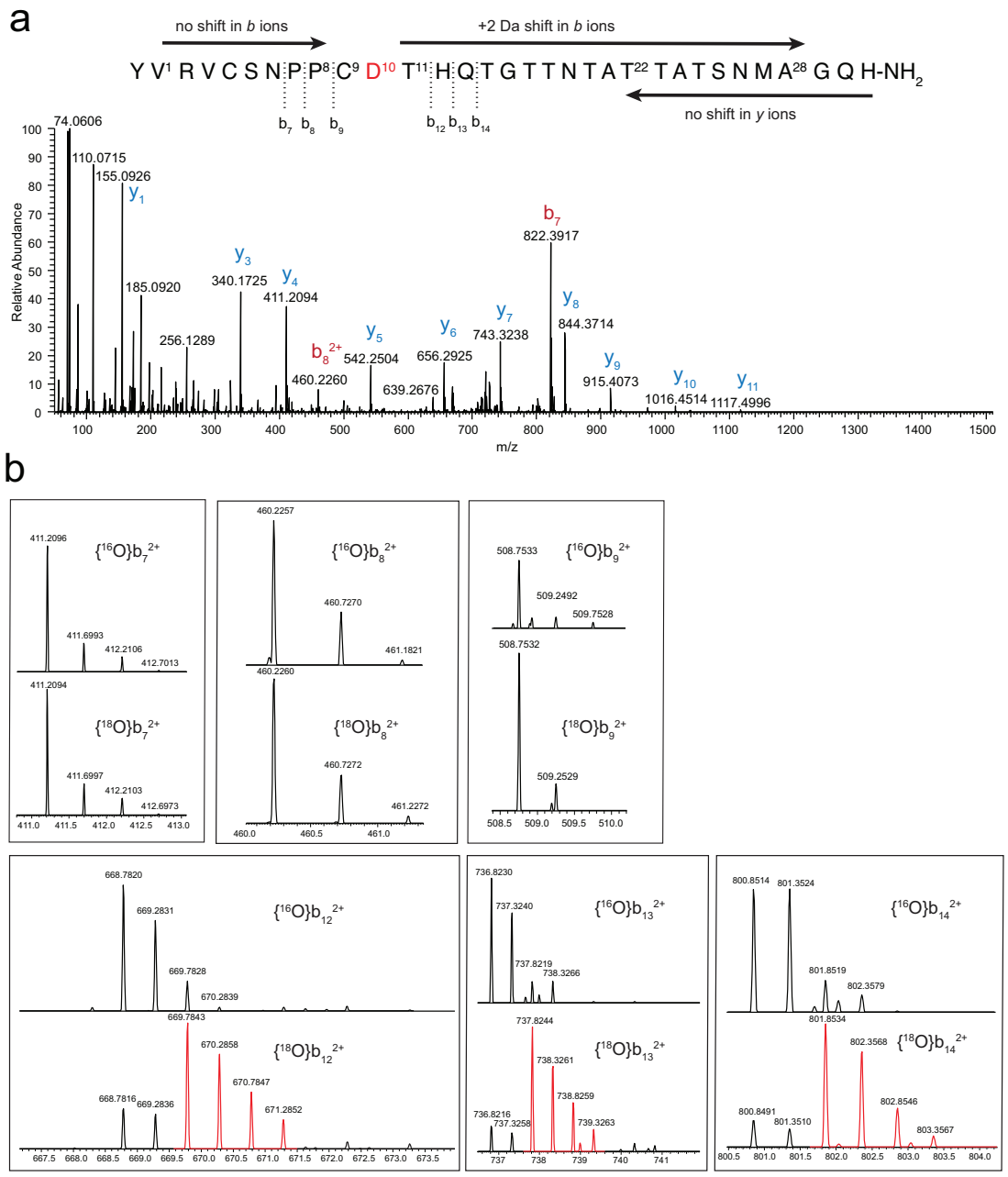
4.4.3 Mechanistic considerations

The fact that D10-32 was entirely consumed, rather than converted to an equilibrium mixture of aspartyl and isoaspartyl peptides, suggested that isoD10-32 was likely a poor

substrate for OGT. In order to test whether it reacted, I conducted a time-course experiment with isoD10-32 and wild-type OGT and compared it to one of D10-32 under identical conditions. While D10-32 was consumed, isoD10-32 was modified by OGT with poor efficiency (Figure 4.22). In addition, the amount of succinimide produced from isoD10-32 was at least an order of magnitude lower than that produced from D10-32 (Figure 4.22).

Though these data show that OGT can glycosylate aspartate, in addition to serine/threonine and glutamate, whether OGT glycosylates any cellular proteins on aspartate is currently unknown. In order to assess the feasibility of aspartate glycosylation, I compared the rates of glycosylation on serine, glutamate and aspartate under identical conditions. Using three nearly identical peptide substrates with different glycosyl acceptor residues (S10-35, D10-32, and E10-32), I measured the rate of starting material disappearance as a proxy for the rate of glycosylation. The serine peptide reacted approximately 5-fold faster than the aspartyl peptide and at least 100-fold faster than the glutamyl peptide (Figure 4.23). Based on structural data supported by measurements of reaction rates for an OGT mutant, OGT (D554N), we previously proposed that glycosylation on glutamate is disfavored by a repulsive interaction with aspartate 554 in OGT [107]. The nucleophile in D10-32 would not approach the D554 carboxylate as closely, which may explain its faster rate of glycosylation compared with E10. This notion is corroborated by previous findings that an E10D mutation enhanced the ability to co-immunoprecipitate an HCF substrate with OGT [16]. Given that glycosylation on aspartate is relatively fast for a model peptide substrate, the possibility exists that OGT may glycosylate some proteins on aspartate.

Figure 4.15: *(On the following page)* HCD MS/MS reveals position of ^{18}O incorporation into D10-32 starting material isomer. (a) 30 eV HCD spectrum of D10-32 S.M. isomer formed by treating D10-32 with wild-type OGT for 90 minutes in ^{18}O water. Major *b* and *y* ions are labeled. The peptide sequence of D10-32 is shown and annotated to highlight the differences observed for *b* and *y* ions when this species is formed in ^{16}O- vs. ^{18}O-water. (b) Selected zooms of *b* ions near the site of incorporation. A full list of ions observed by HCD is provided in Tables 4.1, 4.2, 4.3. Vertical axes are not shown; each ion was plotted with the same vertical scaling for both ^{16}O and ^{18}O samples. Horizontal axes are *m/z*.


a
no shift in b ions
+2 Da shift in b ions
Y V R V C S N P P C D T H Q T G T T N T A T T A T S N M A G Q H-NH2
no shift in y ions
Relative Abundance
m/z
b
{16O}b7 2+
{18O}b7 2+
{16O}b8 2+
{18O}b8 2+
{16O}b9 2+
{18O}b9 2+
{16O}b12 2+
{18O}b12 2+
{16O}b13 2+
{18O}b13 2+
{16O}b14 2+
{18O}b14 2+

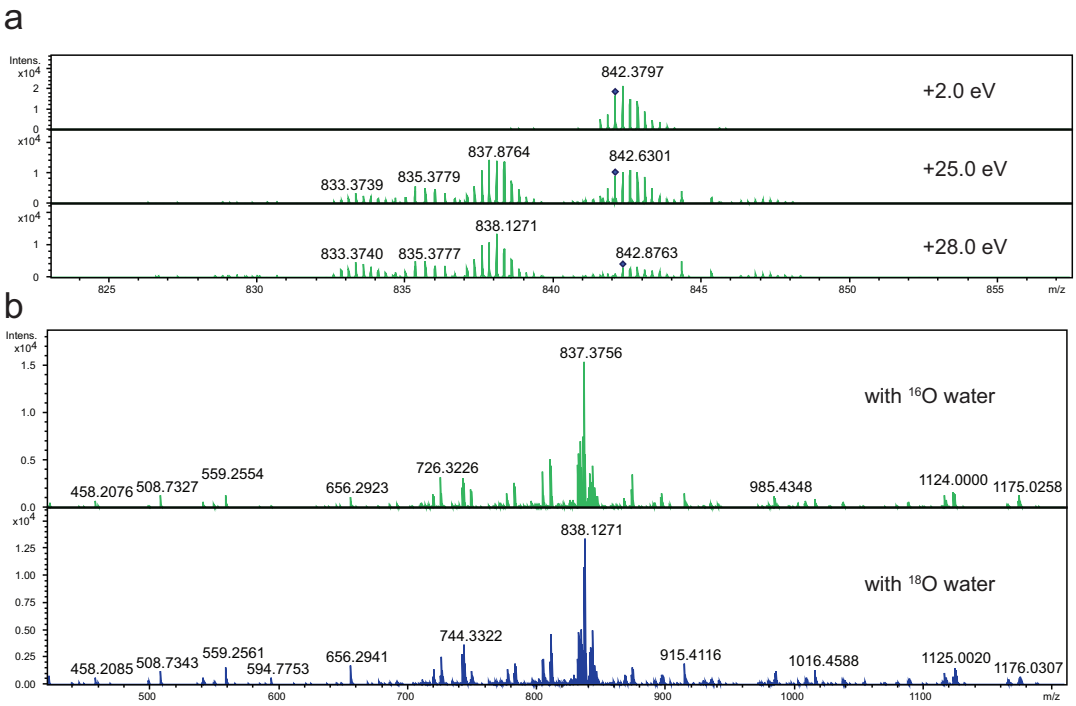


Figure 4.16: CID MS/MS reveals position of ^{18}O incorporation into D10-32 starting material isomer. (a) CID spectra or D10-32 S.M. isomer at 2, 25 and 28 eV show that greater than 25 eV is required for fragmentation of the major 4+ ion (marked with a dark blue diamond). (b) A comparison of D10-32 S.M. isomer formed by treating D10-32 with wild-type OGT for 90 minutes in ^{16}O-water (top) or ^{18}O-water (bottom). Horizontal axes are *m/z*, Vertical axes are normalized. A full list of ions observed by CID is provided in Tables 4.4, 4.5, and 4.6.

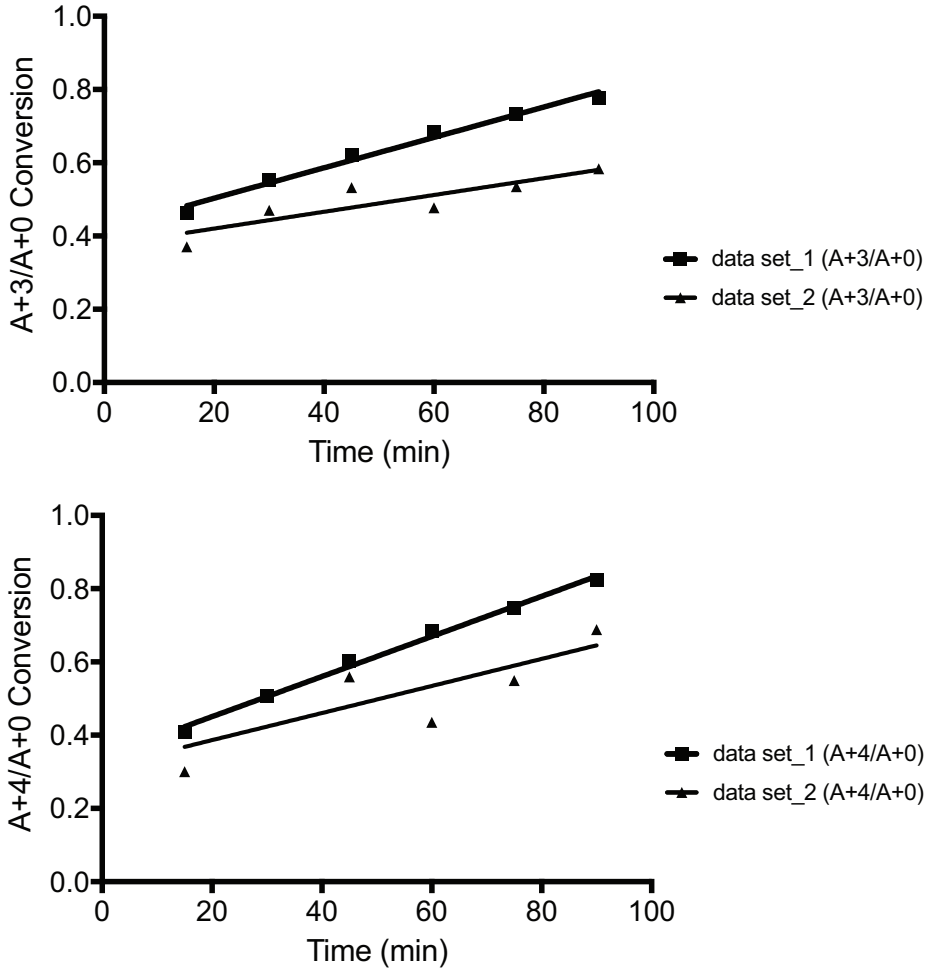


Figure 4.17: Rate of ^{18}O incorporation into D10-32. The rate of succinimide hydrolysis was determined by the rate of ^{18}O incorporation into the starting material and isomer peaks. The ratio of isotopomers at each time point relative to the expected ratio for one molar equivalent of ^{18}O incorporation (at 67 atom%) was used to obtain conversions. Time courses were conducted twice, on separate days, so data sets were plotted independently. The average slope between the four linear regressions affords a k_1 of 0.0039 min^{-1}, corresponding to a $t_{\frac{1}{2}} \approx 177$ min.

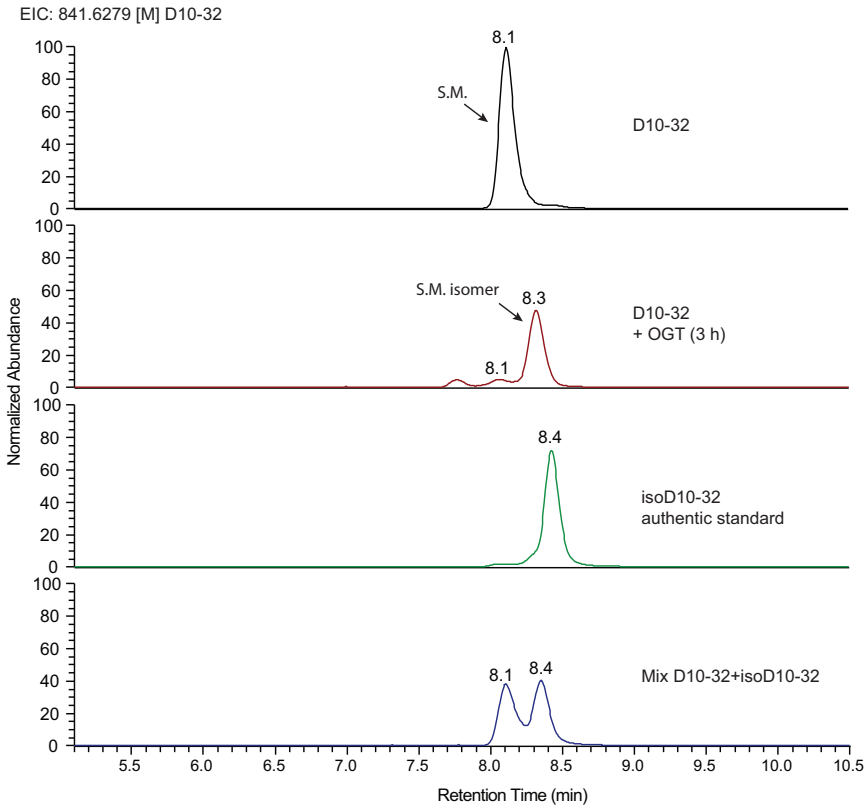


Figure 4.18: The S.M. isomer peak (D10-32) has the same LC retention time as an authentic isoD10-32 peptide. EIC traces corresponding to the D10-32 starting material show that an authentic isoD10-32 peptide has a different retention time than starting material (S.M.), but the same retention time as the S.M. isomer peak that forms upon reaction with OGT. Vertical axes are normalized to 1.15×10^7 counts.

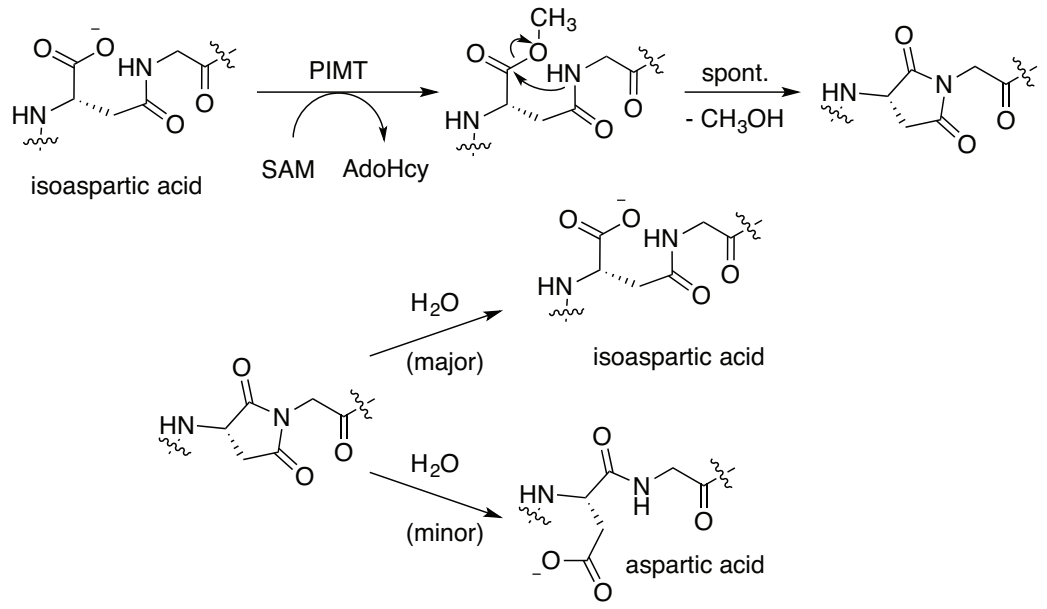


Figure 4.19: PIMT catalyzes isoaspartate methylation, producing a succinimide that hydrolyzes to a mixture of aspartate and isoaspartate. PIMT catalyzes SAM-dependent isoaspartate methylation, which leads to succinimide formation. Subsequent succinimide hydrolysis leads to a mixture of aspartate and isoaspartate. The example shown is for the dipeptide "DG".

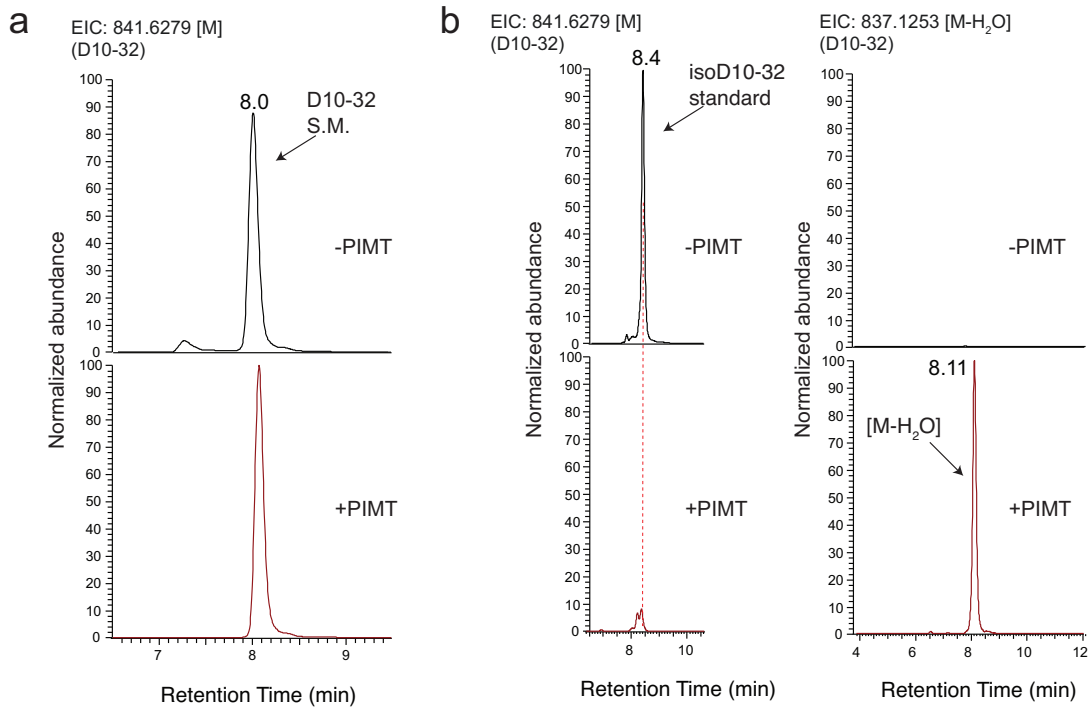


Figure 4.20: PIMT methylates isoaspartate, but not aspartate. (a) D10-32 is not methylated by PIMT; there is no significant anhydropeptide observed for both –PIMT and +PIMT samples. (b) isoD10-32 is methylated by PIMT; the consumption of the isoaspartyl peptide (left) is concomitant with the production of anhydropeptide (right). Vertical axes are normalized to 1.03×10^7 counts (isoD10-32 starting material) and 2.08×10^6 counts (anhydropeptide).

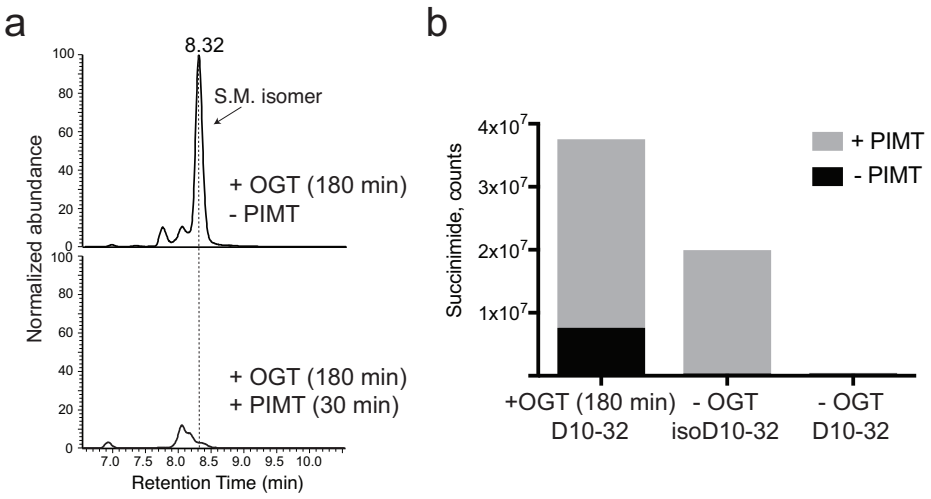


Figure 4.21: The starting material isomer is an isoaspartate-containing peptide. (a) EIC traces for 841.6279 (4+), corresponding to D10-32 starting material and isomer. Traces are after treatment with wild-type OGT alone (top) or with wild-type OGT followed by PIMT (bottom). (b) Normalized succinimide abundance for OGT and PIMT treated D10-32 (left), pure isoD10-32 (middle) or pure D10-32 (right); abundances were measured by integrating the respective EIC traces for succinimide, 837.1253 (4+). The retention time of the anhydropeptide species formed by PIMT and OGT are identical.

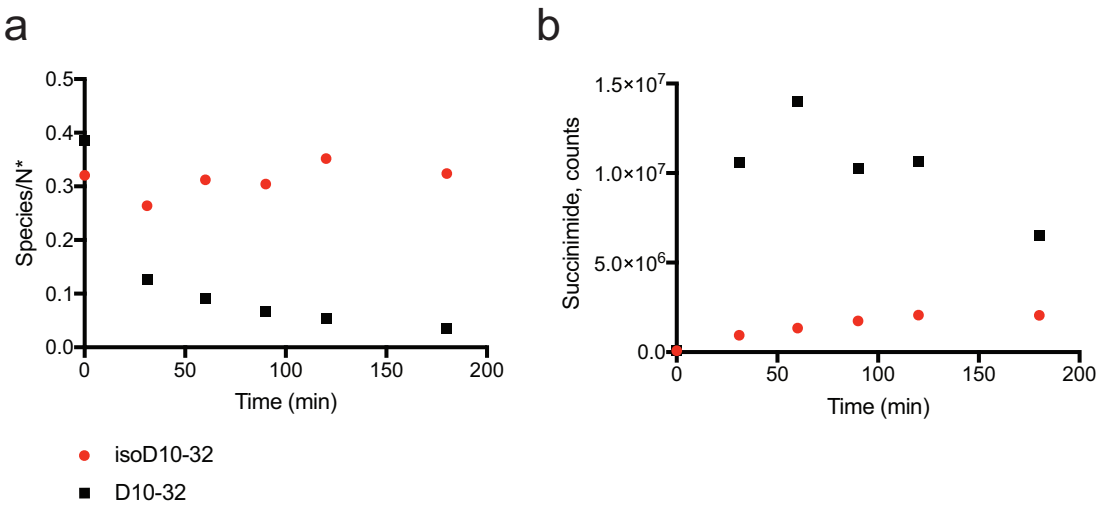
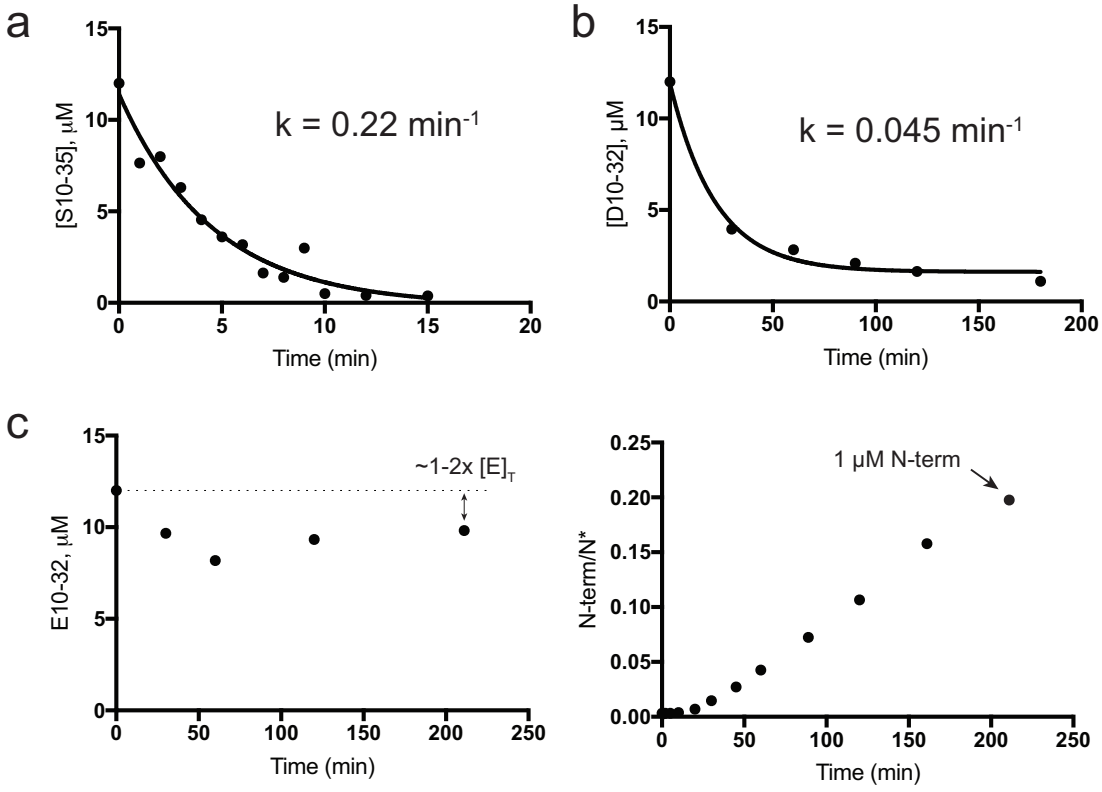


Figure 4.22: isoD10-32 is a poor substrate for OGT. (a) Plot of D10-32 or isoD10-32 abundance, relative to N^*, determined from integrated EIC traces [841.6279 (4+)] and [574.2668 (2+)], respectively. (b) Normalized succinimide abundance, as measured by integration of EIC traces for succinimide, [837.1253 (4+)].

Figure 4.23: *(On the following page)* Comparison of glycosyl acceptor abilities. (a) A plot of S10-35 starting material over time shows rapid consumption when serine is the acceptor. Data was fit with to a one-phase decay (see Equation 3.1). (b) A plot of D10-32 starting material over time shows consumption of an Asp-containing acceptor to be $\approx 5\times$ slower than for a serine-containing acceptor. Reactant concentrations were inferred from fractional abundances relative to an untreated sample, and scaled according to the initial reactant concentration (12 μM). Data was fit with to a one-phase decay (Equation 3.1). (c) A plot of E10-32 over time reveals slow consumption of starting material; after more than 200 minutes, the amount of starting material consumed was on the order of $1 - 2\times$ $[\mathrm{E}]_\mathrm{T}$ (left). Starting material was measured by relative abundance to known amount used in the reaction. The N-terminal cleavage fragment ("N-term") produced in the E10-32 cleavage reaction (right) was quantified by stable isotope dilution (SID) [151] using an isotopically labeled N-term internal standard (N^*, see Table B.1). A calibration curve was generated, and concentrations were determined by extrapolation (Figure A.17). After 200 minutes, the concentration of N-term was measured to be ≈ 1 μM, consistent with the amount of starting material that had been consumed.


a
[S10-35], μM
k = 0.22 min-1
Time (min)
b
[D10-32], μM
k = 0.045 min-1
Time (min)
c
E10-32, μM
~1-2x [E]T
Time (min)
N-term/N*
1 μM N-term
Time (min)

4.5 Conclusions

Until now isoaspartate formation was known only as a non-enzymatic process associated with aging proteins [69, 84, 201, 222]. Our findings show that OGT can catalyze aspartate-to-isoaspartate isomerization in a model substrate, and one might expect that other enzymes that glycosylate aspartate might also promote this process as long as the local sequence permits succinimide formation [38]. While we do not yet know if OGT modifies aspartate in any cellular proteins, we note that PARPs naturally glycosylate aspartic acid residues [46, 96, 169, 247, 275]; this activity may, therefore, lead to succinimide formation in some cases. If so, it is worth considering whether isoaspartate formation is always an unwanted side reaction or may serve other functions [8, 17, 40, 238].

Acknowledgments

I am grateful for support from the National Science and Engineering Research Council (NSERC) of Canada as a PGS-M and PGS-D3 fellowship recipient. This work was funding by a National Institutes of Health grant (R01 GM094263) to Suzanne Walker. I thank David J. Vocadlo (Simon Fraser University, Vancouver, BC, Canada) for providing UDP-5*S*GlcNAc. Sunia A. Trauger, Bogdan Budnik and John Neveu provided mass spectrometry assistance at various times during this work. Additional thanks to Michael B. Lazarus, Sara E. S. Martin and Zebulon G. Levine for their thoughtful discussions pertaining to this work and for critical reading of the *J. Am. Chem. Soc.* manuscript.

4.6 Methods

4.6.1 Reagents and General Methods

UDP-GlcNAc was obtained from Promega as a 100 mM solution in water. THP was obtained from Novagen as a 0.5 M solution in water. Calf alkaline phosphatase (CIP) was obtained from New England Biolabs. ^{18}O-water (97 atom%) was obtained from Cambridge isotope laboratories. PIMT and SAM were obtained from the Isoquant® isoaspartate detection kit (Promega). Peptides were obtained from Biomatik (Cambridge, ON, Canada); their purity was evaluated by HPLC and mass spectrometry, and they were used as received. Lyophilized peptides were dissolved in water to $\approx$ 5 mM and added to reactions according to the final concentration desired. Peptide concentrations were determined by measuring A_{274} (λ_{max}, tyrosine) in water, assuming an extinction coefficient of 1400 $M^{-1} \cdot cm^{-1}$. A complete list of peptides used in this study is provided in Table B.1. As the model peptide employed for these studies was derived from an HCF-1 cleavage substrate peptide, numbering begins from the first valine in the sequence (underlined in Table B.1) in order to remain consistent with the numbering convention used in the HCF-1 literature.

4.6.2 General LC-MS Methods

General measurements of peptides were performed on a q-Exactive Plus FT-mass spectrometer (Thermo Scientific™) equipped with an Ultimate 3000 uHPLC. Reaction mixture components were separated using a 33 minute reverse phase HPLC method (Table B.8); the last half of the LC method was empirically optimized to wash the column and re-equilibrate between runs to maintain column performance. with a Kinetex® C18 (2.6 μm particle size, 100 Å pore size, 2.1 mm ID, 150 mm length) at a constant flow rate of 0.200 mL/min and column temperature of 35 °C. 5 μL of sample was injected, unless otherwise specified.

Detection on the q-Exactive Plus mass spectrometer was performed between 300-2000 *m/z*, using an acquisition target of 3×10^6, maximum IT of 200 ms at a resolution of 70,000 for MS and 35,000 for MS/MS data. For MS/MS analysis, dynamic exclusion was turned off and a global inclusion list was used to maximize S/N for the targeted analysis. A collision energy of 30 eV and an isolation window of 6 *m/z* was used for higher-energy collision-induced dissociation (HCD).

Additional measurements of peptides were collected on a Bruker Impact HD q-ToF mass spectrometer coupled to an Agilent 1290 uHPLC. Sample components were separated using a 45 minute reverse phase HPLC method (Table B.9) with a Waters XBridge Peptide BEH C18 column (5 μm particle size, 300 Å pore size, 1.0 mm ID, 100 mm length) at a constant flow rate of 0.100 mL/min with ambient column temperature. 10 μL of sample was injected, unless otherwise specified. A collision energy of 28 eV and an isolation window of 6 *m/z* was used for collision-induced dissociation (CID). The source and ToF parameters were optimized for the transmission of higher molecular weight ions. A counter current drying gas flow rate of 12 L/min, and a nebulizing gas pressure of 45 bar were used. Funnel 1 was set to 150 V, and funnel 2 was set to 36 V, with a hexapole RF of 300 V_{p-p}. A calibration segment was present at the beginning and end of the run for *m/z* calibration with sodium formate cluster using the Bruker HPC algorithm for the curve fitting.

Unless otherwise stated, extracted ion chromatograms displayed are for the indicated *m/z* $\pm$5 ppm tolerance.

4.6.3 Purification of OGT

Wild-type OGT, OGT (K842M), OGT (D554N) and hOGT4.5 were expressed and purified as outlined in section 2.11.5.

4.6.4 *In vitro* peptide isomerization experiments

Unless specified otherwise, reactions were conducted under the conditions outlined below, where the substrate:enzyme ratio is 10:1.

Isomerization reactions were conducted as follows: 20 μL reaction volume with components having the following final concentrations: 1 mM UDP-GlcNAc, 1 mM THP, 12 μM peptide substrate, 5 μM N* (internal standard; Table B.1), 0.1 U CIP and 1.2 μM OGT; the reaction buffer contained 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT was added last and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After the desired reaction time, reactions were quenched by addition of four volumes (80 μL) of ice-cold methanol supplemented with 0.1% formic acid. Reactions were then centrifuged at 16,100×g for 10 minutes to remove any precipitated protein (4 °C), and transferred to fresh tubes. Samples were concentrated to dryness using a speedvac (T = 30 °C; ≈45–60 minutes). Samples were re-suspended in 100 μL 0.1% formic acid, mixed by vortexing, and centrifuged for 10 minutes at 16,100×g (4 °C) to remove any insoluble material. Samples were then transferred to LC vials, and maintained at 4 °C until they were analyzed.

4.6.5 *In vitro* single turnover (STO) isomerization reactions

Isomerization reactions were conducted as follows: 20 μL reaction volume with components having the following final concentrations: 1 mM UDP-GlcNAc, 1 mM THP, 12 μM peptide substrate, 5 μM N* (internal standard; Table B.1), 0.1 U CIP and 20 μM OGT; the reaction buffer contained 20 mM $MgCl_2$, 20 mM Tris and 150 mM NaCl at pH 7.4. OGT was added last and reactions were mixed by pipetting then incubated at 37 °C in a thermomixer to ensure uniform heating. After the desired reaction time, reactions were

quenched by addition of four volumes (80 μL) of ice-cold methanol supplemented with 0.1% formic acid. Reactions were then centrifuged at 16,100×g for 10 minutes to remove any precipitated protein (4 °C), and transferred to fresh tubes. Samples were concentrated to dryness using a speedvac (T = 30 °C; ≈45–60 minutes). Samples were re-suspended in 100 μL 0.1% formic acid, mixed by vortexing, and centrifuged for 10 minutes at 16,100×g (4 °C) to remove any insoluble material. Samples were then transferred to LC vials, and maintained at 4 °C until they were analyzed.

4.6.6 ^{18}O incorporation experiments

^{18}O experiments were conducted under single turnover conditions, as described in section 4.6.5. Each experimental condition (^{18}O-water & wild-type OGT; ^{16}O-water & wild-type OGT; ^{18}O-water & OGT (K842M) were conducted twice. Stock reagents were diluted into ^{16}O or ^{18}O water, to give ≈ 100 atom% and 67 atom% abundance, respectively.

4.6.7 Isotope modeling

The experimental data was fit according to a least-squares minimization of residuals to determine the proportion of zero, one or two ^{18}O incorporations that best represented the observed population. Analysis was performed using MatLab using a custom written script. The best fit corresponded to 0.64% zero, 96.9% one, and 2.46% two incorporations.

Peak area for individual isotopomers was obtained by EIC integration. Isotopomer abundances were then converted to ratios relative to the monoisotopic peak (A+0). Using Bruker Data Analysis software, the relative ratios of isotopomers expected for the incorporation of zero, one or two ^{18}O atoms (determined at 67 atom% abundance) were calculated. The Web-based server MS-Isotope (UCSF) can also be used to compute these ratios. A

charge envelope corresponding to the quadruply charged species for the D10-32 peptide was modeled at the observed resolution: 38,506 (q-Exactive). ^{18}O incorporation rates (reaction conversion) were measured indirectly as the change in ratio of $\frac{A+n}{A+0}$ over time relative to the ratio expected for the addition of one molar equivalent of ^{18}O into the peptide. In order to ensure that the incorporation of a single molar equivalent of 67 atom% abundance ^{18}O was consistent with experimental data, the observed isotopomer ratios for experimental data (90 minute treatment with wild-type OGT and ^{18}O water) were fit to the expected 67 atom% incorporation model, and residuals were determined.

4.6.8 *In vitro* cleavage/glycosylation experiments

Cleavage (E10-32) and glycosylation (S10-35) reactions were conducted under the same conditions outlined isomerization reactions (see section 4.6.4) wherein the substrate:enzyme ratio is 10:1.

4.6.9 *In vitro* PIMT reactions

PIMT assays were conducted by adding 0.2 μL of SAM (1 mM stock) and 10 μL of PIMT (used as obtained from Promega) directly to isomerization reactions (conducted as described in section 4.6.4). Reactions were incubating for 30 minutes at 30 °C with a thermomixer. As in the case of OGT reactions, reactions were quenched with the addition of four volumes (120 μL) ice-cold methanol supplemented with 0.1% formic acid. Reactions were then centrifuged at 16,100×g for 10 minutes to remove any precipitated protein (4 °C), and transferred to fresh tubes. Samples were concentrated to dryness using a speedvac (T = 30 °C; ≈45–60 minutes). Samples were re-suspended in 100 μL 0.1% formic acid, mixed by vortexing, and centrifuged for 10 minutes at 16,100×g (4 °C) to remove any insoluble

material. Samples were then transferred to LC vials, and maintained at 4 °C until they were analyzed.

4.6.10 MS/MS of reaction-derived isoD10-32 for mapping the site of ^{18}O incorporation

Samples for MS/MS were prepared as described in section 4.6.6. MS^2 spectra examined were for the peak corresponding to isoD10-32 (S.M. isomer) formed after a 90 minute reaction of D10-32 with wild-type OGT in the presence of either ^{16}O or ^{18}O water. Spectra were measured on both q-Exactive (HCD) and q-ToF (CID) mass spectrometers. With both instruments, we isolated the quadruply-charged species 841.6279 *m/z*. With the q-ToF, ion isolation was verified using a 2 eV MS^2 energy. Appreciable parent ion remained at 25 eV collision energy; however, the ion was almost entirely fragmented at 28 eV (See Figure 4.16). There was no incorporation of ^{18}O post-reaction, as determined by comparing samples resuspended in ^{16}O vs. ^{18}O water supplemented with 0.1% formic acid prior to MS-analysis. Data were manually inspected for predicted fragment ions (as determined using the UCSF MS-Product server) within a 5 ppm mass accuracy window.

Chapter 5

Conclusion

> "Isn't it funny how day by day nothing changes,
> but when you look back, everything is different?"
>
> – C.S. Lewis (Prince Caspian)

The works presented in Chapters 2, 3 and 4 illuminate details of the function of OGT *beyond* Ser/Thr glycosylation. Chapter 2 showed us that HCF-1, a transcriptional co-regulator of human cell-cycle progression, is subject to cleavage *directly* by OGT. Further, OGT effects this cleavage using the chemistry employed for Ser/Thr glycosylation; that is, the same active site, and the same co-substrate, UDP-GlcNAc - a common cellular metabolite, which links OGT to cellular metabolism - creating the potential to link cell-cycle regulation with cellular metabolic status. While tempting, this remains to be shown definitively. Furthermore, little is known about local (e.g. within sub-cellular compartments) concentration dynamics for UDP-GlcNAc within a cell, with measurements to date being limited to whole-cells [22]. Notably, the concentration of UDP-GlcNAc required for HCF-1 cleavage *in vitro* is higher than that for Ser/Thr glycosylation, a feature that may be the result of an uncoupling between UDP-GlcNAc consumption and amide bond scission events. Even if cellular UDP-GlcNAc levels vary over a narrow range of concentrations, such an uncoupling would amplify these differences.

From our mechanistic studies into HCF-1 cleavage, we wonder whether there is a biological significance to the pyroglutamate product formed in the cleavage process. Mutation of HCF-1 to replace the proteolytic repeats recognized by OGT with an ectopic protease recognition site (Taspase1) failed to rescue the phenotype of blocking HCF-1 proteolysis [27]. Pyroglutamate residues are found at the N-terminus of many antibodies and secreted proteins/peptides [25, 156, 200, 219]; pyroglutamate formation can occur spontaneously, and while it more commonly occurs at glutamine residues, it can also occur for glutamate [235]. An enzyme, glutaminyl cyclase, has been shown to accelerate this process *in vivo* [156, 212, 219]. Whether the pyroglutamate residue formed during cleavage persists *in vivo*, is unknown, and given the misassignment of the cleavage site initially suggests that at least a portion of HCF-1 isolated from cells no longer bears the pyroglutamate residue. Efforts to prepare an antibody that selectively recognizes the N-terminal pyroglutamate of an HCF cleavage product have so far been unsuccessful. However, obtaining HCF-1 from cells using general HCF-1 antibodies, followed by MS analysis may reveal the relative abundances of HCF-1 polypeptides with and without this modification; cellular HCF-1 has already been examined by MS [184], but whether peptides were observed where the pyroglutamate had been lost has not been described.

With regard to the slow rate for HCF-1 cleavage, we hypothesized that amide bond hydrolysis occurs after dissociation from the enzyme, and suggested that this was the rate-determining step of cleavage. However, another possibility is that dissociation from OGT is the rate-determining step in this process. This notion may be supported by the fact that in Chapter 4, when we compared the relative rates of starting material consumption for cleavage, serine glycosylation and aspartate isomerization, at least *in vitro*, the glutamate-containing acceptor (E10-32) was consumed far slower than S10-35 or D10-32. Curiously, it

appeared as if wild-type OGT was not able to undergo multiple turnovers of HCF-1 cleavage, even when the substrate was supplied in excess. This could be the result of strong product inhibition, where the C-terminal cleavage product may remain bound in the TPRs of OGT and prevent the binding of the next substrate molecule, or an extremely inefficient coupling between glutamate GlcNAcylation and UDP-GlcNAc turnover that results in complete consumption of the co-substrate. Whatever the reason for this, it raises questions about whether OGT effects HCF-1 cleavage in a catalytic or stoichiometric fashion inside cells. It also raises questions about the number of repeats within the full-length HCF protein: is the origin of multiple repeats the result of increasing the probability that cleavage will occur? Are multiple cleavage events needed for HCF-1 to become functional? Or are multiple repeats required to prevent product inhibition?

Having directly observed and isolated a glutamate-linked glycopeptide we cemented our mechanistic picture for the steps involved in HCF-1 cleavage [107]. We were prompted from these studies to examine whether aspartate was a suitable acceptor [109], yet the fact that it reacted was less a surprise than that upon reaction, and forming a glycosyl aspartate, this led to isomerization of the acceptor residue to isoaspartate. In hindsight, the behavior of these glycosyl-glutamate and glycosyl-aspartate species should not have been a surprise at all. Amino acid side chain activation has long been established: numerous examples exist wherein glutamic acids have been shown to be susceptible to activation by peptide coupling reagents, and this activation leads to the formation of pyroglutamate residues or succinimides [12, 11, 19, 35, 60, 143, 185, 190, 229, 263]. In the case of internal pyroglutatamtes, this activation has been shown to predispose the peptide bond to spontaneous hydrolysis at that site [12, 11, 35, 60, 143, 185, 229]. Similar hydrolytic instability has been seen for *N*-acylurea-containing peptides, generated by oxidation of asparagine residues [226]. Fur-

thermore, when a glycosidase is trapped with a reactive sugar fluoride, it, too, became susceptible to backbone amide bond cleavage, presumably through the intermediacy of the same species described in Chapter 3 [129].

One of the real surprises, given this work, is that at present there are no other examples of the use of UDP-GlcNAc as an electrophilic activator in biology. UDP-GlcNAc is one of the most abundant metabolites in cells, and other abundant metabolites like ATP [65], NAD^+ [61], and SAM [65] have all been implicated as co-substrates for electrophilic activation. Does this imply that other glycosyltransferases, or enzymes otherwise annotated as glycosyltransferases, are able to modify glutamic acid or aspartic acid for the purpose of effecting the downstream chemistry that these species are inherently prone to? Is OGT able to glycosylate any other protein substrates on glutamic acid, and as a result promote amide bond cleavage in those substrates? Amide bond cleavage, as a result of glutamate glycosylation, would produce an N-terminal pyroglutamate on the C-terminal cleavage peptide. Unfortunately, at present, a robust method for pyroglutamate enrichment has not been worked out. However, if it were possible to enrich for pyroglutamate residues, one could examine the differences in pyroglutamate-containing peptides under conditions where OGT activity was perturbed (endogenous vs. chemically inhibited vs. OGT overexpression). With regard to aspartate glycosylation, proteomic studies into isoaspartate and succinimide formation have previously been described [2, 127, 242, 277]. Treating protein samples with hydrazine derivatives will label many cellular electrophiles [173], however, using an aldehyde-derivatized resin, only carbonyl-based electrophiles would be retained (due to their ability to form a hydrazone) [2]. One can envision examining cellular succinimide levels under various conditions where OGT activity has been modulated, either chemically or genetically. It may be necessary to employ a PIMT+/– background [201] to observe dif-

ferences in cellular succinimide when OGT activity is perturbed. Alternatively, detection of isoaspartate may be better suited if the expected time-scale for OGT-mediated changes is long. Treating cell lysates first with PIMT (to generate succinimides) followed by hydrazinolysis should allow for enrichment of isoaspartate-containing peptides, indirectly. While it remains to be seen whether OGT is responsible for aspartate glycosylation *in vivo*, efforts to examine this are underway.

The chemical reactivity of glutamate and aspartate linked glycosyl esters is also interesting in the context of protein ADP-ribosylation. Many of the identified ADP-ribosylation sites have been mapped to Asp and Glu residues [46, 96, 169]; it remains to be examined whether any of these proteins are susceptible to backbone cleavage or isoaspartate formation. While the local peptide backbone conformations are likely to play a role in dictating whether a succinimide can form [38], this may simply mean that these chemistries are limited to relatively unstructured regions or regions prone to partial unfolding [113].

Beyond its ability to modify Ser/Thr/Glu/Asp, as described here, it has been reported that Cys is also able to serve as a suitable glycosyl acceptor for OGT [174]. It remains to be seen whether this of modification is of physiological relevance, these studies may suggest a revision of our picture of OGT and the diversity of substrates it may modify. There are other examples of glycosyltransferases which are able to tolerate multiple glycosyl acceptors [157]; however, only PARPs or ADP-ribosyl transferases demonstrate the substrate breadth now seen with OGT [46, 96, 148, 169]. Furthermore, these studies raise many questions as to relative importance of each of these activities for OGT, particularly since the elements of OGT's function that are required for the survival of mammalian cells are still unknown [149]. This is undoubtedly an exciting time to study OGT, and the answers to many of these questions are likely to offer a wealth of new biological insights.

Appendix A

Supplementary Figures

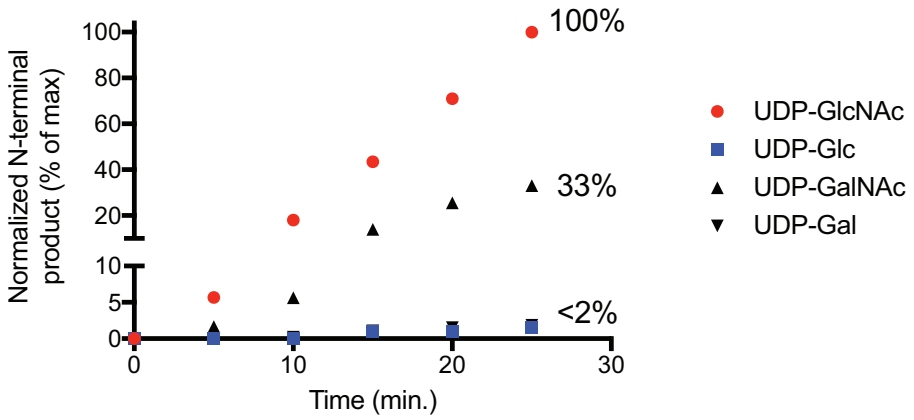


Figure A.1: Role of the nucleotide sugar in E10-32 cleavage. A series of time courses for E10-32 cleavage by wild-type OGT in the presence of different UDP-sugar donors. The N-terminal product ("N-term", m/z 569.2626; 2+) was quantified by integration of the EIC, and normalized by determining the ratio relative to the internal standard, N^*. Peptide sequences are provided in Table B.1. These experiments were conducted with 12 μM E10-32 and 18 μM wild-type OGT.

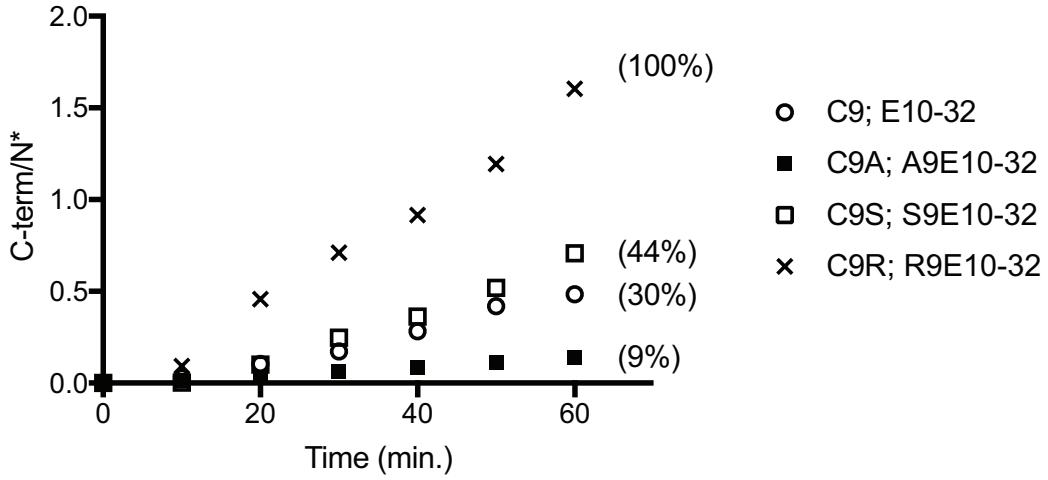


Figure A.2: Relative rates of cleavage for C9X peptides. A series of time courses for X9E10-32 peptide cleavage by OGT (D554N). As all peptides produce the same C-terminal product ("C-term"), relative reaction rates were determined by monitoring the amount of C-term produced over time. EIC traces for "C-term" (m/z 747.6698; 3+) were integrated and normalized by determining the ratio relative to the internal standard, N^*. Peptide sequences are provided in Table B.1. These experiments were conducted with 12 μM E10-32 and 20 μM OGT (D554N).

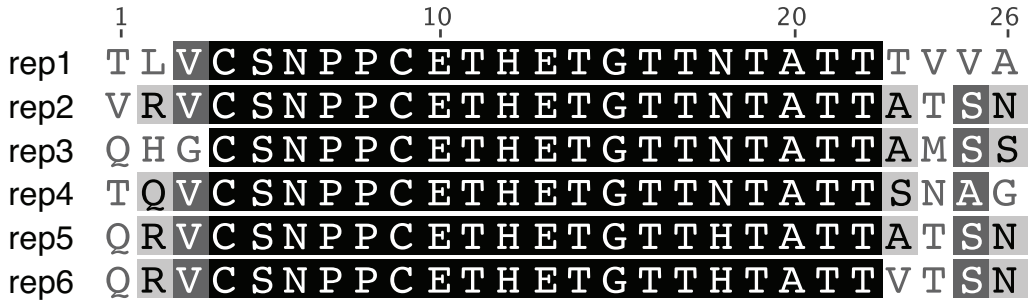


Figure A.3: Primary sequence of HCF proteolytic repeats in *Homo sapiens*. The canonical 26 amino acid sequences for the six repeats are shown [27].

Figure A.4: *(On the following page)* Time course for E10-32 cleavage by wild-type OGT. Shown are: A (*m/z* 845.1319; 4+), B (*m/z* 747.6698; 3+), and C (*m/z* 569.2626; 2+) (labeled as shown in Figure 3.1), as well as two glycopeptide species, labeled BP (byproduct) and Int-1 (intermediate 1) (*m/z* 895.9017; 4+), and an anhydropeptide [M-H_2O], labeled Int-2 (intermediate 2) (*m/z* 840.8784; 4+). There are two peaks for Int-2, indicated by red arrows. These peaks both have the same exact mass and the same fragmentation pattern (see Figure 3.12 for MS/MS spectra). Plots represent a sum of EIC traces for all species. These experiments were conducted with 12 μM E10-32 and 12 μM wild-type OGT. Times indicated are the times between OGT addition and quenching. The y-axes are scaled such that the largest peak in each chromatogram is assigned as 100% abundance. The use of "Relative Abundance (%)" as a vertical axis is not without limitations; each species will not have an identical response signal, so peak heights are not directly comparable. However, this issue also makes a fixed vertical axis undesirable. Graphical representations must be considered with plotted data presented following displayed EICs. Plots of integrated EICs for each species are presented in Figure 3.5.

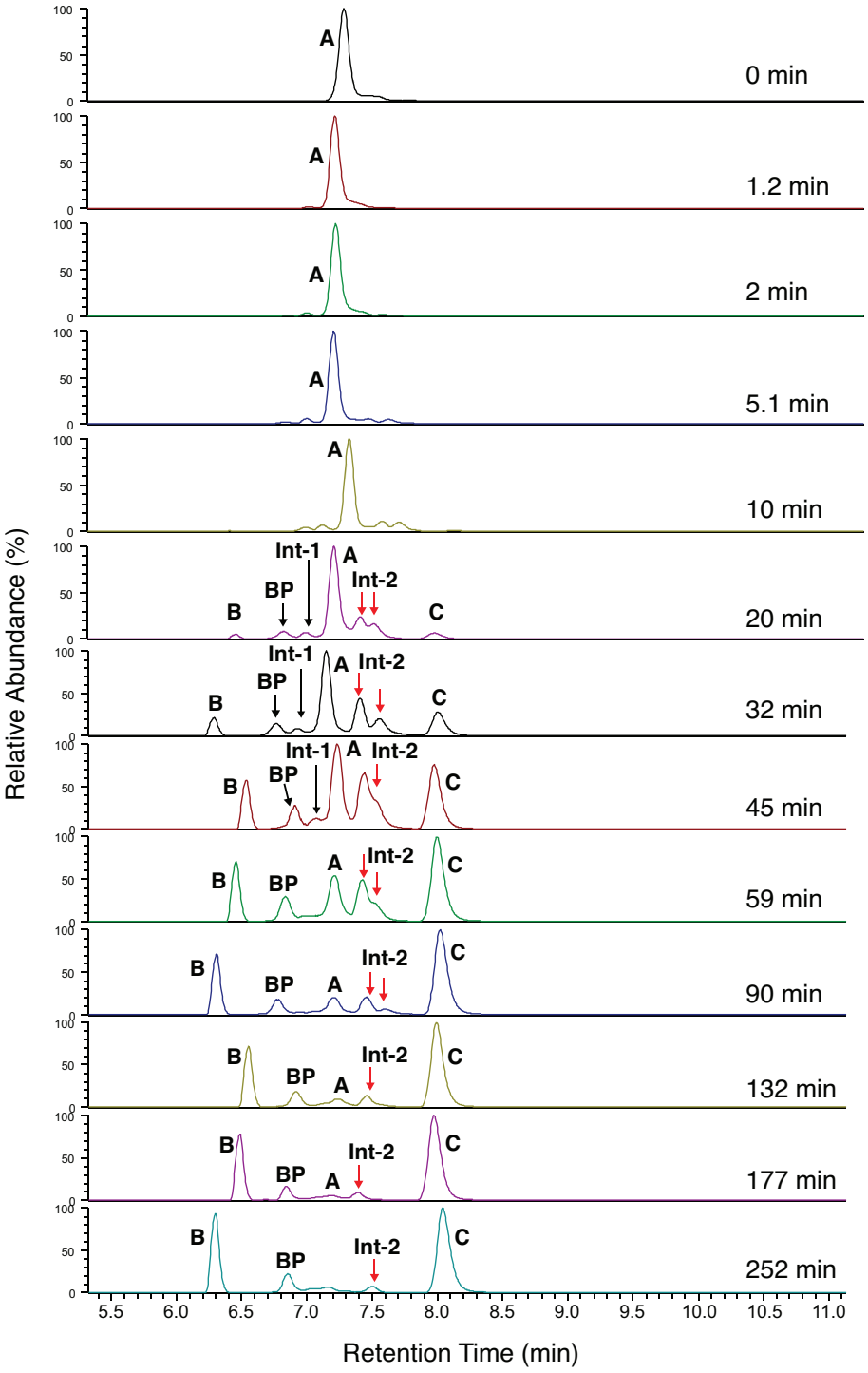
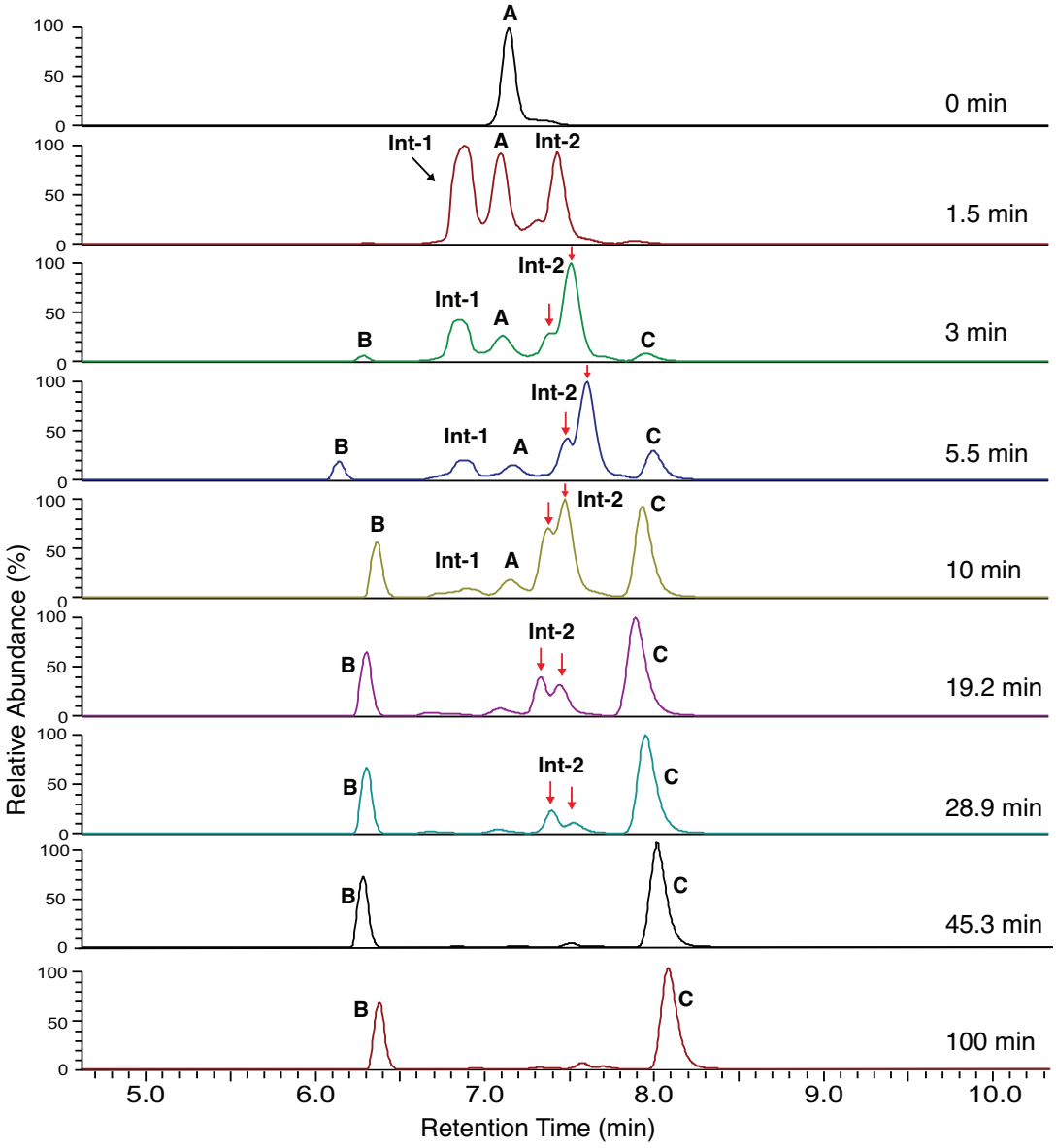

Relative Abundance (%)
Retention Time (min)
0 min
1.2 min
2 min
5.1 min
10 min
20 min
32 min
45 min
59 min
90 min
132 min
177 min
252 min
A
B
BP
C
Int-1
Int-2

Figure A.5: *(On the following page)* Time course for E10-32 cleavage by OGT (D554N), 100 minutes. Shown are: A (*m/z* 845.1319; 4+), B (*m/z* 747.6698; 3+), and C (*m/z* 569.2626; 2+) (labeled as shown in Figure 3.1), as well as a glycopeptide species, Int-1 (intermediate 1) (*m/z* 895.9017; 4+), and an anhydropeptide [M-H_2O], labeled Int-2 (intermediate 2) (*m/z* 840.8784; 4+). There are two peaks for Int-2, indicated by red arrows. These peaks both have the same exact mass and the same fragmentation pattern (see Figure 3.12 for MS/MS spectra). Plots represent a sum of EIC traces for all species. These experiments were conducted with 12 μM E10-32 and 12 μM OGT (D554N). Times indicated are the times between OGT addition and quenching. The y-axes are scaled such that the largest peak in each chromatogram is assigned as 100% abundance. The use of "Relative Abundance (%)" as a vertical axis is not without limitations; each species will not have an identical response signal, so peak heights are not directly comparable. However, this issue also makes a fixed vertical axis undesirable. Graphical representations must be considered with plotted data presented following displayed EICs. Plots of integrated EICs for each species are presented in Figure A.6.


A
0 min
Int-1
A
Int-2
1.5 min
Int-2
Int-1
A
B
C
3 min
Int-2
B
Int-1
A
C
5.5 min
B
Int-2
C
Int-1
A
10 min
B
Int-2
C
Int-1
A
19.2 min
Int-2
B
C
28.9 min
Int-2
B
C
45.3 min
B
C
100 min
B
C
100
50
0
Relative Abundance (%)
5.0
6.0
7.0
8.0
9.0
10.0
Retention Time (min)

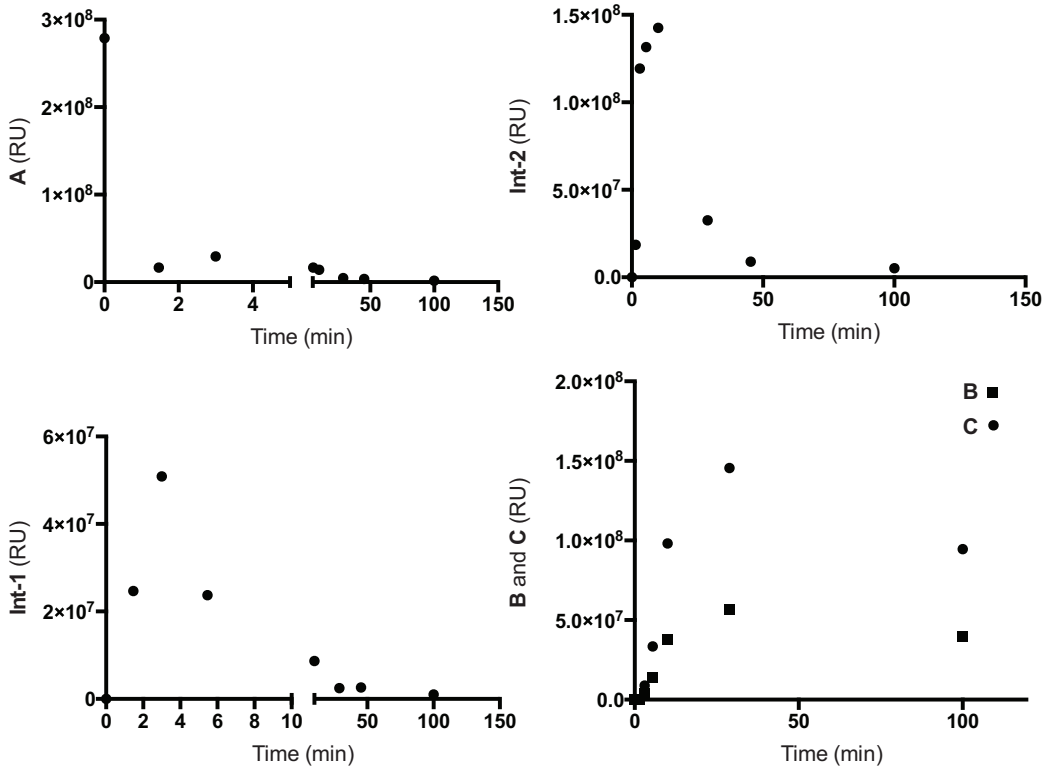


Figure A.6: Graphical representations of the 100 minute time course for cleavage of E10-32 with stoichiometric OGT (D554N). Integration of individual EICs (as shown in Figure A.5) provided quantitation of response units for each species at each time point. Response units are shown for species A (m/z 845.1319; 4+), Intermediate 1 (Int-1, m/z 895.9017; 4+), Intermediate 2 (Int-2, m/z 840.8784; 4+) and products, B (m/z 747.6698; 3+) and C (m/z 569.2626; 2+). The 19 minute time point was excluded from all plots due to lower abundance of each species; the 45 minute time point was excluded in the product plot due to abnormally high signal. All values used are corrected based on the signal of an EIC from a blank run. Integration for Int-2 included both peaks in the EIC (see caption of Figure A.5 for details). No decay fit for A is shown here due to sparse data at early time points.

Figure A.7: *(On the following page)* Time course for E10-32 cleavage by OGT (D554N), 3 minutes. Shown are: A (m/z 845.1319; 4+), B (m/z 747.6698; 3+), and C (m/z 569.2626; 2+), Int-1 (intermediate 1) (m/z 895.9017; 4+), and Int-2 (intermediate 2) (m/z 840.8784; 4+). There is little or no signal for B and C in this time frame. There are two peaks for Int-2, indicated by red arrows. These peaks both have the same exact mass and the same fragmentation pattern (see Figure 3.12 for MS/MS spectra). Plots represent a sum of EIC traces for all species. These experiments were conducted with 3 μM E10-32 and 27 μM OGT (D554N). Times indicated are the times between OGT addition and quenching. The y-axes are scaled such that the largest peak in each chromatogram is assigned as 100% abundance. The use of "Relative Abundance (%)" as a vertical axis is not without limitations; each species will not have an identical response signal, so peak heights are not directly comparable. However, this issue also makes a fixed vertical axis undesirable. Graphical representations must be considered with plotted data presented following displayed EICs. Plots of integrated EICs for each species are presented in Figure A.8.

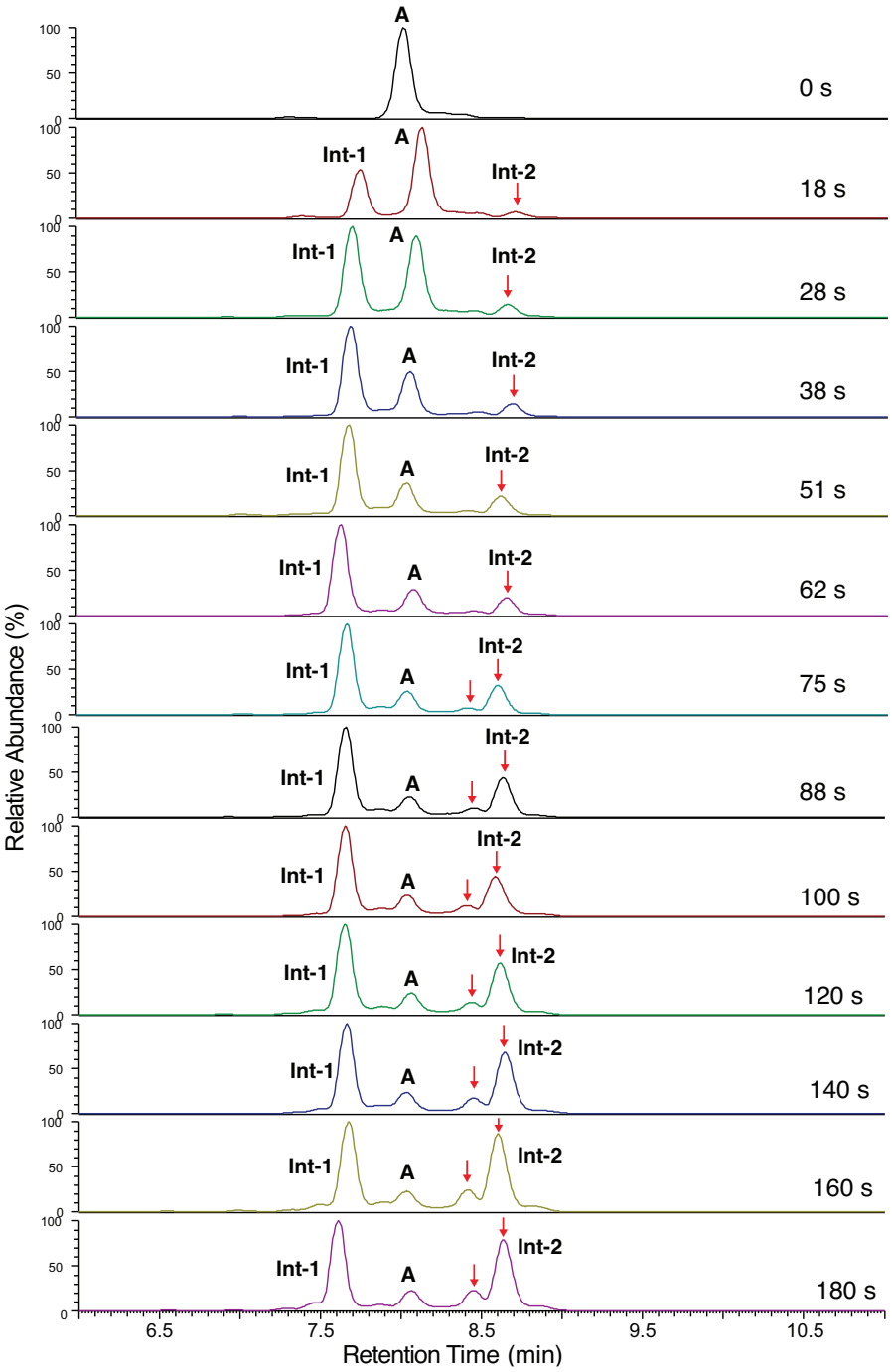

A
Int-1
Int-2
0 s
18 s
28 s
38 s
51 s
62 s
75 s
88 s
100 s
120 s
140 s
160 s
180 s
Relative Abundance (%)
Retention Time (min)
6.5
7.5
8.5
9.5
10.5

Figure A.8: *(On the following page)* Graphical representations of the 3 minute time course for cleavage of E10-32 with excess OGT (D554N). Integration of individual EICs (as shown in Figure A.7) provided quantitation of response units for each species at each time point. For species A (m/z 845.1319; 4+), a first-order exponential decay (of the form shown in Equation 3.1) was used to fit the data. The obtained rate constant is shown. Intermediate 1 (Int-1, m/z 895.9017; 4+) forms rapidly, and towards the end begins to saturate before declining (as seen in longer time courses). Intermediate 2 (Int-2, m/z 840.8784; 4+) forms at the expense of Intermediate 1, and appears to persist at the final time point. The values quantified include both peaks for Int-2, as they both decompose. The products, B (m/z 747.6698; 3+) and C (m/z 569.2626; 2+) form over time; however, in this short time course, the lag-phase between reaction initiation and initial product formation is evident. It appears that over the first minute only Intermediate 1 and 2 have been produced, with little decomposition of Intermediate 2. All values used are corrected based on the signal of an EIC from a blank run.

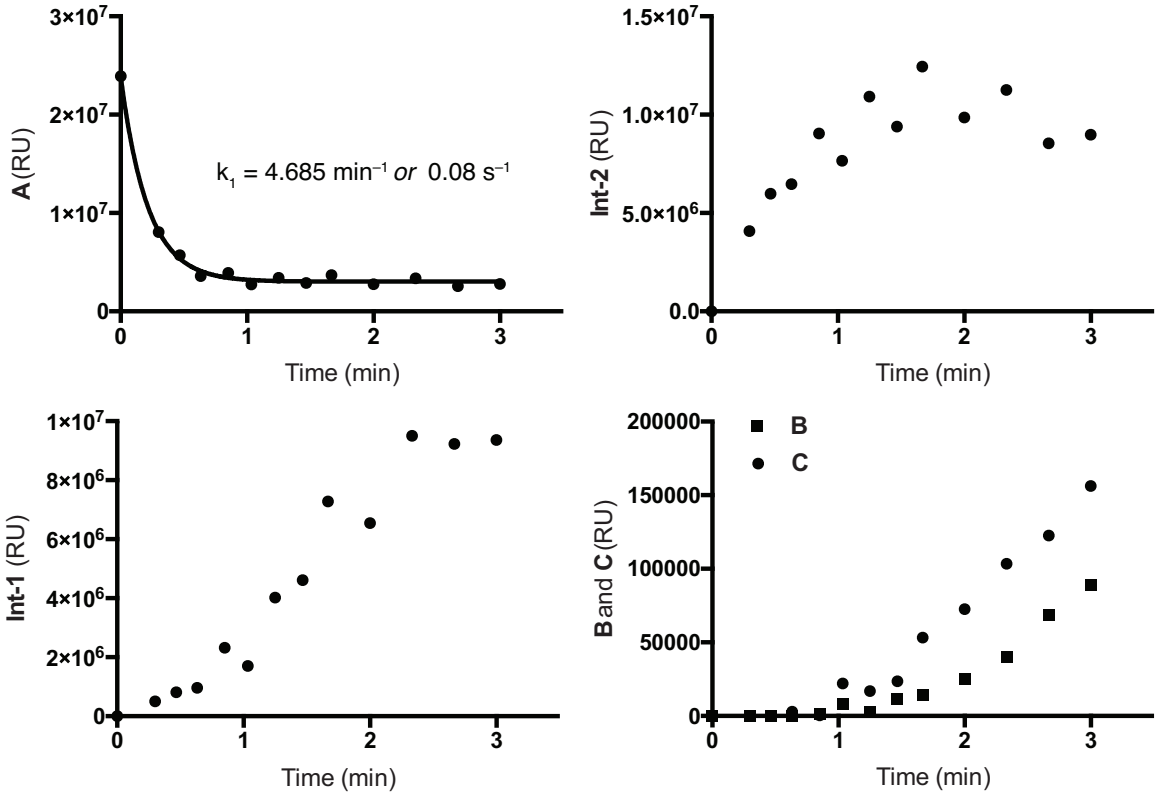
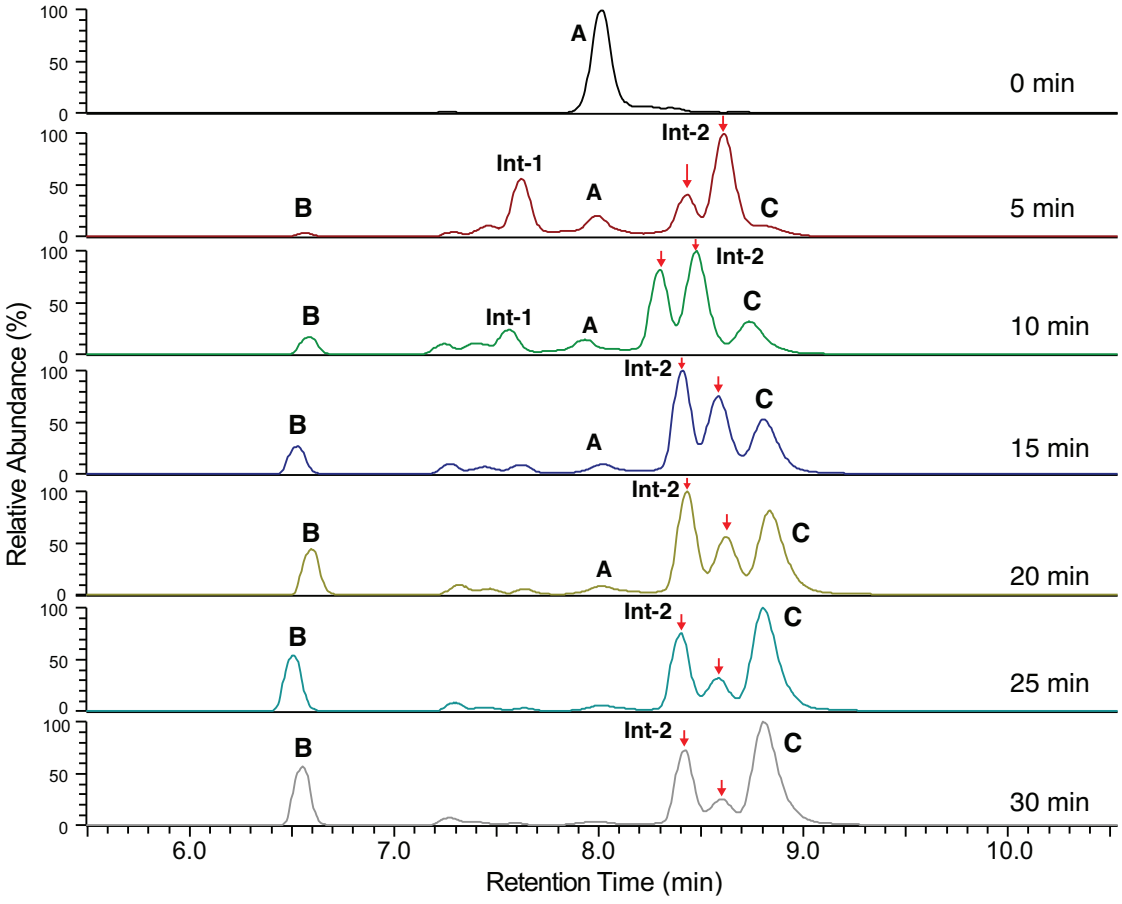

A (RU)
3×10^7
2×10^7
1×10^7
0
k1 = 4.685 min−1 or 0.08 s−1
0
1
2
3
Time (min)
Int-2 (RU)
1.5×10^7
1.0×10^7
5.0×10^6
0.0
Time (min)
Int-1 (RU)
1×10^7
8×10^6
6×10^6
4×10^6
2×10^6
Time (min)
B and C (RU)
200000
150000
100000
50000
B
C
Time (min)

Figure A.9: *(On the following page)* Time course for E10-32 cleavage by OGT (D554N), 30 minutes. LC traces showing EICs for each of: A (*m/z* 845.1319; 4+), B (*m/z* 747.6698; 3+), and C (*m/z* 569.2626; 2+), Int-1 (intermediate 1) (*m/z* 895.9017; 4+), and Int-2 (intermediate 2) (*m/z* 840.8784; 4+). There are two peaks for Int-2, indicated by red arrows. These peaks both have the same exact mass and the same fragmentation pattern (see Figure 3.12 for MS/MS spectra). Plots represent a sum of EIC traces for all species. These experiments were conducted with 3 μM E10-32 and 27 μM OGT (D554N). Times indicated are the times between OGT addition and quenching. The y-axes are scaled such that the largest peak in each chromatogram is assigned as 100% abundance. The use of "Relative Abundance (%)" as a vertical axis is not without limitations; each species will not have an identical response signal, so peak heights are not directly comparable. However, this issue also makes a fixed vertical axis undesirable. Graphical representations must be considered with plotted data presented following displayed EICs. Plots of integrated EICs for each species are presented in Figure A.10.


Relative Abundance (%)
100
50
0
A
0 min
Int-1
Int-2
B
A
C
5 min
10 min
15 min
20 min
25 min
30 min
6.0
7.0
8.0
9.0
10.0
Retention Time (min)

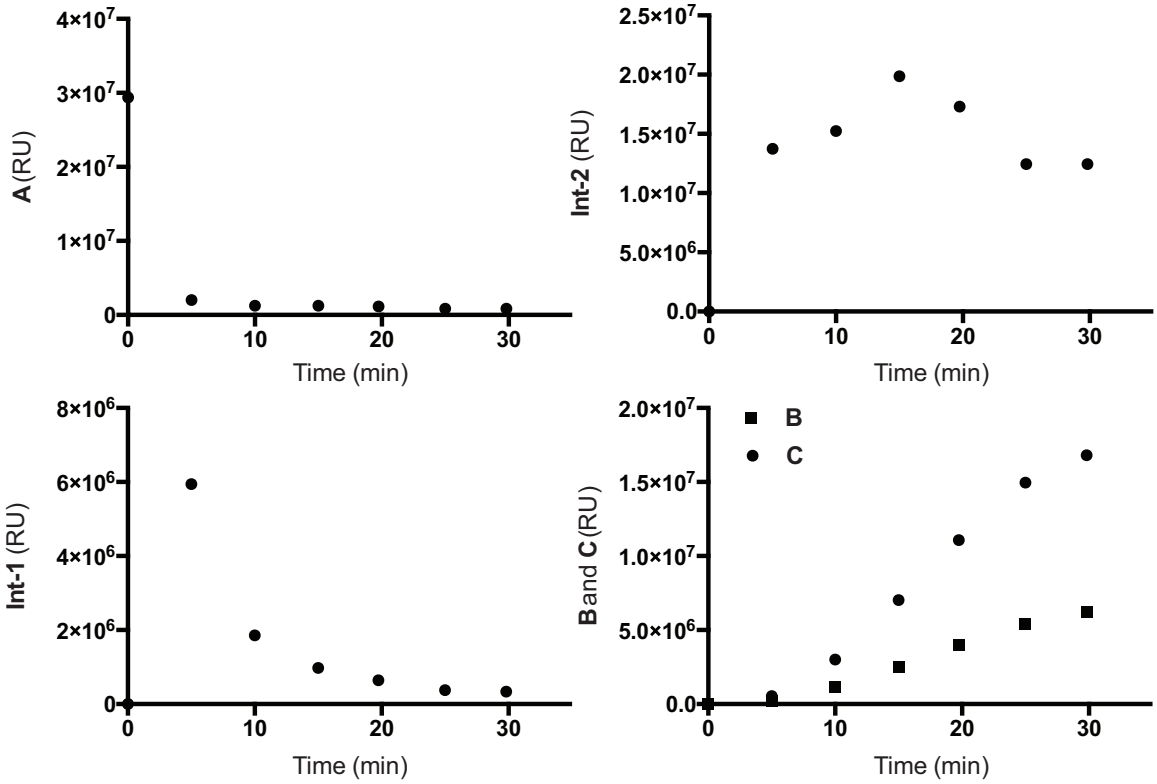


Figure A.10: Graphical representations of the 30 minute time course for cleavage of E10-32 with excess OGT (D554N). Integration of individual EICs (as shown in Figure A.9) provided quantitation of response units for each species at each time point. Time points are widely separated so the changes in response units over time are more informative for some species than others (e.g., the accumulation of products), although the temporal information is clear and consistent with data previously presented. The masses of species are the same as in previous captions. Note the lag between reaction initiation and initial product formation. The values quantified include both peaks for Int-2, as they both decompose. All values used are corrected based on the signal of an EIC for a blank run.

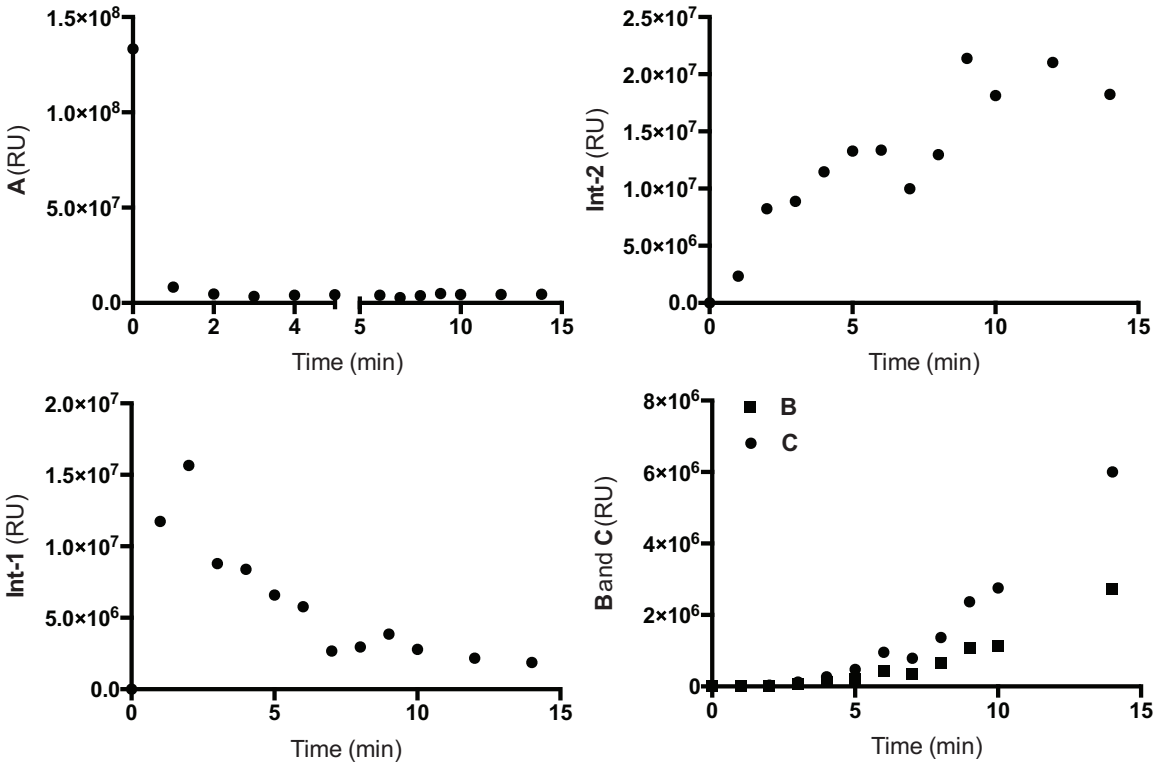


Figure A.11: Graphical representations of the 14 minute time course for cleavage of E10-32 with excess OGT (D554N). Data is based on an intermediate length time course analogous to those shown in Figures A.7 and A.9. Integration of individual EICs provided quantitation of response units for each species at each time. The masses of species are the same as in previous captions. Noteworthy is the decomposition of Int-1, which appears to occur from about 2 minutes to about 7 minutes. Int-2 forms steadily, but appears to reach a maximum between 10 and 15 minutes, consistent with data shown in Figure A.10. The integrated values include both peaks for Int-2. All values used are corrected based on the signal of an EIC for a blank run.

Figure A.12: *(On the following page)* Intermediate 2 is a dehydrated form of E10-32 (A). (a) A comparison of the fragmentation pattern (HCD) for A (top) and Int-2 (bottom) shows many of the same fragment ions, however, the *m/z* range from 600-950 shows key differences. A complete list of fragment ions observed by MS/MS for Int-2 and A are presented in Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, and Tables 3.7, 3.8, 3.9, respectively (with *m/z* errors for fragment ions presented in Figure A.16). A more completely annotated MS/MS spectrum for Int-2 is presented in Figure 3.13. (b) In order to filter out dehydration artifacts resulting from numerous Ser/Thr residues in the C-terminal region of the HCF-short peptide, a side-by-side comparison between A and Int-2 was performed. A zoom of the 600-950 *m/z* range reveals *b* ions are dehydrated in Int-2, but not in A. Based on this comparison, we can localize the dehydration in Int-2 to the $C^{9}E^{10}$ region of the peptide (HCF repeat numbering).

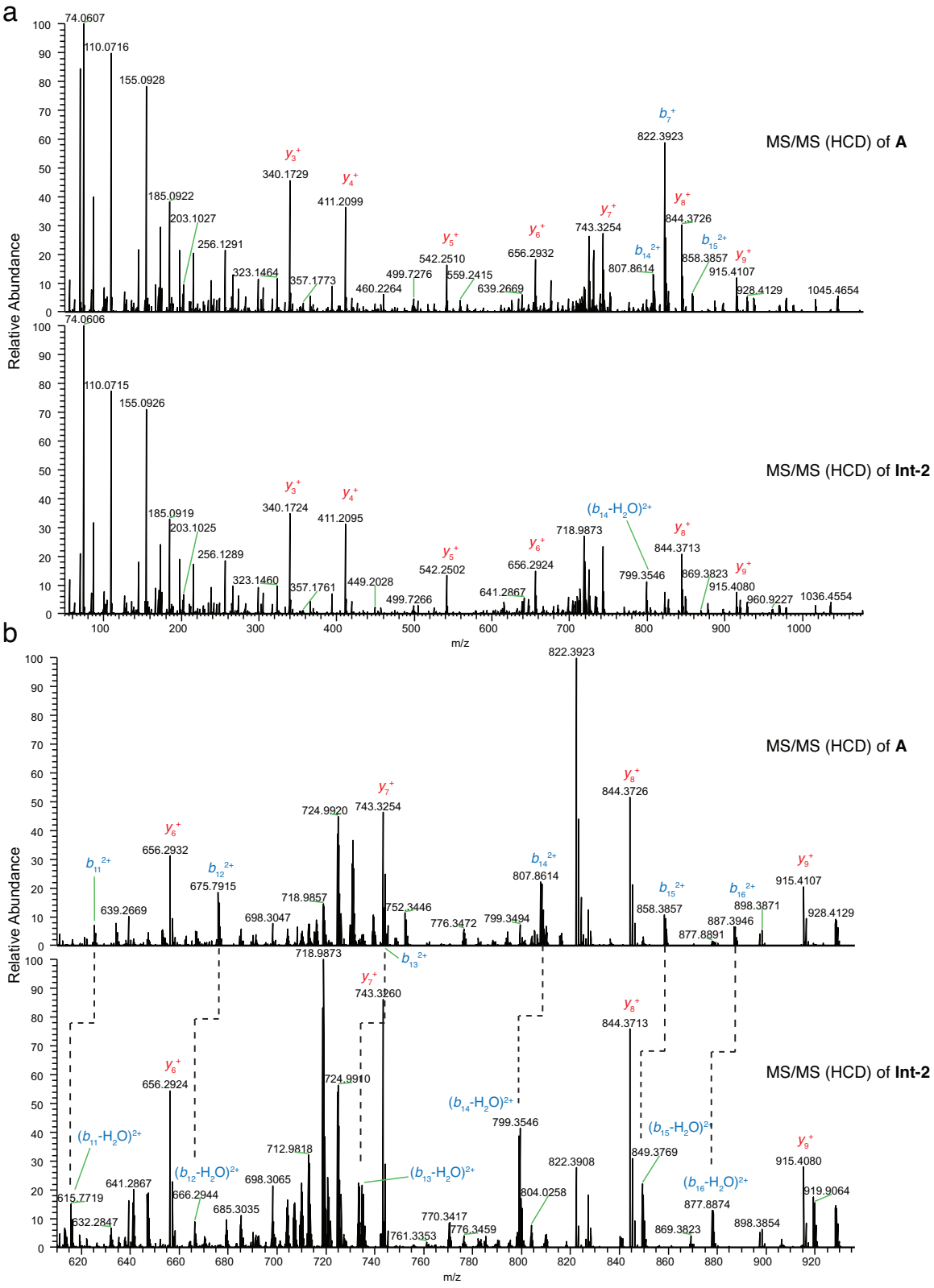

a
MS/MS (HCD) of A
MS/MS (HCD) of Int-2
Relative Abundance
m/z
b
MS/MS (HCD) of A
MS/MS (HCD) of Int-2
Relative Abundance
m/z

Figure A.13: *(On the following page)* Different HCF peptides produce the same intermediates with both wild-type OGT and OGT (D554N). LC traces showing retention times for glycopeptide and anhydropeptide intermediates formed in reactions of three different HCF peptides using either wild-type OGT or OGT (D554N). Traces were selected from time courses such that both anhydropeptides were of similar intensity. For E10-32 and E10V11-32 intermediates the vertical axis is normalized to a fixed intensity for each species; for R9E10-32 intermediates the vertical axis is normalized such that the tallest peak is 100%, this is due to the low relative abundance of wild-type-produced intermediates, making them poorly visualized when at the same scale. The LC trace for the E10V11-32 glycopeptide is not shown because the abundance of glycopeptide intermediate was extremely low and a large quantity of glycopeptide byproduct was observed, making the trace uninformative. (To be more specific: due to these problems, the glycopeptide intermediate for this peptide sequence could only be assigned after it was identified in reactions with the mutant enzyme, whereas all other intermediates identified in the reactions with wild-type OGT were identified without reliance on results for the mutant enzyme.)

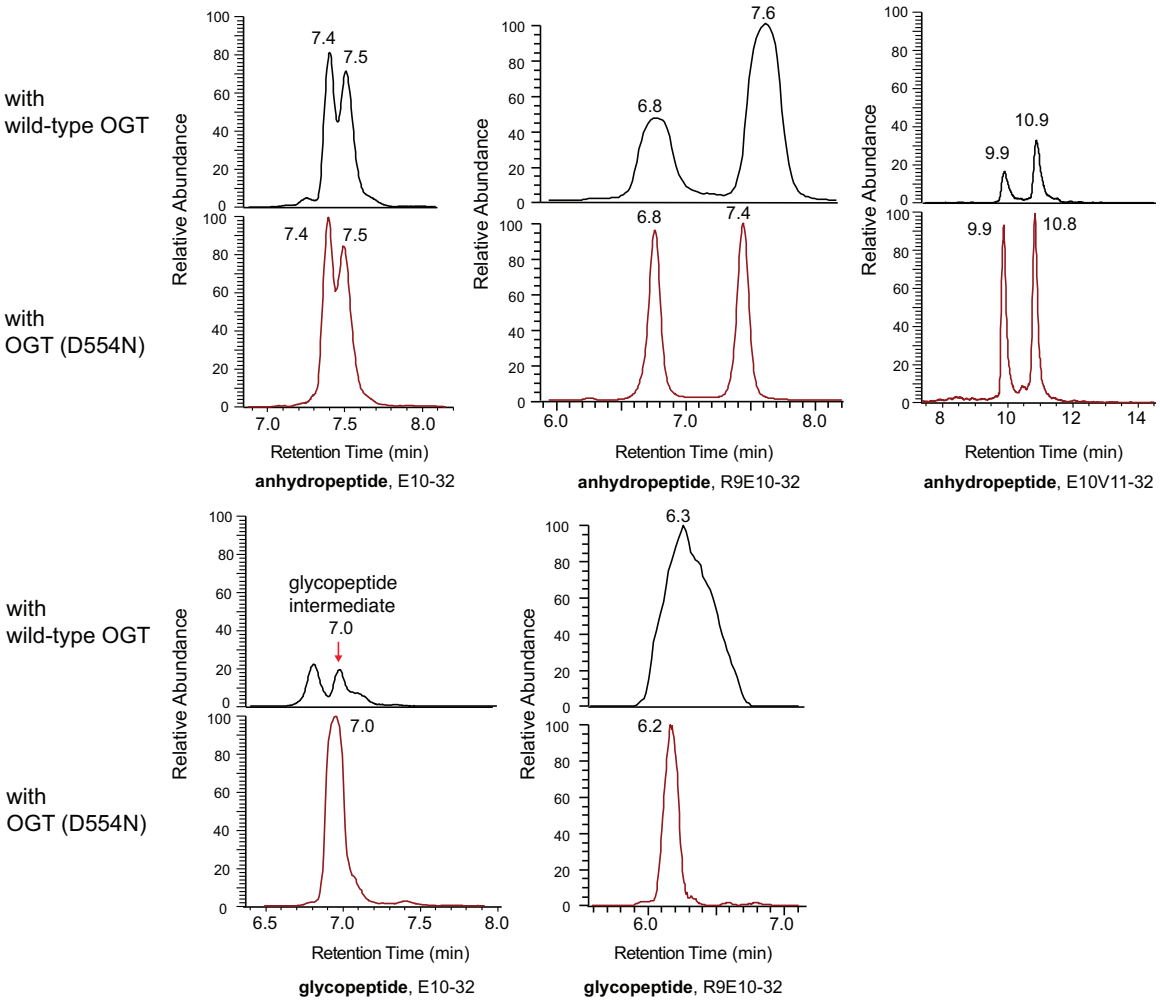
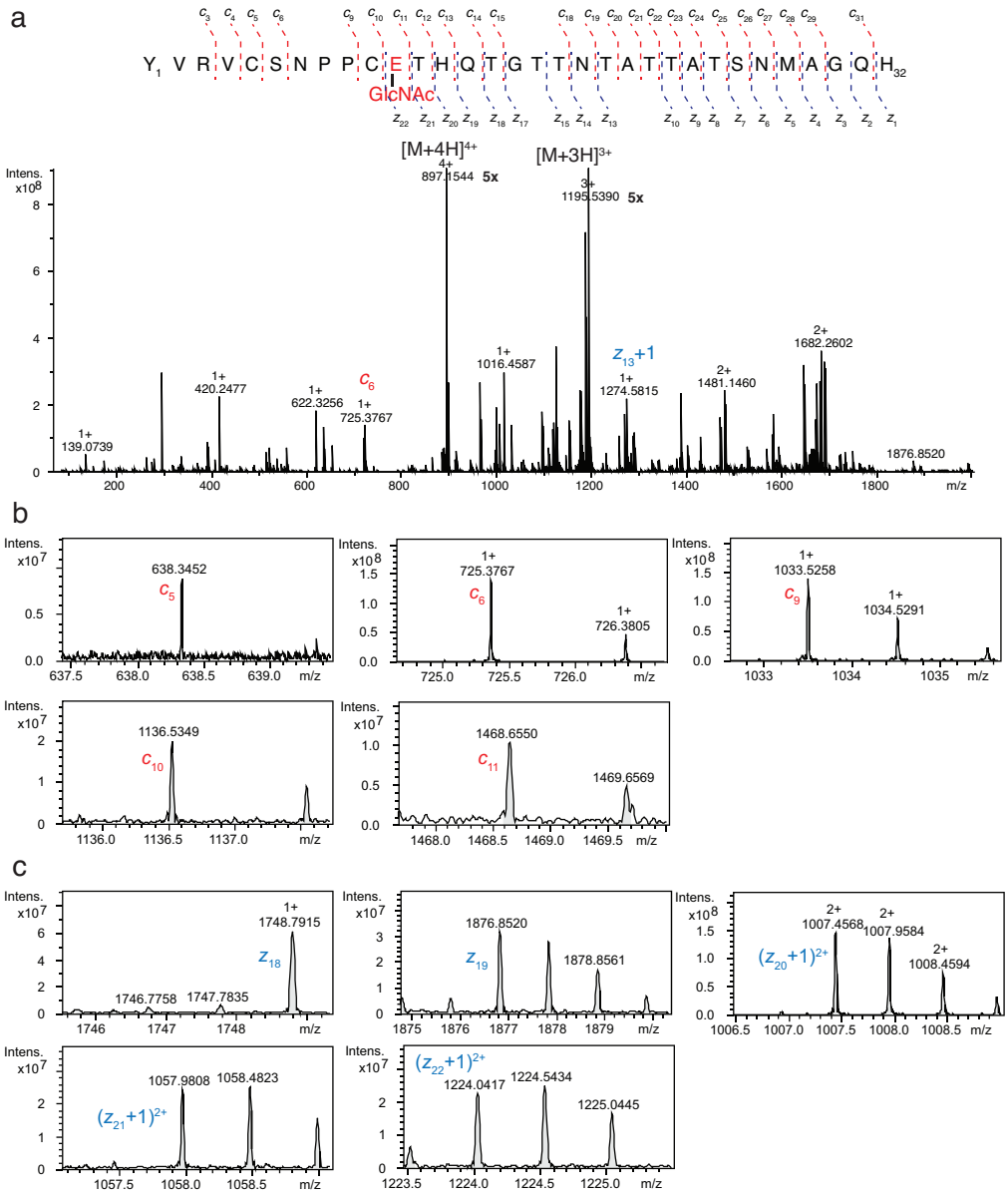

with
wild-type OGT
with
OGT (D554N)
Relative Abundance
7.4
7.5
Retention Time (min)
anhydropeptide, E10-32
6.8
7.6
7.4
anhydropeptide, R9E10-32
9.9
10.9
10.8
anhydropeptide, E10V11-32
glycopeptide
intermediate
7.0
glycopeptide, E10-32
6.3
6.2
glycopeptide, R9E10-32

Figure A.14: *(On the following page)* ECD MS/MS spectrum of Intermediate 1, with zooms. (a) Full ECD spectrum of Intermediate 1 m/z 895.9017 (4+). The inset sequence shows the observed c and z ions graphically. Fragmentation N-terminal to proline is not observed by ECD (as would be the case for the c_7 and c_8 fragments). Further, due to the low efficiency of ECD fragmentation, most c and z ions are present at abundances too low to be meaningfully annotated on this spectrum. (b) Zoomed views of c ions near the glycosylated glutamate residue. (c) Zoomed views of z ions near the glycosylated glutamate residue. A full list of observed theoretical ions for Int-1 is supplied in Tables 3.10, 3.11, 3.12, 3.13. Figures A.15 and 3.17 show additional zooms of ions in the region surrounding the GlcNAc-modified residue.


a
Y1 V R V C S N P P C E T H Q T G T T N T A T T A T S N M A G Q H32
GlcNAc
[M+4H]4+
4+
897.1544 5x
[M+3H]3+
3+
1195.5390 5x
Intens.
x10^8
1+
139.0739
1+
420.2477
1+
622.3256
c6
1+
725.3767
1+
1016.4587
z13+1
1+
1274.5815
2+
1481.1460
2+
1682.2602
1876.8520
m/z
b
Intens.
x10^7
638.3452
c5
637.5 638.0 638.5 639.0 m/z
Intens.
x10^8
1+
725.3767
c6
1+
726.3805
725.0 725.5 726.0 m/z
Intens.
x10^8
1+
1033.5258
c9
1+
1034.5291
1033 1034 1035 m/z
Intens.
x10^7
1136.5349
c10
1136.0 1136.5 1137.0 m/z
Intens.
x10^7
1468.6550
c11
1469.6569
1468.0 1468.5 1469.0 1469.5 m/z
c
Intens.
x10^7
1+
1748.7915
z18
1746.7758
1747.7835
1746 1747 1748 m/z
Intens.
x10^7
1876.8520
z19
1878.8561
1875 1876 1877 1878 1879 m/z
Intens.
x10^8
2+
1007.4568
2+
1007.9584
(z20+1)2+
2+
1008.4594
1006.5 1007.0 1007.5 1008.0 1008.5 m/z
Intens.
x10^7
1057.9808
1058.4823
(z21+1)2+
1057.5 1058.0 1058.5 m/z
Intens.
x10^7
(z22+1)2+
1224.0417
1224.5434
1225.0445
1223.5 1224.0 1224.5 1225.0 m/z

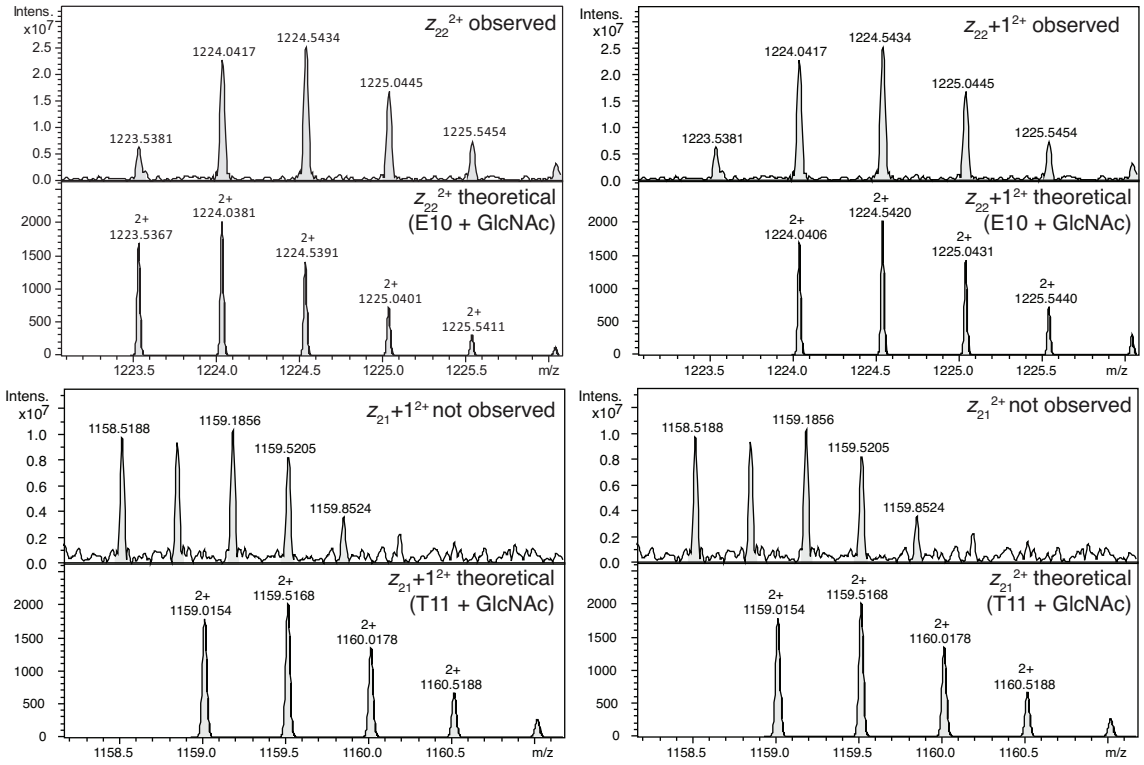


Figure A.15: The GlcNAc moiety in Intermediate 1 is located at E11, not T12. Zoom of z ions 21 and 22 from the ECD spectrum of Intermediate 1. Fragments corresponding to the ${z_{22}}^{2+}$ and $(z_{22}{+}1)^{2+}$ ions for a peptide in which a GlcNAc is modeled at residue E11 are observed; the analogous ions are not observed for a peptide in which the GlcNAc is modeled at T11.

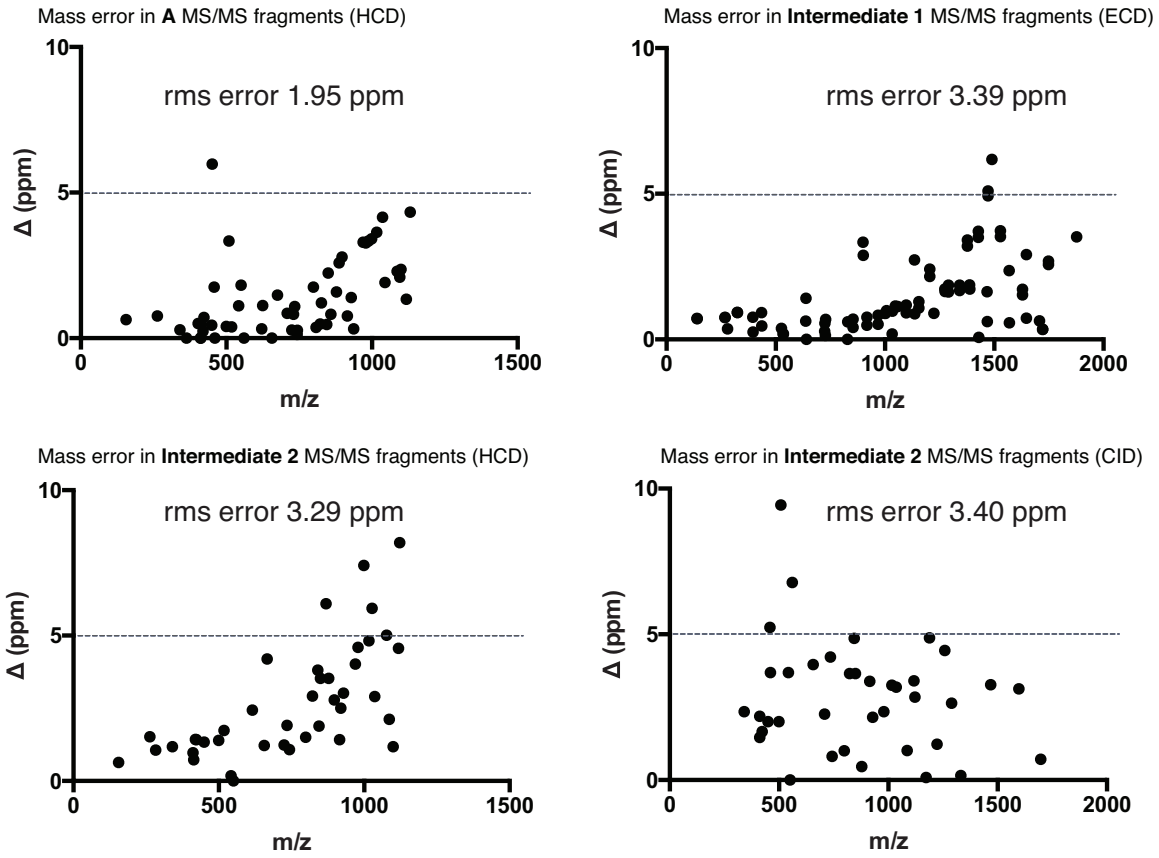
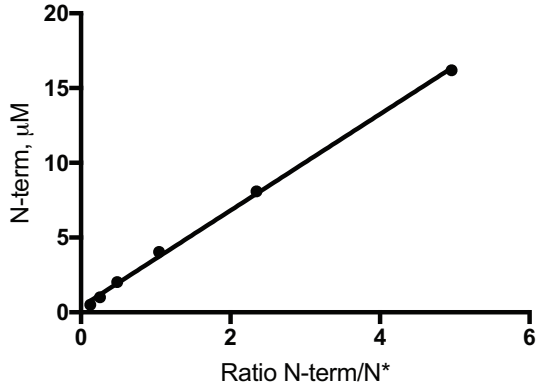


Figure A.16: Errors in MS/MS fragment ions for A, Intermediate 1 and Intermediate 2. A graphical representation of the mass errors (in ppm) between expected and observed fragment ions for each species evaluated by MS/MS in this study. Tabulated data for of all ions observed, for each species, are presented in Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12, and 3.13. The rms error value inset in each graph is the mean for all observed theoretical ions for that species. Rms error was calculated as: $x_{rms} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \ldots + x_n^2)}$. The maximum error tolerated for any ion reported was 10 ppm, however, the vast majority of ions were within 5 ppm of the theoretical m/z.



	[N-term], μM
Best-fit values ± SE	
Slope	3.222 ± 0.06101

	[N-term], μM
Goodness of Fit	
R square	0.9986
Sy.x	0.254

Figure A.17: Calibration curve for SID quantitation of N-term. N-term and N^* were combined at discrete proportions and diluted into E10-32 reaction mixtures (lacking OGT) to account for matrix effects during measurement. Each concentration was measured in duplicate.

Appendix B

Supplementary Tables

Name	Sequence	Purity#
CKII3K	H_2N–KKKYPGGSTPVS‡SANMM–OH	95%
S10-35	H_2N–KKKY$\underline{V}$RVCSNPPCST^{11} ...HETGTTNTATTATSNMAGQH–OH	91%
26mer	H_2N–$\underline{V}$RVCSNPPCET^{11} HETGTTNTATTATSN–OH	98%
C9R-26	H_2N–$\underline{V}$RVCSNPPRET^{11} HETGTTNTATTATSN–OH	96%
C9S-26	H_2N–$\underline{V}$RVCSNPPSET^{11} HETGTTNTATTATSN–OH	96%
C9Y-26	H_2N–$\underline{V}$RVCSNPPYET^{11} HETGTTNTATTATSN–OH	96%
C9A-26	H_2N–$\underline{V}$RVCSNPPAET^{11} HETGTTNTATTATSN–OH	96%
C9NH_2-26	H_2N–$\underline{V}$RVCSNPPA†ET^{11} HETGTTNTATTATSN–OH	96%
A10-26	H_2N–$\underline{V}$RVCSNPPCAT^{11} HETGTTNTATTATSN–OH	98%
Q10-26	H_2N–$\underline{V}$RVCSNPPCQT^{11} HETGTTNTATTATSN–OH	98%
32mer	H_2N–Y$\underline{V}$RVCSNPPCET^{11} ...HETGTTNTATTATSNMAGQH–OH	89 (98)%

37mer	H_2N–Y<u>V</u>RVCSNPPCET^{11} ...HETGTTNTATTATSNMAGQHGCSNP–OH	89%
E10-32	H_2N–Y<u>V</u>RVCSNPPCET^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	98%
A9E10-32	H_2N–Y<u>V</u>RVCSNPPAET^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	96%
R9E10-32	H_2N–Y<u>V</u>RVCSNPPRET^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	95%
S9E10-32	H_2N–Y<u>V</u>RVCSNPPSET^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	98%
E10V11-32	H_2N–Y<u>V</u>RVCSNPPCEV^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	96%
A10-32	H_2N–Y<u>V</u>RVCSNPPCAT^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	97%
D10-32	H_2N–Y<u>V</u>RVCSNPPCDT^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	96%
A9D10-32	H_2N–Y<u>V</u>RVCSNPPADT^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	96%
R9D10-32	H_2N–Y<u>V</u>RVCSNPPRDT^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	98%
S9D10-32	H_2N–Y<u>V</u>RVCSNPPSDT^{11} ...HQTGTTNTATTATSNMAGQH–NH_2	95%

isoD10-32	H_2N–Y$\underline{V}$RVCSNPPCisoDT^{11} ... HQTGTTNTATTATSNMAGQH–NH_2	99%
N-term	H_2N–YVRVCSNPPC–OH	99%
N*	H_2N–YVR*VCSNPPC–OH	99%
HCF-C	H_2N–T^{11}HETGTTNTATTATSN–OH	98%

Table B.1: List of peptides employed in these studies. Peptides were all obtained from Biomatik. Notes: $^{\#}$ parenthetical value is for separate batch upon re-order. ‡ denotes site of principle glycosylation.[142] * denotes heavy Arg: $6\times^{13}$C + $4\times^{15}$N (+10 Da). $\underline{\text{V}}$ denotes the first numbered residue, to maintain convention with HCF-1 literature; Residue X^{11} is labeled for convenience. A† denotes a non-canonical amino alanine residue.

Strain	Construct‡	Plasmid
JCJ-C119	HCF-3R-EEE† (where $X^{\alpha} = X^{\beta} = X^{\gamma} = E$)	pET47b
JCJ-C146	HCF-3R-EAA† (where $X^{\alpha} = E$; $X^{\beta} = X^{\gamma} = A$)	pET47b
JCJ-C147	HCF-3R-AEA† (where $X^{\beta} = E$; $X^{\alpha} = X^{\gamma} = A$)	pET47b
JCJ-C150	HCF-3R-AAE (where $X^{\alpha} = X^{\beta} = A$; $X^{\gamma} = E$)	pET47b

Table B.2: HCF-3R-XXX constructs. †Allele originally made by Michael B. Lazarus/Jiaoyang Jiang. ‡Generic sequence for HCF-1-EAA (C9X) is as follows:

MAHHHHHHHHSAALEVLFQGPGYQDPPGTVTLVCSNPPC$\mathbf{X}^{\alpha}$THETGTTNTATTTVVA

. . . NLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPC$\mathbf{X}^{\beta}$THETGTTN

. . . TATTATSNMAGQHGCSNPPC$\mathbf{X}^{\gamma}$THETGTTNTATTAMSSVGAN.

Each underlined region denotes HCF-1 cleavage repeats one, two and three, respectively.

Strain	Construct‡	Plasmid
JCJ-C169	HCF-3R-EAA (C9A); X = A	pET47b
JCJ-C160	HCF-3R-EAA (C9M); X = M	pET47b
JCJ-C170	HCF-3R-EAA (C9S); X = S	pET47b
JCJ-C159	HCF-3R-EAA (C9T); X = T	pET47b
JCJ-C167	HCF-3R-EAA (C9R); X = R	pET47b
JCJ-C165	HCF-3R-EAA (C9V); X = V	pET47b
JCJ-C166	HCF-3R-EAA (C9D); X = D	pET47b
JCJ-C195	HCF-3R-EAA (C9K); X = K	pET47b
JCJ-C193	HCF-3R-EAA (C9E); X = E	pET47b
JCJ-C196	HCF-3R-EAA (C9Y); X = Y	pET47b

Table B.3: HCF-3R-EAA (C9X) constructs. ‡Generic sequence for HCF-1-EAA (C9X) is as follows:

MAHHHHHHHHSAALEVLFQGPGYQDPPGTVTLVCSNPP**X**ETHETGTTNTATTTVVA

. . . NLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTN

. . . TATTATSNMAGQHGCSNPPCATHETGTTNTATTAMSSVGAN

Strain	Construct‡	Plasmid
JCJ-C163	HCF-3R-EAA (T11M); X = M	pET47b
JCJ-C164	HCF-3R-EAA (T11A); X = A	pET47b
JCJ-C168	HCF-3R-EAA (T11N); X = N	pET47b
JCJ-C162	HCF-3R-EAA (T11I); X = I	pET47b
JCJ-C161	HCF-3R-EAA (T11V); X = V	pET47b
JCJ-C183	HCF-3R-EAA (T11S); X = S	pET47b

Table B.4: HCF-3R-EAA (T11X) constructs. ‡Generic sequence for HCF-1-EAA (T11X) is as follows:

MAHHHHHHHHSAALEVLFQGPGYQDPPGTV<u>TLVCSNPPCE**X**HETGTTNTATTTVVA</u>

. . . NLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTN

. . . TATTATSNMAGQHGCSNPPCATHETGTTNTATTAMSSVGAN

Strain	Construct‡	Plasmid
JCJ-C120	HCF-3R-AAA†; X = A	pET47b
JCJ-C154	HCF-3R-SAA†; X = S	pET47b
JCJ-C209	HCF-3R-DAA; X = D	pET47b
JCJ-C211	HCF-3R-NAA; X = N	pET47b
JCJ-C210	HCF-3R-QAA; X = Q	pET47b

Table B.5: HCF-3R-XAA constructs. †Allele originally made by Michael B. Lazarus/Jiaoyang Jiang. ‡Generic sequence for HCF-1-XAA (E10X) is as follows:

MAHHHHHHHHSAALEVLFQGPGYQDPPGTV<u>TLVCSNPPC**X**THETGTTNTATTTVVA</u>

. . . NLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCATHETGTTN

. . . TATTATSNMAGQHGCSNPPCATHETGTTNTATTAMSSVGAN

Strain	Construct	Residues‡	Plasmid
JCJ-C121	C14	1–56	pET47b
JCJ-C122	C15	1–92	pET47b
JCJ-C131	C16	1–118	pET47b
JCJ-C123	C17	1–30 ‖ 57–151	pET47b
JCJ-C137	C18	1–30 ‖ 57–118	pET47b
JCJ-C132	C19	1–30 ‖ 93–151	pET47b
JCJ-C124	C20	1–30 ‖ 119–151	pET47b

Table B.6: HCF-3R truncation constructs. ‡Numbering corresponds to that shown in B.2 on page 214

Strain	Construct	Genotype	Plasmid
JCJ-C207	Wild-type OGT	ncOGT (23–1041)	pET24b
JCJ-C220	OGT (D554N)	ncOGT (23–1041), D554N	pET24b
JCJ-C219	OGT (D554A)	ncOGT (23–1041), D554A	pET24b
JCJ-C217	OGT (K842M)	ncOGT (23–1041), K842M	pET24b
JCJ-C222	OGT (H498A)	ncOGT (23–1041), H498A	pET24b
JCJ-C221	OGT (H558A)	ncOGT (23–1041), H558A	pET24b
JCJ-C102	hOGT4.5 (Δ9 TPR)†	ncOGT (323–1041)	pET24b
JCJ-C232	hOGT4.5 (D554N)†	ncOGT (323–1041), D554N	pET24b
JCJ-C156	OGT5.5 (Δ8 TPR)†	ncOGT (289–1041)	pET24b
JCJ-C157	OGT6.5 (Δ7 TPR)†	ncOGT (255–1041)	pET24b
JCJ-C186	OGT7.0 (Δ6.5 TPR)	ncOGT (238–1041)	pET24b
JCJ-C187	OGT7.5 (Δ6 TPR)	ncOGT (221–1041)	pET24b
JCJ-C188	OGT8.0 (Δ5.5 TPR)	ncOGT (203–1041)	pET24b
JCJ-C189	OGT8.5 (Δ5 TPR)	ncOGT (187–1041)	pET24b
JCJ-C190	OGT9.0 (Δ4.5 TPR)	ncOGT (169–1041)	pET24b
JCJ-C191	OGT9.5 (Δ4 TPR)	ncOGT (153–1041)	pET24b

Table B.7: List of OGT constructs employed in these studies. †Allele originally made by Michael B. Lazarus. Note: OGT7.0 (Δ6.5 TPR), OGT8.5 (Δ5 TPR) and OGT9.5 (Δ4 TPR) yielded little soluble protein, presumably due to misfolding of the TPR domain.

Step	Time (min)	%A†	%B‡
1	0 – 3	100	0
2	3 – 3.1	90	10
3	3.1 – 17.3	77.5	22.5
4	17.3 – 17.4	0	100
5	17.4 – 22.4	40	60
6	22.4 – 27.4	40	60
7	27.4 – 29.4	100	0
8	29.4 – 33	100	0

Table B.8: LC gradient method (q-Exactive). †Mobile phase "A" is 0.1% formic acid in water. ‡Mobile phase "B" is 0.1% formic acid in acetonitrile.

Step	Time (min)	%A†	%B‡
1	0	97	3
2	2	97	3
3	30	55	45
4	35	30	70
5	35.1	97	3
6	45	97	3

Table B.9: LC gradient method (q-ToF). †Mobile phase "A" is 0.1% formic acid in water. ‡Mobile phase "B" is 0.1% formic acid in acetonitrile.

	hOGT4.5 (D554N):UDP:CKII glycopeptide (ternary product)
Data collection	
Space Group	I121
Cell dimensions	
a,b,c (Å)	98.35 137.19 153.07
α, β, γ (°)	90.00 103.05 90.00
Resolution (Å)	40.93–2.05 (2.09–2.05)
R_{sym} or R_{merge}	0.097 (0.324)
I / σI	5.9 (1.9)
Completeness (%)	95.2 (86.3)
Redundancy	2.7 (2.3)
Refinement	
Resolution (Å)	40.925–2.05
Number of reflections	117417
R_{work} / R_{free}	0.2172 / 0.2374
Number of atoms	
Protein	11093
Ligand/ion	93
Water	772
B-factors	
Protein	32.3
Ligand/ion	21.8
Water	33.6
RMS deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.661

Table B.10: Data collection and refinement statistics for hOGT4.5 (D554N):UDP:CKII glycopeptide ternary structure. Values in parentheses are for the highest-resolution shell. One crystal was used for this structure.

Appendix C

Glossary of Terms

Abbreviation	Definition
AdoHcy	*S*-Adenosyl-L-homocysteine (also SAH)
ADP	Adenosine diphosphate
ADPr	ADP-ribosylation
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
CID	Collision-induced dissociation
CIP	Calf intestinal alkaline phosphatase
Da	Dalton (1 amu)
ECD	Electron-capture dissociation
EIC	Extracted ion chromatogram
ETD	Electron-transfer dissociation
ESI	Electropray ionization
FT	Fourier transform
FWHM	Full width at half maximum
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
Glc	Glucose
Gtf(s)	Glycosyltransferase(s)
HCD	Higher-energy collisional dissociation
HCF-1	Host cell factor-1
HPLC	High Performance Liquid chromatography
HRV3C	Human Rhinovirus 3C [protease]
ICR	Ion cyclotron resonance
IPTG	Isopropyl β-D-1-thiogalactopyranoside

LB	Luria-Bertani broth
LC	Liquid chromatography
MS	Mass spectrometry
MS^n	Tandem mass spectrometry (general)
MS/MS	Tandem mass spectrometry (also MS^2)
MWCO	Molecular weight cut-off
NT-sugar	Nucleotide-sugar
OGT	*O*-linked-*N*-acetylglucosaminetransferase
PAGE	Polyacrylamide gel electrophoresis
PARP	poly(ADP-ribosyl) polymerase
pE	pyroglutamic acid (also pyroGlu)
PIMT	Protein isoaspartyl methyltransferase
PGP	Pyroglutamate aminopeptidase
PMSF	Phenylmethylsulfonyl fluoride
ppm	Parts-per-million
PTM	Post-translational modification
RMS	Root mean squared
S/N	Signal-to-Noise ratio
SAM	*S*-adenosylmethionine
SEC	Size-exclusion chromatography
SDS	Sodium dodecyl sulfate
SID	Stable isotope dilution
STO	Single-turnover
TBS	Tris-buffered saline
THP	Tris(3-hydroxypropyl)phosphine
TIC	Total ion chromatogram
ToF	Time-of-Flight
THP	Tris(3-hydroxypropyl)phosphine
TPR(s)	Tetratricopeptide repeat(s)
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
UDP	Uridine diphosphate
UDP-5*S*GlcNAc	Uridine diphosphate 5-thio-*N*-acetylglucosamine
UDP-Gal	Uridine 5'-diphosphogalactose
UDP-GalNAc	Uridine 5'-diphosphate *N*-acetylgalactosamine
UDP-GlcNAc	Uridine 5'-diphosphate *N*-acetylglucosamine
UDP-Glc	Uridine 5'-diphosphoglucose
uHPLC	Ultra High Performance Liquid chromatography

'U'	Enzyme activity unit
'wt'	Wild-type

Table C.1: List of abbreviated terms

Bibliography

[1] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, **66**, 213 (2010).

[2] J. F. Alfaro, L. A. Gillies, H. G. Sun, S. Dai, T. Zang, J. J. Klaene, B. J. Kim, J. D. Lowenson, S. G. Clarke, B. L. Karger, and Z. S. Zhou. Chemo-Enzymatic Detection of Protein Isoaspartate Using Protein Isoaspartate Methyltransferase and Hydrazine Trapping. *Analytical chemistry*, **80**, 3882 (2008).

[3] R. K. Allan and T. Ratajczak. Versatile TPR domains accommodate different modes of target protein recognition and function. *Cell Stress and Chaperones*, **16**, 353 (2010).

[4] J. Alonso, M. Schimpl, and D. M. F. van Aalten. O-GlcNAcase: Promiscuous Hexosaminidase or Key Regulator of O-GlcNAc Signaling? *The Journal of Biological Chemistry*, **289**, 34433 (2014).

[5] Y. Anraku, R. Mizutani, and Y. Satow. Protein Splicing: Its Discovery and Structural Insight into Novel Chemical Mechanisms. *IUBMB Life*, **57**, 563 (2005).

[6] C. M. apRhys, D. M. Ciufo, E. A. O'Neill, T. J. Kelly, and G. S. Hayward. Overlapping octamer and TAATGARAT motifs in the VF65-response elements in herpes simplex virus immediate-early promoters represent independent binding sites for cellular nuclear factor III. *Journal of Virology*, **63**, 2798 (1989).

[7] A. Ardèvol and C. Rovira. The Molecular Mechanism of Enzymatic Glycosyl Transfer with Retention of Configuration: Evidence for a Short-Lived Oxocarbenium-Like Species. *Angewandte Chemie International Edition*, **50**, 10897 (2011).

[8] A. Artigues, A. Birkett, and V. Schirch. Evidence for the in vivo deamidation and isomerization of an asparaginyl residue in cytosolic serine hydroxymethyltransferase. *The Journal of Biological Chemistry*, **265**, 4853 (1990).

[9] L. Athmer, J. Kindrachuk, F. Georges, and S. Napper. The Influence of Protein Structure on the Products Emerging from Succinimide Hydrolysis. *The Journal of Biological Chemistry*, **277**, 30502 (2002).

[10] K. D. Ballard and S. J. Gaskell. Dehydration of peptide [M+ H]$^+$ ions in the gas phase. *Journal of the American Society for Mass Spectrometry*, **4**, 477 (1993).

[11] A. R. Battersby and J. J. Reynolds. 104. Studies on specific chemical fission of peptide links. Part III. Fission of peptides containing one glutamic acid residue. *Journal of the Chemical Society*, pages 524–530 (1961).

[12] A. R. Battersby and J. C. Robinson. 407. Studies on specific chemical fission of peptide links. Part II. The preparation and hydrolysis of 1-acylpyrrolid-2-ones derived from [small alpha]-glutamyl peptides. *Journal of the Chemical Society*, pages 2076–2084 (1956).

[13] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie. imosflm: a new graphical interface for diffraction-image processing with mosflm. *Acta Crystallographica Section D*, **67**, 271 (2011).

[14] A. D. Becke. Densityfunctional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, **98**, 5648 (1993).

[15] P. Bheda, H. Jing, C. Wolberger, and H. Lin. The Substrate Specificity of Sirtuins. *Annual Review of Biochemistry*, **85**, 405 (2016).

[16] T. Bhuiyan, P. Waridel, V. Kapuria, V. Zoete, and W. Herr. Distinct OGT-Binding Sites Promote HCF-1 Cleavage. *PLOS ONE*, **10**, e0136636 (2015).

[17] B. Biterge, F. Richter, G. Mittler, and R. Schneider. Methylation of histone H4 at aspartate 24 by Protein L-isoaspartate O-methyltransferase (PCMT1) links histone modifications with protein homeostasis. *Scientific Reports*, **4**, 6674 (2014).

[18] G. L. Blatch and M. Lässle. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays*, **21**, 932 (1999).

[19] M. Bodanszky and S. Natarajan. Side reactions in peptide synthesis. II. Formation of succinimide derivatives from aspartyl residues. *The Journal of Organic Chemistry*, **40**, 2495 (1975).

[20] M. R. Bond and J. A. Hanover. A little sugar goes a long way: The cell biology of O-GlcNAc. *The Journal of Cell Biology*, **208**, 869 (2015).

[21] V. S. Borodkin, M. Schimpl, M. Gundogdu, K. Rafie, H. C. Dorfmueller, D. A. Robinson, and D. M. F. van Aalten. Bisubstrate UDP–peptide conjugates as human O-GlcNAc transferase inhibitors. *Biochemical Journal*, **457**, 497 (2014).

[22] R. R. Bosch, M.-J. J. M. Pouwels, P. N. Span, A. J. Olthaar, C. J. Tack, A. R. M. M. Hermus, and C. G. J. F. Sweep. Hexosamines Are Unlikely to Function as a Nutrient-Sensor in 3T3-L1 Adipocytes: A Comparison of UDP-Hexosamine Levels after Increased Glucose Flux and Glucosamine Treatment. *Endocrine*, **23**, 17 (2004).

[23] Y. Bourne and B. Henrissat. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Current Opinion in Structural Biology*, **11**, 593 (2001).

[24] K. Brown, N. Dedi, J. Westcott, S. Scally, T. D. Bugg, P. A. Charlton, and G. M. Cheetham. Crystal structure of the pseudomonas aeruginosa murg: Udp-glcnac substrate complex. *Protein & Peptide Letters*, **20**, 1002 (2013).

[25] Y. Burstein, F. Kantour, and I. Schechter. Partial amino-acid sequence of the precursor of an immunoglobulin light chain containing NH2-terminal pyroglutamic acid. *Proceedings of the National Academy of Sciences*, **73**, 2604 (1976).

[26] C. Butkinaree, K. Park, and G. W. Hart. O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochimica et Biophysica Acta*, **1800**, 96 (2010).

[27] F. Capotosti, S. Guernier, F. Lammers, P. Waridel, Y. Cai, J. Jin, J. W. Conaway, R. C. Conaway, and W. Herr. O-GlcNAc Transferase Catalyzes Site-Specific Proteolysis of HCF-1. *Cell*, **144**, 376 (2011).

[28] A. D. Carlson and R. M. Riggin. Development of Improved High-Performance Liquid Chromatography Conditions for Nonisotopic Detection of Isoaspartic Acid to Determine the Extent of Protein Deamidation. *Analytical Biochemistry*, **278**, 150 (2000).

[29] S. A. Carr, M. E. Hemling, and M. F. Bean. Integration of mass spectrometry in analytical biotechnology. *Analytical Chemistry*, **63**, 2802 (1991).

[30] W. J. Chazin, J. Kördel, E. Thulin, T. Hofmann, T. Drakenberg, and S. Forsén. Identification of an Isoaspartyl Linkage Formed Upon Deamidation of Bovine Calbindin-D9k and Structural Characterization by 2D ^{1}H-NMR. *Biochemistry*, **28**, 8646 (1989).

[31] Q. Chen, Y. Chen, C. Bian, R. Fujiki, and X. Yu. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, **493**, 561 (2012).

[32] X. Chen and F. Turecek. Simple b ions have cyclic oxazolone structures. A neutralization-reionization mass spectrometric and computational study of oxazolone radicals. *Journal of the American Society for Mass Spectrometry*, **16**, 1941 (2005).

[33] W. D. Cheung, K. Sakabe, M. P. Housley, W. B. Dias, and G. W. Hart. O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *The Journal of Biological Chemistry*, **283**, 33935 (2008).

[34] C. F. Chou, A. J. Smith, and M. B. Omary. Characterization and dynamics of O-linked glycosylation of human cytokeratin 8 and 18. *The Journal of Biological Chemistry*, **267**, 3901 (1992).

[35] A. N. Chulin, I. L. Rodionov, L. K. Baidakova, L. N. Rodionova, T. A. Balashova, and V. T. Ivanov. Preparation and reactivity of aminoacyl pyroglutamates. Facile synthesis of 10-membered-ring cyclic dipeptides derived from 1,4-diaminobutyric and glutamic acids. *Journal of Peptide Science*, **11**, 175 (2005).

[36] A. J. Clarke, R. Hurtado-Guerrero, S. Pathak, A. W. Schüttelkopf, V. Borodkin, S. M. Shepherd, A. F. M. Ibrahim, and D. M. F. van Aalten. Structural insights into mechanism and specificity of O-GlcNAc transferase. *The EMBO Journal*, **27**, 2780 (2008).

[37] S. Clarke. Protein carboxyl methyltransferases: two distinct classes of enzymes. *Annual Review of Biochemistry*, **54**, 479 (1985).

[38] S. Clarke. Propensity for Spontaneous Succinimide Formation From Aspartyl and Asparaginyl Residues in Cellular Proteins. *International Journal of Peptide and Protein Research*, **30**, 808 (1987).

[39] E. Conti, M. Uy, L. Leighton, G. Blobel, and J. Kuriyan. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell*, **94**, 193 (1998).

[40] A. Corti and F. Curnis. Isoaspartate-dependent molecular switches for integrin-ligand recognition. *Journal of Cell Science*, **124**, 515 (2011).

[41] P. M. Coutinho, E. Deleury, G. J. Davies, and B. Henrissat. An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology*, **328**, 307 (2003).

[42] J. Cox and M. Mann. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annual Review of Biochemistry*, **80**, 273 (2011).

[43] D. Crich and K. Sana. S_N2-Type Nucleophilic Opening of β-Thiolactones (Thietan-2-ones) as a Source of Thioacids for Coupling Reactions. *The Journal of Organic Chemistry*, **74**, 3389 (2009).

[44] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner. WebLogo: a sequence logo generator. *Genome Research*, **14**, 1188 (2004).

[45] S. Cukierman. Et tu, grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta*, **1757**, 876 (2006).

[46] C. M. Daniels, S.-E. Ong, and A. K. L. Leung. Phosphoproteomic Approach to Characterize Protein Mono- and Poly(ADP-ribosyl)ation Sites from Cells. *Journal of Proteome Research*, **13**, 3510 (2014).

[47] S. Daou and N. Mashtalir. Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 2747 (2011).

[48] A. K. Das, P. W. Cohen, and D. Barford. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *The EMBO Journal*, **17**, 1192 (1998).

[49] G. J. Davies, T. M. Gloster, and B. Henrissat. Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*, **15**, 637 (2005).

[50] N. DeGraan-Weber, J. Zhang, and J. P. Reilly. Distinguishing Aspartic and Isoaspartic Acids in Peptides by Several Mass Spectrometric Fragmentation Methods. *Journal of the American Society for Mass Spectrometry*, **27**, 2041 (2016).

[51] R. Dentin, S. Hedrick, J. Xie, J. I. Yates, and M. Montminy. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science*, **319**, 1402 (2008).

[52] A. DiDonato, P. Galletti, and G. D'Alessio. Selective Deamidation and Enzymatic Methylation of Seminal Ribonuclease. *Biochemistry*, **25**, 8361 (1986).

[53] S. Doll and A. L. Burlingame. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chemical Biology*, **10**, 63 (2015).

[54] H. C. Dorfmueller, V. S. Borodkin, D. E. Blair, S. Pathak, I. Navratilova, and D. M. F. van Aalten. Substrate and product analogues as human O-GlcNAc transferase inhibitors. *Amino Acids*, **40**, 781 (2010).

[55] C. M. Eakin, A. Miller, J. Kerr, J. Kung, and A. Wallace. Assessing analytical methods to monitor isoAsp formation in monoclonal antibodies. *Frontiers in Pharmacology*, **5**, 87 (2014).

[56] P. Edman. Method for determination of the amino acid sequence in peptides. *Acta Chemica Scandinavica*, **4**, 283 (1950).

[57] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of Coot. *Acta Crystallographica Section D*, **66**, 486 (2010).

[58] B. W. Erickson and S. A. Khan. Synthetic lactam and thiolactone models of protein metastable binding sites. *Annals of the New York Academy of Sciences*, **421**, 167 (1983).

[59] J. C. Errey, S. S. Lee, R. P. Gibson, C. Martinez-Fleites, C. S. Barry, P. M. J. Jung, A. C. O'Sullivan, B. G. Davis, and G. J. Davies. Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. *Angewandte Chemie International Edition*, **49**, 1234 (2010).

[60] R. S. Feinberg and R. B. Merrifield. Modification of peptides containing glutamic acid by hydrogen fluoride-anisole mixtures. γ-Acylation of anisole or the glutamyl nitrogen. *Journal of the American Chemical Society*, **97**, 3485 (1975).

[61] J. L. Feldman, K. E. Dittenhafer-Reed, and J. M. Denu. Sirtuin catalysis and regulation. *The Journal of Biological Chemistry*, **287**, 42419 (2012).

[62] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science*, **246**, 64 (1989).

[63] B. S. Fletcher, C. Dragstedt, L. Notterpek, and G. P. Nolan. Functional cloning of SPIN-2, a nuclear anti-apoptotic protein with roles in cell cycle progression. *Leukemia*, **16**, 1507 (2002).

[64] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian '09 Revision D.01. Gaussian Inc., Wallingford CT, 2013.

[65] M. Funabashi, Z. Yang, K. Nonaka, M. Hosobuchi, Y. Fujita, T. Shibata, X. Chi, and S. G. Van Lanen. An ATP-independent strategy for amide bond formation in antibiotic biosynthesis. *Nature Chemical Biology*, **6**, 581 (2010).

[66] T. Furuchi, K. Sakurako, M. Katane, M. Sekine, and H. Homma. The role of protein L-isoaspartyl/D-aspartyl O-methyltransferase (PIMT) in intracellular signal transduction. *Chemistry & Biodiversity*, **7**, 1337 (2010).

[67] M. C. Gambetta and J. Müller. A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma*, **124**, 429 (2015).

[68] G. J. Gatto, B. V. Geisbrecht, S. J. Gould, and J. M. Berg. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nature Structural Biology*, **7**, 1091 (2000).

[69] T. Geiger and S. Clarke. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *The Journal of Biological Chemistry*, **262**, 785 (1987).

[70] T. Gerster and R. G. Roeder. A herpesvirus trans-activating protein interacts with transcription factor OTF-1 and other cellular proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 6347 (1988).

[71] T. M. Gloster and D. J. Vocadlo. Mechanism, Structure, and Inhibition of O-GlcNAc Processing Enzymes. *Current Signal Transduction Therapy*, **5**, 74 (2010).

[72] T. M. Gloster, W. F. Zandberg, J. E. Heinonen, D. L. Shen, L. Deng, and D. J. Vocadlo. Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. *Nature Chemical Biology*, **7**, 174 (2011).

[73] E. D. Goddard-Borger. Enzyme mechanisms: Sugary shears. *Nature Chemical Biology*, **12**, 892 (2016).

[74] J. P. Gogarten, A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, and E. Hilario. Inteins: Structure, Function, and Evolution. *Annual Review of Microbiology*, **56**, 263 (2002).

[75] H. Gómez, I. Polyak, W. Thiel, J. M. Lluch, and L. Masgrau. Retaining Glycosyltransferase Mechanism Studied by QM/MM Methods: Lipopolysaccharyl-α-1,4-galactosyltransferase C Transfers α-Galactose via an Oxocarbenium Ion-like Transition State. *Journal of the American Chemical Society*, **134**, 4743 (2012).

[76] Q. Gong, L. Li, X. Wu, and H. Ma. Pyroglutamate aminopeptidase 1 may be an indicator of cellular inflammatory response as revealed using a sensitive long-wavelength fluorescent probe. *Chemical Science*, **7**, 4694 (2016).

[77] J. D. Goodreid, J. Janetzko, J. P. Santa Maria, K. S. Wong, E. Leung, B. T. Eger, S. Bryson, E. F. Pai, S. D. Gray-Owen, S. Walker, W. A. Houry, and R. A. Batey. Development and Characterization of Potent Cyclic Acyldepsipeptide Analogues with Increased Antimicrobial Activity. *Journal of Medicinal Chemistry*, **59**, 624 (2016).

[78] S. C. Griffith, M. R. Sawaya, D. R. Boutz, N. Thapar, J. E. Katz, S. Clarke, and T. O. Yeates. Crystal structure of a protein repair methyltransferase from Pyrococcus furiosus with its l-isoaspartyl peptide substrate. *Journal of Molecular Biology*, **313**, 1103 (2001).

[79] H. Grossenbacher, W. Märki, M. Coulot, D. Müller, and W. J. Richter. Characterization of succinimide-type dehydration products of recombinant hirudin variant 1 by electrospray tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, **7**, 1082 (1993).

[80] C. Grotthuss. Sur la décomposition de l'eau et des corps q'uelle tient en dissolution à l'aide de l'électricité galvanique. *Annales de Chimie et de Physique*, **58**, 54 (1806).

[81] M. Guilhaus. Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry*, **30**, 1519 (1995).

[82] C. Guinez, J. Lemoine, J.-C. Michalski, and T. Lefebvre. 70-kDa-heat shock protein presents an adjustable lectinic activity towards O-linked N-acetylglucosamine. *Biochemical and Biophysical Research Communications*, **319**, 21 (2004).

[83] V. A. Gupta and A. H. Beggs. Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skeletal Muscle*, **4**, 11 (2014).

[84] B. H. O. Güttler, H. Cynis, F. Seifert, H.-H. Ludwig, A. Porzel, and S. Schilling. A quantitative analysis of spontaneous isoaspartate formation from N-terminal asparaginyl and aspartyl residues. *Amino Acids*, **44**, 1205 (2013).

[85] S. Ha, D. Walker, Y. Shi, and S. Walker. The 1.9 Å crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Science*, **9**, 1045 (2000).

[86] R. S. Haltiwanger, M. A. Blomberg, and G. W. Hart. Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide beta-N-acetylglucosaminyltransferase. *The Journal of Biological Chemistry*, **267**, 9005 (1992).

[87] R. S. Haltiwanger, G. D. Holt, and G. W. Hart. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *The Journal of Biological Chemistry*, **265**, 2563 (1990).

[88] I. Han and J. E. Kudlow. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. *Molecular and Cellular Biology*, **17**, 2550 (1997).

[89] J. A. Hanover. A Versatile Sugar Transferase Makes the Cut. *Cell*, **144**, 321 (2011).

[90] J. A. Hanover, M. W. Krause, and D. C. Love. Post-translational modifications: Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nature Reviews Molecular Cell Biology*, **13**, 312 (2012).

[91] J. A. Hanover, S. Yu, W. B. Lubas, S.-H. Shin, M. Ragano-Caracciola, J. Kochran, and D. C. Love. Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Archives of Biochemistry and Biophysics*, **409**, 287 (2003).

[92] S. Hardivillé and G. W. Hart. Nutrient regulation of gene expression by O-GlcNAcylation of chromatin. *Current Opinion in Chemical Biology*, **33**, 88 (2016).

[93] G. W. Hart, M. P. Housley, and C. Slawson. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature*, **446**, 1017 (2007).

[94] G. W. Hart, C. Slawson, G. Ramirez-Correa, and O. Lagerlöf. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annual Review of Biochemistry*, **80**, 825 (2011).

[95] K. R. Harwood and J. A. Hanover. Nutrient-driven O-GlcNAc cycling - think globally but act locally. *Journal of Cell Science*, **127**, 1857 (2014).

[96] P. O. Hassa, S. S. Haenni, M. Elser, and M. O. Hottiger. Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going? *Microbiology and Molecular Biology Reviews*, **70**, 789 (2006).

[97] G. D. Holt and G. W. Hart. The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc. *The Journal of Biological Chemistry*, **261**, 8049 (1986).

[98] G. D. Holt, C. M. Snow, A. Senior, R. S. Haltiwanger, L. Gerace, and G. W. Hart. Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine. *The Journal of Cell Biology*, **104**, 1157 (1987).

[99] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, **40**, 430 (2005).

[100] Y. Hu, L. Chen, S. Ha, B. Gross, B. Falcone, D. Walker, M. Mokhtarzadeh, and S. Walker. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 845 (2003).

[101] A. H. Huber and W. I. Weis. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell*, **105**, 391 (2001).

[102] R. Hurtado-Guerrero, H. C. Dorfmueller, and D. M. van Aalten. Molecular mechanisms of O-GlcNAcylation. *Current Opinion in Structural Biology*, **18**, 551 (2008).

[103] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931 (2004).

[104] S. P. N. Iyer, Y. Akimoto, and G. W. Hart. Identification and Cloning of a Novel Family of Coiled-coil Domain Proteins That Interact with O-GlcNAc Transferase. *The Journal of Biological Chemistry*, **278**, 5399 (2003).

[105] S. P. N. Iyer and G. W. Hart. Roles of the Tetratricopeptide Repeat Domain in O-GlcNAc Transferase Targeting and Protein Substrate Specificity. *The Journal of Biological Chemistry*, **278**, 24608 (2003).

[106] I. Jancan and M. A. Macnaughtan. Acid dissociation constants of uridine-5'-diphosphate compounds determined by ^{31}phosphorus nuclear magnetic resonance spectroscopy and internal pH referencing. *Analytica Chimica Acta*, **749**, 63 (2012).

[107] J. Janetzko, S. A. Trauger, M. B. Lazarus, and S. Walker. How the glycosyltransferase OGT catalyzes amide bond cleavage. *Nature Chemical Biology*, **12**, 899 (2016).

[108] J. Janetzko and S. Walker. The making of a sweet modification: structure and function of O-GlcNAc transferase. *The Journal of Biological Chemistry*, **289**, 34424 (2014).

[109] J. Janetzko and S. Walker. Aspartate Glycosylation Triggers Isomerization to Isoaspartate. *Journal of the American Chemical Society*, **139**, 3332 (2017).

[110] J. Jiang, M. B. Lazarus, L. Pasquina, P. Sliz, and S. Walker. A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. *Nature Chemical Biology*, **8**, 72 (2011).

[111] M. Jínek, J. Rehwinkel, B. D. Lazarus, E. Izaurralde, J. A. Hanover, and E. Conti. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nature Structural & Molecular Biology*, **11**, 1001 (2004).

[112] B. A. Johnson and D. W. Aswad. Fragmentation of Isoaspartyl Peptides and Proteins by Carboxypeptidase-Y - Release of Isoaspartyl Dipeptides as a Result of Internal and External Cleavage. *Biochemistry*, **29**, 4373 (1990).

[113] B. A. Johnson, J. M. Shirokawa, W. S. Hancock, M. W. Spellman, L. J. Basa, and D. W. Aswad. Formation of isoaspartate at two distinct sites during in vitro aging of human growth hormone. *Journal of Biological Chemistry*, **264**, 14262 (1989).

[114] K. L. Johnson and D. C. Muddiman. A method for calculating 16o/18o peptide ion ratios for the relative quantification of proteomes. *Journal of the American Society for Mass Spectrometry*, **15**, 437 (2004).

[115] E. Julien and W. Herr. Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *The EMBO Journal*, **22**, 2360 (2003).

[116] E. Julien and W. Herr. A Switch in Mitotic Histone H4 Lysine 20 Methylation Status Is Linked to M Phase Defects upon Loss of HCF-1. *Molecular Cell*, **14**, 713 (2004).

[117] V. Kapuria, U. F. Röhrig, T. Bhuiyan, V. S. Borodkin, D. M. F. van Aalten, V. Zoete, and W. Herr. Proteolysis of HCF-1 by Ser/Thr glycosylation-incompetent O-GlcNAc transferase:UDP-GlcNAc complexes. *Genes & Development*, **30**, 960 (2016).

[118] M. Katan, A. Haigh, C. P. Verrijzer, P. C. van der Vliet, and P. O'Hare. Characterization of a cellular factor which interacts functionally with Oct-1 in the assembly of a multicomponent transcription complex. *Nucleic Acids Research*, **18**, 6871 (1990).

[119] Z. Kazemi, H. Chang, S. Haserodt, C. McKen, and N. E. Zachara. O-Linked beta-N-acetylglucosamine (O-GlcNAc) Regulates Stress-induced Heat Shock Protein Expression in a GSK-3 beta-dependent Manner. *The Journal of Biological Chemistry*, **285**, 39096 (2010).

[120] M. Kebede, M. Ferdaoussi, A. Mancini, T. Alquier, R. N. Kulkarni, M. D. Walker, and V. Poitout. Glucose activates free fatty acid receptor 1 gene transcription via phosphatidylinositol-3-kinase-dependent O-GlcNAcylation of pancreas-duodenum homeobox-1. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 2376 (2012).

[121] D. Kern and M. Schutkowski. Rotational Barriers of cis/trans Isomerization of Proline Analogues and Their Catalysis by Cyclophilin. *Journal of the American Chemical Society*, **119**, 8403 (1997).

[122] S. A. Khan and B. W. Erickson. An equilibrium model of the metastable binding sites of alpha 2-macroglobulin and complement proteins C3 and C4. *The Journal of Biological Chemistry*, **257**, 11864 (1982).

[123] S. A. Khan and B. W. Erickson. Synthesis and regioselective hydrolysis of peptides containing an internal residue of pyroglutamic acid. *Journal of the American Chemical Society*, **106**, 798 (1984).

[124] S. A. Khan, J. M. Sekulski, and B. W. Erickson. Peptide models of protein metastable binding sites: competitive kinetics of isomerization and hydrolysis. *Biochemistry*, **25**, 5165 (1986).

[125] N. Khidekel, S. B. Ficarro, E. C. Peters, and L. C. Hsieh-Wilson. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13132 (2004).

[126] G. A. Khoury, R. C. Baliban, and C. A. Floudas. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific Reports*, **1**, 90 (2011).

[127] J. J. Klaene, W. Ni, J. F. Alfaro, and Z. S. Zhou. Detection and Quantitation of Succinimide in Intact Protein via Hydrazine Trapping and Chemical Derivatization. *Journal of Pharmaceutical Sciences*, **103**, 3033 (2014).

[128] F. A. S. Konuklar and V. Aviyente. Modelling the hydrolysis of succinimide: formation of aspartate and reversible isomerization of aspartic acid via succinimide. *Organic & Biomolecular Chemistry*, **1**, 2290 (2003).

[129] M. P. Kötzler and S. G. Withers. Proteolytic Cleavage Driven by Glycosylation. *The Journal of Biological Chemistry*, **291**, 429 (2016).

[130] L. K. Kreppel, M. A. Blomberg, and G. W. Hart. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *The Journal of Biological Chemistry*, **272**, 9308 (1997).

[131] L. K. Kreppel and G. W. Hart. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *The Journal of Biological Chemistry*, **274**, 32015 (1999).

[132] T. M. Kristie, J. L. Pomerantz, T. C. Twomey, S. A. Parent, and P. A. Sharp. The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. *The Journal of Biological Chemistry*, **270**, 4387 (1995).

[133] T. M. Kristie and P. A. Sharp. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV alpha-trans-activator protein. *Genes & Development*, **4**, 2383 (1990).

[134] T. M. Kristie and P. A. Sharp. Purification of the cellular C1 factor required for the stable recognition of the Oct-1 homeodomain by the herpes simplex virus alpha-trans-induction factor (VP16). *The Journal of Biological Chemistry*, **268**, 6525 (1993).

[135] M. Kumari, S. Kozmon, P. Kulhánek, J. Štepán, I. Tvaroška, and J. Koča. Exploring Reaction Pathways for O-GlcNAc Transferase Catalysis. A String Method Study. *The Journal of Physical Chemistry B*, **119**, 4371 (2015).

[136] L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, **77**, 521 (2008).

[137] J. R. Lamb, S. Tugendreich, and P. Hieter. Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends in Biochemical Sciences*, **20**, 257 (1995).

[138] B. D. Lazarus. Recombinant O-GlcNAc transferase isoforms: identification of O-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology*, **16**, 415 (2006).

[139] B. D. Lazarus, M. D. Roos, and J. A. Hanover. Mutational analysis of the catalytic domain of O-linked N-acetylglucosaminyl transferase. *The Journal of Biological Chemistry*, **280**, 35537 (2005).

[140] M. B. Lazarus, J. Jiang, T. M. Gloster, W. F. Zandberg, G. E. Whitworth, D. J. Vocadlo, and S. Walker. Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nature Chemical Biology*, **8**, 966 (2012).

[141] M. B. Lazarus, J. Jiang, V. Kapuria, T. Bhuiyan, J. Janetzko, W. F. Zandberg, D. J. Vocadlo, W. Herr, and S. Walker. HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science*, **342**, 1235 (2013).

[142] M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz, and S. Walker. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature*, **469**, 564 (2011).

[143] W. J. le Quesne and G. T. Young. 106. Amino-acids and peptides. Part VII. The autohydrolysis of glutamyl-peptides. *Journal of the Chemical Society*, page 594 (1952).

[144] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, **37**, 785 (1988).

[145] S. S. Lee, S. Y. Hong, J. C. Errey, A. Izumi, G. J. Davies, and B. G. Davis. Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nature Chemical Biology*, **7**, 629 (2011).

[146] C. Y. Legault. Cylview, 1.0b, 2009.

[147] W. D. Lehmann, A. Schlosser, G. Erben, R. Pipkorn, D. Bossemeyer, and V. Kinzel. Analysis of isoaspartate in peptides by electrospray tandem mass spectrometry. *Protein Science*, **9**, 2260 (2000).

[148] O. Leidecker, J. J. Bonfiglio, T. Colby, Q. Zhang, I. Atanassov, R. Zaja, L. Palazzo, A. Stockum, I. Ahel, and I. Matic. Serine is a new target residue for endogenous ADP-ribosylation on histones. *Nature Chemical Biology*, **12**, 998 (2016).

[149] Z. G. Levine and S. Walker. The Biochemistry of O-GlcNAc Transferase: Which Functions Make It Essential in Mammalian Cells? *Annual Review of Biochemistry*, **85**, 631 (2016).

[150] X. Li, X. Yu, C. E. Costello, C. Lin, and P. B. O'Connor. Top-Down Study of β 2-Microglobulin Deamidation. *Analytical Chemistry*, **84**, 6150 (2012).

[151] D. C. Liebler and L. J. Zimmerman. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry*, **52**, 3797 (2013).

[152] E. Lira-Navarrete, J. Iglesias-Fernández, W. F. Zandberg, I. Compañón, Y. Kong, F. Corzana, B. M. Pinto, H. Clausen, J. M. Peregrina, D. J. Vocadlo, C. Rovira, and R. Hurtado-Guerrero. Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide N-Acetylgalactosaminyltransferase2. *Angewandte Chemie International Edition*, **53**, 8206 (2014).

[153] M. Liu, J. Cheetham, N. Cauchon, J. Ostovic, W. Ni, D. Ren, and Z. S. Zhou. Protein Isoaspartate Methyltransferase-Mediated 18O-Labeling of Isoaspartic Acid for Mass Spectrometry Analysis. *Analytical Chemistry*, **84**, 1056 (2012).

[154] T.-W. Liu, M. Myschyshyn, D. A. Sinclair, S. Cecioni, K. Beja, B. M. Honda, R. D. Morin, and D. J. Vocadlo. Genome-wide chemical mapping of O-GlcNAcylated proteins in Drosophila melanogaster. *Nature Chemical Biology*, **13**, 161 (2017).

[155] X. Liu, L. Li, Y. Wang, H. Yan, X. Ma, P. G. Wang, and L. Zhang. A peptide panel investigation reveals the acceptor specificity of O-GlcNAc transferase. *The FASEB Journal*, **28**, 3362 (2014).

[156] Y. D. Liu, A. M. Goetze, R. B. Bass, and G. C. Flynn. N-terminal Glutamate to Pyroglutamate Conversion in Vivo for Human IgG2 Antibodies. *The Journal of Biological Chemistry*, **286**, 11211 (2011).

[157] C. Lizak, S. Gerber, G. e. l. Michaud, M. Schubert, Y.-Y. Fan, M. Bucher, T. Darbre, M. Aebi, J.-L. Reymond, and K. P. Locher. Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nature Communications*, **4**, 2627 (2013).

[158] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**, D490 (2013).

[159] D. C. Love. Mitochondrial and nucleocytoplasmic targeting of O-linked GlcNAc transferase. *Journal of Cell Science*, **116**, 647 (2002).

[160] W. A. Lubas, D. W. Frank, M. Krause, and J. A. Hanover. O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *The Journal of Biological Chemistry*, **272**, 9316 (1997).

[161] W. A. Lubas and J. A. Hanover. Functional expression of O-linked GlcNAc transferase. Domain structure and substrate specificity. *Journal of Biological Chemistry*, **275**, 10983 (2000).

[162] W. A. Lubas, M. Smith, C. M. Starr, and J. A. Hanover. Analysis of nuclear pore protein p62 glycosylation. *Biochemistry*, **34**, 1686 (1995).

[163] R. L. Luciano and A. C. Wilson. HCF-1 Functions as a Coactivator for the Zinc Finger Protein Krox20. *The Journal of Biological Chemistry*, **278**, 51116 (2003).

[164] Z. Ma and K. Vosseller. O-GlcNAc in cancer biology. *Amino Acids*, **45**, 719 (2013).

[165] M. Mangone, M. P. Myers, and W. Herr. Role of the HCF-1 basic region in sustaining cell proliferation. *PLOS ONE*, **5**, e9020 (2010).

[166] S. A. Marsh, H. E. Collins, and J. C. Chatham. Protein O-GlcNAcylation and cardiovascular (patho)physiology. *The Journal of Biological Chemistry*, **289**, 34449 (2014).

[167] A. G. Marshall and C. L. Hendrickson. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *International Journal of Mass Spectrometry*, **215**, 59 (2002).

[168] S. Marshall, O. Nadeau, and K. Yamasaki. Dynamic actions of glucose and glucosamine on hexosamine biosynthesis in isolated adipocytes: differential effects on glucosamine 6-phosphate, UDP-N-acetylglucosamine, and ATP levels. *The Journal of Biological Chemistry*, **279**, 35313 (2004).

[169] R. Martello, M. Leutert, S. Jungmichel, V. Bilan, S. C. Larsen, C. Young, M. O. Hottiger, and M. L. Nielsen. Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue. *Nature Communications*, **7**, 12917 (2016).

[170] C. Martinez-Fleites, Y. He, and G. J. Davies. Structural analyses of enzymes involved in the O-GlcNAc modification. *Biochimica et Biophysica Acta*, **1800**, 122 (2010).

[171] C. Martinez-Fleites, M. S. Macauley, Y. He, D. L. Shen, D. J. Vocadlo, and G. J. Davies. Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. *Nature Structural & Molecular Biology*, **15**, 764 (2008).

[172] P. März, J. Stetefeld, K. Bendfeldt, C. Nitsch, J. Reinstein, R. L. Shoeman, B. Dimitriades-Schmutz, M. Schwager, D. Leiser, S. Özcan, U. Otten, and S. Ozbek. Ataxin-10 interacts with O-linked beta-N-acetylglucosamine transferase in the brain. *The Journal of Biological Chemistry*, **281**, 20263 (2006).

[173] M. L. Matthews, L. He, B. D. Horning, E. J. Olson, B. E. Correia, J. R. Yates, P. E. Dawson, and B. F. Cravatt. Chemoproteomic profiling and discovery of protein electrophiles in human cells. *Nature Chemistry*, **9**, 234 (2017).

[174] J. C. Maynard, A. L. Burlingame, and K. F. Medzihradszky. Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), A New Post-translational Modification in Mammals. *Molecular & Cellular Proteomics*, **15**, 3405 (2016).

[175] R. Mazars, A. Gonzalez-de Peredo, C. Cayrol, A.-C. Lavigne, J. L. Vogel, N. Ortega, C. Lacroix, V. Gautier, G. Huet, A. Ray, B. Monsarrat, T. M. Kristie, and J.-P. Girard. The THAP-Zinc Finger Protein THAP1 Associates with Coactivator HCF-1 and O-GlcNAc Transferase. *The Journal of Biological Chemistry*, **285**, 13364 (2010).

[176] D. A. McClain. Hexosamines as mediators of nutrient sensing and regulation in diabetes. *Journal of Diabetes and its Complications*, **16**, 72 (2002).

[177] P. N. McFadden and S. Clarke. Conversion of isoaspartyl peptides to normal peptides: implications for the cellular repair of damaged proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 2595 (1987).

[178] S. McNicholas, E. Potterton, K. S. Wilson, and M. E. M. Noble. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D*, **67**, 386 (2011).

[179] V. R. Meyer. *Pitfalls and Errors of HPLC in Pictures*, chapter 2.34, pages 110–111. Wiley, Weinheim, 3rd edition, 2013.

[180] J. Michaud, V. Praz, N. James Faresse, C. K. JnBaptiste, S. Tyagi, F. Schutz, and W. Herr. HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. *Genome Research*, **23**, 907 (2013).

[181] S. Miertuš, E. Scrocco, and J. Tomasi. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chemical Physics*, **55**, 117 (1981).

[182] S. Miertus and J. Tomasi. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chemical Physics*, **65**, 239 (1982).

[183] A. Morin, B. Eisenbraun, J. Key, P. C. Sanschagrin, M. A. Timony, M. Ottaviano, and P. Sliz. Collaboration gets the most out of software. *eLife*, **2**, e01456 (2013).

[184] S. A. Myers, S. Daou, E. B. Affar, and A. Burlingame. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments-a study of the O-GlcNAcylated protein Host Cell Factor C1. *Proteomics*, **13**, 982 (2013).

[185] J. M. Nalbone, N. Lahankar, L. Buissereth, and M. Raj. Glutamic Acid Selective Chemical Cleavage of Peptide Bonds. *Organic Letters*, **18**, 1186 (2016).

[186] H. D. Niall. [36] Automated edman degradation: The protein sequenator. *Methods in Enzymology*, **27**, 942 (1973).

[187] D. Nolte and U. Muller. Human O-GlcNAc transferase (OGT): genomic structure, analysis of splice variants, fine mapping in Xq13.1. *Mammalian Genome*, **13**, 62 (2002).

[188] O. Nørregaard Jensen. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology*, **8**, 33 (2004).

[189] P. B. O'Connor, J. J. Cournoyer, S. J. Pitteri, P. A. Chrisman, and S. A. McLuckey. Differentiation of aspartic and isoaspartic acids using electron transfer dissociation. *Journal of the American Society for Mass Spectrometry*, **17**, 15 (2006).

[190] M. A. Ondetti, A. Deer, J. T. Sheehan, and J. Pluscec. Side reactions in the synthesis of peptides containing the aspartylglycyl sequence. *Biochemistry*, **7**, 4069 (1968).

[191] R. F. Ortiz-Meoz, J. Jiang, M. B. Lazarus, M. Orman, J. Janetzko, C. Fan, D. Y. Duveau, Z.-W. Tan, C. J. Thomas, and S. Walker. A small molecule that inhibits OGT activity in cells. *ACS Chemical Biology*, **10**, 1392 (2015).

[192] S. Özcan, S. S. Andrali, and J. E. L. Cantrell. Modulation of transcription factor function by O-GlcNAc modification. *Biochimica et Biophysica Acta*, **1799**, 353 (2010).

[193] J. Painter and E. A. Merritt. TLSMDweb server for the generation of multi-group TLS models. *Journal of Applied Crystallography*, **39**, 109 (2006).

[194] K. K. Palaniappan and C. R. Bertozzi. Chemical Glycoproteomics. *Chemical Reviews*, **116**, 14277 (2016).

[195] J. Park, F. Lammers, W. Herr, and J.-J. Song. HCF-1 self-association via an interdigitated Fn3 structure facilitates transcriptional regulatory complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 17430 (2012).

[196] J. B. Parker, H. Yin, A. Vinckevicius, and D. Chakravarti. Host cell factor-1 recruitment to E2F-bound and cell-cycle-control genes is mediated by THAP11 and ZNF143. *Cell Reports*, **9**, 967 (2014).

[197] J. L. Pascual-ahuir, E. Silla, and I. Tuñon. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *Journal of Computational Chemistry*, **15**, 1127 (1994).

[198] S. Pathak, J. Alonso, M. Schimpl, K. Rafie, D. E. Blair, V. S. Borodkin, A. W. Schüttelkopf, O. Albarbarawi, and D. M. F. van Aalten. The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nature Structural & Molecular Biology*, **22**, 744 (2015).

[199] G. Pekkurnaz, J. C. Trinidad, X. Wang, D. Kong, and T. L. Schwarz. Glucose Regulates Mitochondrial Motility via Milton Modification by O-GlcNAc Transferase. *Cell*, **158**, 54 (2014).

[200] C. Prasad and A. Peterkofsky. Initiation by methionine of mouse immunoglobulin light chain containing NH-2terminal pyroglutamic acid. *The Journal of Biological Chemistry*, **250**, 171 (1975).

[201] Z. Qin, A. Dimitrijevic, and D. W. Aswad. Accelerated protein damage in brains of PIMT+/- mice; a possible model for the variability of cognitive decline in human aging. *Neurobiology of Aging*, **36**, 1029 (2015).

[202] J. L. Radkiewicz, H. Zipse, and S. Clarke. Accelerated racemization of aspartic acid and asparagine residues via succinimide intermediates: an ab initio theoretical exploration of mechanism. *Journal of the American Chemical Society*, **118**, 9148 (1996).

[203] S. T. Rao and M. G. Rossmann. Comparison of super-secondary structures in proteins. *Journal of Molecular Biology*, **76**, 241 (1973).

[204] J. P. Richard, G. Williams, A. C. O'Donoghu, and T. L. Amyes. Formation and Stability of Enolates of Acetamide and Acetate Anion: An Eigen Plot for Proton Transfer at α-Carbonyl Carbon. *Journal of the American Chemical Society*, **124**, 2957 (2002).

[205] I.-H. Riu, I.-S. Shin, and S.-I. Do. Sp1 modulates ncOGT activity to alter target recognition and enhanced thermotolerance in E. coli. *Biochemical and Biophysical Research Communications*, **372**, 203 (2008).

[206] E. P. Roquemore, M. R. Chevrier, R. J. Cotter, and G. W. Hart. Dynamic O-GlcNAcylation of the small heat shock protein αB-crystallin. *Biochemistry*, **35**, 3578 (1996).

[207] H.-B. Ruan, M. O. Dietrich, Z.-W. Liu, M. R. Zimmer, M.-D. Li, J. P. Singh, K. Zhang, R. Yin, J. Wu, T. L. Horvath, and X. Yang. O-GlcNAc Transferase Enables AgRP Neurons to Suppress Browning of White Fat. *Cell*, **159**, 306 (2014).

[208] H.-B. Ruan, X. Han, M.-D. Li, J. P. Singh, K. Qian, S. Azarhoush, L. Zhao, A. M. Bennett, V. T. Samuel, J. Wu, J. R. Yates, and X. Yang. O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1α stability. *Cell Metabolism*, **16**, 226 (2012).

[209] J. L. Sacoman, R. Y. Dagda, A. R. Burnham-Marusich, R. K. Dagda, and P. M. Berninsone. Mitochondrial O-GlcNAc transferase (mOGT) regulates mitochondrial structure, function and survival in HeLa cells. *The Journal of Biological Chemistry*, **292**, 4499 (2017).

[210] M. M. Savitski, F. Kjeldsen, M. L. Nielsen, and R. A. Zubarev. Hydrogen rearrangement to and from radical z fragments in electron capture dissociation of peptides. *Journal of the American Society for Mass Spectrometry*, **18**, 113 (2007).

[211] G. Scherer, M. L. Kramer, and M. Schutkowski. Barriers to rotation of secondary amide peptide bonds. *Journal of the American Chemical Society*, **120**, 5568 (1998).

[212] S. Schilling, C. Wasternack, and H.-U. Demuth. Glutaminyl cyclases from animals and plants: a case of functionally convergent protein evolution. *Biological Chemistry*, **389**, 1 (2008).

[213] M. Schimpl, X. Zheng, V. S. Borodkin, D. E. Blair, A. T. Ferenbach, A. W. Schüttelkopf, I. Navratilova, T. Aristotelous, O. Albarbarawi, D. A. Robinson, M. A. Macnaughtan, and D. M. F. van Aalten. O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nature Chemical Biology*, **8**, 969 (2012).

[214] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[215] M. Scigelova, M. Hornshaw, A. Giannakopulos, and A. Makarov. Fourier Transform Mass Spectrometry. *Molecular & Cellular Proteomics*, **10**, M111.009431 (2011).

[216] J. W. Scott and J. S. Oakhill. The sweet side of AMPK signaling: regulation of GFAT1. *The Biochemical journal*, **474**, 1289 (2017).

[217] D. L. Shen, T. M. Gloster, S. A. Yuzwa, and D. J. Vocadlo. Insights into O-Linked N-Acetylglucosamine ([0-9]O-GlcNAc) Processing and Dynamics through Kinetic Analysis of O-GlcNAc Transferase and O-GlcNAcase Activity on Protein Substrates. *The Journal of Biological Chemistry*, **287**, 15395 (2012).

[218] Y. Shi. Serine/Threonine Phosphatases: Mechanism through Structure. *Cell*, **139**, 468 (2009).

[219] Y.-P. Shih, C.-C. Chou, Y.-L. Chen, K.-F. Huang, and A. H. J. Wang. Linked Production of Pyroglutamate-Modified Proteins via Self-Cleavage of Fusion Tags with TEV Protease and Autonomous N-Terminal Cyclization with Glutaminyl Cyclase In Vivo. *PLOS ONE*, **9**, e94812 (2014).

[220] S.-H. Shin, D. C. Love, and J. A. Hanover. Elevated O-GlcNAc-dependent signaling through inducible mOGT expression selectively triggers apoptosis. *Amino Acids*, **40**, 885 (2010).

[221] H. Steen and M. Mann. The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology*, **5**, 699 (2004).

[222] R. C. Stephenson and S. Clarke. Succinimide Formation From Aspartyl and Asparaginyl Peptides as a Model for the Spontaneous Degradation of Proteins. *The Journal of Biological Chemistry*, **264**, 6164 (1989).

[223] L. J. Stern, J. H. Brown, T. S. Jardetzky, J. C. Gorga, R. G. Urban, J. L. Strominger, and D. C. Wiley. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*, **368**, 215 (1994).

[224] S. Stern and W. Herr. The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeo domain: evidence for a homeo domain recognition subdomain. *Genes & Development*, **5**, 2555 (1991).

[225] C. L. Stevenson, T. D. Williams, R. J. Anderegg, and R. T. Borchardt. Identification and quantitation of tetrapeptide deamidation products by mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, **10**, 567 (1992).

[226] K. Tanabe, A. Taniguchi, T. Matsumoto, K. Oisaki, Y. Sohma, and M. Kanai. Asparagine-selective cleavage of peptide bonds through hypervalent iodine-mediated Hofmann rearrangement in neutral aqueous solution. *Chemical Science*, **5**, 2747 (2014).

[227] Z. Tao, P. Gao, and H.-w. Liu. Identification of the ADP-Ribosylation Sites in the PARP-1 Automodification Domain: Analysis and Implications. *Journal of the American Chemical Society*, **131**, 14258 (2009).

[228] I. Terashima, A. Koga, and H. Nagai. Identification of deamidation and isomerization sites on pharmaceutical recombinant antibody using H218O. *Analytical Biochemistry*, **368**, 49 (2007).

[229] A. P. Tofteng, K. K. Sørensen, K. W. Conde-Frieboes, T. Hoeg-Jensen, and K. J. Jensen. Fmoc Solid-Phase Synthesis of C-Terminal Peptide Thioesters by Formation of a Backbone Pyroglutamyl Imide Moiety. *Angewandte Chemie International Edition*, **48**, 7411 (2009).

[230] C. R. Torres and G. W. Hart. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *The Journal of Biological Chemistry*, **259**, 3308 (1984).

[231] R. Trapannone, D. Mariappa, A. T. Ferenbach, and D. M. F. van Aalten. Nucleocytoplasmic human O-GlcNAc transferase is sufficient for O-GlcNAcylation of mitochondrial proteins. *Biochemical Journal*, **473**, 1693 (2016).

[232] R. R. Traxinger and S. Marshall. Coordinated regulation of glutamine:fructose-6-phosphate amidotransferase activity by insulin, glucose, and glutamine. Role of hexosamine biosynthesis in enzyme regulation. *The Journal of Biological Chemistry*, **266**, 10148 (1991).

[233] F. Turecek. NC αBond Dissociation Energies and Kinetics in Amide and Peptide Radicals. Is the Dissociation a Non-ergodic Process? *Journal of the American Chemical Society*, **125**, 5954 (2003).

[234] I. Tvaroška, S. Kozmon, M. Wimmerová, and J. Koča. Substrate-Assisted Catalytic Mechanism of O-GlcNAc Transferase Discovered by Quantum Mechanics/Molecular Mechanics Investigation. *Journal of the American Chemical Society*, **134**, 15563 (2012).

[235] D. R. Twardzik and A. Peterkofsky. Glutamic acid as a precursor to N-terminal pyroglutamic acid in mouse plasmacytoma protein (protein synthesis-initiation-immunoglobulins-pyrrolidone carboxylic acid). *Proceedings of the National Academy of Sciences of the United States of America*, **69**, 274 (1972).

[236] S. Tyagi, A. L. Chabes, J. Wysocka, and W. Herr. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Molecular Cell*, **27**, 107 (2007).

[237] K. Vaidyanathan and L. Wells. Multiple Tissue-specific Roles for the O-GlcNAc Post-translational Modification in the Induction of and Complications Arising from Type II Diabetes. *The Journal of Biological Chemistry*, **289**, 34466 (2014).

[238] B. van Loo, H. P. Permentier, J. Kingma, H. Baldascini, and D. B. Janssen. Inactivation of epoxide hydrolase by catalysis-induced formation of isoaspartate. *FEBS Letters*, **582**, 1581 (2008).

[239] A. Varki, R. Cummings, J. Esko, H. Freeze, P. Stanley, C. Bertozzi, G. Hart, and M. Etzler, editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2nd edition, 2009.

[240] E. Verdin. NAD^+ in aging, metabolism, and neurodegeneration. *Science*, **350**, 1208 (2015).

[241] R. Vidya, M. Eggen, S. K. Nair, G. I. Georg, and R. H. Himes. Synthesis of Cryptophycins via an N-Acyl-β-lactam Macrolactonization. *The Journal of Organic Chemistry*, **68**, 9687 (2003).

[242] V. Vigneswara, J. D. Lowenson, C. D. Powell, M. Thakur, K. Bailey, S. Clarke, D. E. Ray, and W. G. Carter. Proteomic identification of novel substrates of a protein isoaspartyl methyltransferase repair enzyme. *The Journal of Biological Chemistry*, **281**, 32619 (2006).

[243] D. J. Vocadlo. O-GlcNAc processing enzymes: catalytic mechanisms, substrate specificity, and enzyme regulation. *Current Opinion in Chemical Biology*, **16**, 488 (2012).

[244] J. Vogel and T. Kristie. The Dynamics of HCF-1 Modulation of Herpes Simplex Virus Chromatin during Initiation of Infection. *Viruses*, **5**, 1272 (2013).

[245] J. L. Vogel and T. M. Kristie. Autocatalytic proteolysis of the transcription factor-coactivator C1 (HCF): a potential role for proteolytic regulation of coactivator function. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9425 (2000).

[246] J. L. Vogel and T. M. Kristie. The novel coactivator C1 (HCF) coordinates multiprotein enhancer formation and mediates transcription activation by GABP. *The EMBO Journal*, **19**, 683 (2000).

[247] S. Vyas, I. Matic, L. Uchima, J. Rood, R. Zaja, R. T. Hay, I. Ahel, and P. Chang. Family-wide analysis of poly(ADP-ribose) polymerase activity. *Nature Communications*, **5**, 4426 (2014).

[248] K. Weber and W. Konigsberg. Amino Acid Sequence of the f_2 Coat Protein. *The Journal of Biological Chemistry*, **242**, 3563 (1967).

[249] C. Weigert, K. Klopfer, C. Kausch, K. Brodbeck, M. Stumvoll, H. U. Häring, and E. D. Schleicher. Palmitate-Induced Activation of the Hexosamine Pathway in Human Myotubes. *Diabetes*, **52**, 650 (2003).

[250] J. M. Wells and S. A. McLuckey. Collision-induced dissociation (CID) of peptides and proteins. *Methods in Enzymology*, **402**, 148 (2005).

[251] L. Wells, K. Vosseller, and G. W. Hart. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science*, **291**, 2376 (2001).

[252] M. Wilhelm, J. Schlegl, H. Hahne, A. M. Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J.-H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, and B. Kuster. Mass-spectrometry-based draft of the human proteome. *Nature*, **509**, 582 (2014).

[253] A. C. Wilson, M. Boutros, K. M. Johnson, and W. Herr. HCF-1 amino- and carboxy-terminal subunit association through two separate sets of interaction modules: involvement of fibronectin type 3 repeats. *Molecular and Cellular Biology*, **20**, 6721 (2000).

[254] A. C. Wilson, K. LaMarco, M. G. Peterson, and W. Herr. The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell*, **74**, 115 (1993).

[255] A. C. Wilson, M. G. Peterson, and W. Herr. The HCF repeat is an unusual proteolytic cleavage signal. *Genes & Development*, **9**, 2445 (1995).

[256] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D*, **67**, 235 (2011).

[257] S. G. Withers and G. J. Davies. Glycobiology: The case of the missing base. *Nature Chemical Biology*, **8**, 952 (2012).

[258] M. Worth, H. Li, and J. Jiang. Deciphering the Functions of Protein O-GlcNAcylation with Chemistry. *ACS Chemical Biology*, **12**, 326 (2017).

[259] J. O. Wrabl and N. V. Grishin. Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *Journal of Molecular Biology*, **314**, 365 (2001).

[260] J. Wysocka and W. Herr. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in Biochemical Sciences*, **28**, 294 (2003).

[261] J. Wysocka, M. P. Myers, C. D. Laherty, R. N. Eisenman, and W. Herr. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes & Development*, **17**, 896 (2003).

[262] P. Xiao and J. P. Capone. A cellular factor binds to the herpes simplex virus type 1 transactivator Vmw65 and is required for Vmw65-dependent protein-DNA complex assembly with Oct-1. *Molecular and Cellular Biology*, **10**, 4974 (1990).

[263] C. C. Yang and R. B. Merrifield. The beta-phenacyl ester as a temporary protecting group to minimize cyclic imide formation during subsequent treatment of aspartyl peptides with HF. *The Journal of Organic Chemistry*, **41**, 1032 (1976).

[264] X. Yang, P. P. Ongusaha, P. D. Miles, J. C. Havstad, F. Zhang, W. V. So, J. E. Kudlow, R. H. Michell, J. M. Olefsky, S. J. Field, and R. M. Evans. Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature*, **451**, 964 (2008).

[265] X. Yang, F. Zhang, and J. E. Kudlow. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell*, **110**, 69 (2002).

[266] W. Yi, P. M. Clark, D. E. Mason, M. C. Keenan, C. Hill, W. A. Goddard, E. C. Peters, E. M. Driggers, and L. C. Hsieh-Wilson. Phosphofructokinase 1 Glycosylation Regulates Cell Growth and Metabolism. *Science*, **337**, 975 (2012).

[267] S. A. Yuzwa and D. J. Vocadlo. O-GlcNAc and neurodegeneration: biochemical mechanisms and potential roles in Alzheimer's disease and beyond. *Chemical Society Reviews*, **43**, 6839 (2014).

[268] N. E. Zachara and G. W. Hart. O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochimica et Biophysica Acta*, **1673**, 13 (2004).

[269] N. E. Zachara and G. W. Hart. O-GlcNAc modification: a nutritional sensor that modulates proteasome function. *Trends in Cell Biology*, **14**, 218 (2004).

[270] N. E. Zachara, N. O'Donnell, W. D. Cheung, J. J. Mercer, J. D. Marth, and G. W. Hart. Dynamic O-GlcNAc Modification of Nucleocytoplasmic Proteins in Response to Stress: A survival response of mammalian cells. *The Journal of Biological Chemistry*, **279**, 30133 (2004).

[271] Z. U. Zargar and S. Tyagi. Role of Host Cell Factor-1 in cell cycle regulation. *Transcription*, **3**, 187 (2014).

[272] N. Zeytuni and R. Zarivach. Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module. *Structure*, **20**, 397 (2012).

[273] F. Zhang, K. Su, X. Yang, D. B. Bowe, A. J. Paterson, and J. E. Kudlow. O-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell*, **115**, 715 (2003).

[274] Q. Zhang, X. Liu, W. Gao, P. Li, J. Hou, J. Li, and J. Wong. Differential Regulation of the Ten-Eleven Translocation (TET) Family of Dioxygenases by O-Linked β- N-Acetylglucosamine Transferase (OGT). *The Journal of Biological Chemistry*, **289**, 5986 (2014).

[275] Y. Zhang, J. Wang, M. Ding, and Y. Yu. Site-specific characterization of the Asp- and Glu-ADP-ribosylated proteome. *Nature Methods*, **10**, 981 (2013).

[276] P. Zhou, Z. Wang, X. Yuan, C. Zhou, L. Liu, X. Wan, F. Zhang, X. Ding, C. Wang, S. Xiong, Z. Wang, J. Yuan, Q. Li, and Y. Zhang. Mixed lineage leukemia 5 (MLL5) protein regulates cell cycle progression and E2F1-responsive gene expression via association with host cell factor-1 (HCF-1). *The Journal of Biological Chemistry*, **288**, 17532 (2013).

[277] J. X. Zhu, H. A. Doyle, M. J. Mamula, and D. W. Aswad. Protein repair in the brain, proteomic analysis of endogenous substrates for protein L-isoaspartyl methyltransferase in mouse brain. *The Journal of Biological Chemistry*, **281**, 33802 (2006).

[278] Y. Zhu, T.-W. Liu, S. Cecioni, R. Eskandari, W. F. Zandberg, and D. J. Vocadlo. O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. *Nature Chemical Biology*, **11**, 319 (2015).

[279] K. O. Zhurov, L. Fornelli, M. D. Wodrich, Ü. A. Laskay, and Y. O. Tsybin. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chemical Society Reviews*, **42**, 5014 (2013).

[280] D. Zibrova, F. Vandermoere, O. Göransson, M. Peggie, K. V. Mariño, A. Knierim, K. Spengler, C. Weigert, B. Viollet, N. A. Morrice, K. Sakamoto, and R. Heller. GFAT1 phosphorylation by AMPK promotes VEGF-induced angiogenesis. *Biochemical Journal*, **474**, 983 (2017).

[281] R. A. Zubarev. Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrometry Reviews*, **22**, 57 (2003).

[282] R. A. Zubarev. Electron-capture dissociation tandem mass spectrometry. *Current Opinion in Biotechnology*, **15**, 12 (2004).

[283] R. A. Zubarev and N. L. Kelleher. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society*, **120**, 3265 (1998).

[284] R. A. Zubarev, N. A. Kruger, E. K. Fridriksson, M. A. Lewis, D. M. Horn, B. K. Carpenter, and F. W. McLafferty. Electron Capture Dissociation of Gaseous Multiply-Charged Proteins Is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. *Journal of the American Chemical Society*, **121**, 2857 (1999).