



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

Year : 2015

## Mechanisms of OGT-mediated HCF-1 protein maturation

Tanja Bhuiyan

Tanja Bhuiyan, 2015, Mechanisms of OGT-mediated HCF-1 protein maturation

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : [urn:nbn:ch:serval-BIB\\_37447C1358252](http://nbn:ch:serval-BIB_37447C1358252)

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



**UNIL** | Université de Lausanne

Faculté de biologie  
et de médecine

**Centre Intégréatif de Génomique**

*Mechanisms of OGT-mediated  
HCF-1 protein maturation*

**Thèse de doctorat ès sciences de la vie (PhD)**

présentée à la

Faculté de biologie et de médecine  
de l'Université de Lausanne

par

**Tanja BHUIYAN**

Biologiste diplômée de l'Université de Stuttgart, Allemagne

**Jury**

Prof. Grégoire Millet, président  
Prof. Winship Herr, directeur de thèse  
Prof. Ivan Stamenkovic, expert  
Prof. Margot Thome-Miazza, experte  
Prof. Chad Slawson, expert

Lausanne, 2015



UNIL | Université de Lausanne  
Faculté de biologie  
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

# Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

<i>Président</i>	Monsieur Prof. Grégoire Millet
<i>Directeur de thèse</i>	Monsieur Prof. Winship Herr
<i>Experts</i>	Madame Prof. Margot Thome
	Monsieur Prof. Chad Slawson
	Monsieur Prof. Ivan Stamenkovic

le Conseil de Faculté autorise l'impression de la thèse de

**Madame Tanja Bhuiyan**

Biologiste diplômée de l' Université de Stuttgart, Allemagne

intitulée

*Mechanisms of OGT-mediated  
HCF-1 protein maturation*

Lausanne, le 10 juillet 2015

pour La Doyenne  
de la Faculté de biologie et de médecine

Prof. Grégoire Millet

# Acknowledgements

First and foremost, I am grateful to those who helped me to complete this work. To my thesis supervisor, **Winship Herr**, for the guidance he offered me. Not only he taught me the practical and theoretical lessons to become a scientist, but also the *values* to become a *good* scientist. To the former and present members of the Herr group, especially **Francesca Capotosti** (I had the honor to take over her project), **Fabienne Lammers** (the good soul of the lab), and **Vaibhav Kapuria** (my five-year collaborator). I also thank **Vincent Zoete** and **Ute Roehrig** (for the molecular modeling), **Patrice Waridel** (for his expertise in proteomics), **Viviane Praz** (for her bioinformatics support), **Nathalie Clerc** (for her constant administrative support), **Tanguy Corre** (for translating text in French), and **Riccardo Sinisi** (for answering me all questions about organic chemistry).

I am particularly thankful to my friends, without whom none of this would have been possible: **Karin Nestorow** (my childhood friend; for “being there”), **Lioba Courth** (for all the campus parties and the years of friendship), **Maria Laura Jeanrenaud** (for her positive energy), **Ilaria Mosconi** (for unbreakable friendship), **Julia Cajan** (for the discussions about life), **Lisa Pollaro** (for her generosity), **Mauro Manassi** (for the trips together and for iMovie), **Friederike Bienert** (for her good heart and some craziness), **Alexandra Svoboda** (for being in my life), **Rosamaria Cannavo** (for her kindness and her friendship), **Iris Finci** (for her help and support), and **Evelina Thunnel** (for being my special Swedish friend).

My greatest thanks go to **Andrea Prunotto** for having walked with me this last path during my thesis, with his good mood and constant support, with his great mind for science and love for life, and with his faith in my abilities.



# Résumé pour un grand public

OGT (O-GlcNAc transferase) est une enzyme polyvalente et unique : elle peut non seulement fixer un sucre sur certaines protéines cellulaires (c'est la glycosylation; une modification de protéines, qui est la fonction normale d'OGT) mais elle est également capable de couper une protéine en deux. OGT est donc aussi une protéase !

La seule protéine connue, coupée par OGT s'appelle HCF-1, qui est un important régulateur de la division cellulaire. HCF-1 peut agir sur des gènes cibles (HCF-1 est donc un régulateur transcriptionnel) et assurer la progression du cycle cellulaire afin que la cellule se divise en deux cellules saines avec un noyau chacune. En effet, si OGT ne coupe pas HCF-1 en deux, HCF-1 ne peut pas effectuer correctement ses fonctions et par la suite, la cellule, au lieu de se diviser, devient une seule cellule –éventuellement cancéreuse– avec deux noyaux. C'est pourquoi, l'étude des mécanismes qui permettent à OGT de couper HCF-1 sans erreurs est fondamentale pour comprendre comment l'intégrité des cellules humaines est maintenue.

Dans cette thèse, j'ai identifié les prérequis de la protéine HCF-1 qui permettent à OGT d'effectuer sa fonction de protéase. Curieusement, parmi des centaines d'acides aminés de HCF-1, j'en ai identifié un seul – un glutamate – qui est crucial pour la coupure. De plus, j'ai identifié d'autres éléments dans la protéine HCF-1 qui promeuvent une coupure efficace. En effet, OGT s'attache à ces éléments de HCF-1 et favorise leur glycosylation. Donc, OGT garantit la maturation d'HCF-1 au travers de différents mécanismes et assure l'intégrité cellulaire.



# Abstract

Post-translational protein modifications are crucial for many fundamental cellular and extracellular processes and greatly contribute to the complexity of organisms. Human HCF-1 is a transcriptional co-regulator that undergoes complex protein maturation involving reversible and irreversible post-translational modifications. Upon synthesis as a large precursor protein, HCF-1 undergoes extensive reversible glycosylation with  $\beta$ -N-acetylglucosamine giving rise to O-linked- $\beta$ -N-acetylglucosamine (O-GlcNAc) modified serines and threonines. HCF-1 also undergoes irreversible site-specific proteolysis, which is important for one of HCF-1's major functions — the regulation of the cell-division cycle. HCF-1 O-GlcNAcylation and site-specific proteolysis are both catalyzed by a single enzyme with an unusual dual enzymatic activity, the O-GlcNAc transferase (OGT). HCF-1 is cleaved by OGT at any of six highly conserved 26 amino acid repeated sequences (HCF-1<sub>PRO</sub> repeats), but the mechanisms and the substrate requirements for OGT-mediated cleavage are not understood. In the present work, I characterized substrate requirements for OGT-mediated cleavage and O-GlcNAcylation of HCF-1. I identified key elements within the HCF-1<sub>PRO</sub>-repeat sequence that are important for proteolysis. Remarkably, an invariant single amino acid side-chain within the HCF-1<sub>PRO</sub>-repeat sequence displays particular OGT-binding properties and is essential for proteolysis. Additionally, I characterized substrate requirements for proteolysis outside of the HCF-1<sub>PRO</sub> repeat and identified a novel, highly O-GlcNAcylated OGT-binding sequence that enhances cleavage of the first HCF-1<sub>PRO</sub> repeat. These results link OGT association and its O-GlcNAcylation activities to HCF-1<sub>PRO</sub>-repeat proteolysis.



# Résumé

La modification post-traductionnelle des protéines est cruciale pour plusieurs processus cellulaires et extra-cellulaires fondamentaux et contribue significativement à la complexité des organismes. HCF-1 chez l'humain est un co-régulateur transcriptionnel qui subit une maturation protéique complexe, impliquant des modifications post-traductionnelles réversibles et irréversibles. Après la synthèse de la protéine comme précurseur, HCF-1 subit une glycosylation extensive et réversible avec  $\beta$ -N-acetylglucosamine qui résulte en sérines et thréonines modifiées par O-linked- $\beta$ -N-acetylglucosamine (O-GlcNAc). HCF-1 subit également une protéolyse irréversible et site-spécifique, importante pour une des fonctions majeures d'HCF1 – la régulation du cycle cellulaire. L'O-GlcNAcylation et la protéolyse d'HCF-1 sont toutes deux catalysées par une seule enzyme avec une peu commune double activité enzymatique, l'O-GlcNAc transférase (OGT). HCF-1 est clivée par OGT sur six séquences de répétition ultra-conservées (nommées HCF-1<sub>PRO</sub> répétitions), chacune consistant en 26 acides aminés. Les mécanismes et les substrats requis pour le clivage effectué par OGT ne sont pas compris. Dans le travail présenté ici, j'ai caractérisé les substrats requis pour le clivage et l'O-GlcNAcylation d'HCF-1 effectués par OGT. J'ai identifié des éléments clefs dans la séquence de l'HCF-1<sub>PRO</sub> répétition qui sont importants pour la protéolyse. De manière remarquable, un seul chaîne d'acide aminé invariant dans la séquence de l'HCF-1<sub>PRO</sub> répétition présente des propriétés de liaison à OGT particulières et est absolument essentiel pour la protéolyse. J'ai également caractérisé les substrats requis pour la protéolyse en dehors de l'HCF-1<sub>PRO</sub> répétition et identifié une nouvelle séquence de liaison à OGT très O-GlcNAcylatée qui augmente le clivage de la première HCF-1<sub>PRO</sub> répétition. Ces résultats lient l'association d'OGT et ses activités d'O-GlcNAcylation à la protéolyse d'HCF-1<sub>PRO</sub> répétition.



# Contents

**Acknowledgements**

**Résumé pour un grand public**

**Abstract**

**Résumé**

**List of Figures.....v**

**Chapter I: Introduction.....1**

I.1 The human proteome is complex ..... 1

I.2 Post-translational modifications.....2

I.2.1 The different forms of PTMs ..... 2

I.2.2 Cross-talk between PTMs..... 4

I.2.3 Evolution and conservation of PTMs ..... 5

I.3 Proteolysis .....5

I.3.1 Human proteases ..... 5

I.3.2 Protease signaling ..... 7

I.3.3 Site-specific proteolysis in cell-cycle progression ..... 7

I.4 Glycosylation .....10

I.4.1 N-linked glycosylation ..... 12

I.4.2 O-linked glycosylation..... 12

I.4.3 Biological role of glycans ..... 12

I.5 O-GlcNAcylation .....13

I.5.1 Cross-talk between O-GlcNAcylation and other PTMs..... 13

I.5.2 The role of O-GlcNAc in transcription ..... 15

I.5.3 UDP-GlcNAc and the role of O-GlcNAc in metabolism ..... 15

I.5.4 Functional mechanisms of O-GlcNAcylation ..... 17

I.5.5 O-GlcNAc identification on proteins..... 17

I.6 OGT.....17

I.6.1 Structure of human OGT ..... 18

I.6.2 Substrate specificity of OGT ..... 18

I.6.3 Mechanism of O-GlcNAcylation..... 21

I.7 HCF-1 .....22

I.7.1 HCF-1 serves as a binding platform for proteins involved in transcription..... 24

I.7.2 HCF-1 proteolytic processing is important for proper cell-cycle progression..... 24

I.7.3 OGT catalyzes site-specific proteolysis of the HCF-1<sub>PRO</sub> repeats ..... 27

I.7.4 Functional versatility of OGT–HCF-1 interactions ..... 29

I.7.5 Evolution of HCF-1 site-specific proteolysis ..... 29

I.7.6 HCF-1 in human disease ..... 31

I.8 Conclusions .....32

I.9 Thesis scope .....	32
<b>Chapter II: HCF-1<sub>PRO</sub>-repeat requirements for OGT-mediated proteolysis .....</b>	<b>33</b>
<b>Introduction .....</b>	<b>33</b>
<b>Results .....</b>	<b>33</b>
II.1 HCF-1 <sub>PRO</sub> -repeat proteolysis requires OGT and intact UDP-GlcNAc .....	34
II.2 E10 is crucial for HCF-1 <sub>PRO</sub> -repeat proteolysis .....	38
II.3 Serine at position 10 of the HCF-1 <sub>PRO</sub> repeat activates HCF-1 <sub>PRO</sub> -repeat O-GlcNAcylation .....	41
II.4 HCF-1 <sub>PRO</sub> -repeat-OGT association is sensitive to UDP-GlcNAc .....	44
II.5 Proper HCF-1 <sub>PRO</sub> -repeat-OGT interactions require intact UDP-GlcNAc .....	47
II.6 E10 in the HCF-1 <sub>PRO</sub> -repeat cleavage region displays unfavorable interactions with the OGT-UDP-GlcNAc complex .....	49
II.7 Strains caused by E10 in the OGT catalytic domain are specific to the glutamate side-chain structure .....	52
II.8 Refinement of the HCF-1 <sub>PRO</sub> -repeat bipartite structure .....	56
II.9 Residue T14 in the threonine region is important for interactions with residues in the OGT TPR domain .....	60
<b>Discussion .....</b>	<b>62</b>
The HCF-1 <sub>PRO</sub> repeat contains a flexible region (Hinge region) .....	62
HCF-1 <sub>PRO</sub> -repeat proteolysis requires intact UDP-GlcNAc .....	62
Proposed initial step for HCF-1 <sub>PRO</sub> -repeat proteolysis .....	63
The analog UDP-5SGlcNAc changes the HCF-1 <sub>PRO</sub> -repeat-OGT binding mode drastically .....	66
The HCF-1 <sub>PRO</sub> -repeat threonine region is important for HCF-1 <sub>PRO</sub> -repeat-OGT association .....	66
A glycosyltransferase coopted as protease? .....	67
<b>Chapter III: Requirements for HCF-1 proteolysis outside of the HCF-1<sub>PRO</sub> repeat .....</b>	<b>68</b>
<b>Introduction .....</b>	<b>68</b>
<b>Results .....</b>	<b>68</b>
III.1 Sequences lying outside of the HCF-1 <sub>PRO</sub> repeat promote cleavage .....	68
III.1.1 Subdivision of HCF-1 <sub>PRO</sub> sequences for deletion analysis .....	71
III.1.2 <i>In vitro</i> HCF-1 cleavage assay with <i>in vitro</i> transcribed and translated substrates .....	73
III.1.3 <i>In vivo</i> HCF-1 cleavage assay in HEK 293 cells .....	75
III.1.4 <i>In vitro</i> HCF-1 cleavage assay with bacterially synthesized substrates and OGT .....	77
III.2 Region II can activate a heterologous HCF-1 <sub>PRO</sub> -repeat substrate for cleavage and O-GlcNAcylation .....	80
III.3 Region II activity on HCF-1 <sub>PRO</sub> -repeat cleavage is sequence specific .....	82
III.4 Region II activity displays positional specificity .....	84
III.5 Region II activity is dependent on Region IV .....	86
III.6 Regions I, II, and III together potentially affect cleavage of HCF-1 <sub>PRO</sub> repeats 1, 2, and 3 .....	88

<b>Discussion .....</b>	<b>90</b>
HCF-1 <sub>PRO</sub> -repeat cleavage can get enhanced by flanking HCF-1 sequences .....	90
Region II activity on HCF-1 <sub>PRO</sub> -repeat cleavage is complex.....	91
HCF-1 cleavage activity correlates with HCF-1 O-GlcNAcylation efficiency .....	92
<b>Chapter IV: HCF-1–OGT-association and O-GlcNAcylation activities and linkage to</b>	
<b>HCF-1<sub>PRO</sub>-repeat cleavage .....</b>	<b>93</b>
<b>Introduction .....</b>	<b>93</b>
<b>Results .....</b>	<b>93</b>
IV.1 Region II enhances HCF-1–OGT association.....	94
IV.2 Region II represents an independent OGT-binding sequence.....	96
IV.3 Efficient Region II–OGT binding requires the OGT TPR domain.....	100
IV.4 Region II contains a cluster of O-GlcNAcylation sites.....	102
IV.5 Region II placed C-terminal of the HCF-1 <sub>PRO</sub> repeat is differentially O-GlcNAcylated .....	105
IV.6 Subdivision of Region II leads to reduced cleavage activities of the sub-regions <i>in vivo</i> .....	107
IV.7 HCF-1 <sub>PRO</sub> -repeat proteolysis does not interfere with Region II O-GlcNAcylation .....	110
IV.8 Region II enhances HCF-1 <sub>PRO</sub> -repeat proteolysis independently of its O-GlcNAcylation	
status .....	112
IV.9 HCF-1 O-GlcNAcylation levels are independent of OGT-mediated HCF-1 <sub>PRO</sub> -repeat	
processing.....	114
<b>Discussion .....</b>	<b>117</b>
Region II represents a novel OGT-binding sequence .....	117
Region II is rich in O-GlcNAcylated serines and threonines.....	118
Region II O-GlcNAcylation and HCF-1 <sub>PRO</sub> -repeat proteolysis are independent .....	119
Proposed model for Region II activity on HCF-1 <sub>PRO</sub> -repeat cleavage .....	120
<b>Chapter V: Conclusions and Perspectives .....</b>	<b>122</b>
V.1 The HCF-1 <sub>PRO</sub> repeats are highly-specific protease recognition sites containing three	
different elements important for cleavage .....	122
The HCF-1 <sub>PRO</sub> -repeat threonine region .....	122
The HCF-1 <sub>PRO</sub> -repeat cleavage region .....	123
The HCF-1 <sub>PRO</sub> -repeat Hinge region .....	124
V.2 A single invariant amino acid side-chain is essential for proteolysis.....	124
V.3 How did OGT-induced cleavage of the HCF-1 <sub>PRO</sub> repeats arise?.....	125
V.4 Efficient HCF-1 <sub>PRO</sub> -repeat cleavage requires HCF-1 sequences flanking the HCF-1 <sub>PRO</sub>	
repeat .....	127
V.5 The two faces of OGT .....	128
V.6 Final remarks .....	130
<b>Materials and Methods.....</b>	<b>131</b>
Plasmid constructs and DNA template preparation.....	131
Cell culture and plasmid transfections .....	132
Bacterial protein expression .....	132

<i>In vitro</i> HCF-1–OGT binding assay.....	133
HCF-1 cleavage and O-GlcNAcylation assays.....	133
Denaturing immunoprecipitation.....	134
Immunoblot and antibodies.....	134
Sample preparation for O-GlcNAcylation analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS).....	135
LC-MS/MS analysis.....	135
Molecular dynamics simulations.....	136
Analysis of crystal structures and molecular dynamics simulations.....	137
<b>References.....</b>	<b>138</b>
<b>Appendix.....</b>	<b>147</b>

# List of Figures

## Chapter I

Figure I-1: Occurrence of experimentally detected post-translational modifications (PTMs).....	3
Figure I-2: Cross-talk between PTMs of the N-terminal tail of histone H3. ....	4
Figure I-3: The two major proteolysis mechanisms of mammalian proteases. ....	6
Figure I-4: Human HCF-1 and MLL undergo proteolytic maturation to regulate the cell cycle. ....	9
Figure I-5: Mammalian glycan biosynthetic pathways. ....	11
Figure I-6: Cross-talk between O-GlcNAcylation and O-phosphorylation. ....	14
Figure I-7: The Hexosamine Biosynthetic Pathway (HBP) and its end product UDP-GlcNAc. ....	16
Figure I-8: Structure of human OGT. ....	20
Figure I-9: Proposed ordered bi-bi kinetic mechanism for OGT O-GlcNAcylation. ....	21
Figure I-10: The VP16 induced complex (VIC) and the viral cycle of Herpes simplex virus (HSV).....	23
Figure I-11: HCF-1 maturation and regulation of the human cell-division cycle. ....	26
Figure I-12: The HCF-1 <sub>PRO</sub> repeats and the proposed bipartite interaction with OGT.....	28
Figure I-13: The HCF protein family in vertebrates and invertebrates. ....	30

## Chapter II

Figure II-1: Structure of HCF-1 proteins. ....	35
Figure II-2: HCF-1 <sub>PRO</sub> -repeat cleavage is sensitive to OGT and UDP-GlcNAc concentrations.....	36
Figure II-3: Chemical structures of UDP-GlcNAc, UDP-5SGlcNAc and UDP. ....	37
Figure II-4: Residue E10 in the HCF-1 <sub>PRO</sub> repeat is crucial for proteolysis. ....	40
Figure II-5: The HCF-1 <sub>PRO</sub> repeat binds analogous to an O-GlcNAcylation substrate in the OGT catalytic domain. ....	43
Figure II-6: HCF-1 <sub>PRO</sub> -repeat–OGT interactions are sensitive to UDP-GlcNAc. ....	46
Figure II-7: The sugar moiety of UDP-GlcNAc in the OGT active site is important for proper HCF-1 <sub>PRO</sub> -repeat–OGT association. ....	48
Figure II-8: The HCF-1 <sub>PRO</sub> -repeat residue E10 displays unfavorable interactions with the OGT–UDP-GlcNAc complex. ....	51
Figure II-9: The E10 side-chain strains HCF-1 <sub>PRO</sub> -repeat–OGT interactions in the presence of UDP-GlcNAc. ....	53
Figure II-10: The E10 side-chain makes unfavorable interactions with the OGT catalytic domain in the presence of UDP-GlcNAc.....	55
Figure II-11: The HCF-1 <sub>PRO</sub> repeat contains a flexible region between the E10 cleavage site and the threonine region. ....	57
Figure II-12: Refinement of the regions within the HCF-1 <sub>PRO</sub> repeat.....	59
Figure II-13: Residue T14 in the threonine region is important for interactions with residues in the OGT TPR domain. ....	61
Figure II-14: Proposed initial step for the HCF-1 <sub>PRO</sub> -repeat cleavage mechanism. ....	65

## Chapter III

Figure III-1: HCF-1 <sub>PRO</sub> -repeat cleavage is context dependent. ....	70
Figure III-2: Sequence conservation of HCF-1 sequence 867-1098. ....	72
Figure III-3: Deletion analysis with substrates synthesized in a wheat-germ extract <i>in vitro</i> transcription and translation system. ....	74
Figure III-4: HCF-1 <sub>PRO</sub> -repeat cleavage analysis with HCF-1rep1 wild-type and deletion substrates synthesized in HEK 293 cells <i>in vivo</i> . ....	76
Figure III-5: <i>In vitro</i> HCF-1 <sub>PRO</sub> -repeat cleavage analysis with substrates and OGT synthesized and purified from <i>E. coli</i> . ....	79
Figure III-6: Region II can activate HCF-1 <sub>PRO</sub> -repeat cleavage in a heterologous context. ....	81
Figure III-7: The effect of Region II on HCF-1 <sub>PRO</sub> -repeat cleavage is sequence specific. ....	83
Figure III-8: Region II C-terminal of the HCF-1 <sub>PRO</sub> repeat inhibits HCF-1 <sub>PRO</sub> -repeat cleavage. ....	85
Figure III-9: The effect of Region II on HCF-1 <sub>PRO</sub> -repeat cleavage is dependent on Region IV. ....	87
Figure III-10: Regions I, II, and III together affect cleavage of a substrate containing HCF-1 <sub>PRO</sub> repeats 1, 2, and 3. ....	89

## Chapter IV

Figure IV-1: Region II enhances HCF-1rep1–OGT association. ....	95
Figure IV-2: Region II is an independent OGT-binding sequence. ....	97
Figure IV-3: Region IV supports Region II–OGT association. ....	99
Figure IV-4: Efficient Region II–OGT binding requires the OGT TPR domain. ....	101
Figure IV-5: Identification of O-GlcNAcylation and phosphorylation sites in the uncleaved HCF-1rep1 precursor. ....	103
Figure IV-6: Region II placed C-terminal of the HCF-1 <sub>PRO</sub> repeat displays an altered O-GlcNAcylation pattern. ....	106
Figure IV-7: Region II dissection: analysis of Region II sub-regions for cleavage and O-GlcNAcylation activities. ....	109
Figure IV-8: The Region II O-GlcNAcylation status is influenced by HCF-1 <sub>PRO</sub> -repeat E10 mutations but not by the absence of HCF-1 <sub>PRO</sub> -repeat cleavage <i>per se</i> . ....	111
Figure IV-9: Region II O-GlcNAcylation is not fundamental for Region II activity on HCF-1 <sub>PRO</sub> -repeat proteolysis. ....	113
Figure IV-10: HCF-1 O-GlcNAcylation levels are independent of OGT-mediated HCF-1 <sub>PRO</sub> -repeat processing. ....	116
Figure IV-11: Proposed model for Region II cleavage enhancement of the first HCF-1 <sub>PRO</sub> repeat. ....	121

## Appendix

Figure Appendix-1: Serine and threonine density in the HCF-1 sequence 867-1071. ....	147
--	-----

# Chapter I :

## *Introduction*

The availability of the human genome sequence has transformed biomedical research over the past decade. The analysis of the human genome has led to the identification of approximately 20,000 protein-coding genes. At first sight, the number of human genes seemed surprisingly low, given that similar numbers have been identified in organisms with much lower complexity, such as in the mustard weed *Arabidopsis thaliana* (approximately 25,000 protein-coding genes) or in the roundworm *Caenorhabditis elegans* (around 20,000 protein-coding genes). How can we assess and explain the complexity of an organism then, if not based on its gene count? In fact, nature has emerged with a plethora of mechanisms beyond the genetic code that have allowed organisms to acquire complexity and adaptability during evolution. One of these mechanisms — post-translational modification — is the major theme of this thesis.

### **I.1 The human proteome is complex**

Each cell in the human body, with few exceptions, contains genetic material and the proteins encoded by these genes are the major players in cellular functions. The entire set of proteins synthesized in a cell, in a tissue or in an organism at a certain time is called the proteome. Unlike the genome, the makeup of a proteome can vary dramatically from cell to cell, as a result of controlled gene transcription patterns, alternative splicing and post-translational modifications. These mechanisms increase immensely the complexity of the human proteome. Some of these mechanisms, such as post-translational modification, carry biological information that is not accessible through genomic or transcriptomic approaches. To study protein structure and function as well as protein–protein interactions, a variety of techniques is available:

- Genetic tools that allow the study of protein function, e.g. by mutational analyses.
- Proteomics, the large-scale analysis of proteins using mass spectrometry approaches that allow the study of proteins and their modifications at the quantitative level.
- X-ray crystallography, nuclear magnetic resonance (NMR) and single-particle electron cryo-electron microscopy (cryo-EM) are powerful techniques to study the structure of proteins, which can unravel how proteins interact with each other at atomic resolution.

- Molecular dynamics, the computational modeling of biomolecular systems, which enable the prediction of structures and related energetics, and are especially valuable for molecules that are unstable and difficult to isolate.

Thus, a combination of the abovementioned techniques allows the investigation of protein post-translational modifications.

## **I.2 Post-translational modifications**

Protein activity and function are regulated at numerous levels of control, including transcription, mRNA processing, and translation. However, once a protein is folded into its tertiary structure, changes in activity and function may still be required. Nature has emerged with a sophisticated mechanism by which the structure and function of proteins can be changed rapidly in response to changes of the environment or internal conditions. These can occur, for example, during the development of an organism.

Post-translational modification (PTM) is a biochemical mechanism by which amino acid residues of a protein are covalently modified. PTMs greatly enhance the informational complexity of an organism by allowing amino acid properties to be altered in response to internal or external conditions. More than 200 different types of PTMs have been identified to date (Minguez et al., 2012), with an ever-increasing number. PTMs are found in all life kingdoms (Beltrao et al., 2013), in Archaea, protists, bacteria and eukaryotes, demonstrating their ubiquitous usefulness.

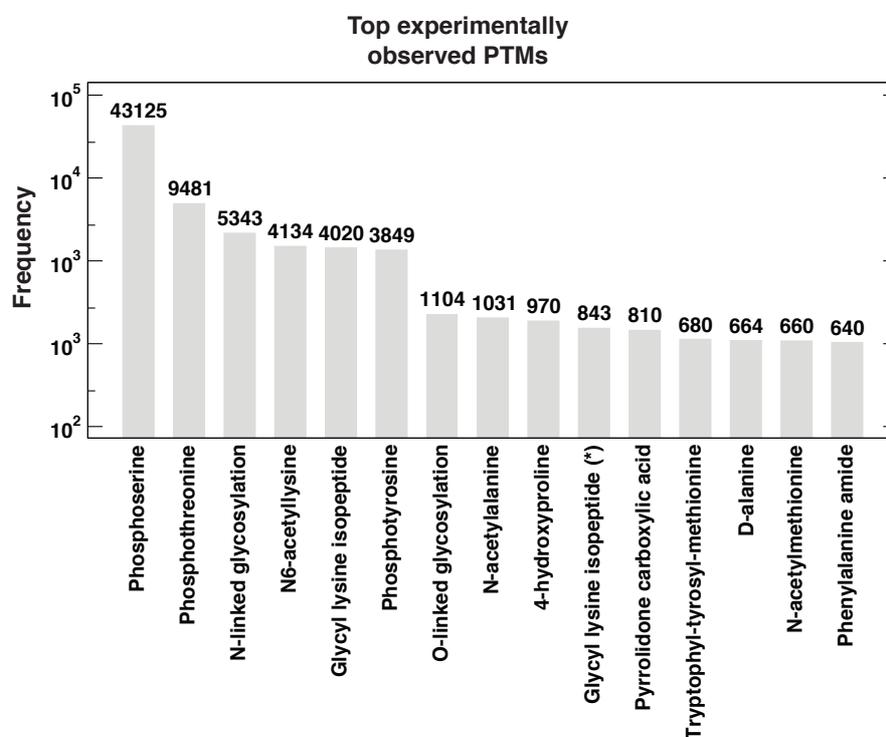
### **I.2.1 The different forms of PTMs**

PTMs can, in general, be categorized into reversible and irreversible (or at least not known to be reversible) modifications. Reversible modifications are more adaptable than irreversible modifications because they can be added or removed according to the needs of the organism. In this section, focus is given on the characteristics of reversible modifications. Particular emphasis on two different types of modifications, irreversible site-specific proteolysis and reversible O-linked glycosylation is given in sections I.3 and I.5, respectively.

Among the most abundant experimentally observed reversible PTMs are phosphorylation, N-linked glycosylation, acetylation and O-linked glycosylation (Khoury et al., 2011; Figure I-1). The most prevalent type of PTM is serine and threonine phosphorylation, but this number probably also reflects the large amount of phosphorylation studies conducted. Protein phosphorylation is a very adaptable modification due to its highly dynamic nature. An individual protein substrate can be phosphorylated or unphosphorylated, so that the population of substrate molecules contains a mixture of both molecular states. Therefore, the relative stoichiometry of the phosphorylated state at the steady state carries information

about the relative activities of the enzymes responsible for phosphorylation or dephosphorylation (Prabakaran et al., 2012).

PTMs can influence protein activity, stability, interactions or localization. It is important to point out that PTMs do not only work in isolation but also in coordination (between PTMs of the same type or of different types), which creates a combinatorial increase of the number of possible molecular states of a protein. One well-studied example of a protein harboring many sites for different types of PTMs is the transcription factor p53. P53 exerts anti-proliferative effects in response to various types of cellular stress, including DNA damage (Brooks and Gu, 2003). Upon DNA damage, this protein can be modified at more than 20 different sites. Thus, an enormous amount of different combinations of these modifications is possible, so that the protein's behavior can be altered in many ways.

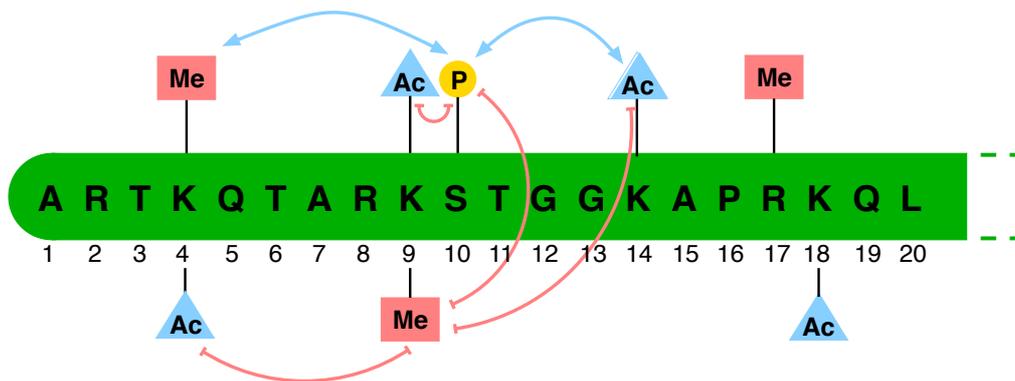


**Figure I-1: Occurrence of experimentally detected post-translational modifications (PTMs), as curated from SwissProt by Khoury et al. (2011). (\*) interchain with G-Cter in ubiquitin.**

### I.2.2 Cross-talk between PTMs

The term “cross-talk” has been widely used to describe interplay, i.e. the dependency of two PTMs of different types on one another. Positive crosstalk has been defined as a condition in which one PTM serves as a signal for the addition or the removal of a second PTM or for the recognition by another protein that carries out a second modification. Negative crosstalk refers to direct competition for modification of a single residue in a protein, or to masking by one modification of a recognition site for a second one (Hunter, 2007). This concerted action of PTMs can occur not only on a single protein, but also between different proteins. These large PTM networks, usually studied by computational proteomic approaches, thus reveal the complexity of this interplay of protein modifications and their respective functions.

A well-described example for cross-talk between different types of PTMs is the N-terminal tail of Histone H3 (Figure I-2). Histone modifications influence each other in a context-dependent manner to facilitate or repress transcription.



**Figure I-2: Cross-talk between PTMs of the N-terminal tail of histone H3.**

This scheme illustrates some of the possible modifications of the first 20 amino acids at the N-terminus of histone H3. Blue and red lines show activating and inhibiting effects on the addition of neighboring modifications. Acetylation and methylation of lysine 4 (H3K4) and lysine 9 (H3K9) are mutually exclusive reactions. P, phosphoryl-; Me, methyl-, Ac, acetyl group. Figure adapted from Alberts (2008).

### **I.2.3 Evolution and conservation of PTMs**

To date, we know the functional role of only a small fraction of PTMs. Among the few well-characterized modifications are phosphorylation and acetylation. Large-scale identification of these PTMs across multiple species, using new mass spectrometry methods, has enabled studies of the evolutionary conservation of the post-translationally modified residues and their functions. Phosphorylation and acetylation appear to be very ancient modifications and it has been suggested that their origins might be related to energy sensing, which have since then been coopted to other functions (reviewed in Beltrao et al., 2013).

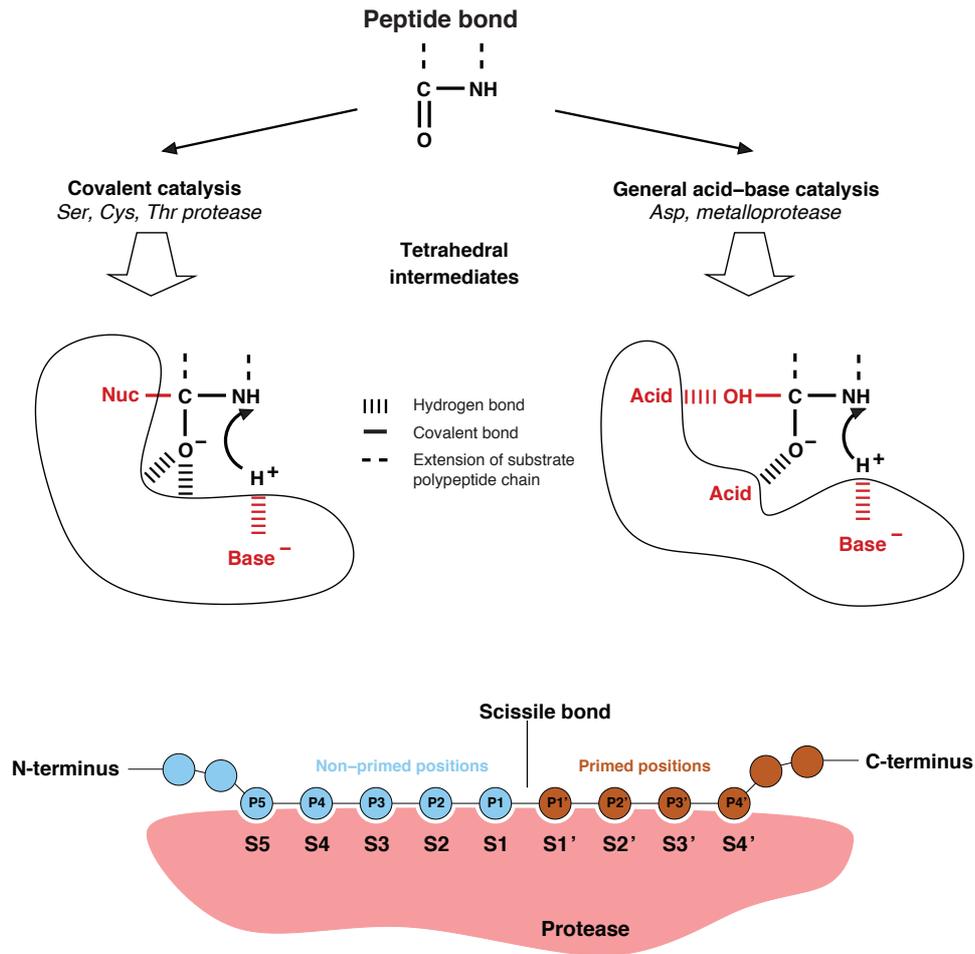
Evolutionary studies have implied that many PTM sites are unlikely to play a biological role and some might change position while retaining function (Beltrao et al., 2013; personal communication at the Gordon Conference “Post-translational modification Networks”, Hong Kong, 2013). Many PTM sites, such as phosphorylation sites, are weakly constrained and are often not conserved (Beltrao et al., 2013). Thus, the study of the biological importance and function of a specific PTM in a protein appears to be very challenging.

## **I.3 Proteolysis**

Proteolysis is the cleavage of proteins by proteolytic enzymes, also called proteases. Proteases are a class of enzymes that hydrolyze peptide bonds in proteins. Like other enzymes, proteases accelerate the reaction by stabilizing the transition state (the highest energy point in a chemical transformation), and thus lower the activation energy for the reaction to occur. The active site of a protease plays a crucial role for the stabilization of the transition state. The active site is a region in the enzyme in which usually the primary substrate-binding site is localized and catalysis occurs. It acts like a template or mold for the substrate and binds specifically to the amino-acid residues next to the scissile bond.

### **I.3.1 Human proteases**

In humans, there are more than 600 proteases, representing approximately 2% of the human genome (Turk et al., 2012). These proteases differ from each other remarkably in a number of characteristics, such as in their size, localization or specificity. To gain an overview, proteases are grouped into families, based on their catalytic mechanism. There are five mammalian protease families: aspartyl, cysteine, metallo-, serine and threonine proteases. During peptide cleavage, a tetrahedral intermediate is formed. The tetrahedral intermediate is the configuration that the substrate should have in the transition state before being transformed into the product; it is thus a prerequisite for peptide-bond cleavage in all types of proteases (Turk, 2006; Drag and Salvesen, 2010; Turk et al., 2012; Figure I-3).



**Figure I-3: The two major proteolysis mechanisms of mammalian proteases.**

Top: A peptide bond can be hydrolyzed via two different mechanisms: Serine (Ser), cysteine (Cys) and threonine (Thr) proteases (on the left) stabilize a tetrahedral intermediate that involves a stable covalent bond to the enzyme's catalytic nucleophile (Nuc), whereas metalloproteases and aspartate (Asp) proteases (on the right) use a non-covalent acid–base mechanism. In covalent catalysis, the nucleophile of the catalytic site is part of an amino acid (usually a histidine) and in general acid–base catalysis, the nucleophile is an activated water molecule. Aspartate or glutamate residues, as well as zinc can serve as acids and bases for these classes of proteases. The arrows indicate the donation of a proton ( $\text{H}^+$ ) that leads to the release of the product C-terminal of the scissile bond. Bottom: Schematic representation of a peptide substrate that binds to a protease. Subsites (numbered S1–Sn) are surfaces in the protease that are able to accommodate a single side-chain of a substrate. The substrate residues they accommodate are numbered P1–Pn. Beginning from the sites on each side of the scissile bond, the amino acid positions are non-primed towards the N-terminus, and primed towards the C-terminus. Figure modified from Turk (2006).

In principal, there are two major routes of proteolysis: Complete proteolysis by the ubiquitin-proteasome system and by the lysosome, which degrade proteins into their amino acid building blocks, ultimately leading to the destruction of proteins (Caballero et al.). In addition to protein degradation related to protein turnover, a second type of proteolysis can occur, which is referred to as limited proteolysis, proteolytic processing or site-specific proteolysis (Goulet and Nepveu, 2004). Site-specific proteolysis, as this type of proteolysis is referred to in this work, occurs when proteins are functionally modified, in many cases activated or inactivated, by highly specific proteases that make a small number of cuts, often only one.

### **I.3.2 Protease signaling**

Proteases have traditionally not been considered as signaling molecules, but this view has changed dramatically, as the role of site-specific proteolysis in a number of fundamental biological processes, for example in development (Schroeter et al., 1998; Yoshida et al., 1996), metabolism (Brown and Goldstein, 1997), immune response (Hailfinger et al., 2011), apoptosis (Patel et al., 1996) or in cell cycle progression (Goulet and Nepveu, 2004; Julien and Herr, 2003) became apparent. The blood coagulation cascade is one of the pioneering examples, demonstrating how site-specific proteolysis regulates an entire cellular process (Davie and Ratnoff, 1964; Macfarlane, 1964). In blood coagulation, the activating signal is passed through a pathway by the sequential activation of protease zymogens (inactive enzyme precursors). Another example for protease signaling is apoptosis, programmed cell death, mediated by caspases, which are proteases synthesized as zymogens (procaspases). When activated by specific signals, the caspase cascade causes cell death to eliminate cells that are damaged or have become superfluous, leading, for example, to the sculpting of the shapes of body parts in the course of development (Berg et al., 2002).

In contrast to other cellular signaling pathways involving reversible PTMs, such as kinase signaling, site-specific proteolysis is, in principle, irreversible. This unidirectional process is responsible to transmit important signals rapidly, such as in caspase-mediated apoptosis or blood coagulation, described above. Moreover, unidirectional signaling is particularly important in cellular processes, such as the cell cycle.

### **I.3.3 Site-specific proteolysis in cell-cycle progression**

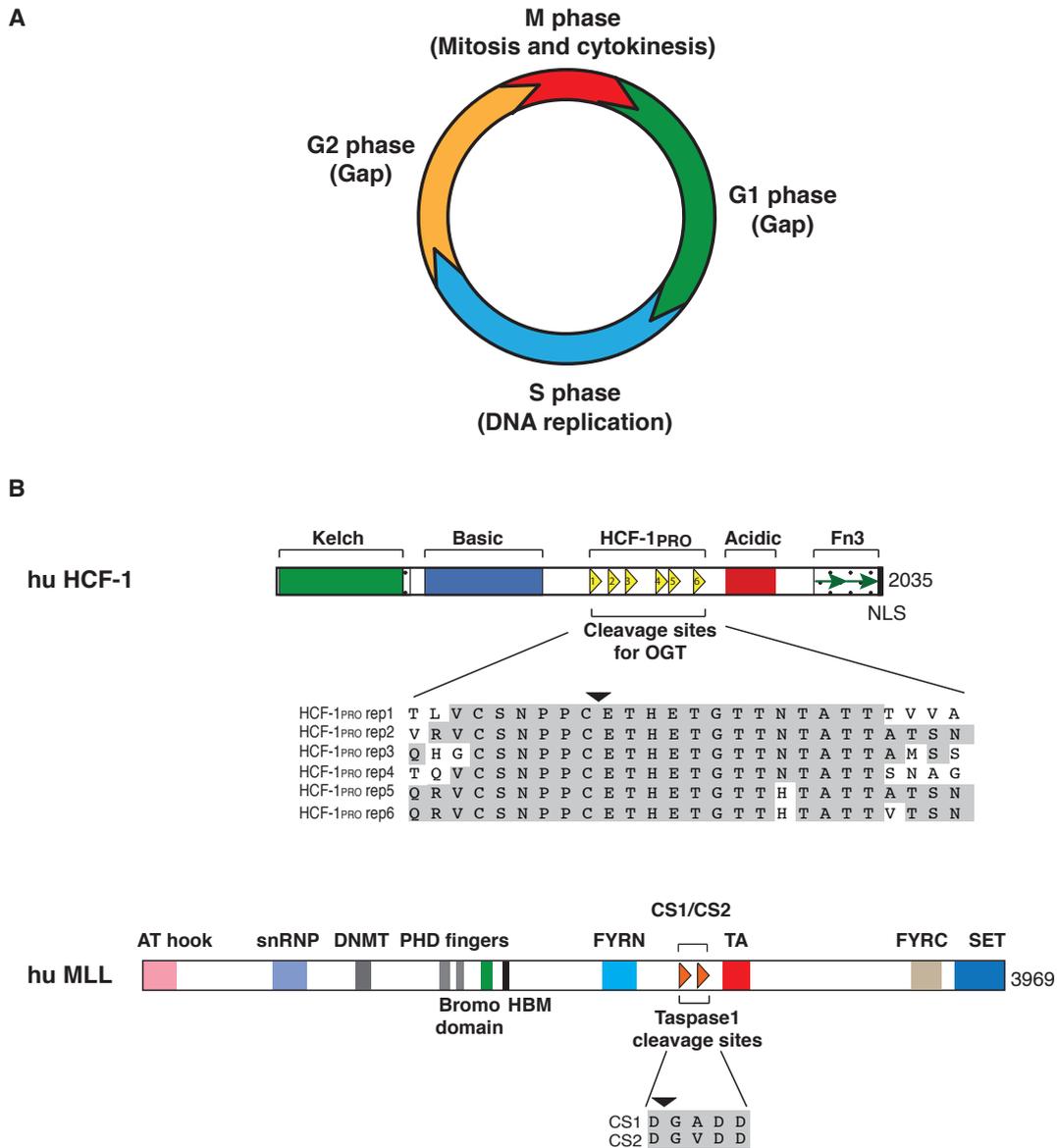
Proteases provide numerous ways of regulating cellular processes throughout the body. The mammalian cell-division cycle can be divided into four main phases, the G1-, S-, G2- and M-phases (Figure I-4 A). The passage from one phase to the next is tightly regulated to ensure that a cell only proceeds to the next phase when cell-cycle processes in the previous phase were completed properly and when environmental conditions are favorable. Two examples of

proteins regulated by site-specific proteolysis to control cell-cycle progression are the trithorax group mixed-lineage leukemia (MLL) family of proteins (Hsieh et al., 2003; Yokoyama et al., 2004) and the herpes simplex virus (HSV) host cell factor 1 (HCF-1) (Kristie et al., 1995; Wilson et al., 1993; Wilson et al., 1995b; Julien and Herr, 2003), which are both synthesized as large precursor proteins and then undergo proteolytic processing to form non-covalently self-associated heterodimers.

MLL (schematically depicted in Figure I-4 B bottom) is cleaved by a threonine protease called Taspase1 (Hsieh et al., 2003) that undergoes autoproteolysis to generate an N-terminal reactive threonine, which is utilized to cleave MLL into its native dimeric form. MLL cleavage by Taspase1 is required to activate its SET domain for full histone methyltransferase activity. Additionally, MLL proteolysis is also required to promote progression through the cell cycle by regulation of genes expressed at the G1/S boundary (Takeda et al., 2006).

HCF-1 (schematically depicted in Figure I-4 B top) is cleaved by the glycosyltransferase O-GlcNAc transferase (OGT; Capotosti et al., 2011), which, to date, is not grouped into any of the above-mentioned protease families, as it does not display homologies with known proteases (Capotosti and Herr, unpublished results) and its mechanism of cleavage has remained largely unknown (see I.6 and I.7.). Nevertheless, it is known that HCF-1 site-specific proteolysis promotes cell-cycle progression through M phase (Julien and Herr, 2003). Interestingly, it has been shown that HCF-1 and MLL interact with each other (Yokoyama et al., 2004) and that HCF-1 and members of the MLL family form multi-protein complexes to regulate G1/S-phase passage (Tyagi et al., 2007; Tyagi and Herr, 2009). Another intriguing point is that MLL and HCF-1 are both cleaved by the same Taspase1 protease in *Drosophila* (Capotosti et al., 2007) but adopted different proteolysis mechanisms in vertebrates.

Hence, although MLL and HCF-1 display many parallels regarding their activation and function, their proteolytic maturation pathways are very different in vertebrates, underscored by their respective protease recognition sequences, which are strikingly distinct (Figure I-4 B). Protease mechanisms thus show great versatility to tightly control fundamental biological processes, such as the cell-division cycle.

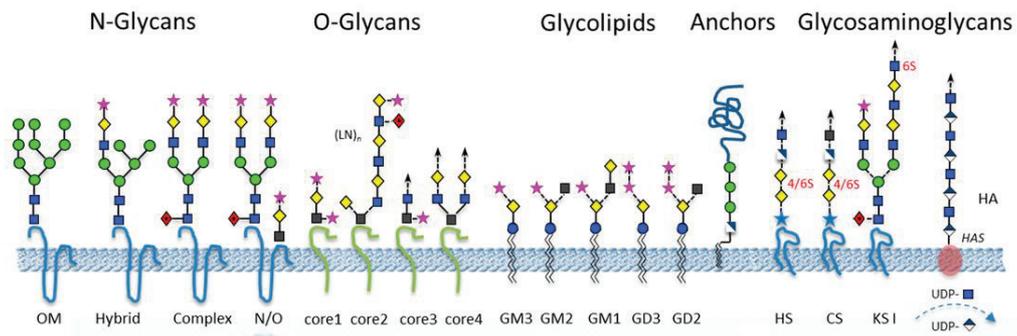


**Figure I-4: Human HCF-1 and MLL undergo proteolytic maturation to regulate the cell cycle.** (A) Schematic representation of the different phases of the mammalian cell-division cycle. (B) Schematic representation of the human HCF-1 and MLL proteins (not to scale). HCF-1 contains a Kelch domain, a region enriched in basic (Basic) or acidic (Acidic) residues, a pair of fibronectin type 2-like repeats (Fn3) and a nuclear localization signal (NLS). HCF-1 is cleaved by OGT at any of six centrally-located 26 amino acid-repeated sequences, called HCF-1<sub>PRO</sub>-repeats (HCF-1<sub>PRO</sub>). MLL contains an AT hook (binds to the minor groove of AT-rich DNA), an snRNP (small nuclear ribonucleic protein) homology domain, a DNA methyltransferase (DNMT) domain, plant homology domain (PHD) fingers, a Bromo domain, an HCF-1 binding motif (HBM), an N- (FYRN) and C- (FYRD) terminal FY-rich self-association domains, a transcriptional activation (TA) domain, and a Su(var)3-9, enhancer-of-zeste, trithorax (SET) domain. MLL is cleaved by Taspase1 at the Taspase cleavage sites CS1 and CS2. The HCF-1 and MLL protease recognition sequences are shown below their respective protein schematics. The arrowheads mark the cleaved peptide bonds.

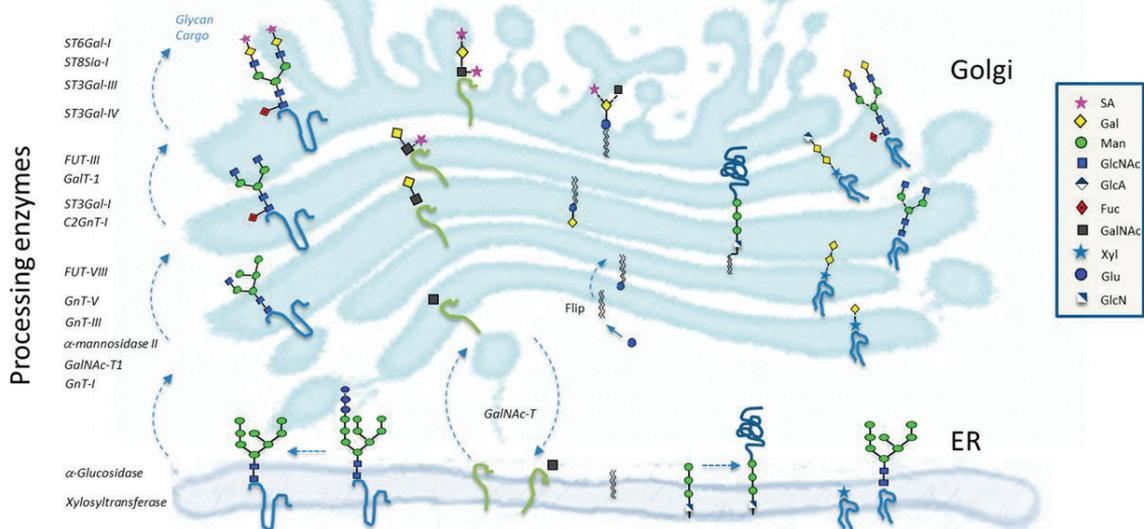
## **I.4 Glycosylation**

Protein glycosylation is the covalent attachment of carbohydrates to amino acids of proteins, giving rise to glycoproteins. A large variety of glycoproteins has been described and found in essentially all living organisms, ranging from eubacteria to eukaryotes (Spiro, 2002). Glycosylation represents one of the most abundant forms of PTMs (Figure I-1), which can be, reversible or irreversible depending on the type of glycosylation. Many glycoproteins are soluble proteins, which are secreted from cells, as well as membrane-bound proteins (Berg et al., 2002). It has been suggested that the attachment of sugar residues onto proteins is one of the most complicated co- or posttranslational modifications that a protein can undergo (Spiro, 2002). In fact, there exists a large variety of glycans (monosaccharides linked via glycosidic bond to each other) that can be attached, either to the amide nitrogen atom in the side chain of asparagine (N-linked glycosylation) or to the oxygen atom in the side chains of serine or threonine (O-linked glycosylation) (Dalziel et al., 2014; see Figure I-5 A for schematics of different glycans).

**A Major classes of glycan**



**B Biosynthesis**



**Figure I-5: Mammalian glycan biosynthetic pathways.**

(A) Schematic representation of a variety of classes of glycans (CS: chondroitin sulphate; HA: hyaluronan; HAS: hyaluronan synthase; HS: heparin and heparin sulphate; KS: keratin sulphate; GM and GD: mono- and disialylated glycosphingolipids, respectively; OM: oligomannose). (B) Biosynthetic pathways of glycans in the endoplasmic reticulum (ER) and Golgi apparatus. Monosaccharide abbreviations: Gal: galactose; Glu: glucose; GlcA: glucuronic acid; GlcN: glucosamine; GlcNAc: N-Acetylglucosamine; Fuc: fucose; SA: sialic acid; Xyl: xylose. From Dalziel et al. (2014). Reprinted with permission from AAAS.

### **I.4.1 N-linked glycosylation**

All N-linked oligosaccharides have in common a pentasaccharide core consisting of three mannose and two N-acetylglucosamine (GlcNAc) residues (Berg et al., 2002). Additional sugars are attached to this core to form the great variety of oligosaccharide patterns found in glycoproteins. In eukaryotes, N-glycosylation begins as a co-translational event in the endoplasmic reticulum (Figure I-5 B), where preassembled blocks of 14 sugars (including two GlcNAc, nine mannose and three glucose moieties) are first added to the nascent polypeptide chain. After removal of some of the glucose and mannose moieties, the protein is transferred to the Golgi apparatus where the glycans lose a variable number of mannose residues and acquire a more complex structure. The glycans are classified as “high” or “oligo” mannose, and as “hybrid” or “complex” (UniProt Consortium, 2015).

### **I.4.2 O-linked glycosylation**

O-linked glycosylation of secreted and membrane-bound proteins takes place in the cis-Golgi compartment after N-glycosylation and folding of the protein (Figure I-5 B). The most common type of O-glycosylation in secreted and membrane-bound mammalian proteins is the addition of reducing terminal N-acetylgalactosamine (GalNAc). The terminal GalNAc residue can be further extended with galactose and/or GlcNAc resulting in eight common core structures, which are often further decorated with the addition of up to three sialic acid residues (UniProt Consortium, 2015).

One type of O-linked glycosylation is the addition of a single N-Acetylglucosamine (GlcNAc) moiety to serine or threonine residues of internal cellular proteins (O-GlcNAcylation; Torres and Hart, 1984). In contrast to other types of O-linked glycosylation, this modification is reversible and highly dynamic (Wells and Hart, 2003). O-GlcNAcylated proteins have been found in a number of eukaryotes including plants and filamentous fungi, but not in *S. cerevisiae* or *S. pombe* (UniProt Consortium, 2015). Interestingly, O-GlcNAcylated proteins can also be found in many viruses that infect eukaryotic cells (Hart et al., 2011). Because of the major importance of this type of modification for the present work, an entire section (I.5) is devoted to this topic.

### **I.4.3 Biological role of glycans**

Glycans exert their biological influence in three ways (Dalziel et al., 2014): First, they have protective and stabilizing functions for glycoproteins. For instance, they maintain their solubility and ensure the correct folding of extracellular domains. Second, they can be targets for recognition by glycan-binding proteins, such as lectins. Third, they can alter the properties of the protein to which they are attached, e.g., growth factors can be modulated by the extent

and type of their glycosylation. Recognition by glycan-binding proteins appears to be of particular importance in cellular communication and cell trafficking, which is crucial for multicellular organisms, for example during development. In human pathologies, glycans play an important role in host-microbe interactions. They can be targets for exogenous pathogens that make use of glycans during early steps in their invasion. In fact, research focused on therapeutic strategies to target protein-glycan interactions is increasingly gaining attention (Dalziel et al., 2014).

## **I.5 O-GlcNAcylation**

O-GlcNAcylation is the monoaddition of  $\beta$ -N-acetylglucosamine (GlcNAc) to the hydroxyl groups of serines or threonines of proteins in the form of O-linked-  $\beta$ -N-acetylglucosamine (O-GlcNAc). O-GlcNAcylation is a reversible PTM that was discovered just 30 years ago (Torres and Hart, 1984). It is the only type of glycosylation that occurs in the nucleus and in the cytosol and is catalyzed in animals by a single enzyme, the O-GlcNAc transferase (OGT). Unlike glycosylation of secreted proteins, as described in I.4, O-GlcNAc is not elongated or further modified. There is also a single enzyme that can remove the O-linked sugar moiety, called O-GlcNAcase (OGA). The addition and removal of O-GlcNAc is highly dynamic and sometimes referred to as O-GlcNAc cycling (Hart et al., 2007). O-GlcNAcylation is generally considered to be more similar to protein phosphorylation than to classical glycosylation in terms of dynamics, localization and function. To date, more than 600 O-GlcNAcylated proteins have been identified, which are involved in almost all types of cellular processes, such as in cell-cycle regulation and division (Slawson et al., 2005; Sakabe and Hart, 2010; Olivier-Van Stichelen, 2012 #171), transcription (Jackson and Tjian, 1988), translation (Zhu et al., 2015), metabolism (Ruan et al., 2012), chromatin remodeling (reviewed in Hanover et al., 2012) and development (Gambetta and Muller, 2014).

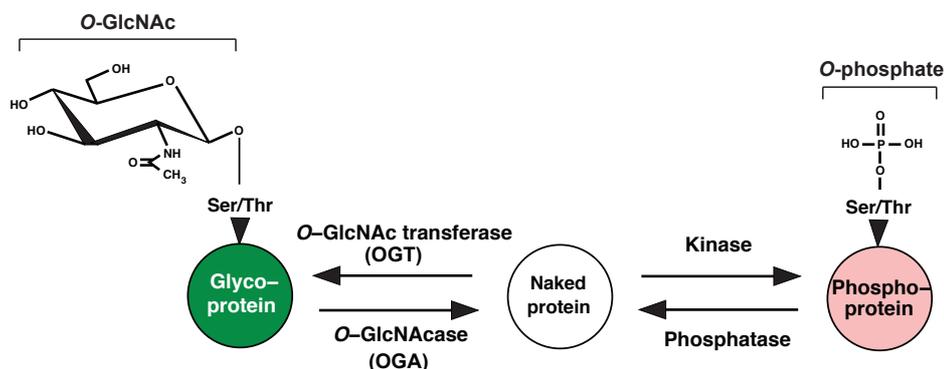
### **I.5.1 Cross-talk between O-GlcNAcylation and other PTMs**

As explained in I.2.2, PTMs of different types can influence each other, what is commonly referred to as cross-talk. Two types of O-GlcNAc cross-talk — with protein phosphorylation, and with protein ubiquitination, respectively — are described here.

Serines and threonines do not only represent sites for O-GlcNAcylation, but also for phosphorylation, and cross-talk between these two PTMs by direct competition for modification sites (Figure I-6) has been described for a number of proteins, including the oncoprotein c-Myc (Kamemura et al., 2002) and the C-terminal domain of RNA polymerase II (Kelly et al., 1993; Comer and Hart, 2001). Occupancies of adjacent sites in proteins by O-GlcNAc and O-phosphate can influence each other as well, as was shown for the tumor suppressor p53 (Yang et al., 2006). One method to establish the existence of cross-talk

between these two modifications is the inhibition of specific kinases, such as GSK-3. Inhibition of GSK-3 causes increased O-GlcNAcylation of many proteins (Wang et al., 2007), suggesting that GSK-3 phosphorylation and O-GlcNAcylation influence each other. Interestingly, the enzymes responsible for the modifications, OGT and GSK-3, are also reciprocally regulated. It has been shown that GSK-3 regulates OGT activity by OGT phosphorylation (Kaasik et al., 2013), demonstrating that O-GlcNAcylation and phosphorylation influence each other not only at the substrate level, but also at the enzyme level. The role and outcome of the cross-talk between O-GlcNAcylation and phosphorylation may vary from substrate to substrate ranging from, for example, circadian proteins (Kaasik et al., 2013) to cell-cycle-control proteins (Wang et al., 2010b). It is also important to point out, however, that O-GlcNAcylation and phosphorylation are usually substoichiometric, suggesting that only particular pools of a protein are cross-regulated (Hunter, 2007). Thus, the physiological importance of cross-talk between these two modifications should be evaluated carefully.

A number of studies proposed cross-talk between O-GlcNAcylation and ubiquitination (reviewed in Ruan et al., 2013). Most of these studies suggest indirect cross-talk via the well-established reciprocal regulation between phosphorylation and ubiquitination (Hunter, 2007). For example, O-GlcNAcylation of p53 inhibits phosphorylation and thereby reduces p53 ubiquitination and degradation (Yang et al., 2006). More direct cross-talk between O-GlcNAcylation and ubiquitination was observed via the deubiquitinase BRCA1-Associated Protein 1 (BAP1), which is involved in the promotion of protein stability. O-GlcNAcylation of specific proteins can lead to the recruitment of BAP1, which deubiquitinates the protein and thereby leads to increased protein stability. One example is the O-GlcNAcylation of the transcription factor PGC-1 $\alpha$ , which is stabilized in this manner to control gluconeogenesis (Ruan et al., 2012).



**Figure I-6: Cross-talk between O-GlcNAcylation and O-phosphorylation.**

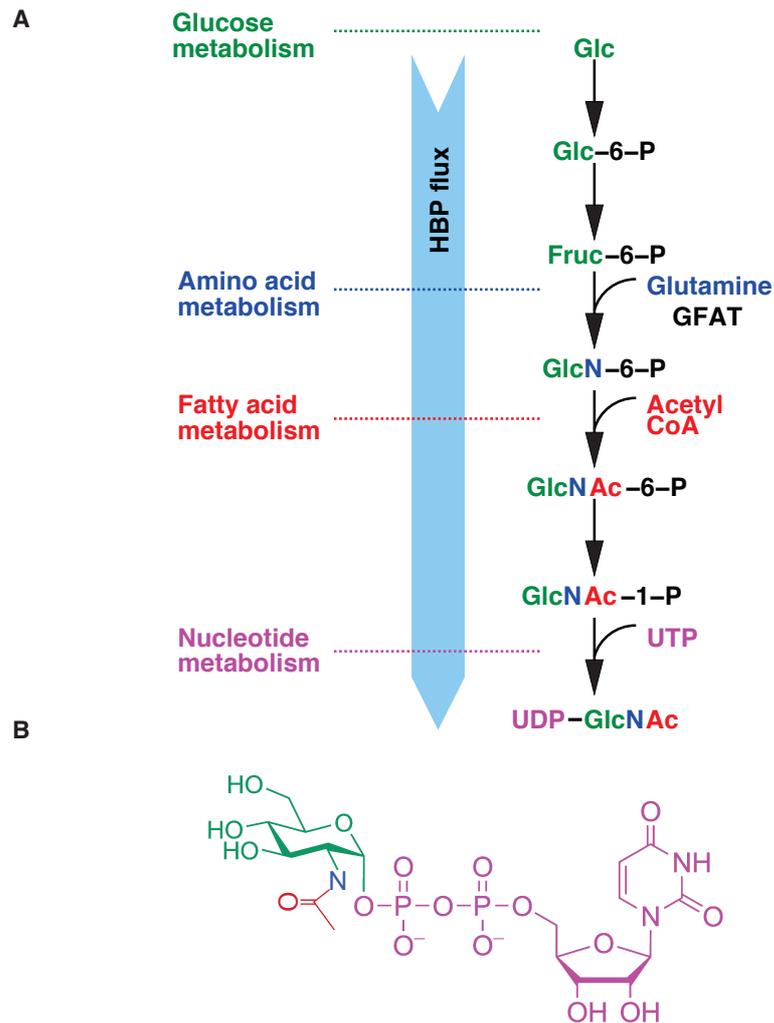
Schematic representation of O-GlcNAc and O-phosphate protein modifications. Whereas O-GlcNAcylation is regulated by a single enzyme (OGT) that attaches N-acetylglucosamine (GlcNAc) to the target protein and a single enzyme (OGA) that can remove it, phosphorylation is regulated by a multitude of kinases and phosphatases. Figure adapted from Hart et al. (2011).

### **I.5.2 The role of O-GlcNAc in transcription**

OGT is not only found in the cytoplasm, but also in the nucleus (Torres and Hart, 1984) where O-GlcNAcylation occurs on nuclear and chromatin-associated proteins (Kelly and Hart, 1989). Remarkably, a plethora of transcription regulating proteins are O-GlcNAcylated, such as p53, RNA polymerase II, TFIIA, and HCF-1 (reviewed in Slawson and Hart, 2011), whereas several studies also reported histones to be modified (Fujiki et al., 2011; Sakabe et al., 2010). In the past, O-GlcNAcylation has been associated with transcriptional repression (Yang et al., 2002). But the majority of recent studies proposed an important role in transcriptional activation in mammalian cells. Both OGT and O-GlcNAcylated proteins have been detected at transcription start-sites (Ranuncolo et al., 2012) of promoters of actively transcribed genes (Deplus et al., 2013; Vella et al., 2013). The mechanism by which OGT influences transcription is not completely understood, but it seems that O-GlcNAcylation of specific target substrates is required. Among the proteins that are known to interact abundantly with OGT at active promoters are the ten-eleven translocation (TET) enzymes (Deplus et al., 2013; Vella et al., 2013) and HCF-1 (Deplus et al., 2013; Dey et al., 2012; reviewed in Gambetta and Muller, 2015).

### **I.5.3 UDP-GlcNAc and the role of O-GlcNAc in metabolism**

OGT uses the high-energy nucleotide sugar UDP-GlcNAc to transfer the GlcNAc moiety onto its O-GlcNAcylation target substrates. Cellular UDP-GlcNAc concentrations fluctuate because *de novo* UDP-GlcNAc synthesis occurs through a broad-range nutrient-sensing pathway: the Hexosamine Biosynthetic Pathway (HBP; Marshall et al., 1991; Figure I-7). The HBP integrates glucose, amino acid, fatty acid, and nucleotide metabolism. As O-GlcNAcylation by OGT depends on UDP-GlcNAc concentrations (Hart et al., 2007), this PTM is widely believed to reflect the intracellular metabolic status. Indeed, O-GlcNAc responds to cellular stimuli including insulin, nutrients and cellular stress (Vosseller et al., 2002; Zachara et al., 2004). Moreover, O-GlcNAcylation regulates metabolic processes, such as transcription during gluconeogenesis (Ruan et al., 2012), glucose production in the liver (Zhang et al., 2014) or lipogenesis via stabilization of the carbohydrate responsive element-binding protein (ChREBP; Guinez et al., 2011). To conclude, protein O-GlcNAcylation and metabolism seem to be deeply interwoven via the HBP. Nevertheless, it is also worth noting that O-GlcNAcylation has been found to be crucial for processes independent of metabolism, such as for development (Gambetta et al., 2009; Gambetta and Muller, 2014). Given that O-GlcNAcylation plays a role for key metabolic processes, it is not surprising that aberrant O-GlcNAcylation has been linked to a spectrum of human diseases, including cancer, diabetes, cardiovascular, and Alzheimer's diseases (reviewed in Hart et al., 2011).



**Figure I-7: The Hexosamine Biosynthetic Pathway (HBP) and its end product UDP-GlcNAc.** (A) Schematic depicting the major steps of the HBP. Glucose (Glc) taken up by the cell or being cleaved from glycogen, enters the HBP and gets activated by a number of metabolic enzymes of the glucose metabolism. Glutamine fructose-6-phosphate amidotransferase (GFAT) catalyzes the addition of a nitrogen atom from the amino acid glutamine onto the sugar, which gets subsequently acetylated, using AcetylCoA, which is a cofactor of fatty acid metabolism. Finally, UDP-GlcNAc is generated, using the high-energy nucleotide UTP of the nucleotide metabolism. (B) Chemical structure of the endproduct of the HBP, UDP-GlcNAc, color-coded according to the scheme in (A).

### **I.5.4 Functional mechanisms of O-GlcNAcylation**

Despite the identification of a plethora of O-GlcNAcylated proteins, relevant for many fundamental cellular processes, the molecular mechanism by which O-GlcNAc modulates protein structure and function is largely unknown. In general, O-GlcNAc modified residues are found on unstructured protein domains and tend to cluster (personal communication with Dr. Tony Hunter, The Salk Institute, USA and Trinidad et al., 2012). Apart from its role in cross-talk with phosphorylation and ubiquitination, described above, O-GlcNAcylation has been described to affect protein enzymatic activity (Yang et al., 2006), protein-protein interactions (Roos et al., 1997), or protein aggregation (Gambetta and Muller, 2014). More recently, a role for O-GlcNAc in protein stabilization of the transcription factor Sp1 during and after translation has been reported (Zhu et al., 2015). Hence, it is difficult to predict how O-GlcNAc affects the molecular properties of a target protein.

### **I.5.5 O-GlcNAc identification on proteins**

If O-GlcNAc modified proteins are so abundant, why has this modification remained relatively unrecognized for so long? A general answer to this question is that O-GlcNAc identification and detection methods have been limited. First, unlike O-phosphate, O-GlcNAc is uncharged and a single O-GlcNAc modified residue does not usually cause an apparent mobility shift on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Only highly O-GlcNAcylated proteins, such as the nuclear pore protein Nup62 or HCF-1, under certain conditions, cause visible shifts on SDS-PAGE. Second, O-GlcNAc is often lost during protein isolation due to the removal of the moiety by OGA and other enzymes during cell lysis. Third, O-GlcNAc is difficult to detect using mass spectrometry analysis because it often occurs at substoichiometric levels and falls off the polypeptide during the ionization process. Recent advances in mass spectrometry, especially detection using the novel electron transfer dissociation (ETD) and higher energy collisional dissociation (HCD) techniques, make it now possible to identify O-GlcNAc on polypeptides with high reliability. Moreover, the recent development of monoclonal antibodies has also improved the ability to specifically detect O-GlcNAcylated proteins.

## **I.6 OGT**

O-GlcNAc transferase (OGT) is the glycosyltransferase that catalyzes the addition of GlcNAc to serines or threonines of nuclear and cytoplasmic proteins. The enzyme is highly conserved in eukaryotes from protozoa to fungi, plants and animals but, interestingly, a related gene in yeast has not been identified. An OGT homolog exists also in bacteria, but there are fundamental structural differences to eukaryotic OGT, particularly in the protein

domains that constitute the active site (Lazarus et al., 2011). OGT belongs to the GT-B superfamily of glycosyltransferases (Gtfs), but is strikingly different from the other members of this family in terms of cellular localization and biological functions. OGT contains two distinct regions, a superhelical tetratricopeptide repeat (TPR) domain (Lubas et al., 1997) and a globular catalytic domain. There are three isoforms of OGT, differing mainly in the number of their N-terminal TPRs (Hanover et al., 2003). The longest form contains 13.5 TPRs and is called ncOGT for nucleocytoplasmic OGT (Figure I-8 A), another form contains only nine TPRs and harbors a mitochondrial target sequence (mOGT), while the shortest form has only 2.5 TPRs (sOGT). The physiological roles of mOGT and sOGT are not well understood, but a catalytic activity for mOGT has been demonstrated (Love et al., 2003).

### **I.6.1 Structure of human OGT**

The human TPR domain contains 13.5 tandemly arrayed TPRs, which are themselves each degenerate 34-amino acid motifs (Jinek et al., 2004) that fold into paired anti-parallel  $\alpha$ -helices and can be found in a variety of proteins, such as protein phosphatase 5 (PP5) and aryl-hydrocarbon-interacting-protein-like 1 (AIPL1). TPRs have been suggested to mediate protein-protein interactions and intracellular localization (D'Andrea and Regan, 2003). A crystal structure (Figure I-8 B) of the human OGT catalytic domain and the last 4.5 TPRs (OGT 4.5) revealed that the catalytic domain is divided into three distinct sub-domains, called N-Cat and C-Cat connected to each other by a third domain, Int-D for Intervening domain (Lazarus et al., 2011). The catalytic domain is connected to the TPR domain via a flexible helix. Whereas OGT uses the cofactor UDP-GlcNAc (I.5.3) to catalyze the sugar transfer to the O-GlcNAcylation substrate, in crystallization conditions, UDP-GlcNAc gets hydrolyzed and results in UDP and GlcNAc, and the latter then dissociates from OGT. Therefore, only UDP is visible in this crystal structure, which binds in a pocket in the C-Cat domain, near the interface with the N-Cat domain.

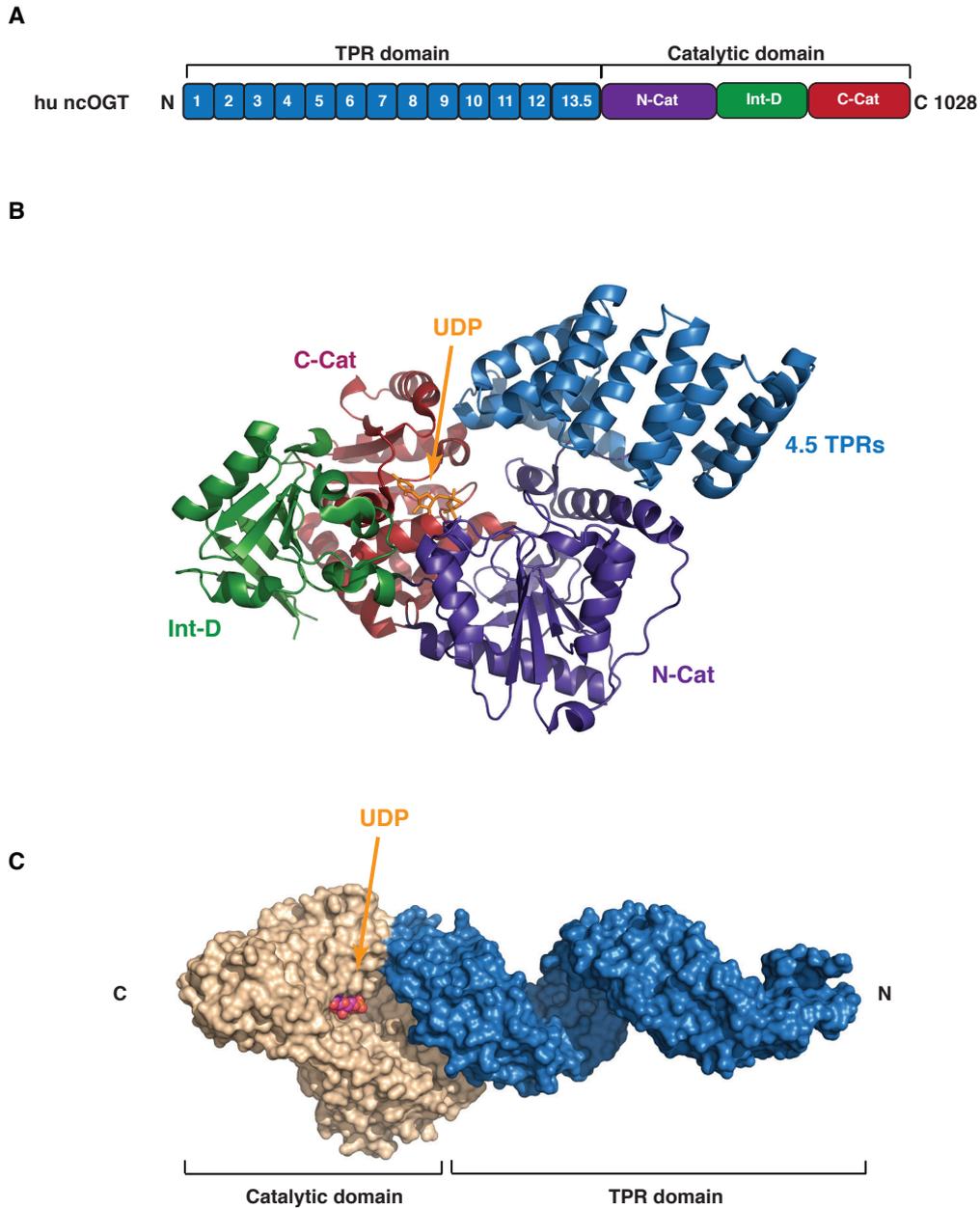
Because the first two TPR units of this crystal structure overlap with the last two TPR units of the previously crystallized human TPR domain (Jinek et al., 2004), it is possible to obtain a structural model of full-length human OGT by homology modeling (Zoete 2011, unpublished results; Figure I-8 C). This model clearly shows the bipartite structure of OGT: The superhelical TPR domain and the globular catalytic domain are structurally — and likely also functionally — different from each other.

### **I.6.2 Substrate specificity of OGT**

OGT O-GlcNAcylates a multitude of substrates, but a consensus sequence has not been identified to date. The prediction of O-GlcNAcylation sites by means of bioinformatics tools, analogous to phosphorylation site prediction, has therefore not been very successful.

Nevertheless, it seems that OGT displays sequence preference for proline or  $\beta$ -branched amino acids (e.g. valine or threonine) flanking the glycoside acceptor residue (Lazarus et al., 2011; Lubas et al., 1997). The reason behind this could be that OGT prefers residues, which enforce an extended conformation surrounding the acceptor residue.

A crystal structure of OGT 4.5 in complex with UDP and a 14-mer peptide substrate of casein kinase II (CK-II) (Lazarus et al., 2011) showed that the CK-II peptide binds in a cleft between the TPR domain and the catalytic domain, and forms contacts to residues in the catalytic domain and to UDP via its amide backbone. Moreover, the cleft is filled with ordered water molecules, which suggests that this adaptable interface can bind to a range of substrates containing different amino acids with different side-chain sizes and polarities, partially explaining how OGT can bind to so many different substrates without a consensus motif. It has also been suggested that OGT targets its substrates via its TPR domain, but how specificity may be achieved is still unclear. A possibility is the formation of multi-protein complexes, in which OGT can bind to adapter proteins that will recruit OGT to its O-GlcNAcylation targets. More generally, available TPR-ligand protein structures of other TPR containing proteins indicate that binding specificity cannot be attributed to a single property. Rather, specificity could be achieved by a combination of factors, such as hydrophobic pockets, residue type, charge and electrostatics between the TPRs and the ligand (Zeytuni et al., 2012).



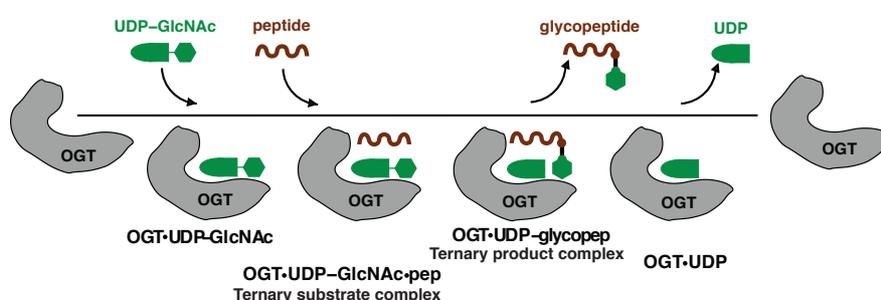
**Figure I-8: Structure of human OGT.**

(A) Schematic of the full-length human nucleocytoplasmic OGT (ncOGT) domain architecture (not to scale). The tetratricopeptide repeat (TPR) domain with 13.5 repeats is colored blue. The catalytic domain is divided in three sub-domains, the N-terminal domain (N-Cat, purple), the intervening domain (Int-D, green) and the C-terminal domain (C-Cat, red). (B) Crystal structure (PDB code 3PE3, (Lazarus et al., 2011)) of human OGT with 4.5 TPRs (OGT 4.5) in complex with UDP. Color-coding as in the scheme in (A). (C) Surface representation of human ncOGT in complex with UDP, modeled after the two structures in Jinek et al. (2004) and Lazarus et al. (2011) (PDB codes 1W3B and 3PE3, respectively) by Dr. Vincent Zoete. The entire catalytic domain is colored in beige and the TPR domain is colored in blue.

### I.6.3 Mechanism of O-GlcNAcylation

The mechanism of the glycosyltransferase reaction by OGT is not fully understood. OGT is proposed to use an “ordered bi-bi” mechanism (Figure I-9; Lazarus et al., 2012). In enzyme kinetics, “bi-bi” according to the Cleland notation (Cleland, 1967), stands for a multisubstrate enzyme reaction involving two substrates (first bi) and two products (second bi). UDP-GlcNAc is proposed to bind first in the catalytic domain, followed by the peptide with the glycoside acceptor residue. The glycosyltransferase reaction causes UDP-GlcNAc hydrolysis, which yields the leaving group UDP that remains transiently bound to the catalytic domain. The GlcNAc moiety is transferred to the peptide (glycopeptide), which is subsequently released from the complex. In the last step, the leaving group UDP also dissociates from OGT, setting the stage for a new reaction cycle. Thus, UDP-GlcNAc is a cosubstrate, but is referred to in this work by the more general term “cofactor”.

There have been numerous attempts in the past to identify a so-called general catalytic base in OGT that could play a key role in the glycosyltransferase mechanism. The catalytic base would be needed to accept the proton from the serine or threonine acceptor hydroxyl. The GlcNAc moiety could then be transferred to the reactive serine or threonine, following the break of the bond between GlcNAc and UDP (the bond between the anomeric carbon atom and the phosphate). Genetic analysis, however, showed that substitution of the basic amino acids in close proximity to the acceptor residue by other amino acids did not inhibit the reaction completely. Based on new OGT structures, the best candidates His 558 and His 498 in the OGT catalytic domain seem to be too far away to execute this function (Lazarus et al., 2012; Schimpl et al., 2012). It has been proposed that interactions between UDP-GlcNAc and the protein substrate facilitate catalysis (active substrate participation; Lazarus et al., 2012). Indeed, in the resolved structural complex, the contact interface between UDP-GlcNAc and the peptide substrate is extensive. Another proposal is that the  $\alpha$ -phosphate of UDP-GlcNAc (Figure I-7 B) might act as a general base (Schimpl et al., 2012).

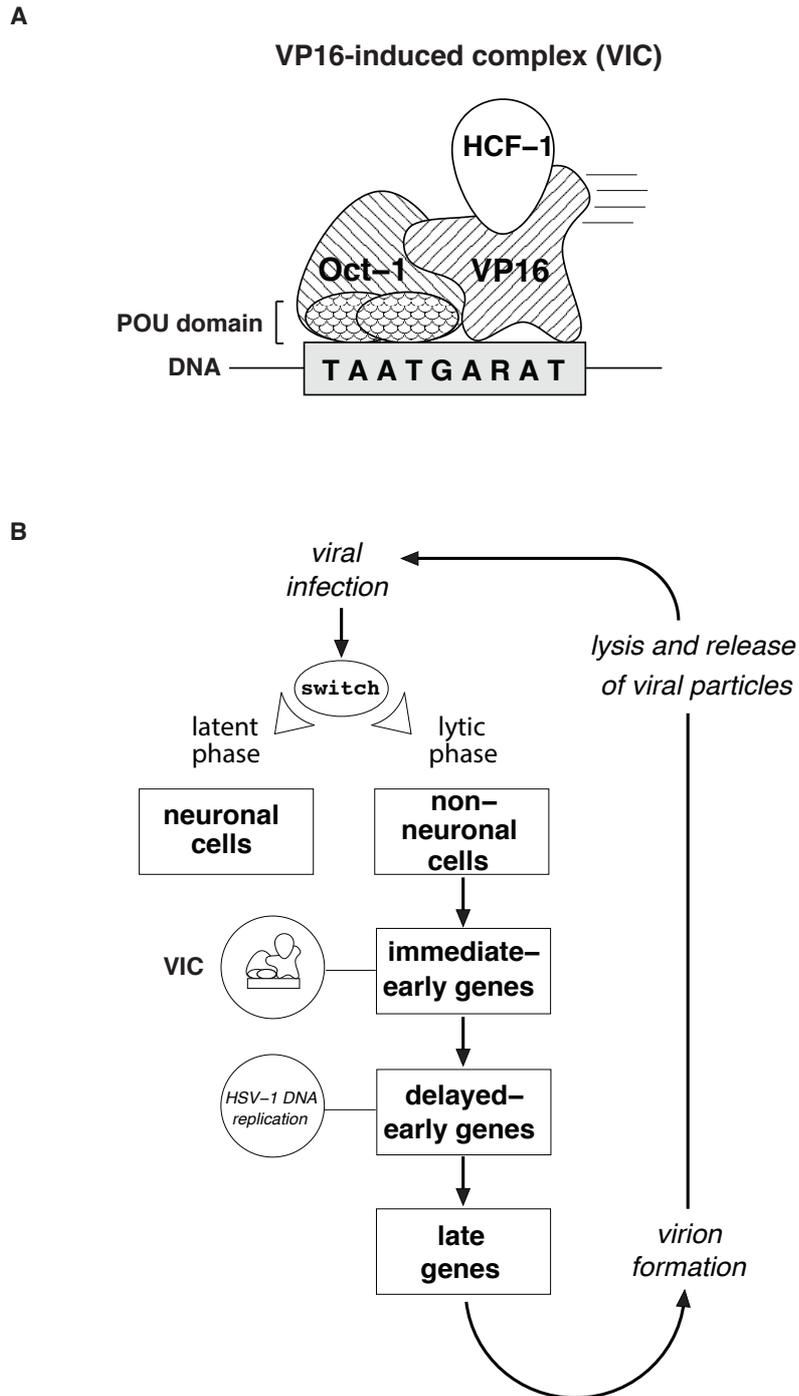


**Figure I-9: Proposed ordered bi-bi kinetic mechanism for OGT O-GlcNAcylation.**

From left to right: OGT binds first the cofactor UDP-GlcNAc, and then the acceptor peptide. After the glycosyl reaction has occurred, the O-GlcNAcyated peptide and UDP get sequentially released and a new reaction cycle can begin. Schematic modified from Lazarus et al. (2012).

## **I.7 HCF-1**

HCF-1 is a transcriptional co-regulator, and evolutionary conserved in animals. It was first discovered as a human host-cell factor for herpes simplex virus (HSV) infection, in which HCF-1 is required to stabilize the formation of the VP16 induced-complex, a multiprotein transcriptional complex, which activates HSV immediate-early gene transcription (Figure I-10; Gerster and Roeder, 1988; Wysocka and Herr, 2003). Subsequent studies have shown that HCF-1 is an important regulator of cell proliferation and chromatin modification states (Goto et al., 1997; Wysocka et al., 2003). To become fully active, HCF-1 must undergo maturation via post-translational modification, which is the subject of this study.



**Figure I-10: The VP16 induced complex (VIC) and the viral life cycle of Herpes simplex virus (HSV).**

(A) The VIC upon HSV infection. The viral protein VP16 binds to HCF-1 and is primed for association with Oct-1 on 'TAATGARAT' regulatory elements present in each immediate-early (IE) gene promoter. Oct-1 associates with DNA via the DNA-binding domain POU (Herr et al., 1988). (B) Schematic representation of the HSV life cycle. The nature of the infected cell dictates the course of infection. Upon primary infection, HSV enters the lytic cycle, which leads to subsequent infection of neurons. In lytic infection, the VIC activates transcription of HSV IE genes, resulting in a cascade of gene transcription, in which early and late genes are transcribed. During latent infection, there is little viral gene transcription. Nevertheless, the virus can be reactivated by stimuli such as stress or UV exposure, which causes renewed episodes of lytic infection at the site of primary infection.

### I.7.1 HCF-1 serves as a binding platform for proteins involved in transcription

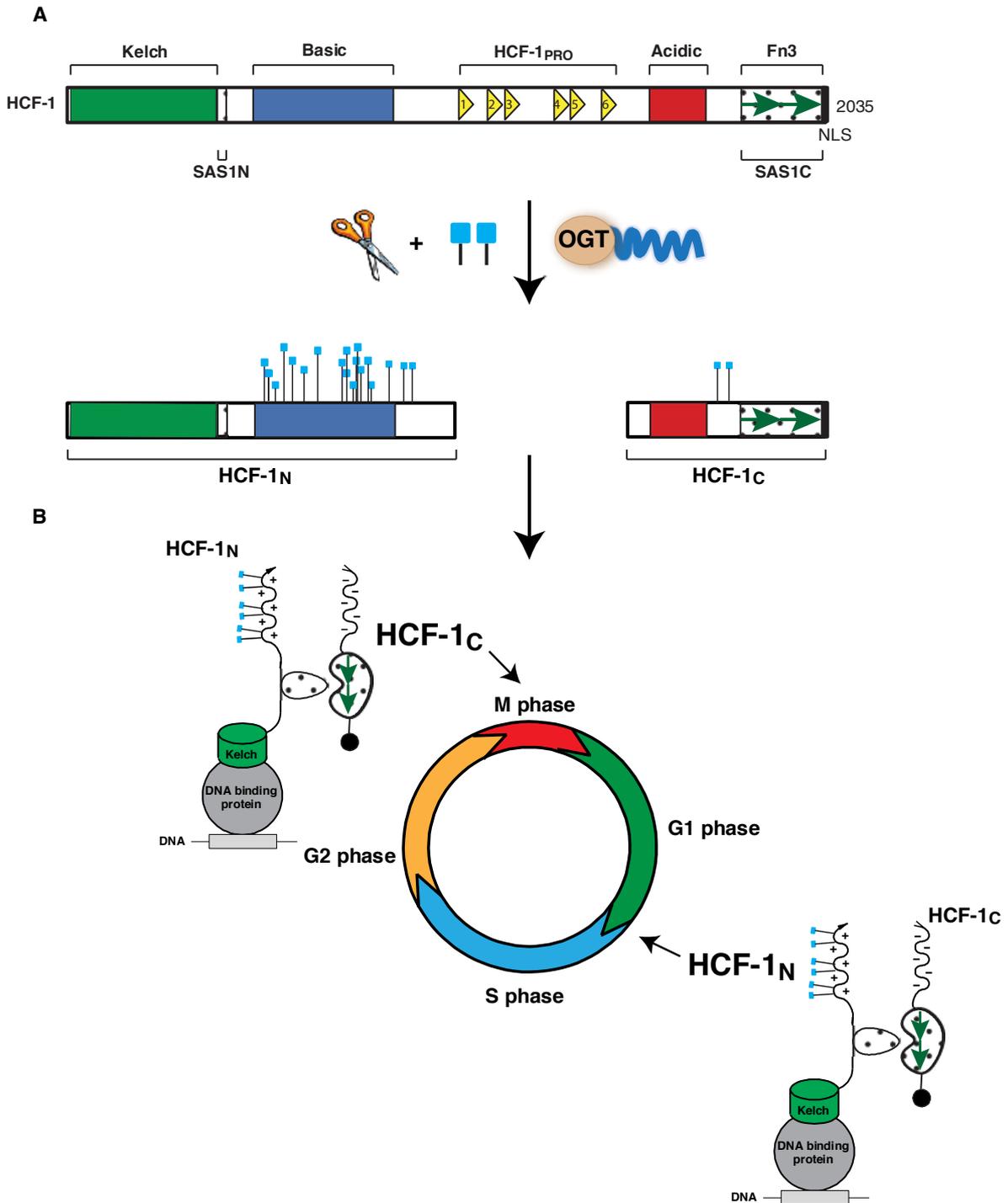
A schematic of the human full-length HCF-1 pre-mature protein is shown in Figure I-11 A. HCF-1 is processed at a series of six centrally-located 26-amino-acid repeats, called HCF-1<sub>PRO</sub> repeats (HCF-1<sub>PRO</sub>) (Wilson et al., 1993; Kristie et al., 1995; Wilson et al., 1995b) to generate two subunits called HCF-1<sub>N</sub> and HCF-1<sub>C</sub>, which remain self-associated. HCF-1 contains several distinguishing regions including (from N-to-C termini): (i) a Kelch domain, a protein interaction domain named after of its similarity to the *Drosophila* protein Kelch, (ii) a SAS1N “self-association” element for association with the C-terminal HCF-1 subunit, (iii) regions with overall basic (called Basic) and acidic (called Acidic) character, separated by the HCF-1<sub>PRO</sub> repeats, (iv) a SAS1C element containing two fibronectin type 3 (Fn3) repeats involved in HCF-1 self-association with the N-terminal subunit (Park et al., 2012), and (v) a carboxyterminal nuclear localization signal (NLS). HCF-1 does not bind DNA by itself, but it can bind to chromatin via association of its conserved Kelch domain with DNA binding proteins, such as transcription factors (Wysocka et al., 2001b). One important cellular function of HCF-1 is the coordination of the passage through the cell cycle (Goto et al., 1997; Julien and Herr, 2003) by association with transcriptional activators or repressors, for example the E2F transcription factor family (Tyagi et al., 2007) and several histone modifying enzymes, such as the mixed lineage leukemia (MLL) histone methyltransferase (Yokoyama et al., 2004) or Sin3 histone deacetylase (Wysocka et al., 2003). Although HCF-1 can selectively interact with transcription activating or repressing enzymes, chromatin immunoprecipitation of HCF-1 complexes revealed that it is present at approximately 5,000 active promoters in HeLa cells (Michaud et al., 2013).

HCF-1 not only interacts with transcriptional regulators, but also with a number of post-translational modification enzymes, involved in chromatin regulation, such as the deubiquitinase BAP1 (Machida et al., 2009; Misaghi et al., 2009; Yu et al., 2010), the protein phosphatase PP1 (Ajuh et al., 2000) or the glycosyltransferase OGT (Wysocka et al., 2003). Thus, HCF-1 acts as a binding platform for a variety of proteins involved in the control of transcription or in the control of chromatin states. The direct functional outcome of these interactions probably depends largely on the choice of HCF-1 binding-partner proteins in these protein complexes.

### I.7.2 HCF-1 proteolytic processing is important for proper cell-cycle progression

HCF-1 is a protein that is both proteolytically processed and O-GlcNAcylated by the glycosyltransferase OGT (Capotosti et al., 2011). O-GlcNAcylation can be detected predominantly on HCF-1<sub>N</sub> (Capotosti et al., 2011; Daou et al., 2011), whereas proteolysis occurs at any of the six HCF-1<sub>PRO</sub> repeats. The resulting heterogeneous collection of N- and

C-terminal subunits remains stably, but non-covalently self-associated (Wilson et al., 1995b) and regulates different phases of the cell-division cycle (Julien and Herr, 2003; Figure I-11 B). While HCF-1<sub>N</sub> is required for G1-to-S-phase transition via association with E2F transcription factors and MLL proteins (Tyagi et al., 2007; Tyagi and Herr, 2009), HCF-1<sub>C</sub> is important for proper M-phase progression via an unknown mechanism. Intriguingly, proteolytic processing is necessary to activate the functions of the C-terminal subunit during M phase, whereas the N-terminal subunit remains functional in the absence of proteolytic processing (Julien and Herr, 2003). A defect in proteolytic processing results in cytokinesis defects and cells that cannot properly exit from the M phase and thus contain two (or more nuclei), which is commonly referred to as “binucleation phenotype”. Binucleated cells are common in cancerous cells, such as HeLa cells, and are therefore a hallmark for cancer, suggesting that HCF-1 function is important to inhibit cancer progression.



**Figure I-11: HCF-1 maturation and regulation of the human cell-division cycle.**

(A) Schematic representation of the human HCF-1 protein (the different domains are described in the text). Upon synthesis, HCF-1 undergoes proteolysis (depicted as scissors) and O-GlcNAcylation (depicted as blue squares) catalyzed by OGT. This maturation process results in hyperglycosylated N-terminal subunits (HCF-1<sub>N</sub>) and hypoglycosylated C-terminal subunits (HCF-1<sub>C</sub>), which remain stably but non-covalently associated via self-association elements (SAS1N and SAS1C). The different subunits regulate distinct phases of the cell-division cycle: HCF-1<sub>N</sub> regulates G1 to S-phase progression and HCF-1<sub>C</sub> regulates the passage through M phase.

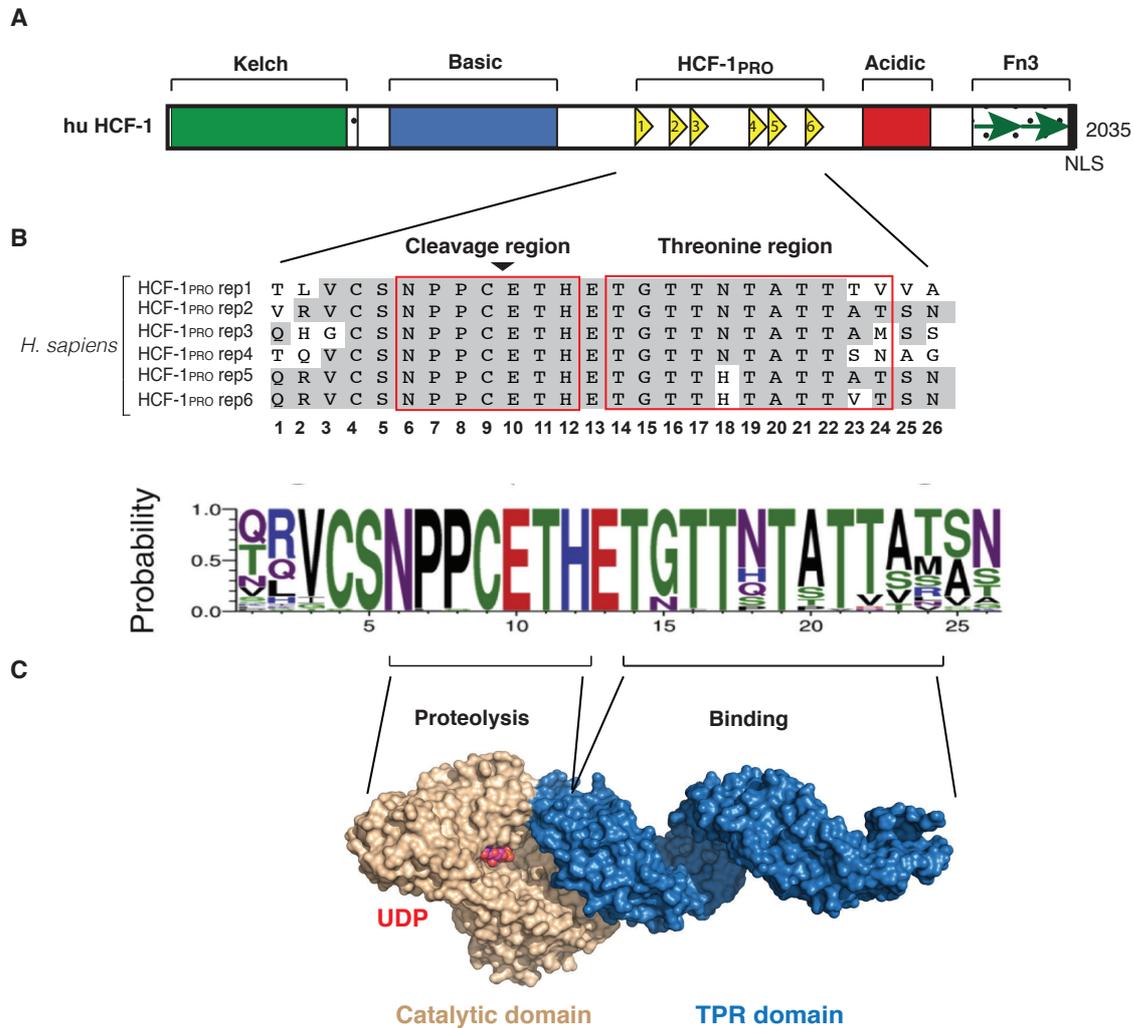
### I.7.3 OGT catalyzes site-specific proteolysis of the HCF-1<sub>PRO</sub> repeats

The human HCF-1 protein contains six HCF-1<sub>PRO</sub> repeats that are not only highly conserved in vertebrate species, but also nearly identical within the HCF-1 protein (Figure I-12 A and B). The HCF-1<sub>PRO</sub> repeats each contain 26 amino acid residues, of which a core sequence of 20 residues displays a remarkably high conservation between species and a high identity between the six repeats (Wilson et al., 1993; Kristie et al., 1995). A stretch of 21 out of 26 residues is recognized and cleaved by OGT (Wilson et al., 1995b; Capotosti et al., 2011). This sequence was further divided into two sub-regions of particular importance for cleavage: A cleavage region consisting of seven amino acids surrounding the cleavage site at glutamate at position 10 (E10) and a threonine region, consisting of a stretch of threonine residues C-terminal of the cleavage region (Capotosti et al., 2011). Recently, the HCF-1<sub>PRO</sub>-repeat cleavage products have been examined *in vitro* by mass spectrometry and the scissile bond mapped between cysteine 9 (C9) and E10 of the repeat (Lazarus et al., 2013; see Chapter II and Appendix). Interestingly, the glutamate at the cleavage site (E10) forms *in vitro* an N-terminal pyroglutamate (an uncommon amino acid derivative; the cyclic lactam of glutamic acid) after cleavage. The physiological role of N-terminal pyroglutamate of HCF-1 polypeptides, if any, is not understood.

The discovery that the glycosyltransferase OGT contains proteolytic activity was unexpected. OGT, as mentioned in I.3.3, has not been grouped into any of the known protease families yet, as OGT's catalytic mechanism for proteolysis has until recently been unknown. Furthermore, to date, HCF-1 has remained the only known substrate for OGT-mediated proteolysis. The HCF-1<sub>PRO</sub> repeats represent an unusually large sequence required for proteolysis, and it had therefore been proposed that the underlying catalytic mechanism for cleavage would be unusual (Wilson et al., 1995b; Capotosti et al., 2007). This hypothesis is underscored by the fact that, in a screen of over 1000 compounds to inhibit the proteolytic activity of HCF-1, only one compound was found, and it showed effective inhibition only at very high concentrations (Capotosti, Bogyo and Herr, unpublished results). Interestingly, it has been shown that a general inhibitor of O-GlcNAcylation, alloxan, inhibits HCF-1 proteolysis and the cofactor for O-GlcNAcylation, UDP-GlcNAc, is required for proteolytic activity (Capotosti et al., 2011). It has thus been suggested that OGT's O-GlcNAcylation activity is important for HCF-1 proteolysis.

OGT has a bipartite protein structure (see I.6.1) and the HCF-1<sub>PRO</sub> repeats display a bipartite sequence architecture. This analogy is intriguing and suggested that OGT interacts with the HCF-1<sub>PRO</sub>-repeat threonine region via its TPR domain to promote binding, whereas the OGT catalytic domain interacts with the HCF-1<sub>PRO</sub>-repeat cleavage region to promote proteolysis (Capotosti et al., 2011; Figure I-12 C). Indeed, consecutive alanine substitutions

of four conserved threonines within the threonine region disrupted OGT–HCF-1 interactions in co-immunoprecipitation assays (Capotosti et al., 2011), supporting the abovementioned model.



**Figure I-12: The HCF-1<sub>PRO</sub> repeats and the proposed bipartite interaction with OGT.**

(A) Schematic representation of the human HCF-1 protein. (B) Sequence alignment of the six human HCF-1<sub>PRO</sub> repeats. The arrowhead indicates the cleaved peptide bond between cysteine at position 9 (C9) and glutamic acid at position 10 (E10). The cleavage region and the threonine region have been defined as described in the text. (C) HCF-1<sub>PRO</sub>-repeat sequence conservation represented as WebLogo, modified from (Capotosti et al., 2011). Alignment of all HCF-1<sub>PRO</sub> repeats from six different species, human, mouse, *Xenopus tropicalis*, *X. laevis*, *Fugu rubripes*, and *Danio rerio*. (C) Proposed bipartite interaction mode of the OGT catalytic domain with the HCF-1<sub>PRO</sub>-repeat cleavage region and of the OGT TPR domain with the HCF-1<sub>PRO</sub>-repeat threonine region (not to scale). OGT model as in Figure I-8 C.

#### I.7.4 Functional versatility of OGT–HCF-1 interactions

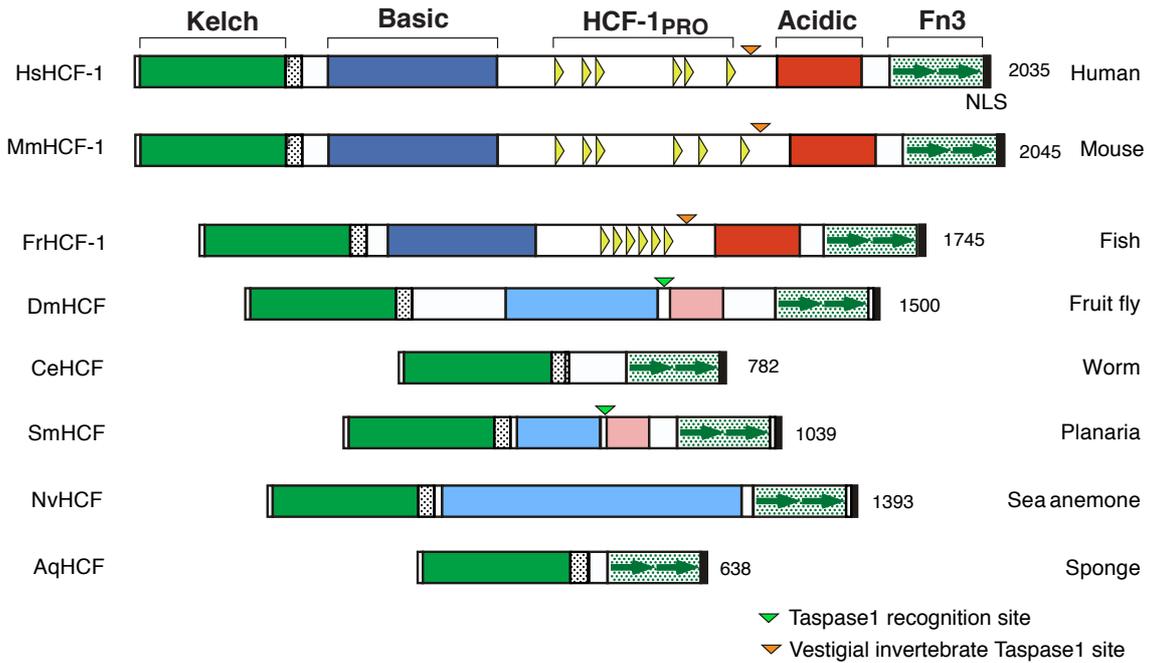
OGT displays functional versatility regarding its substrate HCF-1. First, HCF-1 is one of the main nuclear OGT-binding partners and found to interact with OGT in a number of unbiased OGT interaction assays (Deplus et al., 2013). Indeed, approximately 50% of nuclear OGT is associated with HCF-1 (Daou et al., 2011), and Wilson et al. (1995b) showed that cleavage occurs predominantly in the nucleus. OGT and HCF-1 co-occupy active promoters (Dey et al., 2012), and form multiprotein complexes together, for example, with BAP1 and PGC-1 $\alpha$  (Ruan et al., 2012) to control gluconeogenesis.

Second, O-GlcNAcylation of HCF-1 is extensive within the N-terminal subunit and modified sites in this subunit have been mapped by mass spectrometry (Wang et al., 2010b; Capotosti et al., 2011; Myers et al., 2013). Yet, the role of HCF-1 O-GlcNAcylation, if any, for HCF-1 proteolysis or for its cellular function is still unknown.

Third, it has been shown that HCF-1 proteolysis *per se* is not sufficient to activate HCF-1 functions during the cell cycle because a replacement of the HCF-1<sub>PRO</sub> repeats by Taspase1 cleavage sites leads to effective proteolysis, and yet an HCF-1 mutant binucleation phenotype. It thus seems that OGT mediates HCF-1 maturation in a way that allows HCF-1 to regulate proper cell-cycle progression through the activation of HCF-1<sub>C</sub> (Capotosti et al., 2011). Moreover, OGT — like HCF-1 — plays important roles for cell-cycle progression during M phase and G2/M-phase transition (Slawson et al., 2005; Fong et al., 2012; Sakabe and Hart, 2010; Wang et al., 2010b). Thus, HCF-1–OGT interactions are abundant and may regulate a variety of cellular processes.

#### I.7.5 Evolution of HCF-1 site-specific proteolysis

From insects to humans, HCF proteins undergo proteolytic maturation using different mechanisms (Capotosti et al., 2007; Figure I-13). Interestingly, as mentioned earlier in section I.3.3, it has been shown that *Drosophila* HCF is cleaved by Taspase1, the protease that cleaves MLL (Capotosti et al., 2007), whereas vertebrate HCF-1 proteins are cleaved by OGT. In fact, only vertebrate HCF-1 proteins contain HCF-1<sub>PRO</sub> repeats, while retaining a vestigial, non-functional, invertebrate Taspase1 site, suggesting that there has been a switch during evolution, from Taspase1-mediated HCF cleavage to OGT-mediated HCF cleavage (Capotosti et al., 2007). It thus appears that MLL and HCF-1 proteolytic maturation pathways have diverged during evolution with the emergence of vertebrates (Capotosti et al., 2007). The reason for this is unclear but it is thought that the HCF-1<sub>PRO</sub> repeats, which all lie on one large exon and can nowhere else be found, were integrated into the vertebrate genome by a viral transposition event.



**Figure I-13: The HCF protein family in vertebrates and invertebrates.**

Conserved protein domains are shown in the same color code. HCF proteins in insects contain functional Taspase1 cleavage sites and undergo proteolytic maturation via Taspase1-mediated cleavage. Vertebrate HCF-1 proteins contain HCF-1<sub>PRO</sub> repeats and undergo proteolytic maturation via OGT-mediated cleavage, while retaining a vestigial, non-functional, invertebrate Taspase1 site. Hs: *Homo sapiens*, Mm: *Mus musculus*, Fr: *Fugu rubripes*, Dm: *Drosophila melanogaster*, Ce: *Caenorhabditis elegans*, Sm: *Schmidtea mediterranea*, Nv: *Nematostella vectensis*, Aq: *Amphimedon queenslandica*. Figure by courtesy of Diego Gonzales, Viviane Praz and Shilpi Minocha (former and present members of the Herr laboratory).

### I.7.6 HCF-1 in human disease

HCF-1 is an important regulator of cell division and proliferation (see above) but also appears to have important roles in differentiation and development (Herr laboratory, unpublished results). Apart from the role of the HCF-1 protein in human HSV infection (Figure I-10), the HCF-1 protein and the gene (*HCFC1*) have been implicated in a number of human diseases.

HCF-1 levels correlate with proliferative capacity, and the highest levels are found in rapidly dividing cells such as transformed cell lines (Wysocka et al., 2001b). It has thus been proposed that differentially regulated HCF-1 could play a role in the development of cancer. Indeed, HCF-1 over-expression has been associated with human cancer (Glinsky et al., 2005). Three routes via which HCF-1 could be involved in cancer are the interactions with BAP1, E2F1 or Thanatos-associated protein 11 (THAP11). BAP1 is a putative tumor suppressor in renal carcinoma. The ability of BAP1 to regulate cell proliferation seems to be mediated, at least in part, via interaction with HCF-1 and E2F family members and disruption of BAP1–HCF-1 interactions have been shown to impair BAP1's ability to suppress cell proliferation (Pena-Llopis et al., 2012; Carbone et al., 2013). HCF-1–E2F1 association has been implicated in E2F1-induced apoptosis, and a potential role of HCF-1 in oncogenesis but also tumor suppression has been proposed (Tyagi and Herr, 2009). Another study has involved HCF-1 in human colorectal cancer. In a colorectal cancer cell line, HCF-1–THAP11 interactions are important to regulate transcription and cell growth (Parker et al., 2012).

*HCFC1* resides on the human X chromosome. Missense mutations in the HCF-1 Kelch domain have been associated with an X-linked cobalamin disorder, probably by transcriptional dysregulation of genes involved in cobalamin metabolism (Yu et al., 2013). Furthermore, HCF-1 is implicated in nonsyndromic intellectual disability, a form of mental retardation inherited in an X-linked manner. This disease is caused by a regulatory mutation in the *HCFC1* gene within its 5' untranslated region (UTR) and within the sequence encoding the Kelch domain (Huang et al., 2012).

The role of HCF-1 maturation in disease is not understood, but it is clear that HCF-1 proteolysis is required to regulate the cell cycle and to maintain cell integrity to avoid the formation of potentially cancerous cells. It may therefore only be time before human mutations are shown to affect HCF-1 proteolytic processing.

## I.8 Conclusions

There are, in principle, two types of post-translational modifications (PTMs) that enhance the complexity of an organism: Reversible modifications, being highly adaptable, and irreversible modifications, enabling the transmission of important cellular signals or the destruction of proteins. The human HCF-1 protein is an important regulator of the cell-division cycle and is subject to both reversible and irreversible PTMs. Interestingly, there is a single enzyme, the O-GlcNAc transferase (OGT), which catalyzes both types of PTMs of HCF-1: Site-specific proteolysis and O-GlcNAcylation. The catalytic mechanism for OGT-mediated HCF-1 proteolysis is not understood and the mechanism for OGT O-GlcNAcylation remains unclear.

Whereas site-specific proteolysis is required to activate HCF-1 functions during the cell-division cycle and thereby ensures the maintenance of cell integrity, the role of O-GlcNAcylation for proteolysis or for HCF-1 cellular functions is unknown. To date, HCF-1 is the only protein identified as a substrate for OGT's proteolytic activity, but the substrate requirements remain ill-defined.

## I.9 Thesis scope

The objective of this thesis is the identification of HCF-1 substrate requirements for OGT-mediated proteolytic processing, with the long-term goal to shed light on the OGT cleavage mechanism.

My specific goals in this thesis are:

1. To understand how OGT induces cleavage of the HCF-1<sub>PRO</sub> repeat by an analysis of substrate requirements within this sequence (Chapter II).
2. To identify and characterize potential HCF-1 elements outside of the HCF-1<sub>PRO</sub> repeat that may influence or regulate proteolysis (Chapter III).
3. To investigate the potential role of HCF-1 O-GlcNAcylation in HCF-1 proteolysis, comprising an analysis of O-GlcNAcylated HCF-1 sequences (Chapter IV).

## Chapter II :

# *HCF-1<sub>PRO</sub>-repeat requirements for OGT-mediated proteolysis*

### Introduction

The human HCF-1<sub>PRO</sub> repeats (Figure I-12 B) represent unusually large protease recognition sequences and the mechanism by which OGT cleaves them has remained obscure. Capotosti et al. (2011) identified key elements required for OGT-mediated cleavage: (i) ncOGT (Figure I-8), (ii) a sequence of 21 amino acids within the HCF-1<sub>PRO</sub> repeat comprising the cleavage and threonine regions, and (iii) the donor for O-GlcNAcylation, UDP-GlcNAc (Figure I-7). Within the HCF-1<sub>PRO</sub> repeat, the glutamate at position 10 (E10) was proposed to associate with the OGT catalytic center for cleavage due to its particular OGT-association properties: an E10A mutation causes enhancement of HCF-1–OGT binding with respect to wild-type E10 (Capotosti et al., 2011). In addition, the model of a bipartite interaction between OGT and the HCF-1<sub>PRO</sub> repeat (Figure I-12 C) is intriguing and laid the foundation for further investigations.

In this chapter, I present a mutational analysis of the HCF-1<sub>PRO</sub> repeat and characterize interactions with OGT that promote HCF-1 proteolysis. One part of this project was developed in collaboration with Dr. Vaibhav Kapuria (postdoc in the Herr laboratory) and the laboratory of Dr. Suzanne Walker (Harvard Medical School, Boston, USA) and Dr. David Vocadlo (Simon Fraser University, Canada). Our publication “HCF-1 is cleaved in the active site of O-GlcNAc transferase” (Lazarus et al., 2013) is attached in the Appendix of this thesis (or accessible in Pubmed, PMID: 24311690). Other parts of the project were developed in collaboration with Dr. Vincent Zoete and Dr. Ute Roehrig (both at Swiss Institute of Bioinformatics, Lausanne, Switzerland).

### Results

The results in this chapter can be summarized in six main conclusions: (i) intact UDP-GlcNAc is required for HCF-1<sub>PRO</sub>-repeat cleavage, (ii) the E10 side-chain is essential for HCF-1<sub>PRO</sub>-repeat cleavage, (iii) the HCF-1<sub>PRO</sub> repeat adopts the conformation of an O-GlcNAcylation substrate in the OGT active site, (iv) the E10 side-chain strains interactions in the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex, (v) the HCF-1<sub>PRO</sub>-repeat cleavage and threonine

regions are separated from each other by a flexible linker, called Hinge region and (vi) the threonine region is important for HCF-1<sub>PRO</sub>-repeat–OGT association .

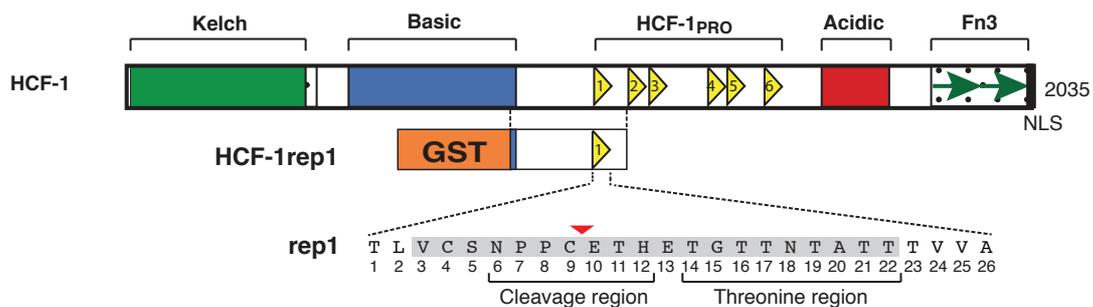
## II.1 HCF-1<sub>PRO</sub>-repeat proteolysis requires OGT and intact UDP-GlcNAc

The six HCF-1<sub>PRO</sub> repeats in the human HCF-1 protein represent functionally equal cleavage sites for OGT. To study OGT-mediated HCF-1 proteolysis *in vitro*, Capotosti et al. (2011) designed a small GST-fusion HCF-1 precursor construct called HCF-1rep1 that comprises HCF-1 residues 867-1071, spanning the first HCF-1<sub>PRO</sub> repeat (Figure II-1). This construct contains a small portion of the HCF-1 Basic region containing several identified O-GlcNAcylation sites (Capotosti et al., 2011) and an uncharacterized stretch of amino acids surrounding HCF-1<sub>PRO</sub> repeat 1. HCF-1rep1 also holds a C-terminal six-fold His-tag for protein purification. As *in vitro* transcription and translation of this construct resulted — next to the synthesis of the protein of interest — in a number of protein degradation products and premature truncated proteins (data not shown), I synthesized the HCF-1rep1 substrate in *E. coli* and purified the translated protein by Nickel-affinity chromatography. The bacterial synthesis and purification of the substrate allowed the study of HCF-1<sub>PRO</sub>-repeat proteolysis without the presence of UDP-GlcNAc or OGT, as is the case in *in vitro* transcription and translation systems (e.g. rabbit reticulocyte lysate).

To establish an *in vitro* HCF-1–OGT cleavage assay with bacterially purified HCF-1 substrates and OGT enzyme, I titrated human ncOGT (from here on called OGT unless stated otherwise; Figure II-2 A) or UDP-GlcNAc (Figure II-2 B) against fixed concentrations (3  $\mu$ M) of the GST–HCF-1rep1 substrate. The reaction was incubated in cleavage buffer (Capotosti et al., 2011) for 16 h at 37°C and the resulting reaction products were resolved by SDS-PAGE, followed by immunoblotting. Antibodies directed towards the N-terminal GST-tag were used to detect the uncleaved precursor protein as well as the N-terminal cleavage product. The intensities of the protein bands on the immunoblots were analyzed and the cleavage efficiencies (percentage of cleaved fragment to total GST–HCF-1rep1 protein in one sample) quantified. These titration experiments revealed that, for optimal cleavage conditions, OGT concentrations of 600 nM (Figure II-2 A) and UDP-GlcNAc concentrations of 0.1 mM (Figure II-2 B) were required. In this *in vitro* cleavage assay system, maximum cleavage of only 40-50% could be observed. In all subsequent experiments presented in this study, I used 600 nM OGT and 1 mM UDP-GlcNAc, unless stated otherwise, to reach the highest cleavage efficiency possible. The reaction products from the titration experiments described above (Figure II-2 A and B) were also examined for O-GlcNAcylation by immunoblot using an antibody directed towards O-GlcNAc modified serines or threonines (data not shown). O-GlcNAcylation of uncleaved HCF-1rep1 protein and N-terminal cleavage products occurred at low concentrations of OGT (76 nM) and UDP-GlcNAc (1  $\mu$ M),

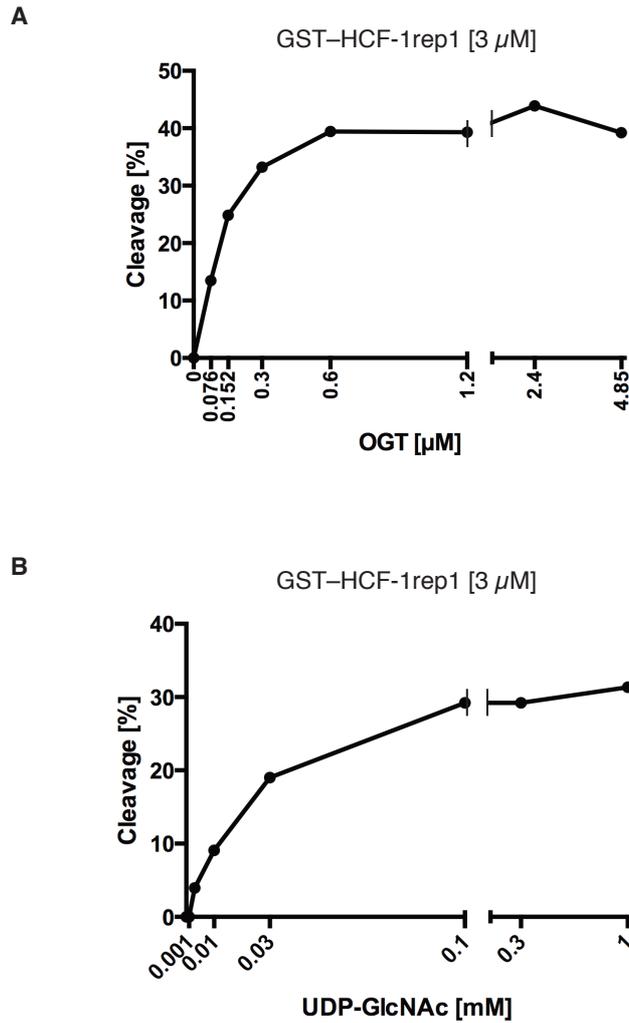
respectively. The  $\alpha$ -O-GlcNAc antibody used is, even at low concentrations, highly sensitive, and thus saturation of the antibody signal is reached in the majority of HCF-1rep1 cleavage reactions. For this reason, quantification of HCF-1rep1 O-GlcNAcylation is not shown.

The experiments above show that both OGT and UDP-GlcNAc are required for HCF-1<sub>PRO</sub>-repeat proteolysis, consistent with previously obtained results (Capotosti et al., 2011). Next, I investigated whether HCF-1rep1 could still be cleaved when the cofactor UDP-GlcNAc in the *in vitro* cleavage assay was substituted by the nucleotide UDP or by an analog called UDP-5SGlcNAc, a kind gift of Dr. David Vocadlo. UDP-5SGlcNAc is an OGT O-GlcNAcylation inhibitor ( $K_i = 8 \mu\text{M}$ ; Gloster et al., 2011), whose sugar ring contains a sulfur atom in place of the endocyclic ring oxygen of GlcNAc, as shown in Figure II-3. UDP-5SGlcNAc does not alter the structure of the OGT catalytic domain and binds to OGT in a manner almost identical to UDP-GlcNAc (Lazarus et al., 2012). The OGT-binding affinity of UDP-5SGlcNAc lies in the same range as the one of UDP-GlcNAc (Gloster et al., 2011 and Schimpl et al., 2012; summarized in Table II-1). Thus equal concentrations of both cofactors were used. Substituting UDP-GlcNAc by UDP (data not shown) or by UDP-5SGlcNAc (experiment performed by Dr. Vaibhav Kapuria; Figure S3 B and C in Lazarus et al., 2013) in the *in vitro* cleavage assay inhibited HCF-1<sub>PRO</sub>-repeat cleavage, suggesting that HCF-1<sub>PRO</sub>-repeat cleavage requires intact UDP-GlcNAc. Intriguingly, not only proteolysis, but also O-GlcNAcylation was inhibited in the presence of UDP or UDP-5SGlcNAc, respectively (data not shown), suggesting that the sugar ring of UDP-GlcNAc plays a crucial role for both HCF-1<sub>PRO</sub>-repeat proteolysis and HCF-1rep1 O-GlcNAcylation.

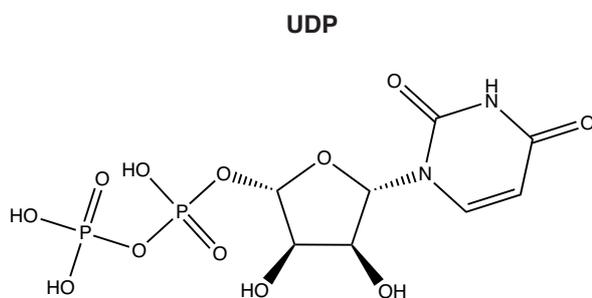
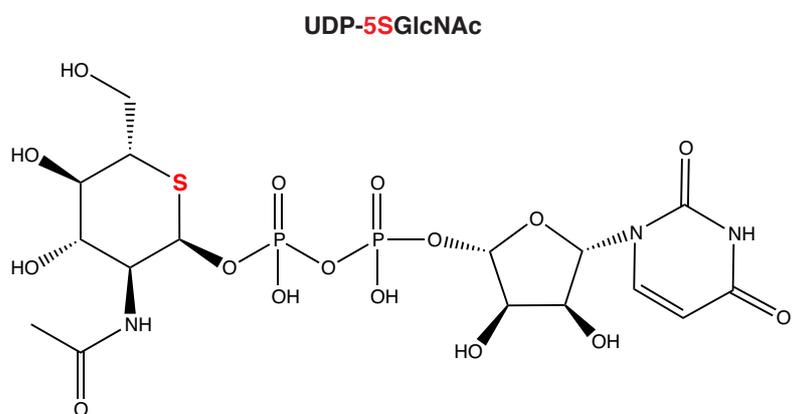
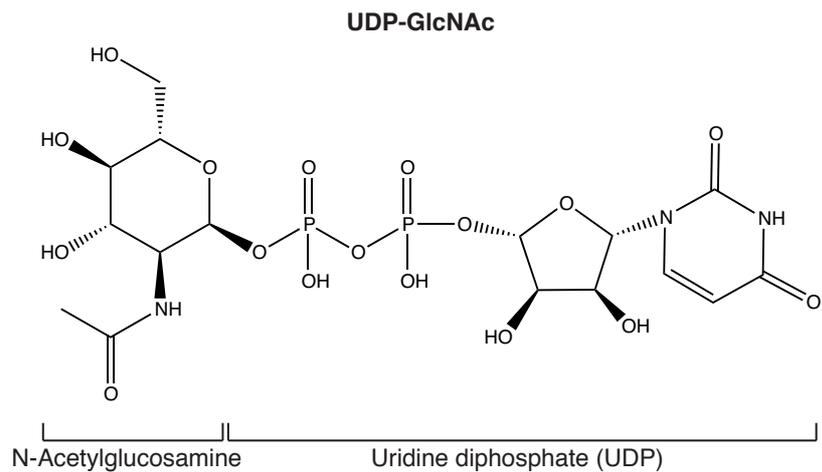


**Figure II-1: Structure of HCF-1 proteins.**

Full-length human HCF-1 and the GST-HCF-1rep1 (HCF-1rep1) construct used to study HCF-1<sub>PRO</sub>-repeat cleavage *in vitro*. The HCF-1<sub>PRO</sub> repeat 1 (rep1) sequence with the cleavage region and threonine region is illustrated. The highly conserved sequence of 20 amino acids (core sequence Kristie et al., 1995) is shaded gray. The red arrowhead indicates the cleaved peptide bond between C9 and E10.



**Figure II-2: HCF-1<sub>PRO</sub>-repeat cleavage is sensitive to OGT and UDP-GlcNAc concentrations.** 3  $\mu\text{M}$  bacterially purified GST-HCF-1rep1 were incubated in the absence or with increasing concentrations of OGT (A) or UDP-GlcNAc (B) in *in vitro* cleavage assays. The concentrations of the remaining components were kept constant (600 nM OGT, 1 mM UDP-GlcNAc, 3  $\mu\text{M}$  GST-HCF-1rep1). GST-HCF-1rep1 precursor and N-terminal cleavage products were visualized by immunoblot using  $\alpha$ -GST antibodies. Bands on immunoblots were analyzed using LI-COR Image Quant quantification software and cleavage quantified as the percentage of cleaved product to total GST-HCF-1rep1 protein (full-length uncleaved plus cleaved product) in the assay.



**Figure II-3: Chemical structures of UDP-GlcNAc, UDP-5SGlcNAc and UDP.**

**Table II-1: OGT-binding affinities of OGT cofactors from Schimpl et al. (2012).**

<b>Cofactor</b>	<b>K<sub>d</sub> [μM]</b>
UDP-GlcNAc	16.1±0.1
UDP-5SGlcNAc	7.5±0.1
UDP	0.54±0.01

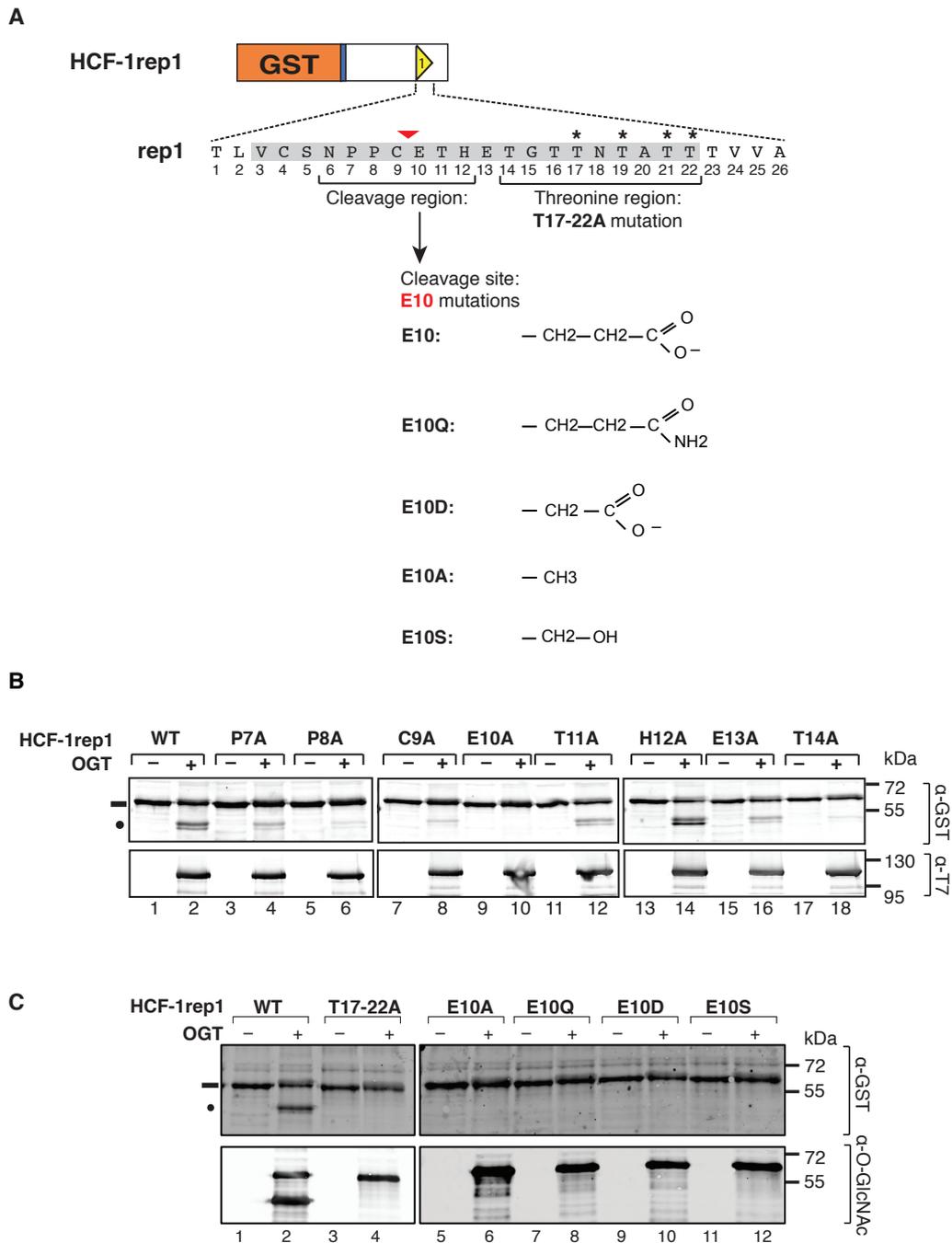
K<sub>d</sub> (dissociation constant) ± standard deviation as determined by Surface Plasmon Resonance.

## II.2 E10 is crucial for HCF-1<sub>PRO</sub>-repeat proteolysis

Previous studies of the HCF-1<sub>PRO</sub> repeat in a heterologous context *in vivo* (Wilson et al., 1995b) and *in vitro* (Capotosti et al., 2011) have shown that the HCF-1<sub>PRO</sub>-repeat cleavage region and the threonine region display sensitivity to alanine mutations. To validate these results using the *in vitro* cleavage assay with bacterially purified substrates and OGT, I analyzed the residues surrounding the cleavage site E10 by an alanine scan (Figure II-4 B): Residues lying N-terminal of the cleavage site at E10 displayed decreased cleavage activity when mutated to alanine, whereas mutation of the residues T11-E13 C-terminal of E10 had less effect on cleavage (compare lanes 3-8 and 11-16 with lanes 1-2). The T14A mutation decreased cleavage dramatically (lanes 17 and 18), presumably by perturbing interactions with the OGT TPR domain. Consistent with this result, interactions of residue T14 with the OGT TPR domain via hydrogen bonds were observed in a crystal structure of human OGT in complex with the HCF-1<sub>PRO</sub>-repeat 2 and UDP-5SGlcNAc (Figure 3 in Lazarus et al., 2013). Importantly, E10 was the only residue, when replaced by alanine, for which HCF-1<sub>PRO</sub>-repeat proteolysis could not be detected. This result suggests that E10 is the only residue in the HCF-1<sub>PRO</sub> repeat that is absolutely required for cleavage.

Because E10, when substituted by alanine, is the only residue in the HCF-1<sub>PRO</sub> repeat that drastically blocks proteolysis, I asked whether substitutions of this glutamate by residues with similar side-chain structures, such as glutamine (E10Q) or aspartic acid (E10D; side-chain structures illustrated in Figure II-4 A) have less drastic effects on proteolysis. To determine whether the mechanism of OGT-mediated proteolysis could be similar to the one of OGT-mediated O-GlcNAcylation, I also generated an E10S mutant. Serine at the position of the cleavage site E10 might potentially provide an O-GlcNAcylation site. I hypothesized that this mutated HCF-1<sub>PRO</sub> repeat could represent an active substrate for O-GlcNAcylation. Additionally, I tested the T17-22A mutant, which was described to inhibit HCF-1<sub>PRO</sub>-repeat cleavage (Capotosti et al., 2011), probably by interrupting the interactions with the OGT TPR domain. Figure II-4 C shows the results of an *in vitro* cleavage assay with the E10 and Threonine-region mutants. The alanine mutations in the threonine region blocked cleavage (lanes 3 and 4), suggesting that binding of the HCF-1<sub>PRO</sub>-repeat threonine region to the OGT TPR domain is important for cleavage. As with the E10A mutation (lanes 5 and 6), the E10 substitutions by glutamine (E10Q, lanes 7 and 8), aspartic acid (E10D, lanes 9 and 10), or serine (E10S, lanes 11 and 12) blocked cleavage, suggesting that even subtle changes in the amino acid side-chain structure of glutamate at this position are detrimental for cleavage. I thus concluded that the proteolysis mechanism of HCF-1<sub>PRO</sub>-repeat cleavage requires the glutamate side chain at position 10. This result is shown in Figure 1 C in Lazarus et al. (2013). The mutation E10S also blocked cleavage. Nevertheless, when O-GlcNAcylation

levels of wild-type (WT) and E10S HCF-1rep1, as determined by  $\alpha$ -O-GlcNAc immunoblot, were compared (Figure II-4 C, lower panel, compare lanes 11 and 12 with lanes 1 and 2), the E10S HCF-1rep1 substrate did not display any difference. I thus inferred that either (i): there are no changes in the HCF-1rep1 O-GlcNAcylation status upon mutation of E10 to serine or (ii): Changes in the HCF-1rep1 O-GlcNAcylation status (i.e., sites of O-GlcNAcylation) cannot be identified by immunoblotting using  $\alpha$ -O-GlcNAc antibodies. To study changes of O-GlcNAcylation levels, proteomic approaches are widely used and represent the method of choice. These studies are described in section II.3 below and in Chapter IV.



**Figure II-4: Residue E10 in the HCF-1<sub>PRO</sub> repeat is crucial for proteolysis.**

(A) Schematic of the HCF-1rep1 precursor construct used to study substrate requirements for HCF-1<sub>PRO</sub>-repeat proteolysis at the cleavage site and in the threonine region. Asterisks indicate alanine mutations in the threonine region; the side-chain structures of the E10 substituents are illustrated below the diagram. (B) *In vitro* HCF-1rep1 cleavage assay. Alanine scan of the residues surrounding the cleavage site at E10. Cleavage efficiencies for each mutant ( $\alpha$ -GST blot, upper panel) and OGT levels ( $\alpha$ -T7 blot, lower panel) in each reaction are shown. (C) *In vitro* HCF-1rep1 cleavage assay with E10 cleavage-site mutants and the Threonine-region mutant T17-22A. HCF-1rep1 cleavage ( $\alpha$ -GST blot, upper panel) and O-GlcNAcylation ( $\alpha$ -O-GlcNAc blot, lower panel) are shown. In (B) and (C), the positions of uncleaved precursor proteins (–) and N-terminal cleavage products (•) are indicated.

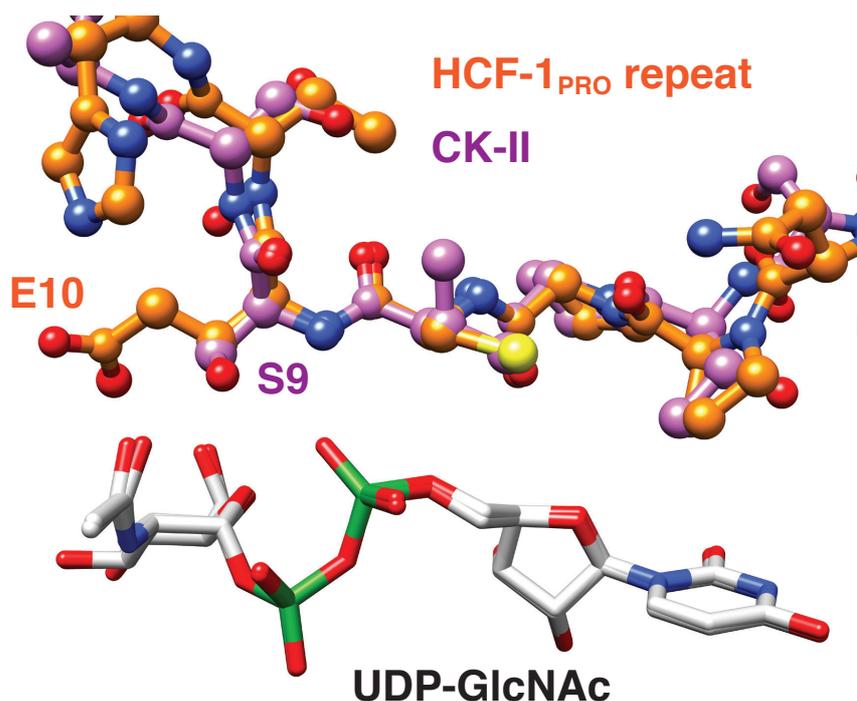
### II.3 Serine at position 10 of the HCF-1<sub>PRO</sub> repeat activates HCF-1<sub>PRO</sub>-repeat O-GlcNAcylation

Lazarus et al. (2011) determined experimentally the structure of the OGT catalytic domain with UDP in complex with an O-GlcNAcylation substrate: A 14-mer peptide YPGGSTPVSSANMM derived from Casein Kinase II (CK-II; the S9 GlcNAc acceptor residue is underlined). This crystal structure (PDB code 3PE4) was highly valuable for molecular modeling studies. Prior to the studies of Lazarus et al. (2013), we performed molecular docking simulations in collaboration with Dr. Vincent Zoete to predict how the HCF-1<sub>PRO</sub> repeat would bind inside the OGT catalytic domain in complex with UDP. The peptide sequence corresponding to HCF-1<sub>PRO</sub> repeat 1 was computationally threaded in place of the CK-II peptide into the OGT catalytic domain. Interestingly, we observed that the HCF-1<sub>PRO</sub> repeat was predicted to bind to the OGT catalytic domain in a manner similar to the CK-II O-GlcNAcylation substrate (data not shown).

In Lazarus et al. (2013), we showed using the experimentally determined OGT–UDP–5SGlcNAc–HCF-1<sub>PRO</sub>-repeat 2 E10Q crystal structure (PDB code 4N3B) for an overlay with an OGT–UDP–5SGlcNAc–CK-II crystal structure (PDB code 4GYG; Lazarus et al., 2012) that the two substrates indeed bind in a similar manner to the OGT catalytic domain. Intriguingly, the E10 cleavage-site residue of the HCF-1<sub>PRO</sub> repeat was superimposed to the CK-II GlcNAc acceptor residue (S9) and their respective side-chains displayed an almost perfect overlap (Figure II-5). Moreover, the two side-chains were located in close proximity to the sugar moiety of UDP-GlcNAc. The HCF-1<sub>PRO</sub>-repeat cleavage substrate thus binds analogous to an O-GlcNAcylation substrate in the OGT active site for O-GlcNAcylation. These results suggest that the HCF-1<sub>PRO</sub> repeat mimics an O-GlcNAcylation substrate in the OGT catalytic domain.

The observations described above led me to test if the E10S mutant, instead of being cleaved (Figure II-4 B, lanes 11 and 12), gets O-GlcNAcylated at E10S within the HCF-1<sub>PRO</sub> repeat. The HCF-1rep1 protein is an excellent O-GlcNAcylation substrate because it contains several O-GlcNAcylation sites in the sequences upstream of the HCF-1<sub>PRO</sub> repeat 1 (mapped in the Basic region in Capotosti et al., 2011). Identification of potential changes of O-GlcNAcylation in the HCF-1<sub>PRO</sub> repeat by immunoblotting is not an adequate method, as the  $\alpha$ -O-GlcNAc antibody signal gets saturated by the numerous O-GlcNAcylation sites lying outside of the HCF-1<sub>PRO</sub> repeat. To circumvent immunoblot analysis, I subjected GST–HCF-1rep1 wild-type and E10S protein preparations from transiently transfected HEK 293 cells to liquid chromatography tandem mass spectrometry analysis (LC-MS/MS) in collaboration with Dr. Patrice Waridel (Protein Analysis Facility, University of Lausanne). This strategy allowed the study HCF-1rep1 O-GlcNAcylation by endogenous OGT. Peptides of the HCF-1rep1

protein were generated by combining trypsin and Glu-C digestion. The identification of O-GlcNAcylated peptides surrounding E10S was performed using the MASCOT™ software (Perkins et al., 1999). In this analysis, a peptide with the sequence VCSNPPCSTHETGTTN (HCF-1<sub>PRO</sub> repeat 1 sequence, E10S underlined) was identified and contained a potential O-GlcNAcylation site, whereas the same peptide derived from wild-type HCF-1rep1 did not contain O-GlcNAcylated sites (data not shown). The exact location of the O-GlcNAcylated residue, however, could not be confirmed in this analysis. To tackle this problem, John Janetzko and Dr. Suzanne Walker used a substrate, which does not contain any O-GlcNAcylated sequences (HCF3R, schematic in Figure 1 A in Lazarus et al., 2013). HCF3R contains HCF-1<sub>PRO</sub> repeats 1, 2 and 3 and the less well-conserved sequences in between the repeats. In an *in vitro* HCF3R–OGT cleavage assay with C14 radiolabeled UDP-GlcNAc, an E10S mutation in one of the three HCF-1<sub>PRO</sub> repeats clearly activated the HCF-1 substrate for O-GlcNAcylation (see Figure S10 in Lazarus et al., 2013), consistent with the LC-MS/MS results. Thus, a replacement of HCF-1<sub>PRO</sub>-repeat residue E10 by serine converts the cleavage substrate into an O-GlcNAcylation substrate. This result suggests that the characteristics of the substrate determine whether to get O-GlcNAcylated or to get cleaved.



**Figure II-5: The HCF-1<sub>PRO</sub> repeat binds analogous to an O-GlcNAcylation substrate in the OGT catalytic domain.**

Overlay of two experimentally determined structures of OGT complexes (close-up view of the OGT catalytic domain). The HCF-1<sub>PRO</sub> repeat (carbons are colored in orange) and the CK-II peptide (carbons are colored in pink) are shown in ball-and-stick representation, lying over UDP-GlcNAc (carbons are colored in gray) in the OGT active site. Atoms are colored as follows: oxygen: red, nitrogen: blue, phosphor: green. For this overlay, the experimentally determined crystal structures of the OGT–UDP-5SGlcNAc–CK-II complex (PDB code 4GYG; Lazarus et al., 2012) and the OGT–UDP-5SGlcNAc–HCF-1<sub>PRO</sub>-repeat 2 E10Q complex (PDB code 4N3B; Lazarus et al., 2013) were used. In the HCF-1<sub>PRO</sub> repeat, E10Q was replaced by glutamate (E10) and in both structures, UDP-5SGlcNAc was replaced by UDP-GlcNAc. E10 of the HCF-1<sub>PRO</sub> repeat and the GlcNAc acceptor residue serine 9 (S9) of CK-II display an almost perfect overlap. Close-up view created with UCSF Chimera software by Dr. Vincent Zoete.

## II.4 HCF-1<sub>PRO</sub>-repeat-OGT association is sensitive to UDP-GlcNAc

In the past, OGT interactions with a single wild-type (WT) HCF-1<sub>PRO</sub> repeat could not be efficiently detected in *in vivo* co-immunoprecipitation assays. To study the properties of HCF-1<sub>PRO</sub>-repeat-OGT association, I established an *in vitro* HCF-1rep1-OGT “pull-down” binding assay that can detect WT HCF-1<sub>PRO</sub>-repeat-OGT interactions. Before performing the OGT-directed pull-down, I pre-incubated bacterially synthesized OGT with bacterially synthesized HCF-1rep1 substrates for 1 h in the presence of 1 mM UDP-GlcNAc and removed an input sample from the mixture. Subsequently, the mixture was incubated with OGT-directed antibody conjugates for 1 h (pull-down). Hence, the total incubation time of OGT and HCF-1rep1 in this assay was only 2 h. The proteins in the immunoprecipitated material and in the input were resolved by SDS-PAGE and OGT and HCF-1rep1 were detected by immunoblot using the respective antibodies. To establish the appropriate temperature for this assay, I tested HCF-1rep1 substrates containing WT or E10A mutated HCF-1<sub>PRO</sub> repeats for OGT binding at different temperatures. The E10A cleavage site mutant inhibits cleavage, and *in vivo* also enhances HCF-1-OGT association (Capotosti et al., 2011). Therefore this mutant served as a positive control for HCF-1rep1-OGT binding. The temperatures ranging from 4°C-42°C for both the pre-incubation and the OGT pull-down influenced HCF-1rep1-OGT binding (data not shown): At low temperatures (4°C and 10°C), OGT interactions with HCF-1rep1 WT and E10A were detectable. However, the E10A mutation did not cause enhancement of OGT binding, demonstrating that pull-down assays performed at 4°C and 10°C do not faithfully reflect *in vivo* HCF-1-OGT binding. Pull-down assays performed at 20°C, 30°C or 37°C faithfully reflected *in vivo* HCF-1-OGT binding and displayed weak WT HCF-1rep1-OGT binding and enhanced E10A HCF-1rep1-OGT binding. Pull-down assays performed at 42°C caused precipitation of the proteins in the assay. Thus, a temperature of 20°C was used for all subsequent pull-down assays. At temperatures below 37°C, HCF-1<sub>PRO</sub>-repeat cleavage is reduced (Capotosti and Herr, unpublished results) and allows to study HCF-1<sub>PRO</sub>-repeat-OGT binding without the occurrence of prominent cleavage, which would reduce the full-length WT binding substrate.

As HCF-1<sub>PRO</sub>-repeat cleavage is sensitive to UDP-GlcNAc concentrations (Figure II-2 B), I investigated whether HCF-1<sub>PRO</sub>-repeat-OGT binding displays similar sensitivity to UDP-GlcNAc. I tested OGT association of HCF-1rep1 WT and E10A at a range of different UDP-GlcNAc concentrations, using the established pull-down assay described above. I also tested the HCF-1rep1 T17-22A mutant in this assay (Figure II-6 A), to investigate whether alanine mutations in the threonine region disrupt OGT binding. Figure II-6 B shows the results of this HCF-1rep1-OGT binding assay in the absence or presence of increasing concentrations of UDP-GlcNAc (0-1.0 mM). In the absence of UDP-GlcNAc, HCF-1rep1 WT and E10A bound

indistinguishably well to OGT (lanes 1 and 2, panel a) but the mutant T17-22A bound more weakly to OGT (lane 3, panel a). With increasing concentrations of UDP-GlcNAc, both WT- and T17-22A-OGT binding decreased, whereas E10A-OGT binding did not decrease (compare lanes 2, 5, 8 and 12, panel a). These results suggest that UDP-GlcNAc inhibits OGT binding of HCF-1<sub>PRO</sub> repeats containing the wild-type cleavage site E10 (WT and T17-22A). In fact, in the presence of 1 mM UDP-GlcNAc, a clear difference between weak WT and strong E10A binding was observed (compare lanes 4 and 5, 7 and 8, 11 and 12, panel a) and binding of the T17-22A mutant (lane 13, panel a) was almost undetectable.

I concluded that OGT-bound UDP-GlcNAc inhibits or strains HCF-1<sub>PRO</sub>-repeat-OGT association. The effect of UDP-GlcNAc is even stronger when mutations in the HCF-1<sub>PRO</sub>-repeat threonine region are present (T17-22A). Hence, not only HCF-1<sub>PRO</sub>-repeat cleavage but also HCF-1<sub>PRO</sub>-repeat-OGT association is sensitive to UDP-GlcNAc. Intriguingly, maximum inhibition of WT HCF-1<sub>PRO</sub>-repeat-OGT association coincides with maximum WT HCF-1<sub>PRO</sub>-repeat cleavage in the presence of 1 mM UDP-GlcNAc (compare to cleavage of HCF-1<sub>rep1</sub> with 1 mM UDP-GlcNAc in Figure II-2 B). I thus hypothesized that the inhibition of HCF-1<sub>PRO</sub>-repeat-OGT association by UDP-GlcNAc might cause cleavage of the HCF-1<sub>PRO</sub> repeat.

The results obtained with the T17-22A mutant show that the alanine mutations in the threonine region indeed impair HCF-1<sub>PRO</sub>-repeat-OGT association because T17-22A HCF-1<sub>PRO</sub>-repeat-OGT binding was consistently less efficient than WT HCF-1<sub>PRO</sub>-repeat-OGT binding (Figure II-6 B). Results obtained in the presence of 1 mM UDP-GlcNAc (lanes 10-13) are shown in Figure 3 E in Lazarus et al. (2013). This result is in concordance with an examination of the OGT-UDP-5SGlcNAc-HCF-1<sub>PRO</sub>-repeat crystal structure, which shows that the threonines in the HCF-1<sub>PRO</sub>-repeat threonine region make interactions with the OGT TPR domain (Lazarus et al., 2013). I concluded, that the HCF-1<sub>PRO</sub>-repeat threonine region contributes largely to HCF-1<sub>PRO</sub>-repeat-OGT binding via interactions with the OGT TPR domain.



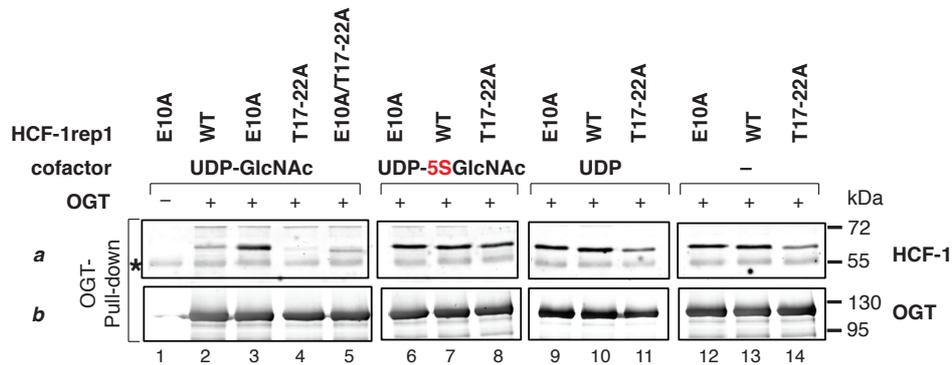
## II.5 Proper HCF-1<sub>PRO</sub>-repeat-OGT interactions require intact UDP-GlcNAc

As HCF-1<sub>PRO</sub>-repeat-OGT association is sensitive to UDP-GlcNAc, I asked next whether this association displays similar sensitivity to UDP-5SGlcNAc. As described above, UDP-5SGlcNAc inhibits HCF-1<sub>rep1</sub> cleavage and O-GlcNAcylation and was shown to bind to OGT like UDP-GlcNAc would, without altering the conformation of the active site (Lazarus et al., 2012). To test if UDP-5SGlcNAc changes the HCF-1<sub>PRO</sub>-repeat-OGT binding mode, I performed HCF-1<sub>rep1</sub>-OGT pull-down assays (described in II.4) in the presence of either UDP-GlcNAc, UDP-5SGlcNAc or UDP. Figure II-7 shows an *in vitro* HCF-1<sub>rep1</sub>-OGT binding assay with HCF-1<sub>rep1</sub> constructs containing the wild-type (WT) HCF-1<sub>PRO</sub> repeat or the E10A- or T17-22A-mutated HCF-1<sub>PRO</sub> repeats. In one instance (Figure II-7, lane 5), I included a double mutant containing both, the E10A mutation and the T17-22A mutations (E10A/T17-22A) in the assay. This double mutant is not cleaved by OGT (data not shown). As expected, in the presence of UDP-GlcNAc, the WT HCF-1<sub>PRO</sub> repeat displayed weak OGT binding (lane 2, panel a), whereas the E10A mutation (lane 3, panel a) enhanced, and the T17-22A mutations abrogated binding (lane 4, panel a) with respect to WT. The double mutant associated more strongly to OGT than the mutant containing the T17-22A mutations alone (compare lanes 4 and 5, panel a), suggesting that the enhancing effect of E10A counteracts the binding defect caused by the mutations in the HCF-1<sub>PRO</sub>-repeat threonine region. This result also shows that the HCF-1<sub>PRO</sub>-repeat threonine region is important for HCF-1<sub>PRO</sub>-repeat-OGT association, even in the presence of the E10A mutation.

In the presence of UDP-5SGlcNAc, WT HCF-1<sub>PRO</sub>-repeat-OGT binding increased with respect to interactions in the presence of UDP-GlcNAc (compare lane 7 to lane 2). Remarkably, also the T17-22A mutant associated more strongly with OGT than in the presence of UDP-GlcNAc (compare lane 4 to lane 8). In fact, all three HCF-1<sub>rep1</sub> constructs bound with the same efficiency to OGT (compare lanes 6-8). These results show that UDP-5SGlcNAc in the OGT active site alters the HCF-1<sub>PRO</sub>-repeat-OGT binding mode drastically. In the presence of UDP, WT and E10A HCF-1<sub>rep1</sub> (lanes 9 and 10) associated strongly and indistinguishably to OGT. The association of the T17-22A mutant was slightly weaker than WT and E10 OGT association (lane 11), suggesting that, also in the presence of UDP, the threonine region is important for OGT association. In the absence of any cofactor, the WT HCF-1<sub>PRO</sub>-repeat as well as the E10A and the T17-22A mutants bound to OGT in a manner similar to binding in the presence of UDP (compare lanes 9-11 to 12-14).

In summary, these results show that not only HCF-1<sub>PRO</sub>-repeat cleavage (see II.1), but also the inhibiting effect of UDP-GlcNAc on HCF-1<sub>PRO</sub>-repeat-OGT association is dependent on the sugar moiety of UDP-GlcNAc because UDP and UDP-5SGlcNAc caused changes of the HCF-1<sub>PRO</sub>-repeat-OGT binding mode. Hence, proper HCF-1<sub>PRO</sub>-repeat-OGT

interactions and HCF-1<sub>PRO</sub>-repeat cleavage are dependent on the presence of intact UDP-GlcNAc. The results obtained with UDP-5SGlcNAc, however, were surprising: A sole replacement of the oxygen ring atom in the sugar by sulfur inhibits HCF-1<sub>PRO</sub>-repeat cleavage and changes the HCF-1<sub>PRO</sub>-repeat–OGT binding mode drastically. In the presence of UDP-5SGlcNAc, the T17-22A mutant associates more strongly to OGT than in any other tested condition (UDP-GlcNAc, UDP, no cofactor), but the reasons for this are currently unknown. It is possible that the oxygen atom of the sugar ring plays a role in HCF-1<sub>PRO</sub>-repeat–OGT binding and proteolysis.



**Figure II-7: The sugar moiety of UDP-GlcNAc in the OGT active site is important for proper HCF-1<sub>PRO</sub>-repeat–OGT association.**

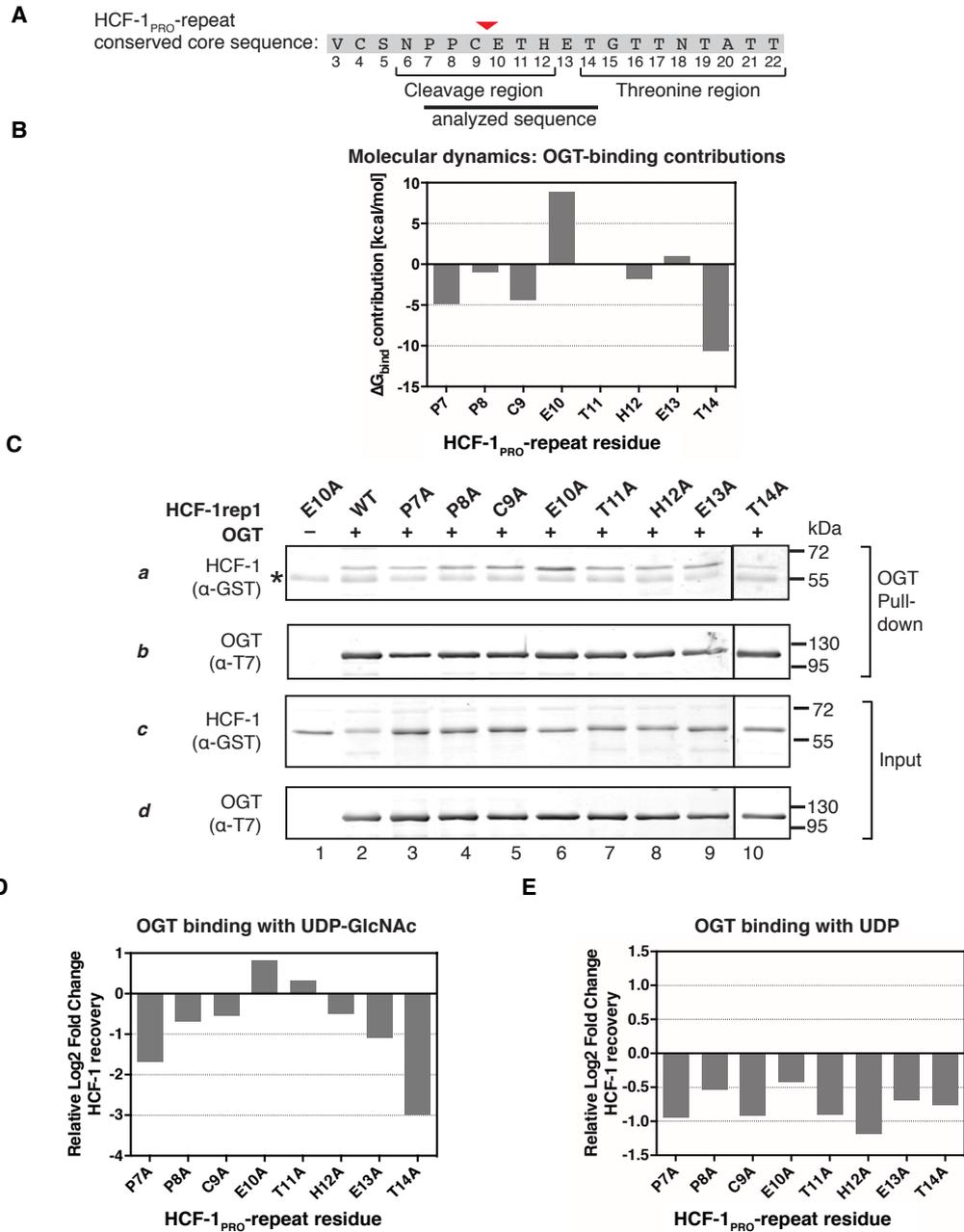
*In vitro* HCF-1rep1–OGT binding assay in the presence of 1 mM UDP-GlcNAc (lanes 1-5), UDP-5SGlcNAc (lanes 6-8), UDP (lanes 9-11), or in the absence of cofactor (lanes 12-14). Shown are panels of the OGT pull-down (panels a and b; 100 % of total pull-down). Immunoblotting with antibodies directed to GST or to the T7 epitope was used to detect GST–HCF-1rep1 and OGT, respectively. IgG heavy chain (\*).

## II.6 E10 in the HCF-1<sub>PRO</sub>-repeat cleavage region displays unfavorable interactions with the OGT–UDP-GlcNAc complex

E10 in the HCF-1<sub>PRO</sub>-repeat cleavage region displays particular OGT-binding properties in the presence of UDP-GlcNAc, the reasons for this, however, remained unclear (see II.4 and II.5 above). In Lazarus et al. (2013) we reported that the HCF-1<sub>PRO</sub>-repeat threonine region binds stably to the OGT TPR region through a network of hydrogen bonds contacting the threonine side-chains and backbones. The binding mode of the HCF-1<sub>PRO</sub>-repeat cleavage region to OGT, however, was less well understood, as binding of the wild-type HCF-1<sub>PRO</sub>-repeat to OGT in crystallization conditions ultimately leads to proteolysis. A replacement of E10 by E10Q in the HCF-1<sub>PRO</sub> repeat and substitution of UDP-GlcNAc by UDP-5SGlcNAc prevented proteolysis. In the OGT–UDP-5SGlcNAc–HCF-1<sub>PRO</sub>-repeat E10Q crystal structure (PDB code 4N3B), the E10Q side-chain is located in close proximity to the sugar ring of UDP-5SGlcNAc (see Figure II-5; Lazarus et al., 2013). To clarify the role of the wild-type HCF-1<sub>PRO</sub>-repeat cleavage region for HCF-1<sub>PRO</sub>-repeat proteolysis, Dr. Vincent Zoete replaced the glutamine side-chain at position 10 by wild-type glutamate and the UDP-5SGlcNAc cofactor by UDP-GlcNAc. Based on this model, individual side-chain contributions to OGT binding for residues P7 to T14 (schematic in Figure II-8 A) were estimated using the Molecular Mechanics – Generalized Born Surface Area (MM-GBSA) approach (Figure II-8 B). The simulations predicted that residues T11-E13 of the HCF-1<sub>PRO</sub> repeat do not contribute, whereas T14 makes a large and favorable contribution to OGT binding. Interestingly, these calculations showed that the E10 residue is the only residue in the analyzed region that displays highly unfavorable interactions with the OGT–UDP-GlcNAc complex.

Next, I tested the HCF-1rep1 mutants containing individual HCF-1<sub>PRO</sub>-repeat alanine substitutions from P7A to T14A in an *in vitro* HCF-1rep1–OGT binding assay in the presence of UDP-GlcNAc (Figure II-8 C). Strikingly, only the E10A mutant within the analyzed region enhanced HCF-1rep1–OGT association considerably, as shown by immunoblot (Figure II-8 C, compare lane 6 to adjacent lanes, panel a). The quantification of the bands on this immunoblot showed this result even more clearly (Figure II-8 D). I concluded that the E10 residue is not only unique for its cleavage properties, but also for its OGT-binding properties within the analyzed region. This is consistent with the molecular dynamics analysis in Figure II-8 B that predicted E10 to be the only residue with largely unfavorable contributions to OGT binding. In contrast, the replacement of UDP-GlcNAc by UDP in the *in vitro* HCF-1rep1–OGT binding assay did not promote increased HCF-1rep1 recovery by the E10A mutation (Figure II-8 E), suggesting that the unfavorable effect of E10 is dependent on the sugar moiety of UDP-GlcNAc. Thus, both molecular dynamics computational analysis and physical

interaction assays argue that the E10 residue is the only residue in the HCF-1<sub>PRO</sub> repeat that makes unfavorable interactions with the OGT–UDP-GlcNAc complex. This is in agreement with results obtained in sections II.4 and II.5, where a negative effect by UDP-GlcNAc, but not by UDP, in HCF-1<sub>PRO</sub>-repeat–OGT association was observed.



**Figure II-8: The HCF-1<sub>PRO</sub>-repeat residue E10 displays unfavorable interactions with the OGT–UDP-GlcNAc complex.**

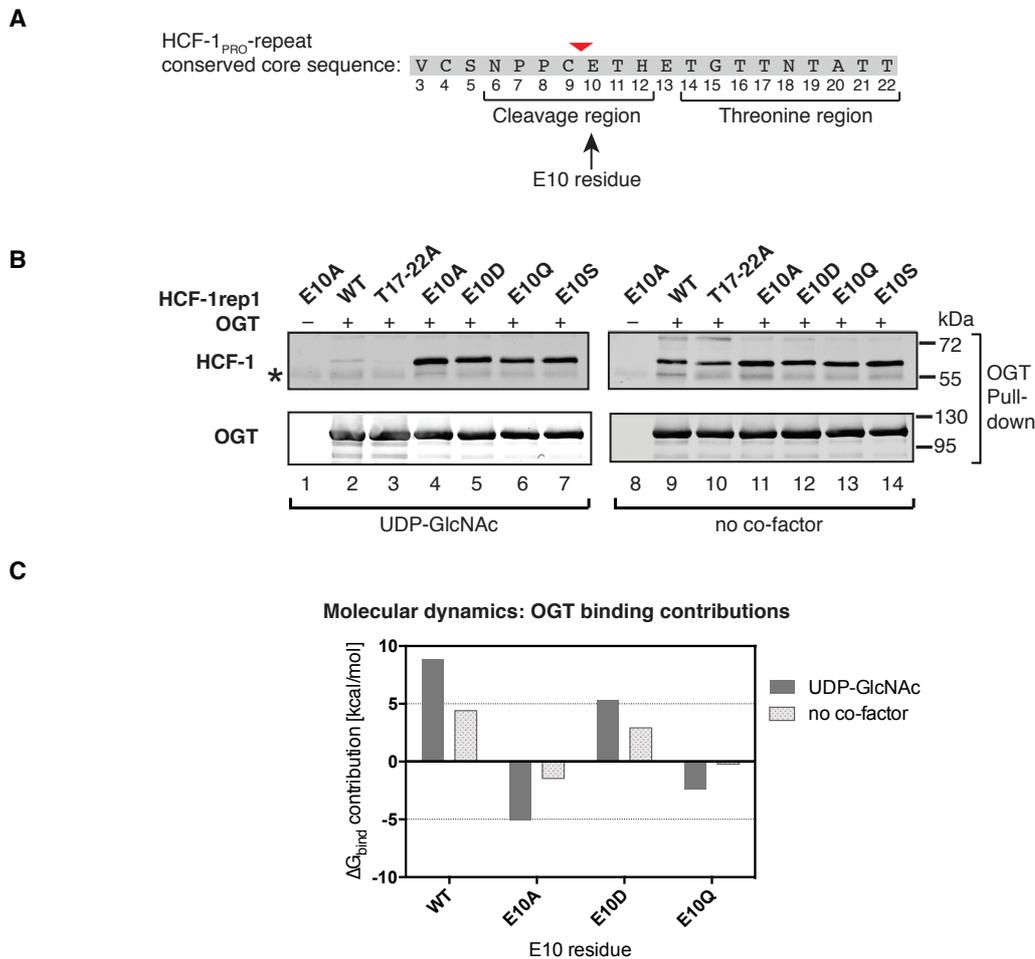
(A) The HCF-1<sub>PRO</sub>-repeat conserved core sequence is shown. The red arrowhead indicates the cleaved peptide bond and the analyzed sequence is highlighted. (B) E10 displays an unfavorable contribution to OGT binding in a molecular dynamics analysis. Calculated contributions of HCF-1<sub>PRO</sub>-repeat residues P7, P8, C9, E10, T11, H12, E13 and T14 to the binding free energy,  $\Delta G_{\text{bind}}$ , for the association between the HCF-1<sub>PRO</sub>-repeat and OGT. Negative values correspond to favorable contributions to binding. (C) *In vitro* HCF-1rep1–OGT pull-down assay in the presence of UDP-GlcNAc with HCF-1rep1 constructs containing individual alanine substitutions in the HCF-1<sub>PRO</sub>-repeat (P7A–T14A). Shown are 100 % of OGT pull-down (panels a and b) and 11 % of the input (panels c and d). Immunoblotting with antibodies directed to GST or to the T7 epitope was used to detect GST–HCF-1rep1 and OGT, respectively. IgG heavy chain (\*). (D) Quantified HCF-1 recovery from the pull-down assay in (C) represented as log<sub>2</sub> fold change relative to wild-type. The immunoblot was analyzed using LI-COR Image Quant quantification software and HCF-1 recovery was calculated as pull-down (values from panel a) over total input (values from panel c extrapolated to 100% input). (E) Quantified HCF-1 recovery from an HCF-1rep1–OGT pull-down assay in the presence of UDP (immunoblot not shown). Quantification as described in (D).

## II.7 Strains caused by E10 in the OGT catalytic domain are specific to the glutamate side-chain structure

The observation that E10 is the only residue within the HCF-1<sub>PRO</sub> repeat displaying particular OGT-binding properties in the presence of UDP-GlcNAc, led me to ask if this property is dependent on the side-chain structure of the glutamate. I therefore performed *in vitro* binding assays with an enlarged set of E10 mutants described earlier (Figure II-4). Figure II-9 B shows HCF-1<sub>rep1</sub>-OGT binding in the presence (left panels) or absence (right panels) of UDP-GlcNAc. As expected, in the presence of the cofactor, the construct containing wild-type (WT) E10 bound weakly to OGT (lane 2, upper panel), the T17-22A mutant displayed even more reduced binding (lane 3, upper panel), and the E10A mutant displayed enhanced binding (lane 4, upper panel). The E10 to aspartate (E10D), as well as the E10 to glutamine (E10Q) mutants, enhanced OGT binding similarly to E10A, albeit the similarity of the aspartate and glutamine side-chains to the WT glutamate side-chain (lanes 5 and 6, upper panel). The E10 to serine (E10S) mutant was tested because, unlike the E10D and E10Q mutants, the E10S HCF-1<sub>PRO</sub> repeat represents an active substrate for O-GlcNAcylation (Lazarus et al., 2013). In this binding assay, also the E10S mutant enhanced association to OGT (lane 7), suggesting that the inhibitory effect of the E10 side-chain on OGT association is specific to the glutamate side-chain, which is an active cleavage substrate. When UDP-GlcNAc was omitted from the assay (right-hand panels), binding of the E10 mutants was not noticeably altered (lanes 11-14), whereas OGT binding of the HCF-1<sub>PRO</sub> repeats containing wild-type E10 (WT and T17-22A) increased considerably (compare lanes 9 and 10 to lanes 2 and 3). These results suggest that the glutamate at position 10 interacts unfavorably with the OGT-UDP-GlcNAc complex. This interaction is specific to the glutamate side-chain structure, as side-chains with similar charge and structure at the same position alter the binding mode.

To understand the mechanism through which E10 causes an unfavorable interaction with the OGT-UDP-GlcNAc complex, Dr. Vincent Zoete estimated the OGT-binding contributions of E10, E10A, E10D and E10Q side-chains in the presence or absence of UDP-GlcNAc by molecular dynamics (Figure II-9 C). WT E10 was predicted to display the most unfavorable OGT interaction among the tested side-chains. E10D displayed more favorable interactions than E10, presumably because the similarly negatively charged aspartate side-chain is one carbon atom shorter than that of glutamate, and thus causes less structural impairments in the OGT-UDP-GlcNAc complex. The neutral side-chains of E10A and E10Q were both predicted to display favorable interactions with OGT. When UDP-GlcNAc was omitted from the OGT active site in the molecular dynamics simulation, the contributions to the binding free energy of the unfavorable contacts (E10 and E10D) decreased, suggesting

that the strains, occurring in the presence of UDP-GlcNAc, were removed. These results are in line with the observations in the pull-down assay in Figure II-9 B. Taking both results together, I concluded that the E10 side-chain in the active site of OGT causes specific strains in the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex.

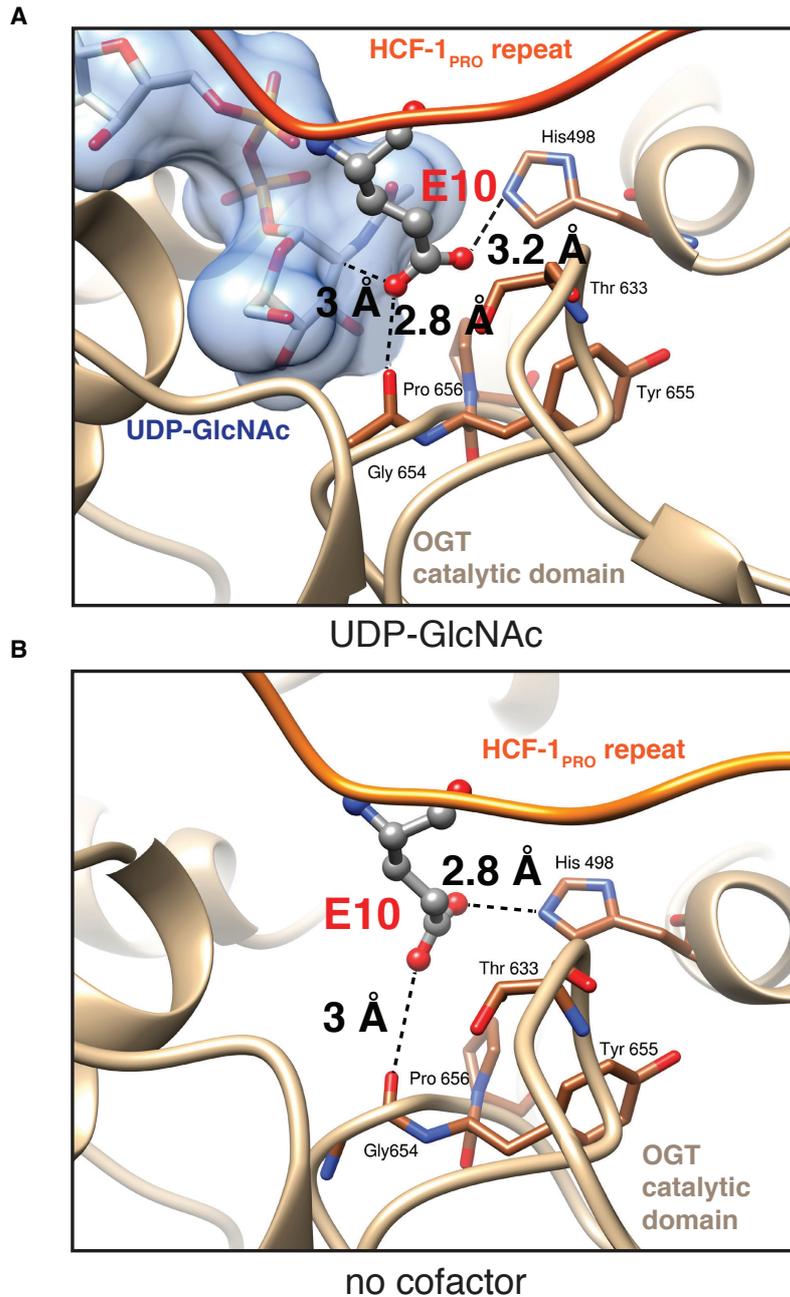


**Figure II-9: The E10 side-chain strains HCF-1<sub>PRO</sub>-repeat–OGT interactions in the presence of UDP-GlcNAc.**

(A) The HCF-1<sub>PRO</sub>-repeat conserved core sequence is shown. The red arrowhead indicates the cleaved peptide bond and the E10 cleavage site residue is highlighted. (B) The side-chain of E10 interacts with UDP-GlcNAc in the OGT active site. *In vitro* HCF-1–OGT pull-down assay with an enlarged set HCF-1rep1 mutants. Constructs with wild-type (WT, lanes 2 and 9), mutated E10 residues (lanes 4-7; lanes 11-14) and T17-T22A mutations (lanes 3 and 10) were incubated in the presence (left panels) or absence (right panels) of UDP-GlcNAc. Immunoblotting with antibodies directed to GST or to the T7 epitope was used to detect GST–HCF-1rep1 and OGT, respectively. IgG heavy chain (\*). (C) E10 makes unfavorable contributions to OGT binding. Calculated contributions of the wild-type (WT, E10), alanine (E10A), aspartate (E10D) or glutamine (E10Q) residue at position 10 to OGT binding. The binding free energy,  $\Delta G_{\text{bind}}$ , for the association between the E10 residue and OGT is shown. Negative values correspond to favorable contributions to binding.

To understand the E10–OGT interactions better, we examined the location of the HCF-1<sub>PRO</sub>-repeat E10 side-chain and UDP-GlcNAc in the OGT active site of the crystal structure (PDB code 4N3C, Lazarus et al., 2013). Consistent with the above-described results, we noticed that the E10 side-chain is located in close proximity (2.8 Å) of the carbonyl oxygen of glycine 654 in the OGT catalytic domain, potentially causing an unfavorable electrostatic repulsion (Figure II-10 A). The E10 carboxylate functional group is maintained in this unfavorable position by the glucose moiety of UDP-GlcNAc. When we performed a molecular dynamics simulation without UDP-GlcNAc (representative snapshot along the trajectory shown in Figure II-10 B), the E10 residue can change its position to prevent the unfavorable interaction with glycine 654 and appears to form a favorable interaction with a nitrogen atom in the imidazole ring of histidine 498 in the OGT catalytic domain.

Together, these results suggest that UDP-GlcNAc inhibits HCF-1<sub>PRO</sub>-repeat–OGT association by imposing that the E10 side-chain be in an unfavorable position within the complex. This incompatibility may cause strains in the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex, which, I hypothesize, initiates or facilitates HCF-1<sub>PRO</sub>-repeat proteolysis. It is thus very likely that the HCF-1<sub>PRO</sub>-repeat threonine region serves as high-affinity bindings interface for the OGT TPR domain and allows for binding of the OGT catalytic domain to the HCF-1<sub>PRO</sub>-repeat cleavage region. Moreover, these results elucidate the reasons for the observed enhancement of OGT association by the HCF-1<sub>PRO</sub> repeat E10A mutation. E10A enhances OGT association because the strains between the E10 side-chain and the OGT–UDP-GlcNAc complex were removed. Other amino acid side-chains such as the ones of aspartate, glutamine or serine have similar effects because they display less unfavorable interactions in the OGT complex. I thus propose that the mechanism of OGT-mediated HCF-1 proteolysis involves strains originating from the cleavage substrate.



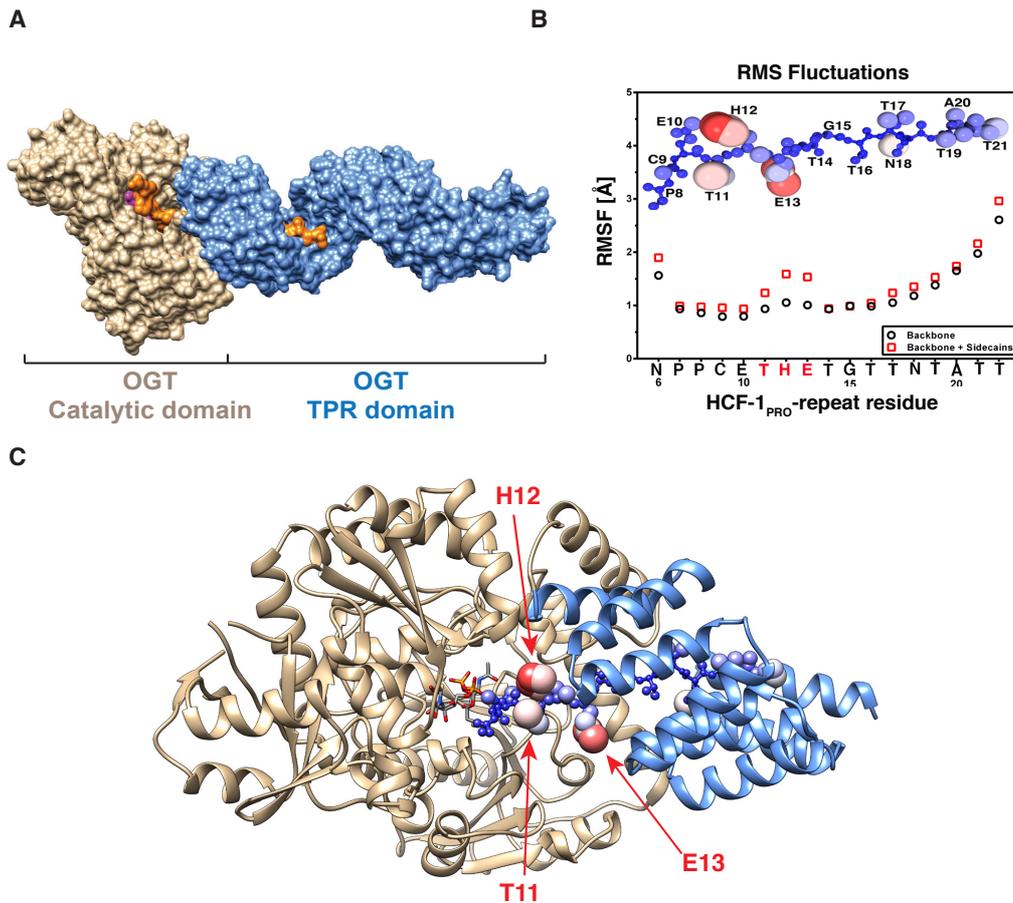
**Figure II-10: The E10 side-chain makes unfavorable interactions with the OGT catalytic domain in the presence of UDP-GlcNAc.**

(A) Close-up view (4N3C crystal structure; Lazarus et al., 2013) of the OGT active site with the HCF-1<sub>PRO</sub> repeat and UDP-GlcNAc. The deprotonated E10 oxygen atom exhibits an unfavorable interaction with the backbone carbonyl of OGT glycine 654 and is located in close proximity of the carbon atom 2 of the sugar moiety of UDP-GlcNAc. (snapshot by Dr. Vincent Zoete). (B) Snapshot from a molecular dynamics simulation (based on the 4N3B structure; Lazarus et al., 2013) of the OGT active site in complex with the HCF-1<sub>PRO</sub> repeat without UDP-GlcNAc. The displayed frame is representative of the average distances sampled along the simulations. In (A) and (B), the E10 side-chain is shown in ball and stick representation (carbons: gray, nitrogen: blue, oxygens: red), and dashed lines indicate distances between atoms.

## II.8 Refinement of the HCF-1<sub>PRO</sub>-repeat bipartite structure

The crystal structure of the OGT–UDP-5SGlcNAc–HCF-1<sub>PRO</sub>-repeat 2 E10Q complex (PDB code 4N3B; Lazarus et al., 2013) was invaluable for my follow-up studies. The structure allowed the creation of an authentic model of full-length OGT bound to UDP-GlcNAc and the HCF-1<sub>PRO</sub>-repeat wild-type peptide using computational molecular dynamics (Figure II-11 A).

Capotosti et al. (2011) divided the HCF-1<sub>PRO</sub> repeat into two regions, termed cleavage region and threonine region (Figure I-12). Using bacterially synthesized HCF-1rep1 substrates for robust *in vitro* cleavage assays, I confirmed that HCF-1<sub>PRO</sub>-repeat cleavage can be inhibited by substitutions of individual HCF-1<sub>PRO</sub>-repeat residues by alanine (see Figure II-4 B). The residues P7–E10 are important for cleavage because they exhibit reduced or no HCF-1<sub>PRO</sub>-repeat cleavage when replaced by alanine. In contrast, alanine mutations of residues T11, H12 and E13 did not display a strong effect on cleavage suggesting that: (i) these residues are not important for HCF-1<sub>PRO</sub>-repeat cleavage or (ii) alanine substitutions of these residues do not affect structural properties that are important for cleavage. To test the latter hypothesis, we determined the flexibility of the HCF-1<sub>PRO</sub>-repeat structure bound to OGT. A crystal structure represents only one possible conformation that a protein complex can adopt, but thermal fluctuations of a given system can provide a more realistic, dynamic picture of the complex and can assess flexibility or rigidity of structures. Dr. Ute Roehrig performed molecular dynamics simulations of the HCF-1<sub>PRO</sub>-repeat 2 (residues 5-24) in complex with full-length human OGT and UDP-GlcNAc in water at room temperature to obtain thermal fluctuations of the system under ambient conditions (based on structure 4N3B; E10Q replaced by E10, UDP-5SGlcNAc replaced by UDP-GlcNAc using molecular dynamics). The Root-Mean-Square Fluctuations (RMSF), i.e., how much each atom fluctuates around its average position, were determined and are illustrated in Figure II-11 B by a graphical representation of obtained RMSF values and in Figure II-11 C by color and size coding of the HCF-1<sub>PRO</sub> repeat embedded in the OGT structure. Residues T11, H12 and E13 of the HCF-1<sub>PRO</sub> repeat are characterized by higher structural fluctuations than observed in the rest of the peptide, suggesting that T11–E13 are not tightly bound to OGT and display structural flexibility when in complex with OGT.

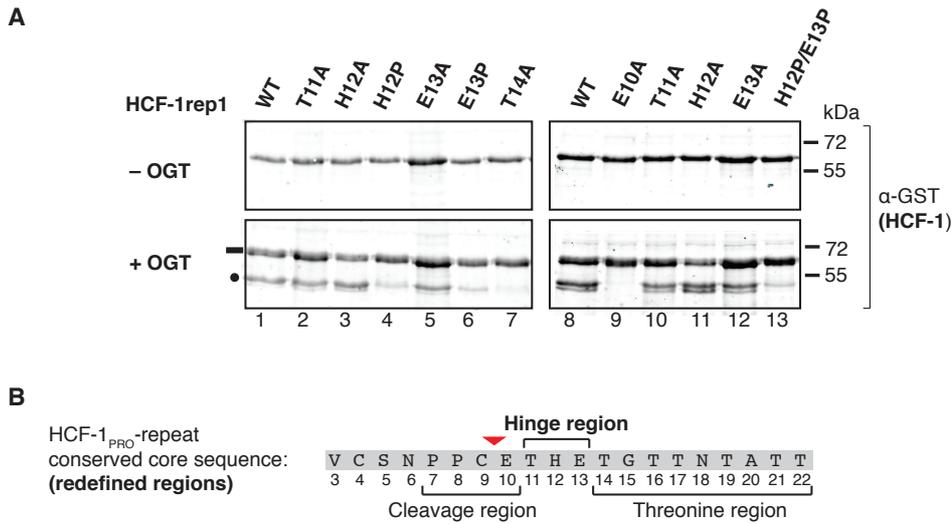


**Figure II-11: The HCF-1<sub>PRO</sub> repeat contains a flexible region between the E10 cleavage site and the threonine region.**

(A) Structure of full-length human OGT in complex with wild-type HCF-1<sub>PRO</sub> repeat 2 (orange) and UDP-GlcNAc (in purple, behind the HCF-1<sub>PRO</sub> repeat, not entirely visible). The OGT catalytic domain is colored in beige, the 13.5 TPR domain is colored in blue. The structure of the OGT TPR region (Jinek et al., 2004) has been assembled with the structure of OGT containing 4.5 TPRs in complex with the HCF-1<sub>PRO</sub>-repeat 2 E10Q and UDP-5SGlcNAc (Lazarus et al., 2013) by molecular modeling. UDP-5SGlcNAc has been replaced by UDP-GlcNAc and the E10Q HCF-1<sub>PRO</sub> repeat has been replaced by the wild-type repeat (Dr. Vincent Zoete). (B) RMSF of the backbone (black open circles) and the full peptide (red open squares) of the HCF-1<sub>PRO</sub> repeat. Residues 11-13 of the HCF-1<sub>PRO</sub> repeat are characterized by higher structural fluctuations. Inset: The structure of the HCF-1<sub>PRO</sub> repeat, with atom size and color given by the respective RMSF (rigid=blue, small; flexible=red, large). (C) Structure of the HCF-1<sub>PRO</sub> repeat, with atom size and color given by the respective RMSF as in (B), embedded in the OGT 4.5 TPR structure with UDP-GlcNAc. Simulations in (B) and (C) performed by Dr. Ute Roehrig.

To examine if the structural flexibility of the residues T11, H12 and E13 is important for HCF-1<sub>PRO</sub>-repeat cleavage, I substituted the residues predicted to display the highest flexibility, H12 and E13, individually or simultaneously by proline. Proline is known to increase the rigidity of peptides due to its cyclic structure. Thus, replacing flexible residues by proline may result in a more rigid structure. Figure II-12 A shows the result of an *in vitro* cleavage assay with HCF-1<sub>PRO</sub>-repeat alanine and proline mutations. I confirmed that individual replacement of H12 or E13 by alanine did not affect HCF-1<sub>PRO</sub>-repeat cleavage (left-hand panels, lanes 3 and 5, respectively). Individual replacement of H12 and E13 by proline, however, decreased HCF-1<sub>PRO</sub>-repeat cleavage (left-hand panels, lanes 4 and 6, respectively) with respect to wild-type HCF-1<sub>PRO</sub>-repeat cleavage (left-hand panels, lane 1). The double mutant containing simultaneous H12P and E13P mutations displayed even less cleavage activity (right-hand panels, lane 6) with respect to the mutants containing single proline mutations, suggesting that the flexibility of H12 and E13 is important for cleavage. Nevertheless, it cannot be ruled out that the effect of proline on the structure of the HCF-1<sub>PRO</sub> repeat might have an influence on HCF-1<sub>PRO</sub>-repeat cleavage.

OGT contains a so-called hinge region between the TPRs and the catalytic domain (Lazarus et al., 2011), and this OGT hinge region surrounds HCF-1<sub>PRO</sub>-repeat residues T11–E13 in the OGT–HCF-1<sub>PRO</sub>-repeat complex (Lazarus et al., 2013). Analogous to the OGT hinge region, we propose that the HCF-1<sub>PRO</sub> repeat also contains a conserved, flexible hinge consisting of residues T11, H12, and E13. Therefore we redefined the HCF-1<sub>PRO</sub>-repeat regions as follows: cleavage region (residues P7–E10), Hinge region (residues T11, H12, and E13) and threonine region (T14–T22; see Figure II-12 B).



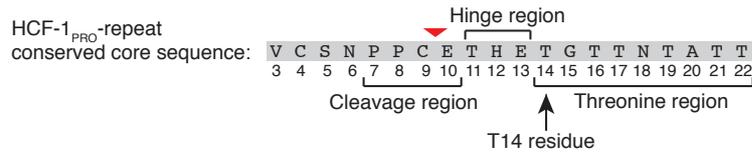
**Figure II-12: Refinement of the regions within the HCF-1<sub>PRO</sub> repeat.**

(A) *In vitro* cleavage assay of HCF-1rep1 constructs containing the wild-type (WT) HCF-1<sub>PRO</sub> repeat or alanine or proline mutations of selected residues within the HCF-1<sub>PRO</sub> repeat. The positions of uncleaved precursor proteins (–) and N-terminal cleavage products (•) are indicated. (B) The redefined regions of the HCF-1<sub>PRO</sub> repeat: The cleavage region has been trimmed to residues P7–E10. The threonine region comprises residues T14–T22. In between the cleavage and the threonine regions lies the Hinge region comprising residues T11–E13. The residues in the Hinge region are characterized by higher structural flexibility with respect to the adjacent regions when the HCF-1<sub>PRO</sub> repeat is bound to OGT.

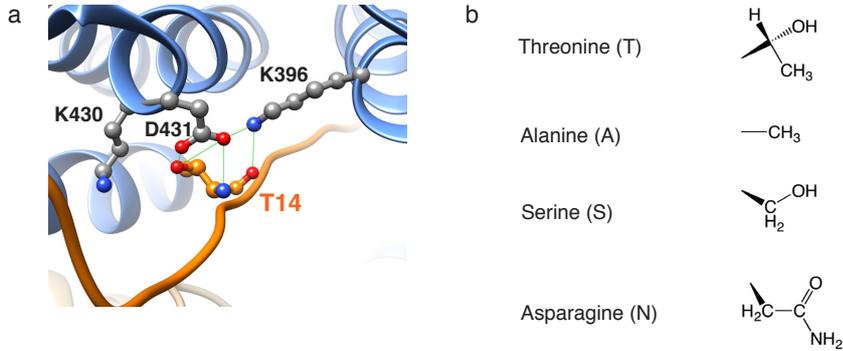
## II.9 Residue T14 in the threonine region is important for interactions with residues in the OGT TPR domain

My studies of HCF-1<sub>PRO</sub>-repeat–OGT association showed that residue T14 in the HCF-1<sub>PRO</sub>-repeat threonine region contributes largely to OGT binding (see Figure II-8 C and D). This finding led me to focus on T14 interactions with the OGT TPR domain (Figure II-13 A). Dr. Vaibhav Kapuria performed mutational analyses in the human OGT protein to identify residues required for HCF-1 proteolysis. Alanine substitutions of TPR residues D431 and K396, but not K430, greatly reduced HCF-1<sub>rep1</sub> cleavage in an *in vitro* cleavage assay (Dr. Vaibhav Kapuria, data not shown). An examination of the location of these residues in the OGT–HCF-1<sub>PRO</sub>-repeat crystal structure (PDB code 4N3B) revealed that D431 and K396, but not K430, are located in close proximity to T14 of the HCF-1<sub>PRO</sub> repeat (Figure II-13 B, a). D431 and K396 are predicted to form hydrogen bonds to the hydroxyl group and to the peptide backbone of T14. This examination was in line with the finding that a T14A mutation inhibits OGT association, as the alanine side-chain cannot form hydrogen bonds. To determine the specificity of the hydrogen bonds formed between D431 and K396 and the T14 side-chain and backbone, I replaced T14 by two different amino acids: Serine (S), which like threonine contains a hydroxyl group capable of forming hydrogen bonds but lacks a methyl group, and asparagine (N), which is known to form strong hydrogen bonds through its amide group (see Figure II-13 B, b for side-chain structures). As a replacement of T14 by alanine leads to a reduction of cleavage as shown in Figure II-4 B, replacement of T14 by serine or asparagine could reactivate cleavage. Figure II-13 C shows the result of an *in vitro* cleavage assay with the described mutants. As expected, the T14A mutation inhibited HCF-1<sub>PRO</sub>-repeat cleavage strongly (Figure II-13 C, lane 2, lower panel). Substitution by serine (lane 3) or asparagine (lane 4) did not reactivate T14A cleavage, suggesting that the side-chain structure of T14 i.e., its additional methyl group compared to serine is crucial for the contacts to residues in the OGT TPR domain. I concluded that the hydrogen bonds formed between T14 and residues in the OGT TPR domain (D431 and K396) are highly specific. As other threonines in the HCF-1<sub>PRO</sub>-repeat threonine region are also predicted to contribute to OGT TPR domain binding through a network of hydrogen bonds (Lazarus et al., 2013), it is conceivable that also these interactions are highly specific.

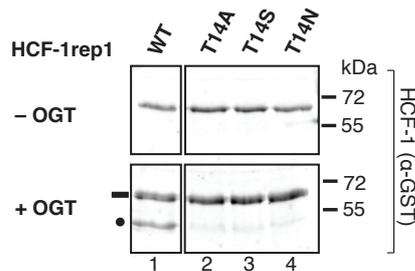
A



B



C



**Figure II-13: Residue T14 in the threonine region is important for interactions with residues in the OGT TPR domain.**

(A) The HCF-1<sub>PRO</sub> repeat with its redefined regions is shown. The red arrowhead indicates the cleaved peptide bond. In this experiment, focus is given on residue T14 in the threonine region. (B) a, Structural snapshot of the HCF-1<sub>PRO</sub>-repeat-OGT crystal structure 4N3B (Lazarus et al., 2013) around T14 of the HCF-1<sub>PRO</sub> repeat (orange). Helices of the OGT TPR region are colored in blue. Hydrogen bonds are illustrated as light blue lines. b, Side-chain structures of the amino acids threonine, alanine, serine and asparagine to study the role of T14 in OGT TPR binding. (C) *In vitro* cleavage assay with HCF-1rep1 constructs containing wild-type (WT, lane 1), alanine (lane 2), serine (lane 3) or asparagine (lane 4) substitutions at position T14 in the HCF-1<sub>PRO</sub> repeat. The positions of uncleaved precursor proteins (-) and N-terminal cleavage products (\*) are indicated.

## Discussion

### The HCF-1<sub>PRO</sub> repeat contains a flexible region (Hinge region)

OGT-mediated HCF-1<sub>PRO</sub>-repeat cleavage requires a large recognition sequence of 21 amino acids within the HCF-1<sub>PRO</sub> repeat (Capotosti et al., 2011). Molecular dynamics analyses of eight residues within this sequence in the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex predicted higher structural flexibility of HCF-1<sub>PRO</sub> repeat residues T11, H12 and E13 with respect to adjacent residues (Figure II-11). It has been proposed that OGT opens its substrate-binding cleft by a hinge-like motion around a pivot point, called hinge region, between the TPR domain and the catalytic domain (Lazarus et al., 2011). Residues T11, H12 and E13, when bound to OGT, are surrounded by this OGT hinge region, thus their flexibility is likely to be caused by the flexible OGT hinge. This hypothesis is in agreement with both molecular dynamics analyses and *in vitro* HCF-1–OGT pull-down assays showing that T11, H12 and E13 do not contribute to OGT binding (Figure II-8 B and D). The flexibility of T11, H12 and E13 (together called HCF-1<sub>PRO</sub>-repeat Hinge region) appears to be important for HCF-1<sub>PRO</sub>-repeat cleavage (Figure II-12 A). The HCF-1<sub>PRO</sub>-repeat Hinge region might be important to accommodate the HCF-1<sub>PRO</sub> repeat inside the OGT enzyme by providing flexibility in the area formed between the OGT hinge and the HCF-1<sub>PRO</sub> repeat. Indeed, it has been suggested that the movement of the OGT hinge region might be required to accommodate large substrates (Lazarus et al., 2011), such as is HCF-1. Nevertheless, residues T11, H12 and E13 are highly conserved (like their adjacent residues) arguing for a function of these residues beyond simply contributing to flexibility.

### HCF-1<sub>PRO</sub>-repeat proteolysis requires intact UDP-GlcNAc

UDP-GlcNAc is the donor substrate for OGT's glycosyltransferase activity, but it is also required for HCF-1<sub>PRO</sub>-repeat proteolysis. Indeed, the efficiency of HCF-1<sub>PRO</sub>-repeat cleavage increases with increasing concentrations of UDP-GlcNAc (Figure II-2 B). Two reasons for this requirement could be: (i) UDP-GlcNAc allosterically activates the OGT active site for proteolysis. Allosteric modulators (proteins or small molecules) can bind to a region of the enzyme that does not participate directly in substrate recognition and processing (allosteric site). The allosteric site can enhance substrate to product transition through conformational changes in the enzyme (Drag and Salvesen, 2010). Evidence for allosteric activation by UDP-GlcNAc could not be found, given that UDP-5SGlcNAc does not allow for cleavage despite its similar OGT-binding properties (Lazarus et al., 2012). Another possibility is that (ii) UDP-GlcNAc could actively participate in the reaction mechanism occurring during proteolysis and would thus not be only a cofactor, but also a cosubstrate for the reaction

(UDP-GlcNAc is also a cosubstrate for the O-GlcNAcylation reaction, see section I.6.3). A cosubstrate is a cofactor that is utilized or transformed during the enzymatic reaction, and is therefore a substrate. The second, preferred hypothesis is consistent with the fact that UDP and UDP-5SGlcNAc inhibit proteolysis.

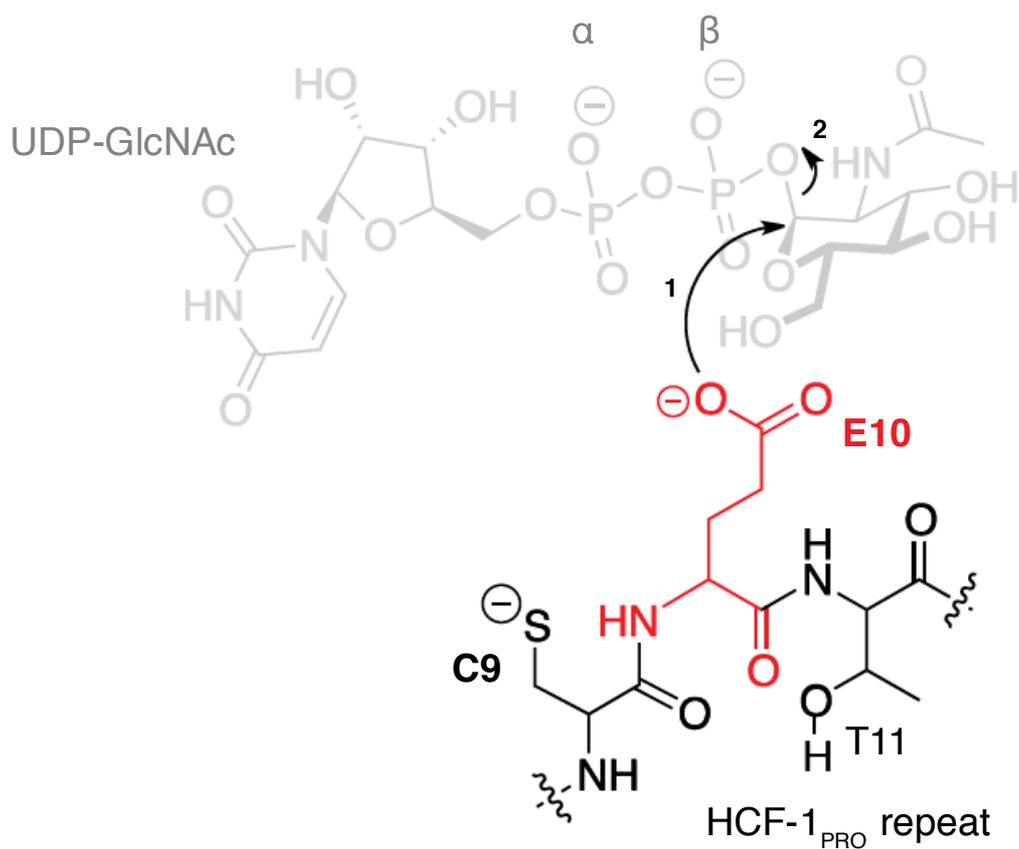
The HCF-1<sub>PRO</sub>-repeat residue E10 is the only residue in the HCF-1<sub>PRO</sub>-repeat cleavage region, which exhibits inhibitory properties for OGT association (Figure II-8). The presence of UDP-GlcNAc in the OGT–UDP-GlcNAc–HCF-1rep1 complex interfered with binding of the E10 side-chain to OGT (Figures II-6 and II-7). Subsequent experiments showed that the E10 side-chain forms unfavorable contacts with residues in the OGT catalytic domain. The unfavorable position of E10 in the OGT active site is stabilized by UDP-GlcNAc, and becomes less unfavorable in the absence of UDP-GlcNAc (Figures II-9 and II-10). UDP in the OGT active site does not cause unfavorable interactions of the E10 side-chain, suggesting that the sugar moiety of UDP-GlcNAc is crucial for this effect (Figure II-8 E). Why could a strained interaction between a cleavage substrate and its enzyme be important for proteolysis? In fact, a release from the substrate strains in the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex would be energetically favored. Thus, I propose that the energy required for E10 binding is used to favor HCF-1<sub>PRO</sub>-repeat proteolysis, which causes the release from the strains.

It has been proposed that E10 lies in the catalytic center for cleavage (Capotosti et al., 2011), and on the basis of the present data, this hypothesis is likely true. It is also evident now why the E10A mutation enhances HCF-1–OGT association. Physical interaction assays and molecular dynamics simulations with E10A, E10Q and E10D substitutions (Figure II-9) clearly showed that alanine — but also residues with slightly different side-chain structures with respect to glutamate — “enhance” HCF-1<sub>PRO</sub>-repeat–OGT association because they display less unfavorable interactions within the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex. In other words, they simply “fit better” inside the OGT catalytic domain.

### **Proposed initial step for HCF-1<sub>PRO</sub>-repeat proteolysis**

The HCF-1<sub>PRO</sub>-repeat E10 residue is unique within the HCF-1<sub>PRO</sub>-repeat cleavage region for its cleavage and binding properties, and its side-chain structure is crucial for HCF-1<sub>PRO</sub>-repeat proteolysis (Figure II-4 C). Lazarus et al. (2013) showed that after proteolysis, the glutamate side-chain E10 forms an amino-terminal pyroglutamate on the cleaved HCF-1<sub>PRO</sub>-repeat polypeptide, supporting a model in which the E10 side-chain participates in the proteolysis mechanism. The pyroglutamate is likely the endproduct of the *in vitro* cleavage reaction, however, the reactions occurring during cleavage before pyroglutamate formation are unknown.

The HCF-1<sub>PRO</sub> repeat adopts the conformation of an O-GlcNAcylation substrate in the OGT active site and the E10 side-chain is located in close proximity to the UDP-GlcNAc sugar moiety. Moreover, replacement of E10 by serine activated the HCF-1<sub>PRO</sub> repeat for O-GlcNAcylation but blocked proteolysis (Lazarus et al., 2013). Based on these data, in Lazarus et al. (2013) we proposed a model for the initial step of the cleavage reaction (Figure II-14): The anomeric carbon atom of UDP-GlcNAc could undergo a nucleophilic attack by the E10 side-chain. Consequently, the bond between the anomeric carbon atom and the oxygen of the  $\beta$ -phosphate would break, leading to hydrolysis of UDP-GlcNAc and, perhaps, to a transient glutamylester (O-GlcNAcylated E10 side-chain). This proposed initial mechanism is also in agreement with the observation that UDP-5SGlcNAc inhibits cleavage: sulfur is less electronegative than oxygen and would activate the anomeric carbon atom less efficiently for a nucleophilic attack (Dr. Vincent Zoete and Dr. Ute Roehrig, personal communications). To support this model, Dr. Vincent Zoete has performed molecular dynamics simulations of the cleavage mechanism. The simulation along the trajectory can be seen in a short movie (please follow the link <http://www.molecular-modelling.ch/SN2-like.mov>). The simulation shows that the proposed initial step for the cleavage reaction (Figure II-14) is thermodynamically possible: Dr. Vincent Zoete engineered movements of the E10 side-chain equivalent to an activation of the cleavage reaction. Following this activation, indeed, an elongation of the bond between the anomeric carbon atom and the oxygen of the  $\beta$ -phosphate of UDP-GlcNAc can be observed. Consequently, the anomeric carbon atom moves closer towards E10. The resulting elongated bond is too long to be considered a covalent bond and this indicates hydrolysis of UDP-GlcNAc and possible O-GlcNAcylation of E10. As discussed above, this initial step for the cleavage reaction could be triggered by the substrate strains originating from E10 in the OGT catalytic domain.



**Figure II-14: Proposed initial step for the HCF-1<sub>PRO</sub>-repeat cleavage mechanism.**

Arrow 1: The anomeric carbon atom of UDP-GlcNAc undergoes a nucleophilic attack by the deprotonated glutamate side-chain at position 10 of the HCF-1<sub>PRO</sub> repeat (E10). Arrow 2: Consequently, the bond between the anomeric carbon atom and the oxygen of the β-phosphate elongates and breaks, leading to UDP-GlcNAc hydrolysis and, perhaps, to transiently O-GlcNAcylated glutamate. The α- and β-phosphates of UDP-GlcNAc are indicated. Figure adapted from Lazarus et al. (2013).

### **The analog UDP-5SGlcNAc changes the HCF-1<sub>PRO</sub>-repeat–OGT binding mode drastically**

When I probed HCF-1<sub>PRO</sub>-repeat–OGT association with UDP-5SGlcNAc instead of UDP-GlcNAc, I obtained startling results (Figure II-7). Whereas strong OGT association of wild-type or E10A HCF-1<sub>PRO</sub> repeats was observed, I also observed strong OGT association of an HCF-1<sub>PRO</sub> repeat containing the T17-22A mutations. As these mutations normally disrupt OGT–HCF-1<sub>PRO</sub>-repeat association in the presence of UDP, UDP-GlcNAc or in the absence of cofactor, this result was unexpected. The reasons for this effect are unclear, but given that UDP-5SGlcNAc inhibits HCF-1 O-GlcNAcylation and proteolysis, the following scenario could be possible: The E10 side-chain could have been “locked” in the OGT catalytic domain by the events that would normally lead to proteolysis. On the basis of our proposed model (see above), the E10 side-chain could attack the anomeric carbon atom of UDP-5SGlcNAc. Contrary to promoting hydrolysis of the bond between the anomeric carbon atom and the phosphate, UDP-5SGlcNAc cannot be (or can merely be) hydrolyzed (Gloster et al., 2011). This could lead to a covalent intermediate of E10 bound to UDP-5SGlcNAc in the OGT catalytic domain. The covalent intermediate could promote strong interactions between the HCF-1<sub>PRO</sub> repeat and OGT, even in the presence of the T17-22A mutations.

### **The HCF-1<sub>PRO</sub>-repeat threonine region is important for HCF-1<sub>PRO</sub>-repeat–OGT association**

In Lazarus et al. (2013) we showed that the HCF-1<sub>PRO</sub>-repeat threonine region stably interacts with the OGT TPR region. This observation was consistent with the negative effects of the HCF-1<sub>PRO</sub>-repeat threonine region mutant T17-22A on OGT binding and HCF-1<sub>PRO</sub>-repeat cleavage (Figure II-4 B). Thus, the threonine region represents the main OGT-binding sequence within the HCF-1<sub>PRO</sub> repeat. The cleavage region, instead, forms a binding interface with UDP-GlcNAc in the OGT catalytic domain (Lazarus et al., 2013). These results were in agreement with the prediction of a bipartite OGT–HCF-1<sub>PRO</sub>-repeat interaction mode (Capotosti et al., 2011 and section I.7.3). It is very likely that the stable interactions between the HCF-1<sub>PRO</sub>-repeat threonine region and the OGT TPR region are required to accommodate the HCF-1<sub>PRO</sub>-repeat cleavage region correctly inside the OGT catalytic domain. Perhaps, without these strong interactions, the HCF-1<sub>PRO</sub>-repeat cleavage region would not bind to OGT because of the unfavorable interactions between the E10 side-chain and residues in the OGT catalytic domain (discussed above).

Residue T14 within the threonine region is a representative for the threonines within the threonine region because it displays large and favorable interactions with residues in the OGT TPR domain via hydrogen bonds (Figures II-8 B and D and Figure II-13). The T14–TPR

domain interactions are specific because T14 cannot be replaced by a residue with related side-chain structure (serine) or by a residue, which can form strong hydrogen bonds (asparagine, Figure II-13). It is not yet understood how OGT recognizes and binds to its multitude of O-GlcNAcylation substrates (see section I.6.2), but it has been debated that O-GlcNAcylation substrates might bind to the OGT TPR domain analogous to how the HCF-1<sub>PRO</sub>-repeat threonine region binds to the TPR domain (Lazarus et al., 2013). A search in the Swissprot protein data bank, using the Basic Local Alignment Search Tool (BLAST; Johnson et al., 2008), did not reveal sequence motifs similar to the HCF-1<sub>PRO</sub>-repeat threonine region in any characterized protein other than in HCF-1 proteins. I therefore propose that the interactions between the HCF-1<sub>PRO</sub>-repeat threonine region and the OGT TPR domain are of highly specific nature, to allow for cleavage in the HCF-1<sub>PRO</sub>-repeat cleavage region. Conventional O-GlcNAcylation substrates might therefore interact differently with the OGT TPR domain.

### **A glycosyltransferase coopted as protease?**

The glycosyltransferase OGT is an unusual protease. The substrate requirements for proteolysis within the HCF-1<sub>PRO</sub> repeat include multiple elements: (i) The HCF-1<sub>PRO</sub>-repeat cleavage region with the essential residue E10, (ii) the HCF-1<sub>PRO</sub>-repeat flexible Hinge region, and (iii) the HCF-1<sub>PRO</sub>-repeat threonine region for OGT binding. The HCF-1<sub>PRO</sub> repeat binds in the OGT active site analogous to a CK-II O-GlcNAcylation substrate. Interestingly, both the HCF-1<sub>PRO</sub>-repeat and the CK-II substrate, contain a proline residue two positions N-terminal of the E10 cleavage site or of the GlcNAc acceptor site (S9), respectively. This suggests that both substrates could have similar motifs for the association with the OGT catalytic domain. In fact, it seems that the HCF-1<sub>PRO</sub> repeat mimics an O-GlcNAcylation substrate and that it uses OGT and UDP-GlcNAc to induce its cleavage. I propose that OGT is a bona fide glycosyltransferase that has been coopted by HCF-1 for cleavage of the HCF-1<sub>PRO</sub> repeats. The substrate itself determines the type of post-translational modification catalyzed in the OGT active site: A single amino acid side-chain at position 10 of the HCF-1<sub>PRO</sub> repeat can either promote proteolysis (glutamate) or O-GlcNAcylation (serine).

UDP-GlcNAc and particularly the GlcNAc moiety are essential for HCF-1<sub>PRO</sub>-repeat proteolysis. As introduced in section I.5.3, intracellular UDP-GlcNAc levels are proposed to reflect the cellular nutrient status via the hexosamine biosynthetic pathway. It is thus possible that HCF-1 coopted OGT as its protease to couple cellular nutrient levels to one of HCF-1's major functions — the regulation of cell-cycle progression.

## Chapter III :

### *Requirements for HCF-1 proteolysis outside of the HCF-1<sub>PRO</sub> repeat*

#### Introduction

HCF-1<sub>PRO</sub>-repeat cleavage has been probed in the past using small HCF-1 precursor constructs containing one to three HCF-1<sub>PRO</sub> repeats. The HCF-1<sub>PRO</sub> repeat was either in its natural HCF-1 context (e.g., HCF-1rep1, HCF3R, HCF-1rep123 in Capotosti et al., 2011 and Lazarus et al., 2013) or in a heterologous context within the C-terminal region of the Oct-1 POU transcription factor (Wilson et al., 1995b). In all these cases, HCF-1<sub>PRO</sub>-repeat cleavage could be observed using SDS-PAGE followed by immunoblot analysis or Coomassie staining. To create a heterologous HCF-1<sub>PRO</sub>-repeat construct that can be synthesized and purified from *E. coli*, Capotosti and Herr (unpublished results) designed a smaller construct called POUrep2, which contains the HCF-1<sub>PRO</sub> repeat 2 embedded between the two structured domains of the Oct-1 POU DNA-binding domain (POU-homeo domain and POU-specific domain; Herr et al., 1988). Surprisingly, in contrast to the other heterologous HCF-1rep constructs, the POUrep2 construct was inactive for cleavage *in vitro* using *in vitro* substrates synthesized in rabbit reticulocyte translation extracts (Capotosti and Herr, unpublished results).

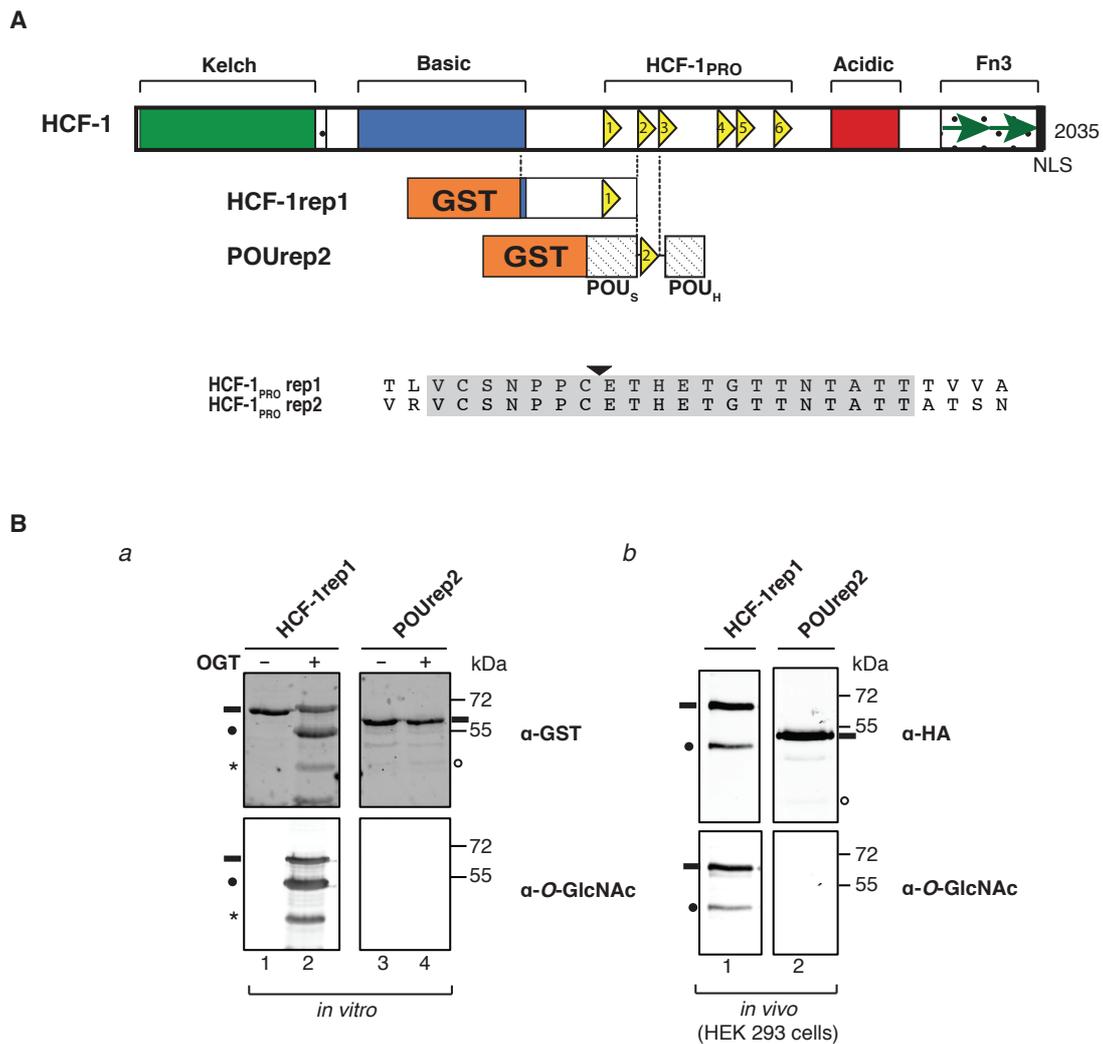
Having identified requirements within the HCF-1<sub>PRO</sub> repeat for cleavage (see Chapter II), I investigated here what HCF-1 sequences lying outside of the HCF-1<sub>PRO</sub> repeat might be required for cleavage.

#### Results

##### III.1 Sequences lying outside of the HCF-1<sub>PRO</sub> repeat promote cleavage

The POUrep2 construct (Figure III-1 A) did not display cleavage activity in previous experiments with *in vitro* translated proteins (Capotosti and Herr, unpublished results). I confirmed this result in an *in vitro* cleavage assay using bacterially synthesized precursors and in an *in vivo* cleavage assay. The bacterial assay was performed as described in Chapter II. For the *in vivo* cleavage assay, vectors encoding the HCF-1rep1 or POUrep2 constructs were transiently expressed in HEK 293 cells and resulting uncleaved precursor and N-terminal cleavage products were purified via an N-terminal HA-epitope tag. Cleavage

was then assayed by immunoblot using  $\alpha$ -HA or  $\alpha$ -GST antibodies. Whereas the HCF-1rep1 wild-type construct was cleaved by OGT *in vitro* (Figure III-1 B a, lanes 1 and 2, upper panel) and *in vivo* (Figure III-1 B b, lane 1, upper panel), the POUrep2 construct displayed only poor, if any, cleavage activity in these two assay systems (a, lanes 3 and 4; b, lane 2, upper panel). When O-GlcNAcylation of the protein precursors and cleaved products was probed by  $\alpha$ -O-GlcNAc immunoblot, contrary to HCF-1rep1, which displayed strong O-GlcNAcylation (a, lanes 1 and 2; b, lane 1, lower panel), the POUrep2 precursor protein did not display O-GlcNAcylation (a, lanes 3 and 4; b, lane 2, lower panel). These results suggest that HCF-1<sub>PRO</sub>-repeat proteolysis is context dependent. In the heterologous context of the POU domain, the HCF-1<sub>PRO</sub> repeat is not efficiently cleaved, indicating that there might be additional HCF-1 sequences lying outside of the HCF-1<sub>PRO</sub> repeat that are important for cleavage. Moreover, the lack of O-GlcNAcylation of the POUrep2 construct might be related to the poor cleavage efficiency of this construct.

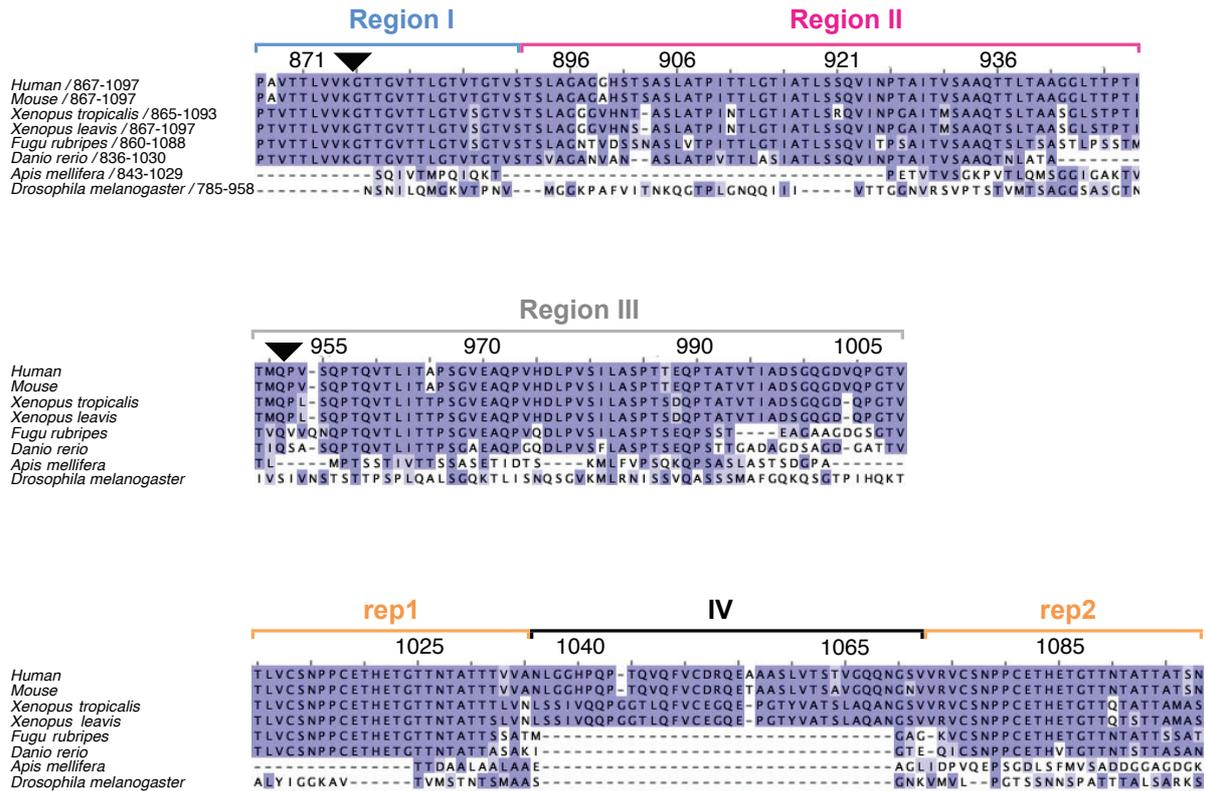


**Figure III-1: HCF-1<sub>PRO</sub>-repeat cleavage is context dependent.**

(A) Schematic representation of the HCF-1 full-length protein and the GST-fusion constructs used in this experiment. The HCF-1rep1 construct is described in Chapter II. POUrep2 contains GST fused to HCF-1<sub>PRO</sub> repeat 2 embedded in the variable linker between the Oct-1-derived POU-specific (POU<sub>S</sub>) and POU-homeo (POU<sub>H</sub>) domains. The sequence alignment of the HCF-1<sub>PRO</sub> repeats 1 and 2 below shows that their 20 amino acid HCF-1<sub>PRO</sub>-repeat core sequences (Kristie et al., 1995) are identical. The black arrowhead indicates the cleaved peptide bond. (B) HCF-1<sub>PRO</sub>-repeat cleavage is context dependent. (a) HCF-1<sub>PRO</sub>-repeat cleavage in an *in vitro* assay with bacterially synthesized substrates and OGT and (b) in an *in vivo* assay with proteins from transiently transfected HEK 293 cells followed by immunoprecipitation via an N-terminal HA-tag. Cleavage and O-GlcNAcylation were visualized by immunoblot using, respectively, α-GST (a, lanes 1–4, top panel) and α-HA (b, lanes 1 and 2 top panel) antibodies for cleavage, and α-O-GlcNAc antibodies (a and b, bottom panels) for O-GlcNAcylation. The positions of uncleaved precursor proteins (–), prominent (●) or faint (○) cleavage products and background C-terminal precursor truncations (\*) are indicated.

### III.1.1 Subdivision of HCF-1rep1 sequences for deletion analysis

To identify HCF-1rep1 sequences outside of the HCF-1<sub>PRO</sub> repeat that might affect cleavage, I divided the residues upstream of the HCF-1<sub>PRO</sub> repeat 1 into three regions: Region I (867-891), comprising identified O-GlcNAcylation sites in Capotosti et al. (2011) and Region II (892-949) and Region III (950-1009), which contain about the same number of residues between Region I and the HCF-1<sub>PRO</sub> repeat, 58 and 60, respectively. Remarkably, these three regions are conserved in vertebrate species, in which HCF-1 cleavage is mediated by OGT, but not in invertebrate species, in which HCF cleavage is mediated by Taspase1 (Figure III-2). Figure III-2 also shows the high degree of conservation of the HCF-1<sub>PRO</sub> repeats 1 (rep1) and 2 (rep2) among vertebrate species, and the less well-conserved stretch in between them, which I called Region IV. Region IV displays a lower degree of conservation, mainly because this sequence is absent in fish HCF-1 proteins. I thus generated deletion constructs containing only one of the three Regions I-III (+I, +II, +III) or constructs lacking only one of the three regions ( $\Delta$ I,  $\Delta$ II,  $\Delta$ III). Additionally, I deleted Region I, II, and III all together ( $\Delta$ I.II.III). The HCF-1<sub>PRO</sub> repeat and the non-conserved Region IV (36 residues) were retained in all constructs (a schematic illustration of all deletion constructs is shown in Figure III-3 A). I analyzed these deletion constructs in three different cleavage assay systems, in order to accurately determine the effect of Regions I, II, III, and later also IV on HCF-1<sub>PRO</sub>-repeat cleavage. The HCF-1<sub>PRO</sub>-repeat mutations E10A and T17-22A, described in Chapter II, were used as negative controls for cleavage.

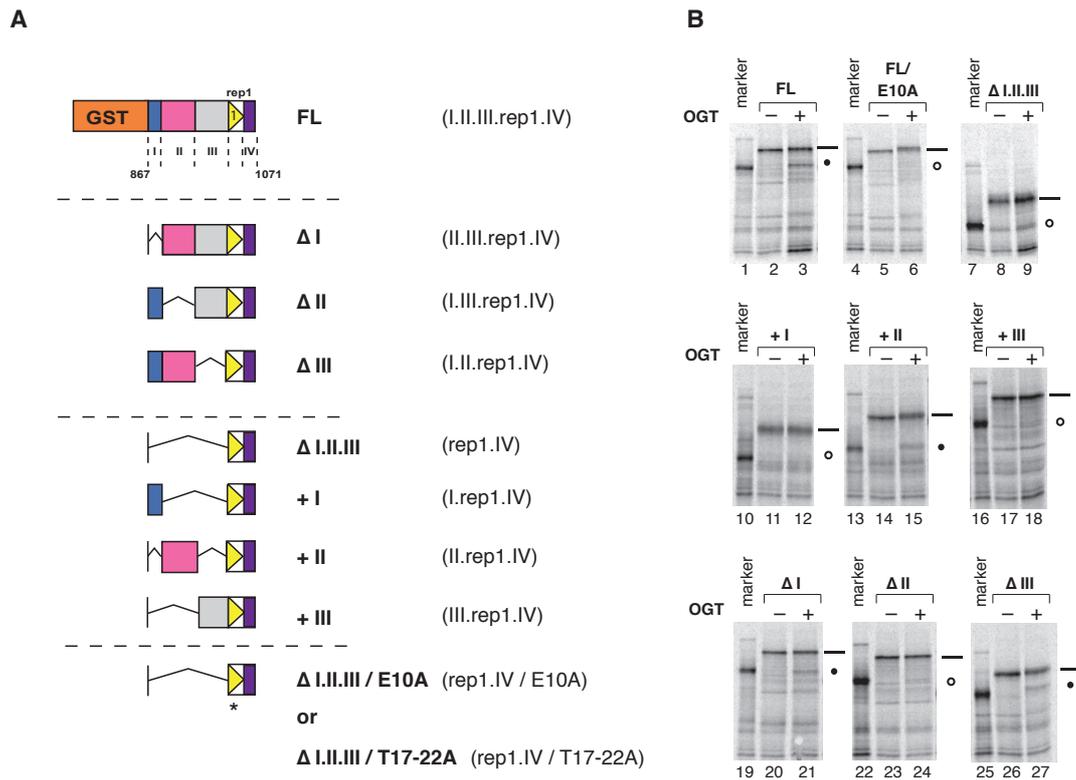


**Figure III-2: Sequence conservation of HCF-1 sequence 867-1098.**

Sequence conservation among vertebrate and invertebrate species of the HCF-1 sequences 867-1098. Six different vertebrate species where HCF-1 is cleaved by OGT: Human, Mouse, *Xenopus tropicalis*, *Xenopus leavis*, *Fugu rubripes* and *Danio rerio* were aligned with two invertebrate species where HCF is cleaved by Taspase1: *Apis mellifera* and *Drosophila melanogaster*, using the Jalview bioinformatics tool (Waterhouse et al., 2009). Residues are colored in blue according to conservation following the Blosum62 score. Regions I, II, III, IV, and the HCF-1<sub>PRO</sub> repeats 1 (rep1) and 2 (rep2) are indicated, and the residues were numbered according to the human HCF-1 sequence. The black arrowheads indicate exon boundaries in the human gene *HCFC1* encoding HCF-1.

### III.1.2 *In vitro* HCF-1 cleavage assay with *in vitro* transcribed and translated substrates

To test whether the deletions described above (Figure III-3 A) have an effect on HCF-1<sub>PRO</sub>-repeat cleavage, I synthesized the substrates by *in vitro* transcription and translation and subjected them to an *in vitro* cleavage assay (Figure III-3 B), as described previously (Capotosti et al., 2011). To distinguish cleavage products from background bands resulting from *in vitro* synthesis of the polypeptides, I developed a PCR-strategy that allowed the synthesis of markers corresponding to the same molecules as the N-terminal cleavage products for each deletion construct (lanes 1, 4, 7, 10, 13, 16, 19, 22 and 25, respectively). In summary, the cleavage assay of HCF-1rep1 full-length (FL) resulted in a slightly slower migrating, uncleaved precursor (lane 3, indicated by –), indicative of O-GlcNAcylation, and in a cleaved N-terminal product (lane 3, indicated by •). The analysis of the negative control for cleavage, the HCF-1rep1 E10A mutant, revealed no cleavage product at the indicated size (lanes 4-6). The  $\Delta$ I.II.III construct displayed little, if any, cleavage (lanes 7-9), suggesting that the HCF-1<sub>PRO</sub> repeat is not efficiently cleaved in the absence of sequences lying N-terminal of the repeat. The loss-of-function analysis of constructs lacking only one of the three regions ( $\Delta$ I,  $\Delta$ II,  $\Delta$ III, lanes 19-27), as well as the gain-of-function analysis of constructs containing only one of the three regions (+I, +II, +III, lanes 10-18) suggested that, among all three regions, Region II is the most active region in promoting HCF-1<sub>PRO</sub>-repeat cleavage.



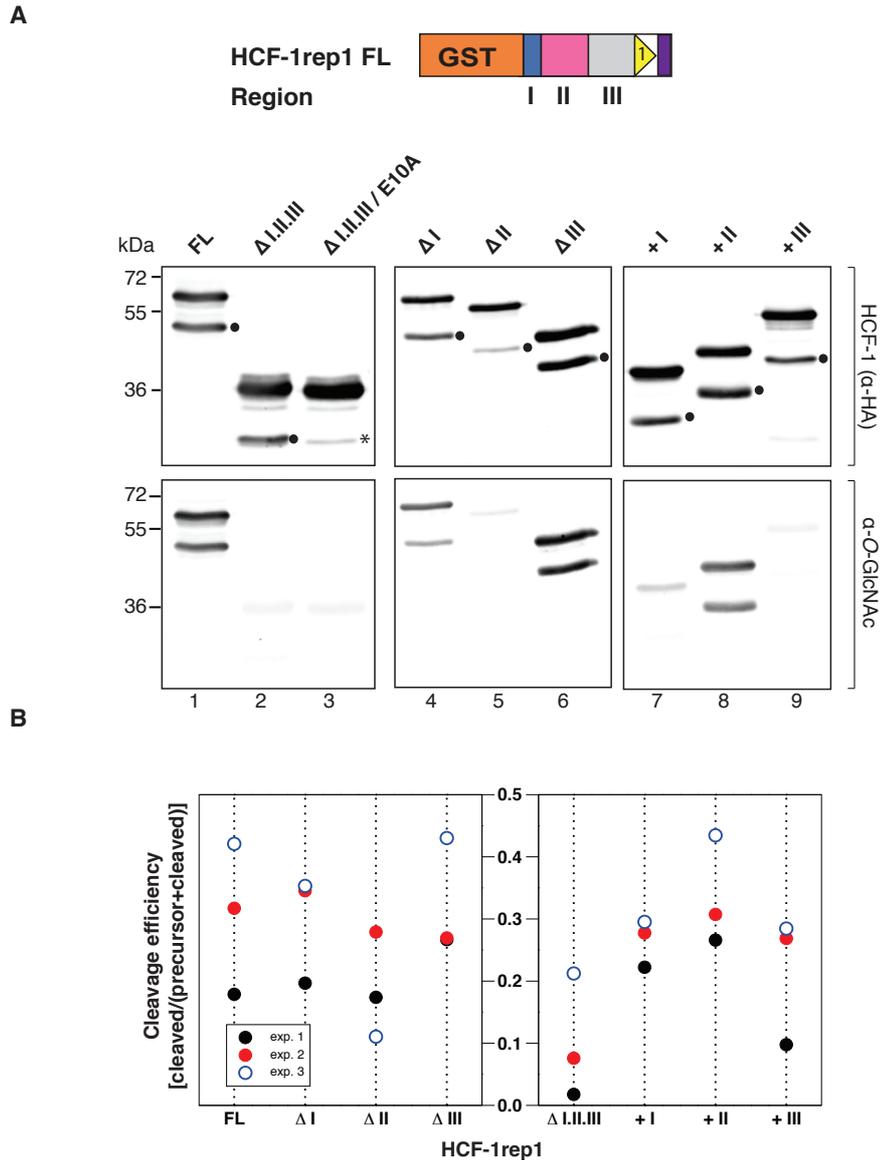
**Figure III-3: Deletion analysis with substrates synthesized in a wheat-germ extract *in vitro* transcription and translation system.**

(A) Schematic of the HCF-1rep1 deletion constructs used in this study. Constructs ΔI, ΔII, and ΔIII lack Regions I, II or III, respectively. ΔI.II.III is a deletion of Regions I, II, and III together and was also generated with the E10A or T17-22A mutations that inhibit cleavage. Constructs +I, +II, and +III contain Regions I, II or III, respectively. The non-conserved Region IV (36 residues) was retained in all constructs. (B) Region II promotes HCF-1<sub>PRO</sub>-repeat cleavage *in vitro*: Cleavage assay of *in vitro* synthesized deletion constructs (illustrated in A) in wheat-germ extract in the presence (+) or absence (-) of human OGT, produced and purified from insect cells (Capotosti et al., 2011). Proteins were resolved by SDS-PAGE and visualized by autoradiography. The markers indicate the size of the N-terminal cleavage products. The positions of uncleaved precursor proteins (-), and prominent (\*) or faint (o) cleavage products are indicated.

### III.1.3 *In vivo* HCF-1 cleavage assay in HEK 293 cells

To extend the aforementioned *in vitro* studies, I analyzed cleavage *in vivo* using the cleavage assay described in section III.1. Figure III-4 A shows an *in vivo* cleavage assay with the full-length and the HCF-1rep1-deletion constructs. As the amount of protein synthesis cannot be controlled *in vivo*, cleavage activities of these constructs can only be assessed when the ratios of precursor versus cleavage product for each sample are compared. Consistent with the results obtained *in vitro*, the presence or absence of the three regions together affected cleavage efficiency (FL vs.  $\Delta$ I.II.III constructs, lanes 1-3, upper panel). The loss-of-function analysis revealed that, among the three regions, the lack of Region II had the strongest negative effect on cleavage (lanes 4-6, upper panel). The addition of each of the three regions individually enhanced HCF-1<sub>PRO</sub>-repeat cleavage (lanes 7-9, upper panel) with respect to the  $\Delta$ I.II.III construct (lane 2, upper panel), with Region II, offering the greatest enhancement. Probing the immunoblot with  $\alpha$ -O-GlcNAc antibodies (lower panel) revealed that HCF-1rep1 constructs containing Region I or Region II were O-GlcNAcylated. The presence of Region II generated the strongest O-GlcNAcylation signal (constructs FL,  $\Delta$ I,  $\Delta$ III and +II), suggesting that Region II activates the HCF-1 precursor for O-GlcNAcylation, perhaps by containing O-GlcNAcylation sites. These results indicate a correlation between the cleavage activity of HCF-1rep1 deletion constructs and their O-GlcNAcylation status.

The results of the *in vivo* HCF-1rep1 cleavage assay displayed variability among independent experiments, particularly in reference to the  $\Delta$ I.II.III and  $\Delta$ III constructs (Figure III-4 B), due to constant protein synthesis and processing in cells. Nevertheless, these results led me to conclude that, although all three regions display cleavage-enhancement activity, Region II has the most pronounced activity on HCF-1<sub>PRO</sub>-repeat proteolysis.



**Figure III-4: HCF-1<sub>PRO</sub>-repeat cleavage analysis with HCF-1rep1 wild-type and deletion substrates synthesized in HEK 293 cells *in vivo*.**

(A) Region II promotes cleavage *in vivo*. HEK 293 cells were transfected with expression vectors encoding HCF-1rep1 FL (full-length) or deletion constructs (see Figure III-3 A). Proteins were immunoprecipitated by an N-terminal HA-epitope tag and assayed for cleavage by visualization via  $\alpha$ -HA-epitope-tag immunoblot. The positions of cleavage products (\*) and non-specific background bands (\*) are indicated. (B) *In vivo* cleavage activities of HCF-1rep1-deletion constructs from three independent experiments (exp.1, exp. 2, exp. 3). Cleavage activities of HCF-1rep1 constructs were assayed as in (A). Bands on immunoblots were analyzed using LI-COR Image Quant quantification software and the cleavage efficiencies quantified as ratio of cleaved product to total GST-HCF-1rep1 protein (full-length plus cleaved fragment) in the assay.

### III.1.4 *In vitro* HCF-1 cleavage assay with bacterially synthesized substrates and OGT

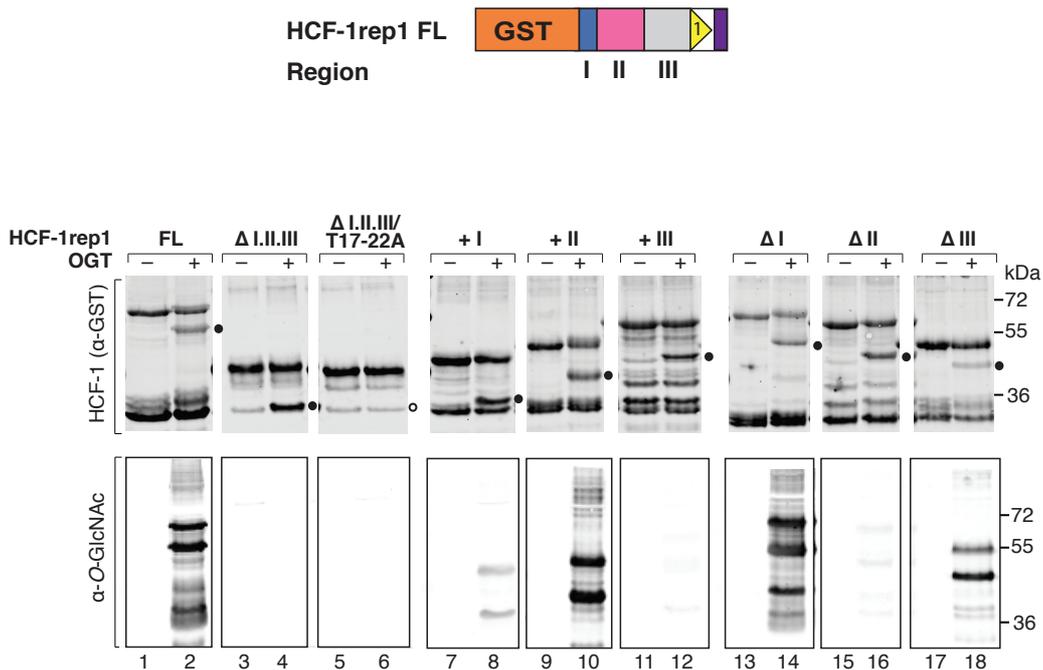
To complement the studies with wheat-germ extract *in vitro* translated precursors and the *in vivo* studies, I performed *in vitro* cleavage assays with bacterially synthesized precursors and OGT. Figure III-5 A shows the results of such an *in vitro* cleavage incubated with bacterially purified OGT in cleavage buffer for 16 h as described in Capotosti et al. (2011). Surprisingly, unlike in the experiment with *in vitro* translated substrates and in the *in vivo* experiment described above, in this assay, only a minor difference in cleavage activity between the HCF-1rep1 FL construct (lanes 1 and 2, upper panel) and the construct lacking all three regions together ( $\Delta$ I.II.III, lanes 3 and 4, upper panel) could be observed. And yet, the negative control for cleavage, construct  $\Delta$ I.II.III containing the T17-22A mutations within the HCF-1<sub>PRO</sub>-repeat ( $\Delta$ I.II.III/T17-22A), was not cleaved (lanes 5 and 6, upper panel), suggesting that the conditions of the *in vitro* cleavage assay were appropriate to identify cleavage activities. The gain-of-function analysis (lanes 7-12, upper panel), as well as the loss-of-function analysis (lanes 13-18, upper panel) did not reveal any particular effect of Region II on HCF-1<sub>PRO</sub>-repeat cleavage in comparison with the effects of Regions I or III. In contrast, probing the immunoblot with  $\alpha$ -O-GlcNAc antibodies (lower panel), revealed the same pattern as in the *in vivo* O-GlcNAcylation analysis above (Figure III-4 A, lower panel) with Region II promoting the most prominent O-GlcNAcylation of HCF-1rep1 proteins (Figure III-5, lower panel, lanes 1 and 2, 9 and 10, 13 and 14, 17 and 18).

The aforementioned studies represent a single endpoint analysis, in which initial rates of cleavage were not determined. Such an analysis can obscure differences between enzymatic or substrate activities if the rates reach a plateau over time. Thus, to test if any difference in HCF-1<sub>PRO</sub>-repeat cleavage activation promoted by Region I, II or III can be observed after shorter incubation periods with OGT, I examined the cleavage efficiencies of a subset of constructs in an 8 h *in vitro* cleavage assay time course (Figure III-5 B). The rate of cleavage for the  $\Delta$ I.II.III HCF-1rep1 construct was substantially lower than for the FL construct. The +II construct displayed slightly more activity than FL, whereas +III displayed activity similar to  $\Delta$ I.II.III. The difference was particularly evident at the earlier time points i.e., after 1 and 2 hours. These results indicate a robust role for Region II in HCF-1<sub>PRO</sub>-repeat cleavage enhancement. I concluded that HCF-1rep1 cleavage enhancement by regions N-terminal of the first HCF-1<sub>PRO</sub> repeat occurs in a time-dependent manner. Thus, I analyzed cleavage activities of HCF-1rep1 constructs after 4 or 8 hours of incubation with OGT in all subsequent *in vitro* cleavage assays, unless stated otherwise.

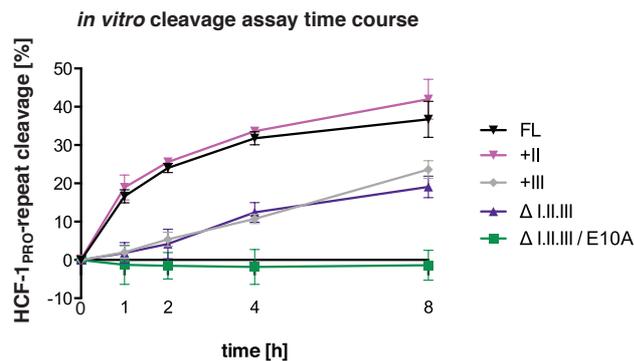
To conclude, the analysis of HCF-1rep1 cleavage in three different cleavage-assay systems (*in vitro* transcription and translation, *in vivo* and *in vitro* with bacterially synthesized substrates) showed that results obtained in these three systems are consistent with each

other. The *in vitro* cleavage assay with synthesized substrates in wheat-germ extract proved to be useful to assay cleavage activities of a large set of substrates in a short time and suggested that Region II has enhancing effects on HCF-1<sub>PRO</sub>-repeat cleavage. The *in vivo* cleavage assay was crucial to understand if similar activities can be observed *in vivo*, mediated by endogenous OGT. The *in vitro* cleavage assay with bacterially synthesized substrates and OGT was essential to assay activities of the different regions on HCF-1<sub>PRO</sub>-repeat cleavage at the early time points of the cleavage reaction. Strikingly, this assay revealed that the initial rates of cleavage are affected, suggesting that the cleavage-enhancing regions affect HCF-1<sub>PRO</sub>-repeat cleavage before reaching a plateau. Based on these data, I conclude that Region II enhances HCF-1<sub>PRO</sub>-repeat cleavage.

A



B

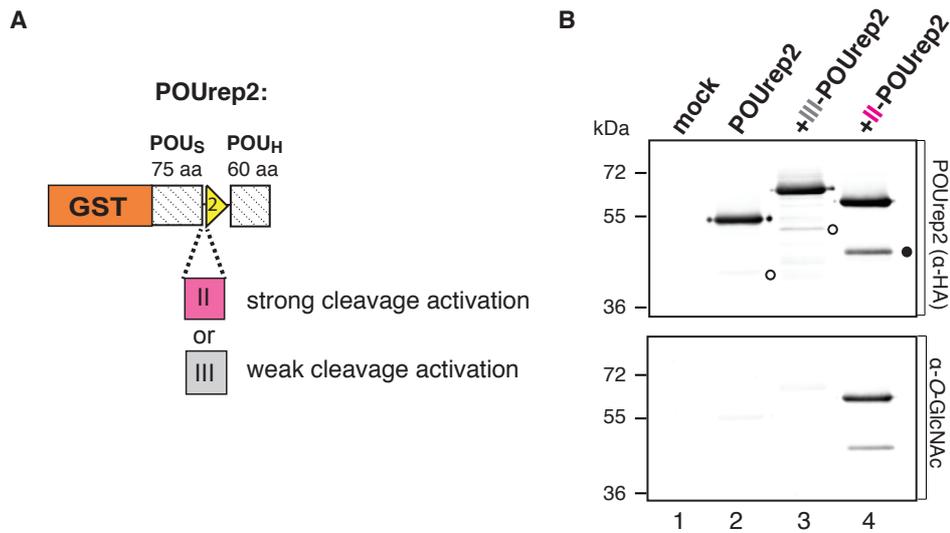


**Figure III-5: *In vitro* HCF-1<sub>PRO</sub>-repeat cleavage analysis with substrates and OGT synthesized and purified from *E. coli*.**

(A) Bacterially synthesized substrates and OGT were incubated for 16 h and products resolved by SDS-PAGE. HCF-1 or O-GlcNAcylated proteins were visualized by immunoblot using  $\alpha$ -GST- (top panel) or  $\alpha$ -O-GlcNAc-directed (bottom panel) antibodies, respectively. The positions of prominent (•) or faint (○) cleavage products are indicated. (B) Region II augments the rate of HCF-1<sub>PRO</sub>-repeat cleavage *in vitro*. Cleavage efficiency over time during a bacterial *in vitro* cleavage assay of selected HCF-1rep1 constructs. HCF-1rep1 constructs were incubated with OGT for 0 to 8 h and precursor and resulting N-terminal cleavage products were resolved by SDS-PAGE and analyzed for cleavage by  $\alpha$ -GST-immunoblot. Cleavage efficiencies were quantified as the percentage of cleaved product to uncleaved and cleaved HCF-1rep1 protein for each sample, using LI-COR Image Quant software. Shown are the means and standard deviations of three independent experiments.

### III.2 Region II can activate a heterologous HCF-1<sub>PRO</sub>-repeat substrate for cleavage and O-GlcNAcylation

To test whether Region II can activate the HCF-1<sub>PRO</sub> repeat in the POUrep2 construct (see Figure III-1), I inserted Region II or Region III upstream of the HCF-1<sub>PRO</sub> repeat 2 in POUrep2 (Figure III-6 A). As shown in an *in vivo* cleavage assay (Figure III-6 B), POUrep2 displayed weak, if any, cleavage activity (lane 2, upper panel, open circle). Inserting Region III upstream of the HCF-1<sub>PRO</sub> repeat in POUrep2 (+III-POUrep2) weakly enhanced cleavage (lane 3, upper panel, open circle), whereas inserting Region II (+II-POUrep2) activated strong HCF-1<sub>PRO</sub>-repeat cleavage (lane 4, upper panel, filled circle). O-GlcNAcylation of the uncleaved and cleaved POUrep2 constructs was assayed using  $\alpha$ -O-GlcNAc antibodies (lower panel). Whereas POUrep2 and +III-POUrep2 displayed weak, if any, apparent O-GlcNAcylation, +II-POUrep2 displayed prominent O-GlcNAcylation, demonstrating that Region II can activate the POUrep2 construct not only for proteolysis, but also for O-GlcNAcylation. These results suggest that Region II can enhance HCF-1<sub>PRO</sub>-repeat cleavage in an HCF-1 context, as well as in a heterologous context. As Region II and Region III are essentially of the same size, the activation by Region II is unlikely to be simply a spacing effect between the HCF-1<sub>PRO</sub> repeat and the POU<sub>S</sub> domain. To investigate the role of Region II in HCF-1<sub>PRO</sub>-repeat proteolysis, I performed genetic analyses of Region II, presented in the following sections. The link between Region II cleavage-enhancement activity and O-GlcNAcylation activity is addressed in Chapter IV.

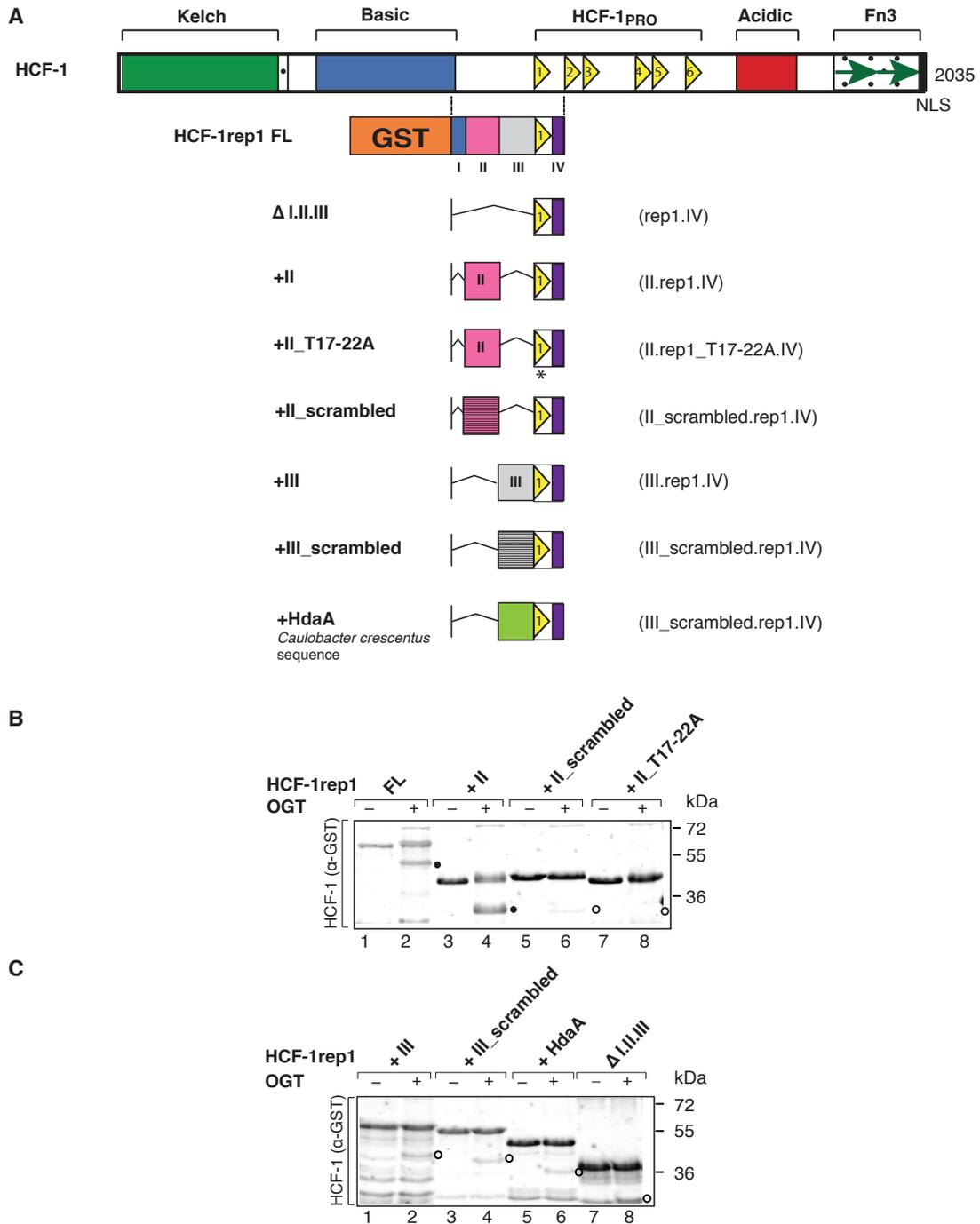


**Figure III-6: Region II can activate HCF-1<sub>PRO</sub>-repeat cleavage in a heterologous context.** (A) Schematic of the GST-fusion construct POUrep2 containing HCF-1<sub>PRO</sub> repeat 2 (rep2), embedded in the linker between the POU<sub>S</sub> and POU<sub>H</sub> domains. Region II or Region III were inserted N-terminal of rep2. (B) *In vivo* cleavage activities in HEK 293 cells transiently transfected with transfection medium (mock) or POUrep2-encoding vectors. Precursors and N-terminal cleavage products were purified via immunoprecipitation of an N-terminal HA-epitope tag. Cleavage (upper panel) and O-GlcNAcylation (lower panel) were detected using  $\alpha$ -HA- and  $\alpha$ -O-GlcNAc antibodies, respectively. The positions of prominent (•) and faint (○) cleavage products are indicated.

### III.3 Region II activity on HCF-1<sub>PRO</sub>-repeat cleavage is sequence specific

Region II represents an HCF-1<sub>PRO</sub>-repeat cleavage enhancer sequence. To test the sequence specificity of Region II's cleavage-enhancing activity, I replaced the wild-type Region II sequence in the HCF-1rep1 +II construct by a scrambled Region II sequence (see schematics in Figure III-7 A). As shown in Figure III-7 B, in a bacterial *in vitro* HCF-1 cleavage assay, Region II\_scrambled (lanes 5 and 6) showed little cleavage activation potential when compared to wild-type Region II (lanes 3 and 4), suggesting that the effect of Region II is sequence specific and neither caused only by its size nor by its amino acid composition. To test whether Region II, in isolation from Region I and III, could enhance proteolysis of an HCF-1<sub>PRO</sub>-repeat threonine region mutant, I introduced the four HCF-1<sub>PRO</sub>-repeat threonine mutations (T17-22A, see Chapter II) in construct +II (+II\_T17-22A, lanes 7 and 8). This mutant disrupts HCF-1<sub>PRO</sub>-repeat–OGT TPR interactions and, consistent with this property, cleavage of this mutant was inhibited, showing that defective HCF-1<sub>PRO</sub>-repeat–OGT TPR interactions cannot be rescued by Region II activity.

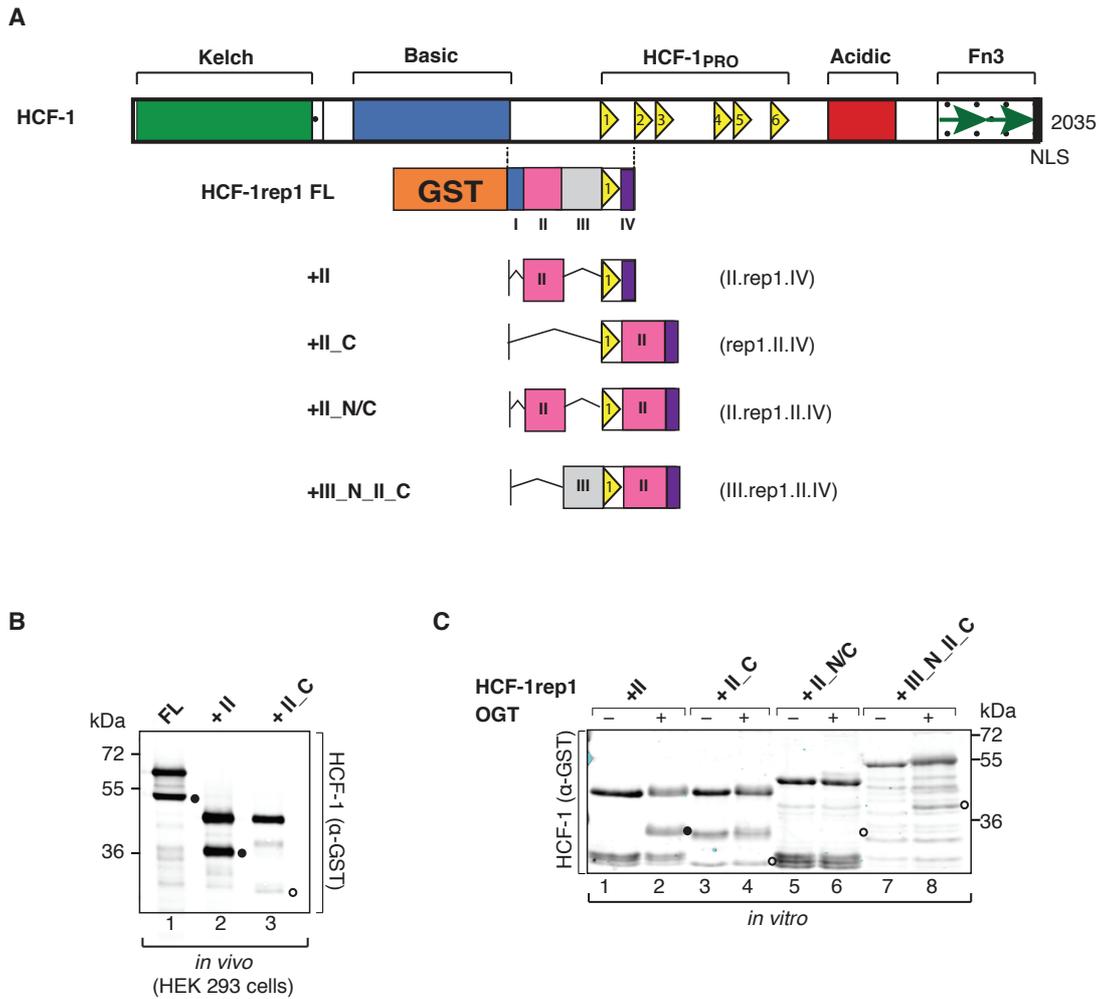
The Region III sequence displays little cleavage activation potential (section III.1). To test if genetic manipulations of Region III elicit similar effects on cleavage, as observed for Region II manipulations, I tested a scrambled Region III sequence and an unrelated heterologous sequence of the same size (derived from the replication protein HdaA of the bacterium *Caulobacter crescentus*) in an *in vitro* cleavage assay (Figure III-7 C). Neither the scrambled Region III sequence (lanes 3 and 4) nor the heterologous sequence (lanes 5 and 6) altered cleavage efficiency as compared to +III (lanes 1 and 2). Indeed, they all displayed  $\Delta$ I.II.III activity (lanes 7 and 8). Thus, Region III, unlike Region II, possesses little, if any, HCF-1<sub>PRO</sub>-repeat cleavage enhancement activity. Together, these results show that efficient HCF-1<sub>PRO</sub>-repeat cleavage, in the context of HCF-1rep1, depends on the Region II sequence and proper HCF-1<sub>PRO</sub>-repeat–OGT interactions.



**Figure III-7: The effect of Region II on HCF-1<sub>PRO</sub>-repeat cleavage is sequence specific.** (A) Schematic representation of the different HCF-1rep1 recombinants used in this experiment. Bacterial *in vitro* cleavage assay (incubation for 4 h at 37°C) of different Region II (B) or (C) Region III HCF-1rep1 constructs. HCF-1 precursor proteins and N-terminal cleavage products were visualized by immunoblot using α-GST-directed antibodies. In (B) and (C), the positions of prominent (\*) and faint (o) cleavage products are indicated.

### III.4 Region II activity displays positional specificity

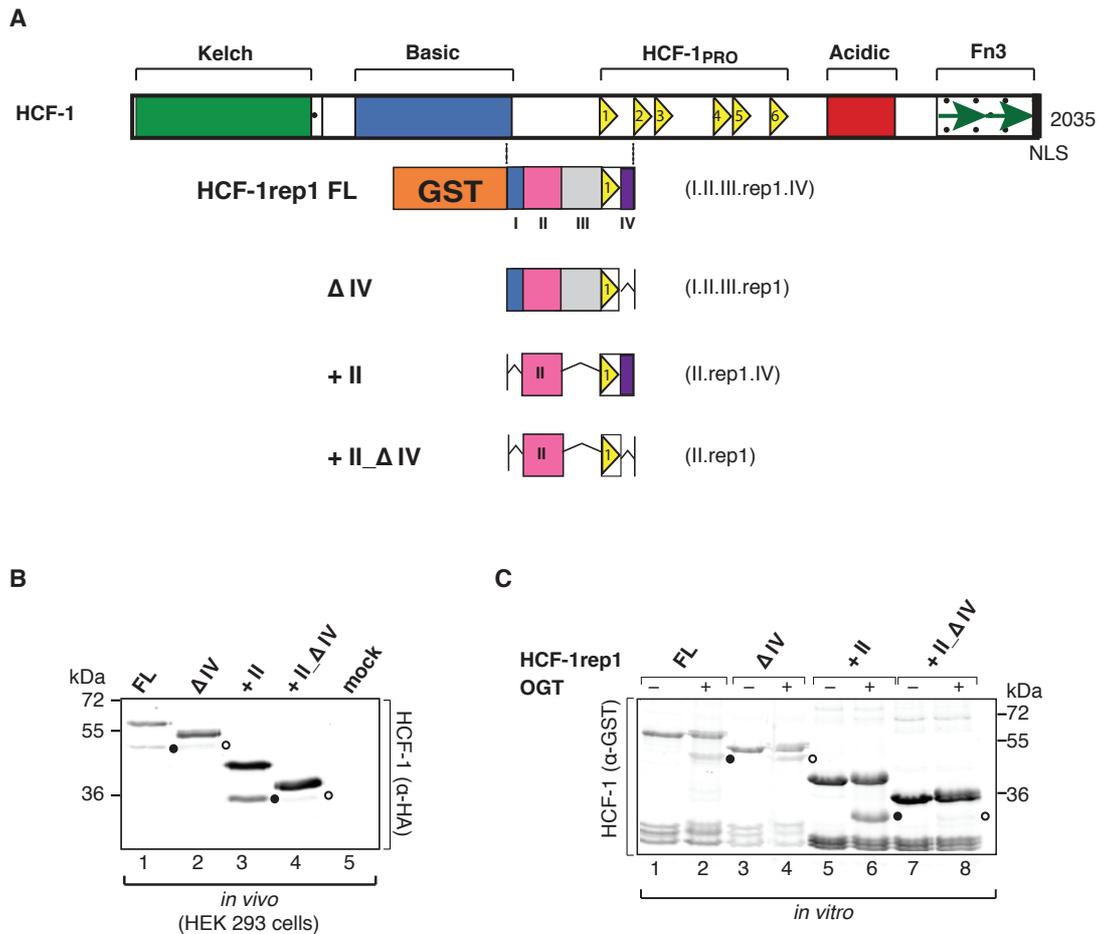
The aforementioned studies established that HCF-1<sub>PRO</sub>-repeat the cleavage-enhancement activity of Region II is sequence specific. To further characterize the activity of Region II, I generated a set of mutants containing Region II placed C-terminal of the HCF-1<sub>PRO</sub> repeat (Figure III-8 A). Figure III-8 B shows an *in vivo* cleavage assay. Whereas HCF-1rep1 FL and +II displayed prominent cleavage (lanes 1 and 2, respectively), the HCF-1rep1 +II\_C mutant (lane 3) displayed only little, if any, cleavage activity, suggesting that the position of Region II is important for its activity on cleavage. Because Region II is inactive at the C-terminus of the HCF-1<sub>PRO</sub> repeat, I tested whether reinserting Region II at the N-terminus of the HCF-1<sub>PRO</sub> repeat would rescue cleavage activity (construct +II\_N/C). I also inserted Region III N-terminal of the HCF-1<sub>PRO</sub>-repeat (construct +III\_N\_II\_C) to control for a potential size effect of the N-terminal Region II sequence. Figure III-8 C shows a bacterial *in vitro* cleavage assay with the constructs described above. As expected, and in concordance with the *in vivo* cleavage experiments (Figure III-8 B), +II displayed prominent cleavage (lanes 1 and 2), whereas +II\_C displayed only little, if any, cleavage activity (lanes 3 and 4). Placing Region II N-terminal of the HCF-1<sub>PRO</sub> repeat in construct +II\_C did not rescue cleavage (lanes 5 and 6), suggesting that Region II at the C-terminus actively inhibits HCF-1<sub>PRO</sub>-repeat cleavage. Replacing the N-terminal Region II by Region III (+III\_N\_II\_C), surprisingly, reactivated cleavage to some extent (lanes 7 and 8) when compared to the activities of +II\_C or +II\_N/C. This result was unexpected, as isolated Region III had lower activity on HCF-1<sub>PRO</sub>-repeat cleavage than Region II. These results highlight that HCF-1<sub>PRO</sub>-repeat cleavage is context-dependent and that Region II's activity on cleavage is complicated being sensitive to its position relative to the HCF-1<sub>PRO</sub> repeat.



**Figure III-8: Region II C-terminal of the HCF-1<sub>PRO</sub> repeat inhibits HCF-1<sub>PRO</sub>-repeat cleavage.** (A) Schematic representation of the HCF-1 full-length protein and the HCF-1rep1 recombinants used in this experiment. (B) *In vivo* cleavage assay. HEK 293 cells were transfected with the indicated constructs and subjected to N-terminal  $\alpha$ -HA-epitope tag immunoprecipitation. The uncleaved precursor protein and N-terminal cleavage products were visualized by  $\alpha$ -GST immunoblot. (C) *In vitro* cleavage assay with 4 h incubation time. Resulting uncleaved precursor and N-terminal cleavage products were analyzed for cleavage by  $\alpha$ -GST immunoblot. In (B) and (C), the positions of prominent (•) and faint (○) cleavage products are indicated.

### III.5 Region II activity is dependent on Region IV

Region II displays not only sequence, but also positional specificity for HCF-1<sub>PRO</sub>-repeat cleavage enhancement. As the latter indicates a context-specific activity of Region II, I tested whether the less well-conserved Region IV, consisting of 36 amino acids C-terminal of the HCF-1<sub>PRO</sub> repeat 1, is required for Region II activity. I therefore deleted Region IV in the HCF-1<sub>PRO</sub> rep1 FL or in the +II construct, resulting in constructs  $\Delta$ IV and +II\_ $\Delta$ IV (Figure III-9 A). In an *in vivo* cleavage assay (Figure III-9 B), the FL construct (lane 1) displayed evident cleavage activity, whereas the  $\Delta$ IV construct (lane 2) displayed less activity. This result suggests that Region IV might play a role for HCF-1<sub>PRO</sub>-repeat cleavage. Consistent with this observation, +II\_ $\Delta$ IV (lane 4) also displayed reduced cleavage with respect to +II (lane 3). These results were more evident in an *in vitro* cleavage assay (Figure III-9 C). These findings indicate that the sequence C-terminal of the first HCF-1<sub>PRO</sub> repeat, Region IV, is important for Region II's cleavage-enhancement activity.



**Figure III-9: The effect of Region II on HCF-1<sub>PRO</sub>-repeat cleavage is dependent on Region IV.**

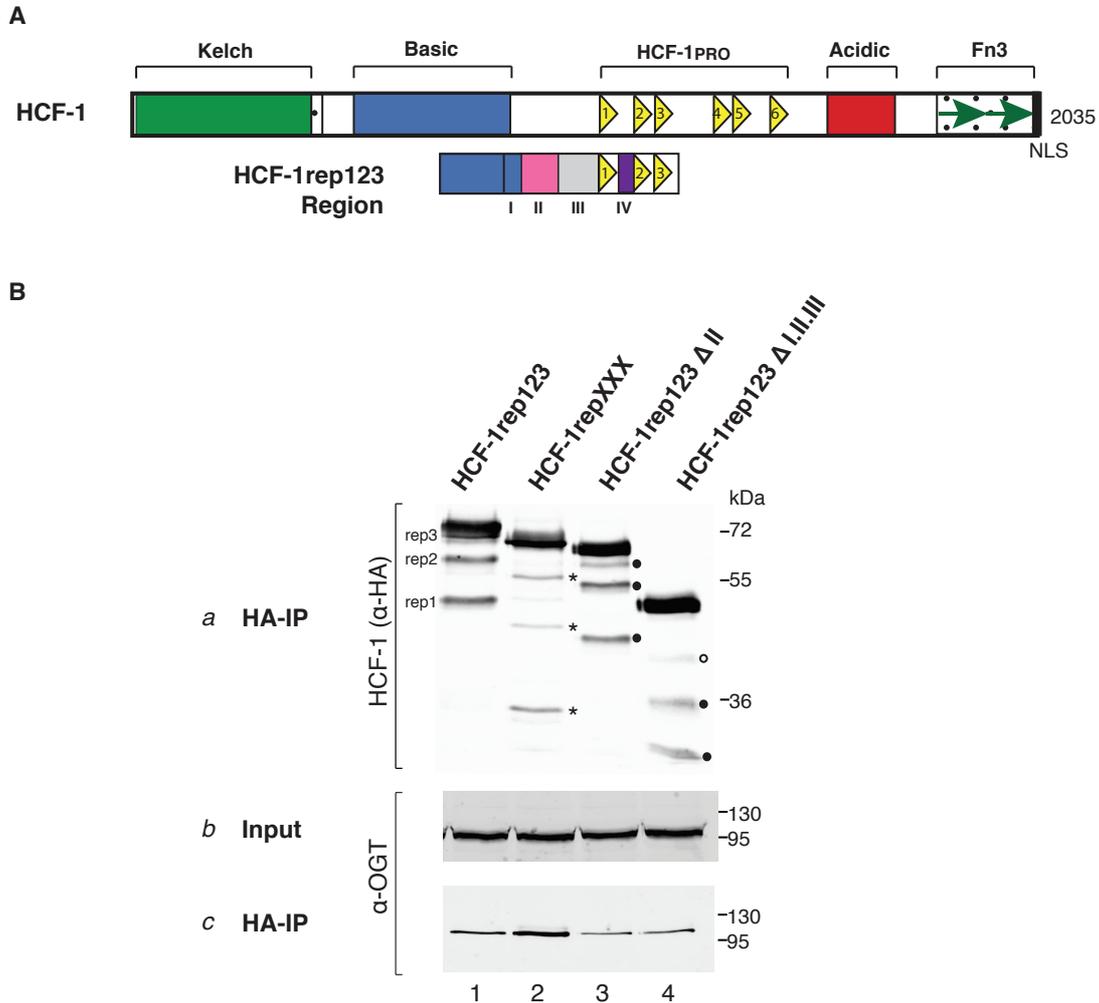
(A) Schematic representation of the HCF-1 full-length protein and the HCF-1rep1 recombinants used in this experiment. (B) *In vivo* cleavage assay. HEK 293 cells were transfected with transfection medium (mock) or with the indicated constructs and subjected to N-terminal  $\alpha$ -HA-epitope tag immunoprecipitation. The uncleaved and cleaved products were visualized by  $\alpha$ -HA immunoblot. (C) *In vitro* cleavage assay with 4 h incubation time. Resulting precursor and N-terminal cleavage products were analyzed for cleavage by  $\alpha$ -GST immunoblot. In (B) and (C), the positions of prominent (•) and faint cleavage products (○) are indicated.

### III.6 Regions I, II, and III together potentially affect cleavage of HCF-1<sub>PRO</sub> repeats 1, 2, and 3

I have identified activities of conserved sequences lying N-terminal of the first HCF-1<sub>PRO</sub> repeat, in particular Region II, on cleavage of the HCF-1<sub>PRO</sub> repeat 1. As each of the six HCF-1<sub>PRO</sub> repeats represents a functional cleavage site for OGT and preference for cleavage of a particular repeat could not be determined (Wilson et al., 1995b), I tested whether the sequences comprising Regions I, II, and III could also affect cleavage of the HCF-1<sub>PRO</sub> repeats 2 and 3. To this end, I used an HCF-1 substrate called HCF-1rep123, which contains half of the HCF-1 Basic Region, Regions I–IV, and the first three HCF-1<sub>PRO</sub> repeats (Capotosti et al., 2011; Capotosti et al., 2007; Figure III-10 A). I created deletion constructs HCF-1rep123  $\Delta$ II and  $\Delta$ I.II.III. Because the HCF-1rep123 protein is insoluble during bacterial synthesis and purification, I subjected HCF-1rep123 and its deletion constructs to an *in vivo* cleavage assay. Figure III-10 B shows an OGT co-immunoprecipitation assay with wild-type or mutant HCF-1rep123 constructs transiently expressed in HEK 293 cells. HCF-1rep123 is cleaved at all three HCF-1<sub>PRO</sub> repeats (lane 1, panel a) and associates with OGT (lane 1, panel c). The same construct containing E10A mutations within all three repeats (HCF-1repXXX) did not display cleavage (lane 2, panel a), but enhanced OGT association (lane 2, panel c), as expected (see Chapter II and Capotosti et al., 2011). Deletion of Region II (HCF-1rep123  $\Delta$ II) did not cause any detectable decrease of cleavage activity of HCF-1<sub>PRO</sub> repeats 1, 2 and 3 (lane 3, panel a), but decreased OGT association when compared to wild-type HCF-1rep123 (compare lane 3 to lane 1, panel c). Deletion of Regions I, II, and III together, slightly decreased HCF-1<sub>PRO</sub>-repeat cleavage. The strongest decrease of cleavage was observed for HCF-1<sub>PRO</sub> repeat 3 (lane 4, panel a, open circle). Interestingly, this construct also displayed decreased OGT association (lane 4, panel c).

I concluded that deletion of Region II alone, in the more natural context of the HCF-1rep123 precursor construct, does neither affect cleavage of HCF-1<sub>PRO</sub> repeat 1, nor of HCF-1<sub>PRO</sub> repeats 2 and 3 in a transient *in vivo* assay. Deleting Regions I, II, and III together, however, slightly affects the cleavage efficiency of the three HCF-1<sub>PRO</sub> repeats. In transient *in vivo* cleavage assays, the simultaneous synthesis and processing of HCF-1 precursor proteins impedes the determination of differences between the activities of different regions on cleavage (III.1.3). It is thus possible that Regions I, II, and III also play a role for the enhancement of HCF-1<sub>PRO</sub>-repeat cleavage in the HCF-1rep123 context or in the HCF-1 full-length context. OGT co-immunoprecipitation revealed that deletion of Region II decreases OGT association with the HCF-1 precursor HCF-1rep123, suggesting that Region II can associate with OGT. This result was surprising, as in previous studies (Capotosti et al., 2011;

Daou et al., 2011) the main OGT-interaction domain was attributed to the HCF-1<sub>PRO</sub> repeats. The role of Region II in OGT association is addressed in Chapter IV.



**Figure III-10: Regions I, II, and III together affect cleavage of a substrate containing HCF-1<sub>PRO</sub> repeats 1, 2, and 3.**

(A) Schematic representation of full-length HCF-1 and the precursor protein HCF-1rep123 (amino acids 686-1166) containing Regions I, II, III, and IV as illustrated. (B) *In vivo* cleavage assay and OGT co-immunoprecipitation. HEK 293 cells were transfected with the wild-type HCF-1rep123 construct (lane 1), and HCF-1rep123 constructs containing mutated, non-cleavable HCF-1<sub>PRO</sub> repeats (HCF-1repXXX; X=E10A, lane 2), or lacking either Region II (HCF-1rep123 ΔII, lane 3) or Regions I, II, and III together (HCF-1rep123 ΔI.II.III, lane 4). Transiently synthesized proteins were subjected to native N-terminal α-HA-epitope tag immunoprecipitation (HA-IP, panel a). Co-immunoprecipitated material was probed for OGT association via α-OGT immunoblot (HA-IP, c). The uncleaved and N-terminal cleavage products were visualized by α-HA immunoblot. The positions of N-terminal cleaved products at the three repeats (rep1, rep2, rep3) and prominent (•) and faint (○) cleavage products are indicated. (\*) Non-specific bands.

## Discussion

### HCF-1<sub>PRO</sub>-repeat cleavage can get enhanced by flanking HCF-1 sequences

HCF-1<sub>PRO</sub>-repeat proteolysis is a slow reaction compared to other proteolytic reactions. For instance, complete casein digestion by trypsin occurs in the range of minutes (Fraser and Powell, 1950). In contrast, maximum HCF-1<sub>PRO</sub>-repeat cleavage (approximately 50 %) is reached only after 8 h of incubation with OGT. HCF-1<sub>PRO</sub>-repeat cleavage is even less efficient in the absence of sequences flanking the first HCF-1<sub>PRO</sub> repeat (Regions I, II, and III), as an HCF-1rep1 construct lacking these regions (construct  $\Delta$ I.II.III) was cleaved only at an efficiency of approximately 20 % after 8 h (Figure III-5 B). Of all regions, Region II, a 58 amino acid sequence rich in serines and threonines, displayed most prominent cleavage-enhancement activity. As cleavage enhancement occurs prominently at the early time points of the reaction, I propose that Region II promotes cleavage in a time-dependent manner.

Region II cleavage enhancement shows similarities with the activity of transcriptional enhancers. Transcriptional enhancers are cis-acting DNA regulatory elements that can up-regulate (i.e. enhance) the transcription of target genes. In many cases, they can be found at a large distance to their target genes. Moreover, they are proposed to form loops, a property that might enable them to interact with their targets (reviewed in Plank and Dean, 2014). Region II at the protein level, lies at a distance of 70 amino acids from the HCF-1<sub>PRO</sub> repeat 1 cleavage site and enhances HCF-1<sub>PRO</sub>-repeat cleavage — a parallel to enhancement of transcription by transcriptional enhancers at the DNA level. Nevertheless, it is not known if Region II can form loops to interact with the HCF-1<sub>PRO</sub>-repeat cleavage site from this distance. Transcriptional enhancers can recruit transcription factors. Interestingly, in one experiment, OGT recruitment by Region II was indicated by an *in vivo* co-immunoprecipitation assay (Figure III-10). Region II–OGT association is addressed in Chapter IV.

Regions I, II, and III are required for efficient HCF-1<sub>PRO</sub>-repeat proteolysis. This raises the question, why an HCF-1 substrate lacking Regions I–III, the HCF3R construct (schematic in Figure 1 A in Lazarus et al., 2013), is effectively cleaved? HCF3R contains the first HCF-1<sub>PRO</sub> repeat and a stretch of C-terminal sequences covering Region IV and HCF-1<sub>PRO</sub> repeats 2 and 3 and does not undergo substantial O-GlcNAcylation (Lazarus et al., 2013). HCF3R was incubated with OGT and UDP-GlcNAc for 5 h, and resulting reaction products were resolved by SDS-PAGE (Lazarus et al., 2013). In this experiment, cleavage of HCF-1<sub>PRO</sub> repeat 3 was less efficient than cleavage of the other two HCF-1<sub>PRO</sub> repeats. This is consistent with results from an *in vivo* cleavage assay with HCF-1rep123 constructs, in which the absence of Regions I–III decreased cleavage of the third HCF-1<sub>PRO</sub> repeat (Figure III-10).

Nevertheless, to compare cleavage activities between the HCF3R and HCF-1rep1 constructs, they have to be examined in parallel in the same experiment, but this has not been done yet. In summary, efficient HCF-1<sub>PRO</sub>-repeat cleavage is promoted by sequences flanking the first HCF-1<sub>PRO</sub> repeat.

### **Region II activity on HCF-1<sub>PRO</sub>-repeat cleavage is complex**

To determine the properties of Region II's cleavage-enhancement activity, I performed genetic analyses of the Region II sequence. Placing Region II C-terminal of the HCF-1<sub>PRO</sub> repeat inhibited cleavage (Figure III-8) and removal of the non-conserved Region IV also impaired Region II activity (Figure III-9). I conclude that Region II activity is complex and dependent on its HCF-1 context. One explanation for this context dependence could be that the sequences surrounding Region II contribute to Region II activity via their structural properties. The HCF-1rep1 sequences comprising residues 867-1071 are, however, predicted to be unstructured. In fact, the online bioinformatics tool PSIPRED (Jones, 1999), predicted very little secondary structure and thus the potential to form loops in Region II (Dr. Ute Roehrig, unpublished results). Thus, structural impairments caused by the HCF-1rep1 secondary structure are unlikely. Moreover, placing Region II in between two structured domains, the POU-specific and the POU-homeo domain, does not impair Region II activity (Figure III-6). Nevertheless, it remains to be determined whether HCF-1-OGT complexes possess structural properties that can influence Region II activity and HCF-1<sub>PRO</sub>-repeat proteolysis.

Interestingly, a scrambled Region II sequence did not activate HCF-1<sub>PRO</sub>-repeat cleavage (Figure III-7), indicating sequence specificity for Region II activity. It is therefore possible that Region II allosterically activates HCF-1<sub>PRO</sub>-repeat cleavage by binding to the enzyme, OGT. As described in the discussion of Chapter II, allosteric effectors can bind to a region of the enzyme that does not participate directly in substrate recognition and processing (allosteric site). The allosteric site can enhance substrate to product transition through conformational changes in the enzyme. Indeed, allosteric activation by proteins derived from nucleosomes has been described for polycomb repressive complex 2 (PRC2) activity (Yuan et al., 2012). In this case, an amino acid stretch derived from histone H3 enhances PRC2 activity through binding to one of the PRC2 subunits. A scrambled histone H3 peptide, however, does not display this activity, similarly to what I observed using the Region II scrambled sequence. PRC2 is a large protein complex quite distinct from OGT as it is composed of four different proteins that interact with each other (oligomeric complex). In fact, most allosteric enzymes are oligomeric. In contrast, OGT consists of only one polypeptide chain that folds into two structurally distinct domains, the catalytic domain and

the TPR domain (see Figure I-8). It thus remains to be solved if OGT is an allosteric enzyme and if Region II can act as an allosteric effector.

### **HCF-1 cleavage activity correlates with HCF-1 O-GlcNAcylation efficiency**

Immunoblot of HCF-1 proteins using  $\alpha$ -O-GlcNAc antibodies after *in vitro* or *in vivo* cleavage assays suggested that Region II contains or activates a number of residues for O-GlcNAcylation (Figures III-4–III-6). These results suggest that Region II activity might be linked to O-GlcNAcylation activity. Nevertheless, it is important to keep in mind that, in the context of the HCF3R construct, O-GlcNAcylation does not appear to be required for proteolysis (Lazarus et al., 2013). I thus conclude that HCF-1 O-GlcNAcylation correlates with HCF-1<sub>PRO</sub>-repeat cleavage efficiency, but that this correlation does probably not reflect causality.

## Chapter IV :

### *HCF-1-OGT-association and O-GlcNAcylation activities and linkage to HCF-1<sub>PRO</sub>-repeat cleavage*

#### Introduction

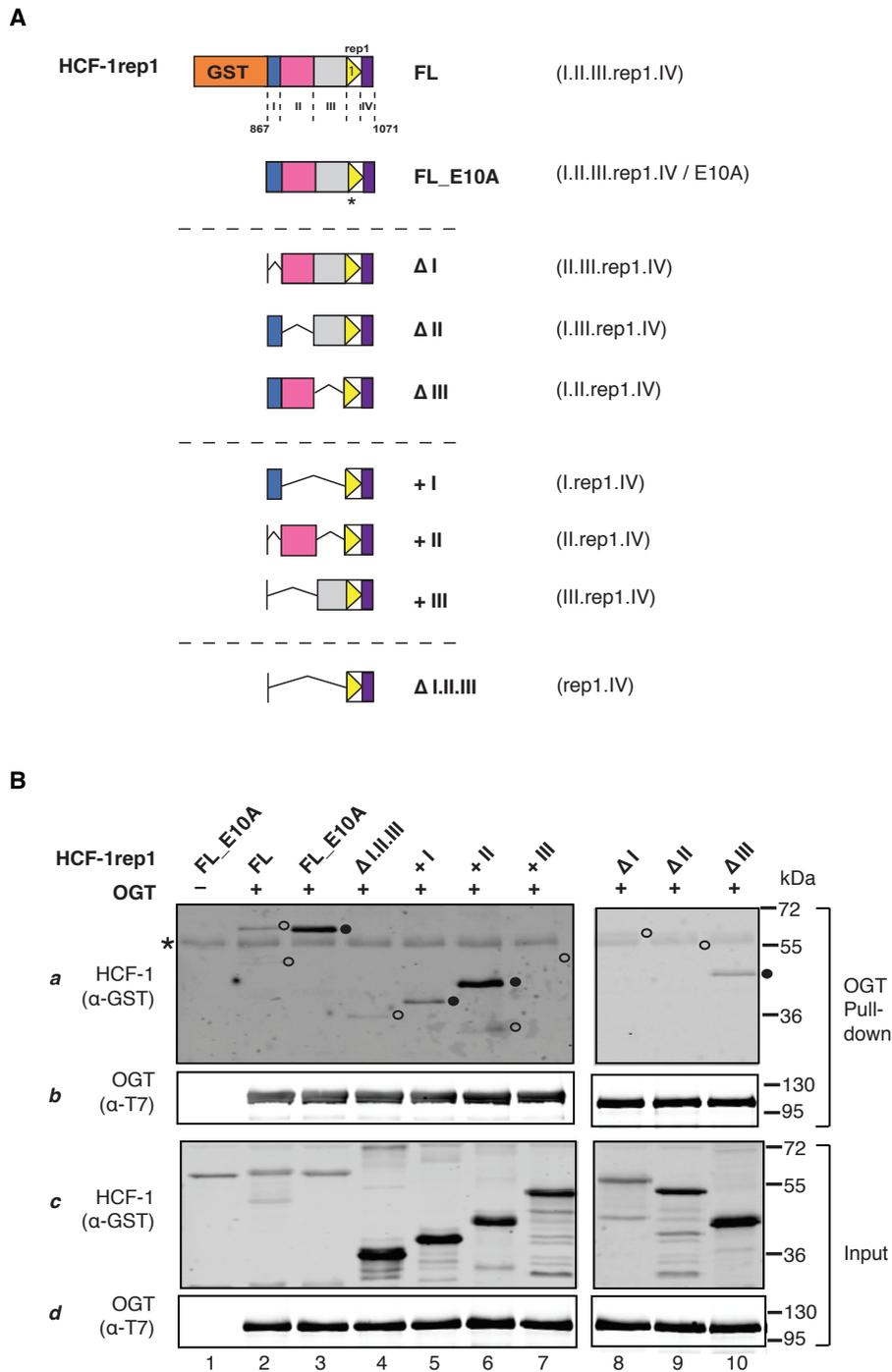
I identified a complex activity of an HCF-1 sequence called Region II, which lies N-terminal of the first HCF-1<sub>PRO</sub> repeat. This region enhances HCF-1<sub>PRO</sub>-repeat cleavage when placed N-terminal of the HCF-1<sub>PRO</sub> repeat. Since, in the aforementioned studies, Region II cleavage activity correlates with O-GlcNAcylation and OGT association, I dissected these three activities further to understand the mechanism of Region II cleavage enhancement. In this chapter, I present mutational, protein–protein association, and proteomic analyses of Region II. Eventually, I propose a hypothetical model that aims to explain Region II activity. This part of the project was developed in collaboration with Dr. Patrice Waridel (Protein Analysis Facility, University of Lausanne).

#### Results

##### IV.1 Region II enhances HCF-1–OGT association

*In vivo* co-immunoprecipitation studies with HCF-1 constructs lacking the cleavage-enhancer sequence Region II indicated that the latter is involved in HCF-1–OGT association (Figure III-10). As Region II influenced HCF-1<sub>PRO</sub>-repeat cleavage neither by its size nor by its amino acid composition (Figure III-7), I investigated whether Region II represents a sequence-specific OGT-binding site. I assayed HCF-1–OGT association in an *in vitro* HCF-1rep1–OGT binding assay, as described in Chapter II. Figure IV-1 B, panel a, shows HCF-1 recovery from this OGT-directed pull-down assay with a set of HCF-1rep1-deletion substrates (illustrated in Figure IV-1 A). As expected, full-length wild-type HCF-1rep1 (FL, lane 2) bound more weakly to OGT than the E10A mutant (FL\_E10A, lane 3). The  $\Delta$ I.II.III deletion construct displayed weak OGT association (lane 4), correlating with its low cleavage efficiency (see Chapter III). The +II construct bound strongly to OGT in comparison with +I or +III (compare lane 6 to lanes 5 and 7), correlating with strong HCF-1<sub>PRO</sub>-repeat cleavage enhancement by Region II (Chapter III). Surprisingly, Region II without Regions I and III (lane 6) displayed higher OGT affinity than in the context of its surrounding regions in FL (lane 2), indicating inhibitory effects of Region I and/or III on OGT binding. When assaying constructs containing

individual deletions of Region I ( $\Delta I$ , lane 8), Region II ( $\Delta II$ , lane 9) or Region III ( $\Delta III$ , lane 10), only weak OGT association, if any, could be detected in the absence of Region I (lane 8) or Region II (lane 9). The absence of Region III, however, promoted moderate OGT binding (lane 10), suggesting that Region III has inhibitory effects on Region II–OGT association. These highly reproducible results show that Region II enhances HCF-1–OGT association. Region II might therefore contain an OGT-binding site.

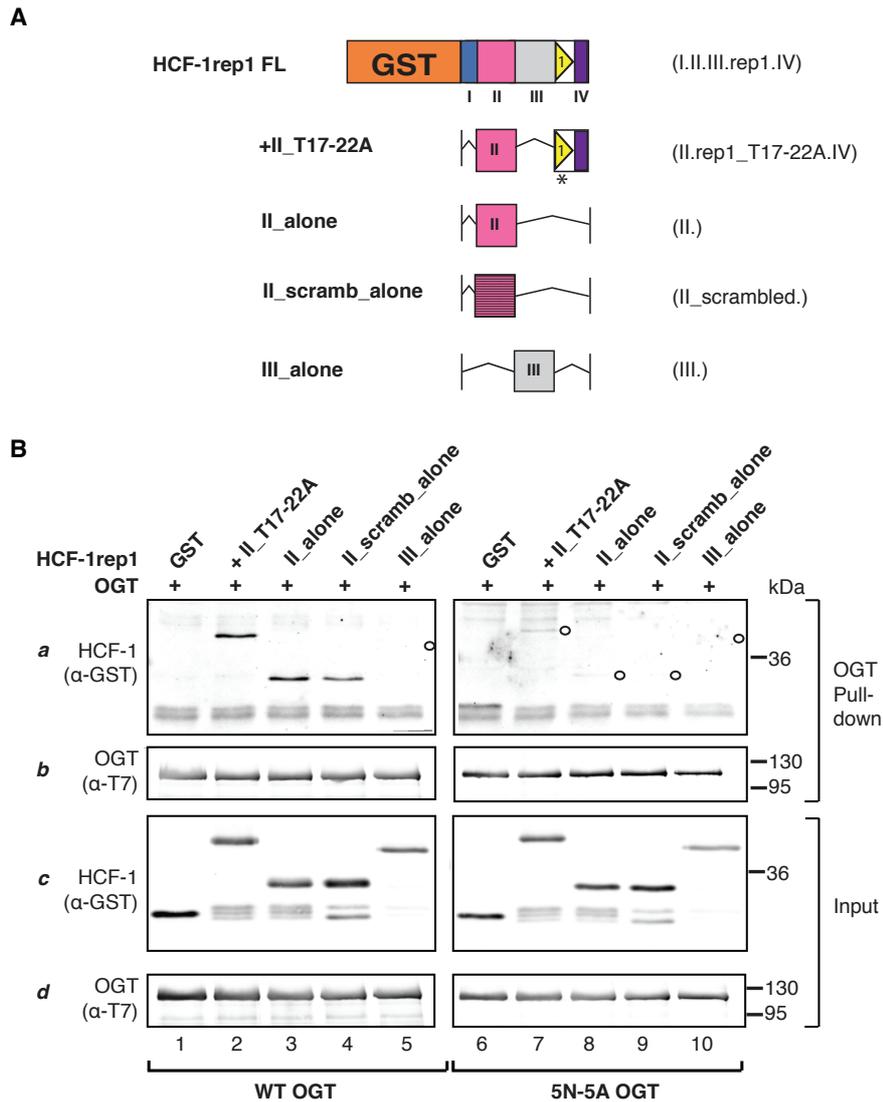


**Figure IV-1: Region II enhances HCF-1rep1-OGT association.**

(A) Schematic of the full-length (FL) HCF-1rep1 precursor construct and the HCF-1rep1-deletion constructs used in this experiment. (B) *In vitro* HCF-1-OGT binding assay with HCF-1rep1 constructs containing WT, or E10A HCF-1<sub>PRO</sub>-repeats (lanes 1-3) or with deletion constructs either containing (lanes 4-7) or lacking Regions I, II or III (lanes 8-10), respectively, in the presence of UDP-GlcNAc. Shown are 100 % of OGT pull-down (panels a and b) and 11 % of the input (panels c and d). Immunoblotting with antibodies directed to GST or to the T7 epitope was used to detect GST-HCF-1rep1 and OGT, respectively. The positions of HCF-1 proteins displaying strong (•) or weak (○) OGT binding are indicated. IgG heavy chain (\*).

## IV.2 Region II represents an independent OGT-binding sequence

As the HCF-1<sub>PRO</sub> repeat is an OGT-binding site (Chapter II and Lazarus et al., 2013), I investigated whether Region II can bind OGT independently of the HCF-1<sub>PRO</sub> repeat and whether the binding mode is similar to the HCF-1<sub>PRO</sub>-repeat-OGT interaction. To test this in an *in vitro* HCF-1-OGT binding assay, I used two main mutational strategies: First, I mutated the HCF-1<sub>rep1</sub> substrate to obtain a set of constructs either containing an OGT-binding-defective HCF-1<sub>PRO</sub> repeat (+II\_T17-22A) or the Region II or Region III sequences alone (II\_alone or III\_alone). Additionally, I engineered a scrambled Region II sequence in the absence of additional HCF-1 sequences (II\_scramb\_alone) to assess if Region II displays sequence-specific binding to OGT (Figure IV-2 A). The second strategy allowed me to understand if the OGT TPR domain is involved in Region II binding. For this purpose, I used the HCF-1 cleavage and O-GlcNAcylation compromised 5N-5A OGT mutant (kindly provided by Dr. Vaibhav Kapuria), containing alanine mutations of five conserved asparagines, which mediate interactions with the HCF-1<sub>PRO</sub> repeat (Lazarus et al., 2013). Figure IV-2 B shows an HCF-1-OGT binding assay with the described constructs in an assay with wild-type OGT (WT) or 5N-5A OGT. The constructs containing the Region II sequence associated strongly with WT OGT (lanes 2 and 3, panel a). Although a scrambled Region II sequence did not enhance HCF-1<sub>PRO</sub>-repeat proteolysis (see Chapter III), this scrambled sequence bound to OGT (lane 4), suggesting that, contrary to Region II's cleavage-enhancement activity, Region II-OGT binding is not dependent on the Region II sequence, but on its amino acid composition. The Region III sequence did not bind to OGT (lane 5), correlating with its weak cleavage-activation potential, and further showing that Region III-OGT association is specific. When the same assay was performed with 5N-5A OGT (right panel a), OGT binding of the Region II sequences was dramatically reduced (lanes 7-9), suggesting a role of the OGT TPR domain in Region II recognition and binding.

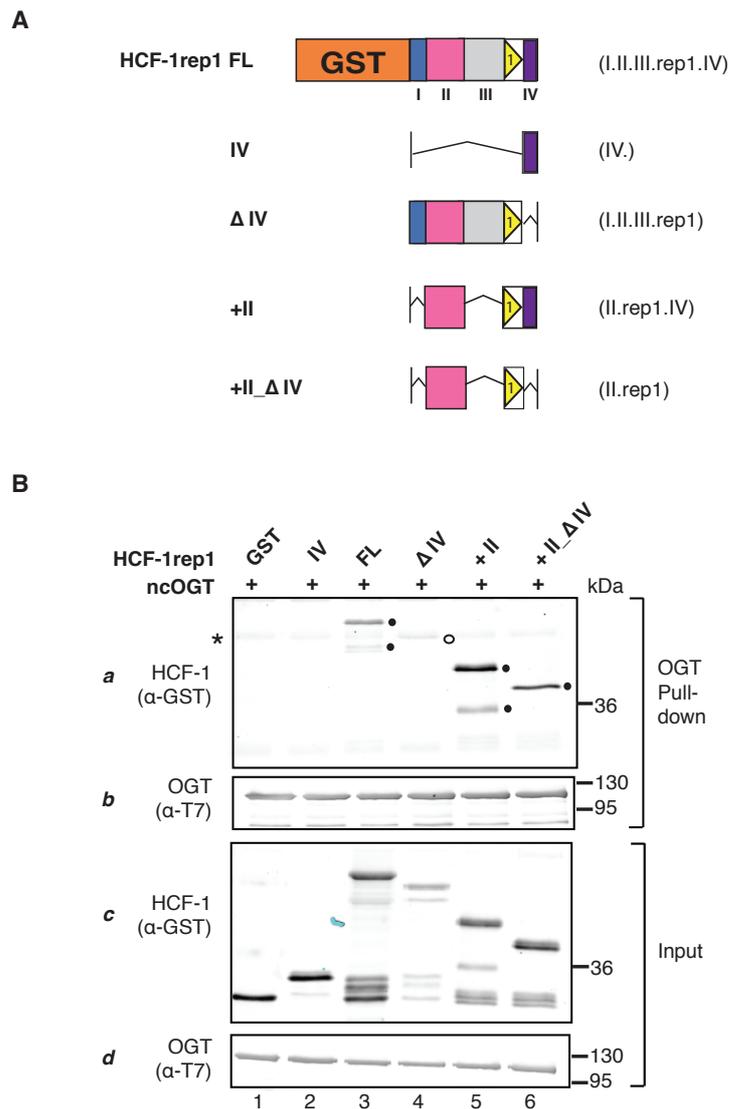


**Figure IV-2: Region II is an independent OGT-binding sequence.**

(A) Schematic representation of the HCF-1rep1 recombinants used in this experiment. (B) *In vitro* HCF-1-OGT binding assay in the presence of UDP-GlcNAc. HCF-1rep1 constructs containing Region II N-terminal of a binding defective HCF-1<sub>PRO</sub> repeat (+II\_T17-22A, lanes 2 and 7), or GST-fusion constructs containing Region II alone (lanes 3 and 8) or a scrambled Region II sequence alone (lanes 4 and 9) were tested for binding with wild-type (WT) OGT (left panel) or with the OGT TPR mutant 5N-5A (right panel). The construct III\_alone (lanes 5 and 10) was used as a negative control for OGT binding. Shown are 100 % of OGT pull-down (panels a and b) and 11 % of the input (panels c and d). Immunoblotting with antibodies directed to GST or to the T7 epitope was used to detect GST-HCF-1rep1 and OGT, respectively. The positions of HCF-1 proteins displaying weak (○) OGT binding are indicated.

Because Region IV interferes with Region II activity for cleavage (Figure III-9), I asked next, whether the non-conserved Region IV also interferes with Region II–OGT binding and in this case, if Region IV represents an independent OGT–binding site. To test this, I engineered an HCF-1 construct containing only Region IV (IV), and I used the deletion constructs containing or lacking Region IV that were tested for cleavage activity in Chapter III (schematics in Figure IV-3 A). Region IV itself does not associate efficiently with OGT (Figure IV-3 B, lane 2, panel a). Deleting Region IV from the HCF-1<sub>rep1</sub> FL construct ( $\Delta$ IV) decreased OGT association (lane 4) with respect to FL–OGT association (lane 3). Consistent with this result, deleting Region IV from the +II construct (+II\_ $\Delta$ IV) reduced Region II–OGT binding efficiency (lane 6) when compared to +II (lane 5). These results suggest that Region IV does not represent an efficient OGT-binding sequence on its own, but it rather cooperates with the Region II sequence to promote HCF-1–OGT association.

Summing up these results, I conclude that the Region II cleavage-enhancer sequence binds autonomously to OGT, representing an independent OGT-binding sequence. As I was not able to detect efficient OGT binding to a single wild-type HCF-1<sub>PRO</sub> repeat in the absence of Region II (Figure IV-1 B,  $\Delta$ I.II.III, lane 4), I conclude that Region II, next to the six HCF-1<sub>PRO</sub> repeats, is one major OGT-binding site of the HCF-1 protein. Furthermore, Region II–OGT interactions are mediated by OGT's TPR domain, suggesting that the TPRs can recognize not only the HCF-1<sub>PRO</sub>-repeat threonine region, but also other HCF-1 sequences lying outside of the HCF-1<sub>PRO</sub> repeat.

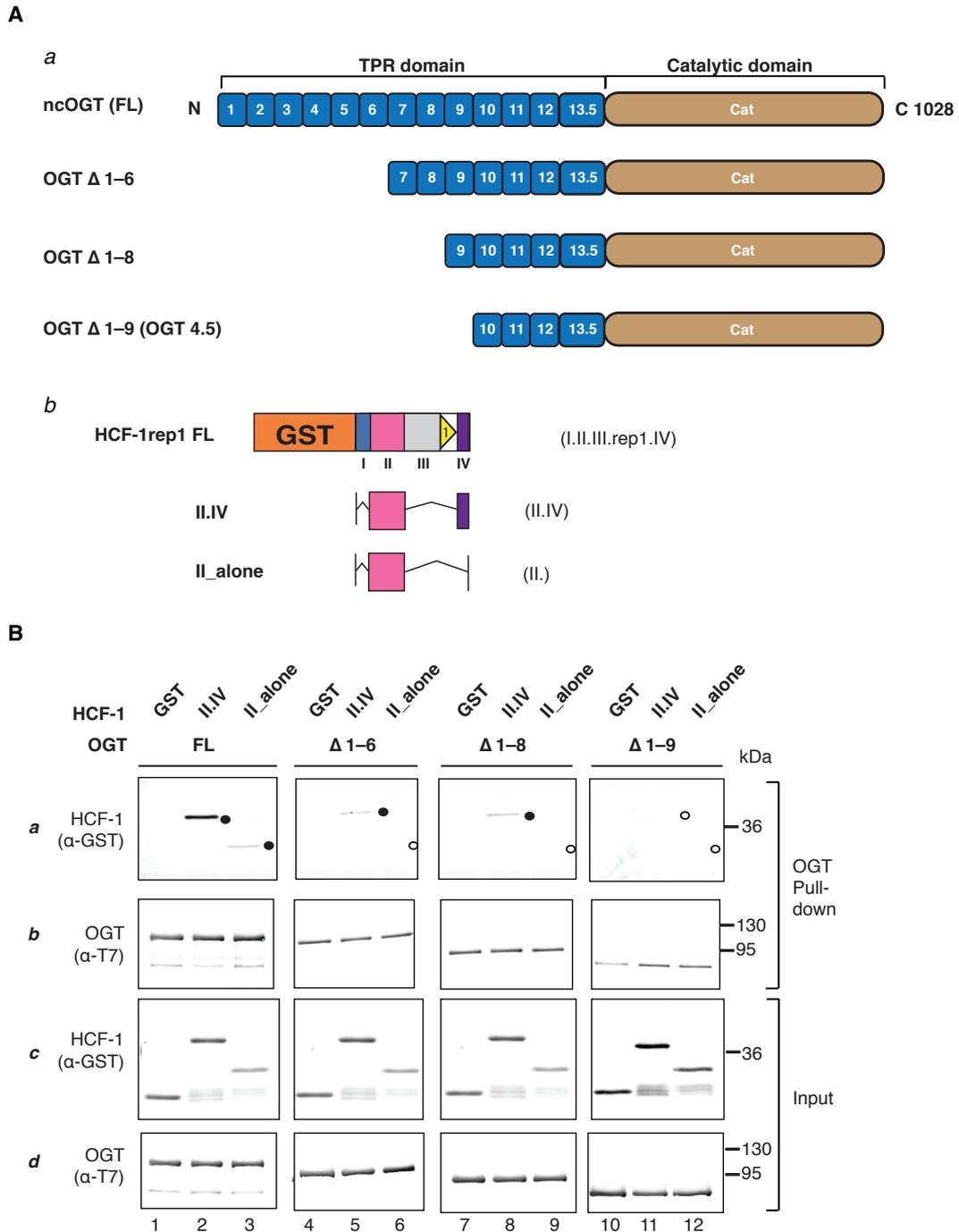


**Figure IV-3: Region IV supports Region II-OGT association.**

(A) Schematics of the full-length (FL) HCF-1rep1 precursor recombinants used in this experiment. (B) *In vitro* HCF-1-OGT binding assay in the presence of UDP-GlcNAc. Shown are 100 % of OGT pull-down (panels *a* and *b*) and 11 % of the input (panels *c* and *d*). Immunoblotting with antibodies directed to GST or to the T7 epitope was used to detect GST-HCF-1rep1 and OGT, respectively. The positions of HCF-1 proteins displaying strong (•) or weak (◦) OGT binding are indicated. IgG heavy chain (\*). The protein band corresponding to HCF-1rep1\_ΔIV (lane 4) co-migrates with the IgG heavy chain.

### IV.3 Efficient Region II–OGT binding requires the OGT TPR domain

The results from the binding experiments above indicated that the OGT TPRs are involved in Region II recognition and binding (Figure IV-2). Furthermore, Region II–OGT binding appears to be supported by Region IV. Therefore I tested which OGT TPRs are required to bind to Region II, and if Region IV is involved in TPR binding. For this binding experiment, I used three OGT TPR mutants (kindly provided by Dr. Vaibhav Kapuria), containing N-terminal deletions of TPRs 1–6, 1–8 or 1–9 (schematically illustrated in Figure IV-4 A; *a*). I used HCF-1 constructs either containing only Region II (II\_alone) or Region II and Region IV (II.IV; Figure IV-4 A; *b*). In general, in *in vitro* HCF-1–OGT binding assays, the bands resulting from GST-directed antibodies (to detect HCF-1 binding) on immunoblots need to be analyzed. This signal, however, is often obscured by the presence of HCF-1 O-GlcNAcylation sites, making the detection of HCF-1 proteins difficult. To assay OGT–HCF-1 binding without the interference of O-GlcNAcylation, I omitted UDP-GlcNAc from the *in vitro* binding assay shown in Figure IV-4 B. Whereas the construct II.IV bound strongly to full-length ncOGT (FL; lane 2, panel *a*), OGT binding efficiency decreased when either OGT TPRs 1–6 ( $\Delta$ 1–6, lane 5) or 1–8 ( $\Delta$ 1–8, lane 8) were removed. OGT binding was not detectable in an assay with OGT containing a deletion of TPRs 1–9 ( $\Delta$ 1–9, lane 11, panel *a*). In contrast, II\_alone–OGT binding was only detected when assayed with FL ncOGT (lane 3, panel *a*) and could not be detected with any of the TPR deletion mutants (lanes 6, 9 and 12). I conclude that efficient Region II association with OGT requires a large proportion of the OGT TPR domain. Region IV can provide stability to Region II–OGT association, consistent with results described in section IV.2 above.



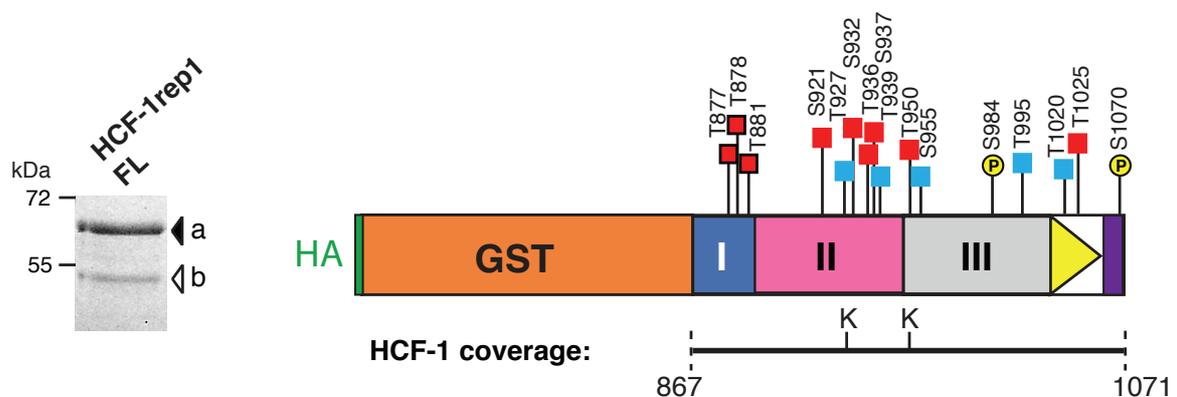
#### IV.4 Region II contains a cluster of O-GlcNAcylation sites

HCF-1 is known to be a highly O-GlcNAcylated protein, and O-GlcNAcylation sites, predominantly in the N-terminal subunit, have been identified in several studies (Capotosti et al., 2011; Myers et al., 2013; Wang et al., 2010a; Daou et al., 2011). Analysis for O-GlcNAcylation by immunoblot of the HCF-1rep1 precursor protein revealed that Region II promotes HCF-1 O-GlcNAcylation (Chapter III). As an analysis via immunoblot using  $\alpha$ -O-GlcNAc antibodies does not reveal the exact positions of the modifications, I subjected the HCF-1rep1 protein to liquid chromatography tandem mass spectrometry (LC-MS/MS) followed by O-GlcNAcylation analysis in collaboration with Dr. Patrice Waridel. O-GlcNAcylation sites in an HCF-1 precursor construct comprising HCF-1 residues 686 to 1166 (HCF-1rep123) have previously been reported (Capotosti et al., 2011). The HCF-1 sequence immediately upstream of the first HCF-1<sub>PRO</sub> repeat, spanning Region II, was not covered by the peptides generated by combined trypsin and Glu-C digestion in this analysis. Thus, O-GlcNAcylation sites could be identified in Region I, but not in Region II and parts of Region III. Since the lack of trypsin and Glu-C sites was the major hurdle to achieve full peptide coverage, I engineered trypsin cleavage sites (lysines) by mutagenesis at two different positions within the HCF-1 sequence (A933K and M951K). The mutations had no effect on HCF-1rep1 cleavage and O-GlcNAcylation levels, as determined by immunoblot (data not shown).

I purified transiently synthesized HCF-1rep1 full-length (FL) with the engineered trypsin cleavage sites from HEK 293 cells and analyzed the precursor uncleaved band (Figure IV-5, band a), as well as the N-terminal cleavage product (band b) for O-GlcNAcylation. The identified O-GlcNAcylation sites of the HCF-1rep1 uncleaved precursor and the N-terminal cleavage product were nearly identical (Table IV-1), suggesting that HCF-1<sub>PRO</sub>-repeat proteolysis has minor effects on the HCF-1rep1 O-GlcNAcylation pattern. Figure IV-5 illustrates schematically the results of the O-GlcNAcylation analysis of the HCF-1rep1 uncleaved precursor protein. Because of the new trypsin cleavage sites, full peptide coverage of the HCF-1 sequence was achieved, and the previously identified confident O-GlcNAcylation sites (red squares; Capotosti et al., 2011) in Region I were confirmed. Additionally, a cluster of four confident (red squares) and two potential (blue squares) O-GlcNAcylation sites in Region II and only one confident and two potential O-GlcNAcylation sites in Region III were identified. The HCF-1<sub>PRO</sub> repeat contained two O-GlcNAcylated sites in the threonine region. These results confirm the O-GlcNAcylation pattern observed by immunoblot of the HCF-1rep1 deletion constructs: Region I contains some, Region III only a few, and Region II contains several (six) O-GlcNAcylation sites. The O-GlcNAcylated

residues in Region II are concentrated in the C-terminal half of Region II and form a cluster, typical for highly O-GlcNAcylated proteins (Trinidad et al., 2012).

Since HCF-1 is phosphorylated (Wysocka et al., 2001a; Dephoure et al., 2008; Olsen et al., 2010) and cross-talk between O-GlcNAcylation and phosphorylation has been observed in previous studies (reviewed in Hart et al., 2011), we also searched for phosphorylation sites in the LC-MS/MS analysis (Figure IV-5, yellow circles). We confirmed a phosphorylation site in Region III at S984 (a potential glycogen synthase kinase-3 (GSK-3) phosphorylation site; Myers et al., 2013) and identified a novel phosphorylation site at S1070 within Region IV. In an *in vivo* cleavage assay, both an alanine substitution of S984 to inhibit phosphorylation and an aspartate substitution of S984 to constitutively mimic phosphorylation did not alter O-GlcNAcylation levels or cleavage efficiencies of the HCF-1rep1 substrates, as detected by immunoblot (data not shown). These results indicate that phosphorylation of S984 in Region III interferes neither with HCF-1<sub>PRO</sub>-repeat cleavage nor HCF-1rep1 O-GlcNAcylation. Curiously, however, whereas we detected phosphorylation of S984 in the HCF-1rep1 precursor and cleaved products, we only detected S984 O-GlcNAcylation in the cleaved product (Table IV-1).



**Figure IV-5: Identification of O-GlcNAcylation and phosphorylation sites in the uncleaved HCF-1rep1 precursor (see also Table IV-1).**

(Left) The full-length (FL) HCF-1rep1 precursor and the N-terminal cleavage product were purified from HEK 293 lysates via  $\alpha$ -HA-epitope immunoprecipitation and visualized by Coomassie staining. The uncleaved precursor (a) and the N-terminal cleavage band (b) were analyzed for O-GlcNAcylation and phosphorylation by LC-MS/MS. (Right) Schematic representation of identified HCF-1rep1 O-GlcNAcylation and phosphorylation sites in the uncleaved HCF-1rep1 precursor. The entire HCF-1 sequences covered by the LC-MS/MS analysis (867-1071) and the engineered trypsin cleavage sites A933K and M951K are indicated below the diagram. Red and blue squares indicate confident or potential O-GlcNAcylated residues, respectively. Squares surrounded in black indicate identified sites in Capotosti et al. (2011). Yellow circles indicate confident phosphorylation sites.

**Table IV-1: Identification of O-GlcNAcylation and phosphorylation sites of the *in vivo* synthesized uncleaved HCF-1rep1 protein and the N-terminal HCF-1rep1 cleavage product.**

Modification	HCF-1rep1 uncleaved	HCF-1rep1 cleaved	Reference
O-GlcNAcylation	T877	T877	Capotosti et al., 2011
O-GlcNAcylation	T878	T878	Capotosti et al., 2011
O-GlcNAcylation	T881	T881	Capotosti et al., 2011
O-GlcNAcylation	N/D	S901, T902, S903, T905 (ambiguous localization)	novel
O-GlcNAcylation	N/D	T918	novel
O-GlcNAcylation	N/D	S920	novel
O-GlcNAcylation	S921	S921	novel
O-GlcNAcylation	T927	N/D	novel
O-GlcNAcylation	S932	S932	novel
O-GlcNAcylation	T936	T936	novel
O-GlcNAcylation	T937	T937	novel
O-GlcNAcylation	T939	T939	novel
O-GlcNAcylation	T950	T950	novel
O-GlcNAcylation	S955	S955	novel
O-GlcNAcylation	N/D	S980	novel
Phosphorylation	S984	S984	Myers et al., 2013
O-GlcNAcylation	N/D	S984	novel
Phosphorylation	T986/T987 (ambiguous localization)	T986/T987 (ambiguous localization)	novel
O-GlcNAcylation	T995	T995	novel
O-GlcNAcylation	T1020	N/A	novel
O-GlcNAcylation	T1025	N/A	novel
Phosphorylation	S1070	N/A	novel

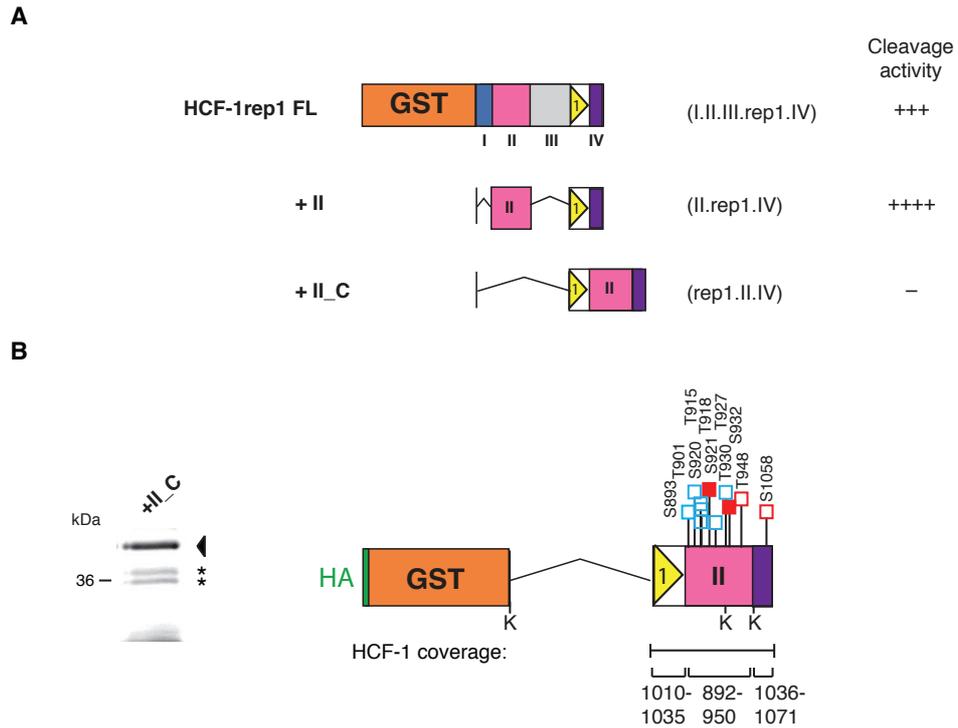
Residues in red and black are confident O-GlcNAcylation and phosphorylation sites (Mascot score > 23 & probability of localization > 70 %) respectively, and residues in blue are potential O-GlcNAcylation sites (Mascot score 14-22 or probability of localization 50-70%).

N/A, not applicable; N/D, not detected

## IV.5 Region II placed C-terminal of the HCF-1<sub>PRO</sub> repeat is differentially O-GlcNAcylated

Since the insertion of Region II C-terminal of the HCF-1<sub>PRO</sub> repeat (construct HCF-1rep1 +II\_C) inhibited HCF-1<sub>PRO</sub>-repeat cleavage (Figure III-8), I investigated whether the C-terminal Region II would still be O-GlcNAcylated and in this case, if the O-GlcNAcylated residues would be identical to Region II located N-terminal of the HCF-1<sub>PRO</sub> repeat (schematics of the constructs in Figure IV-6 A). To ensure full HCF-1 sequence coverage for mass spectrometry analyses of construct HCF-1rep1 +II\_C, I engineered two new trypsin cleavage sites (lysines, one of which lies in the HCF-1 sequence, see Figure IV-6 B) and retained the A933K mutation (described in IV.4 above). The HCF-1 precursor protein was synthesized in HEK 293 cells, purified (Figure IV-6 B, uncleaved band) and analyzed by LC-MS/MS. We obtained full peptide coverage of the HCF-1 sequence, and O-GlcNAcylation sites were identified, as schematically illustrated in Figure IV-6 B. Region II C-terminal of the HCF-1<sub>PRO</sub> repeat contained a total of 10 confident and potential O-GlcNAcylation sites. These are four more sites with respect to the number of sites mapped in the N-terminal Region II in the HCF-1rep1 FL context (Figure IV-5). Moreover, only two of these 10 O-GlcNAcylated residues (S921 and T927) are in common with sites in the N-terminal Region II sequence. In summary, the number of O-GlcNAcylation sites of C-terminal Region II increased, and the O-GlcNAcylation pattern changed, with respect to N-terminal Region II. One additional O-GlcNAcylation site was found in Region IV (S1058).

I conclude that the HCF-1<sub>PRO</sub>-repeat cleavage-enhancer Region II not only loses its activity when placed C-terminal of the HCF-1<sub>PRO</sub> repeat, but also displays an altered O-GlcNAcylation pattern, suggesting that aberrant O-GlcNAcylation might cause loss of Region II activity. Moreover, these results show that HCF-1 O-GlcNAcylation is context sensitive and does not require strong sequence signals.



**Figure IV-6: Region II placed C-terminal of the HCF-1<sub>PRO</sub> repeat displays an altered O-GlcNAcylation pattern.**

(A) Schematics of HCF-1rep1 recombinants analyzed for their cleavage activities (evaluated in Figure III-8). (B) Left-hand-side: HCF-1rep1 +II\_C uncleaved precursor (arrowhead) was purified from HEK 293 lysates via  $\alpha$ -HA-epitope immunoprecipitation, visualized by Coomassie staining and analyzed for O-GlcNAcylation by LC-MS/MS. (\*) Non-specific background bands. Right-hand-side: Schematic representation of identified O-GlcNAcylation sites in +II\_C. HCF-1 sequences covered by the LC-MS/MS analysis and the engineered trypsin cleavage sites (G to K mutation in the GST sequence, and HCF-1 A933K and G1038K) are indicated below the diagram. Red and blue squares indicate confident or potential O-GlcNAcylation sites, respectively. Filled squares indicate sites that were also identified in the N-terminal Region II sequence (HCF-1rep1 FL analysis Figure IV-5); open squares indicate novel O-GlcNAcylation sites that are unique to the +II\_C construct.

## IV.6 Subdivision of Region II leads to reduced cleavage activities of the sub-regions *in vivo*

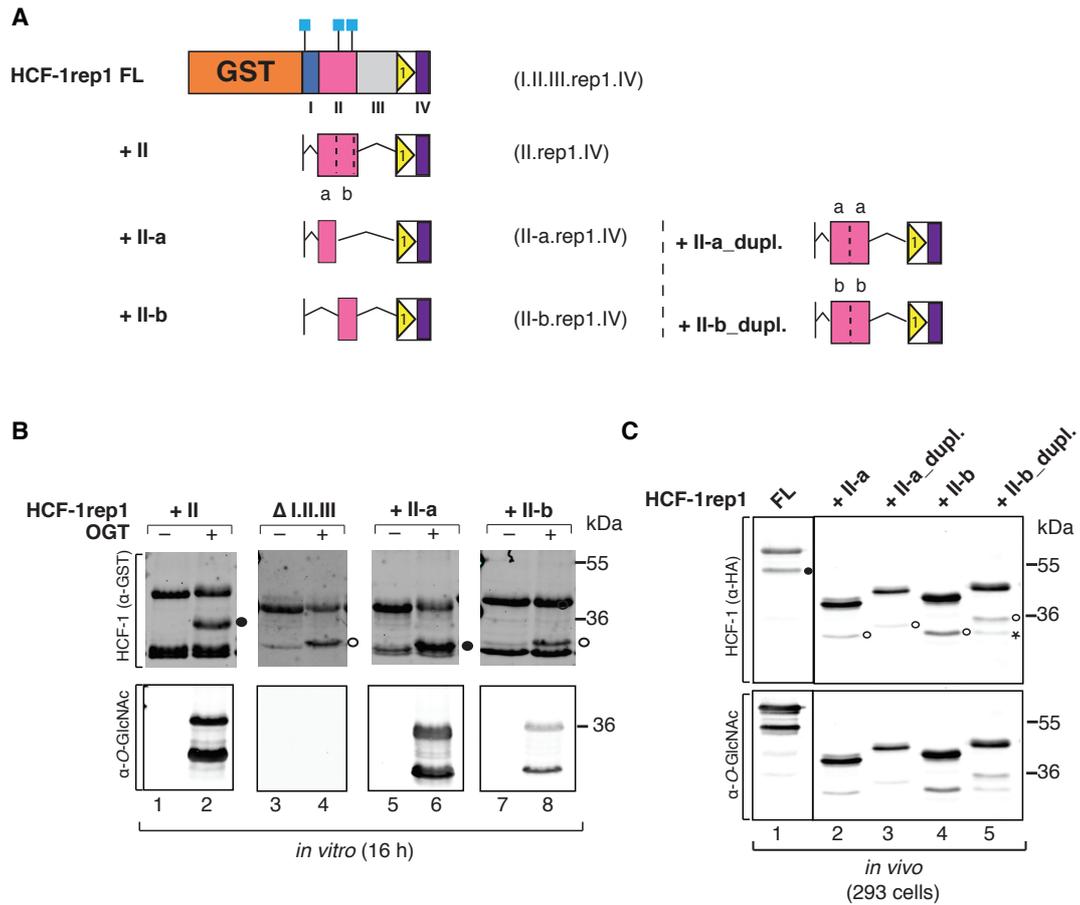
As the Region II cleavage-enhancer sequence consists of 58 amino acids, I investigated whether its activity could be attributed to a smaller sub-region. Mass spectrometry analysis of the HCF-1rep1 precursor protein revealed that Region II contains concentrated O-GlcNAcylation sites at the C-terminus of its sequence (Figure IV-5). I thus divided Region II into a 23 amino acid N-terminal half, called Region II-a, and a 25 amino acid C-terminal half, called Region II-b, covering the concentrated O-GlcNAcylation sites. I created HCF-1rep1 deletion constructs containing either Region II-a or Region II-b (constructs +II-a or +II-b, Figure IV-7 A) and I tested them for cleavage and O-GlcNAcylation activities in a 16 hours *in vitro* cleavage assay (Figure IV-7 B). The +II-a construct displayed prominent cleavage activity (lanes 5 and 6) in comparison with  $\Delta$ I.II.III (lanes 3 and 4) and with +II (lanes 1 and 2), which displayed weak and prominent cleavage activities, respectively. The +II-b construct displayed weak cleavage activity (lanes 7 and 8) similarly to  $\Delta$ I.II.III. It is important to point out that cleavage activities in 16 h *in vitro* cleavage assays are elevated (see Figure III-5). *In vitro* cleavage assays for 4 h showed that cleavage and O-GlcNAcylation activities of constructs +II-a and +II-b were generally weaker (data not shown). Nevertheless, data obtained after 4 h or 16 h incubation gave similar results. Unexpectedly, construct +II-a displayed strong (lanes 5 and 6, lower panel), whereas +II-b displayed little O-GlcNAcylation activity (lanes 7 and 8, lower panel), suggesting that the II-a sequence is hyper- and the II-b sequence is hypo-O-GlcNAcylation when these sequences are isolated from one another. This result was surprising, since I expected the II-b sequence to retain its O-GlcNAcylation sites as in the FL context after the deletion of the surrounding sequences.

I investigated further whether the above-described constructs display similar activities in an *in vivo* cleavage assay (Figure IV-7 C). Contrary to results obtained *in vitro*, in the *in vivo* assay, construct +II-a (lane 2, upper panel) displayed weak cleavage activity when compared to FL HCF-1rep1 (lane 1). Moreover, the +II-b construct (lane 4) displayed slightly more cleavage activity than +II-a, once the ratios between uncleaved and cleaved products were compared (lanes 2 and 4). However, +II-b displayed less cleavage activity than the FL construct (lanes 1 and 4), suggesting that subdivision of Region II results in a loss of cleavage-enhancement activity (note that the *in vivo* activities of FL and +II are similar, see Figure III-4). In contrast to the *in vitro* results, O-GlcNAcylation levels of constructs +II-a and +II-b were similar (compare lanes 2 and 4, lower panel). Duplicating the II-a and II-b sequences (constructs +II-a\_dupl. and +II-b\_dupl.) resulted in similar cleavage and O-GlcNAcylation activities in comparison to constructs containing only a single II-a or II-b sequence. This result suggests that duplications of Region II-a and Region II-b, respectively,

do not display full Region II (not shown in this experiment) or FL activity (compare lanes 3 and 5 and lanes 2 and 4 with lane 1).

The *in vivo* results are in stark contrast with the results obtained *in vitro*. In fact, this was the only analysis, in which I observed great differences between *in vitro* and *in vivo* activities of tested HCF-1 constructs (see Table Appendix-1). I concluded that, in this case, the results obtained *in vivo* reflect the properties of Region II sub-regions more accurately than the *in vitro* data. When I isolated the HCF-1rep1 +II-b protein after bacterial *in vitro* synthesis, I noticed that the yield, compared to the HCF-1rep1 +II-a protein, was much lower. This is indicative of defective protein folding and/or precipitation. Moreover, custom-synthesized Region II-b peptide was insoluble in a range of different solvents (data not shown), supporting the hypothesis that Region II-b solubility is limited. These observations suggest that the *in vitro* cleavage and O-GlcNAcylation activities of the +II-b construct (Figure IV-7 B) have to be interpreted carefully.

The *in vivo* results indicate that splitting the Region II sequence into two sub-regions results in sequences displaying lower cleavage and O-GlcNAcylation activities than the entire Region II sequence. If the individual activities of Region II-a and Region II-b sum up to Region II activity remains unclear because *in vivo* cleavage assays are not suitable to assess cleavage efficiencies from a single experiment in a quantitative manner (described in section III.1). Nevertheless, these results indicate that Region II does not contain a sub-region displaying full Region II activity. Because of potential misfolding of the +II-b construct during bacterial *in vitro* synthesis, I did not assay OGT-binding activities of Regions II-a and II-b in an *in vitro* HCF-1–OGT binding assay, leaving unresolved whether the Region II sub-regions display differential OGT-binding activities.



**Figure IV-7: Region II dissection: analysis of Region II sub-regions for cleavage and O-GlcNAcylation activities.**

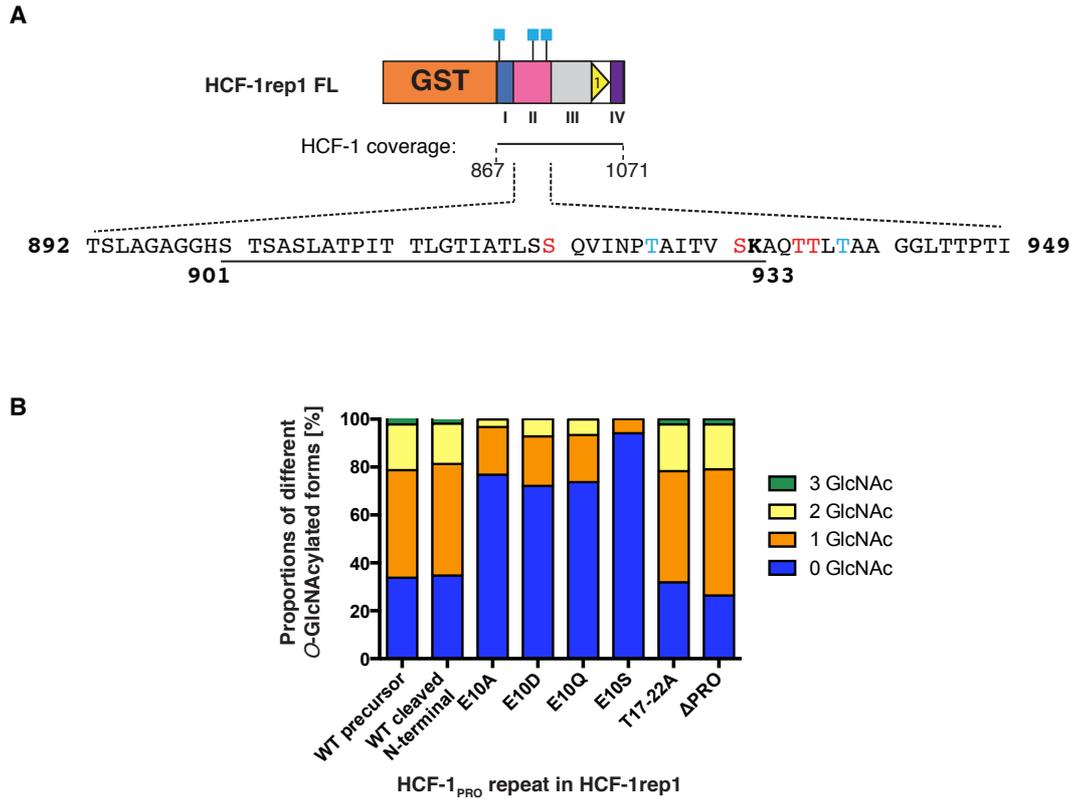
(A) Schematics of HCF-1rep1 recombinants. *In vivo* O-GlcNAcylation sites are schematized as blue squares. (B) *In vitro* cleavage assay with 16 h incubation time. Resulting precursor and N-terminal cleavage products were analyzed for cleavage and O-GlcNAcylation by α-GST and α-O-GlcNAc immunoblot, respectively. (C) *In vivo* cleavage assay with proteins from transiently transfected HEK 293 cells, followed by immunoprecipitation via an N-terminal HA-epitope tag. Cleavage and O-GlcNAcylation were detected by immunoblot using the indicated antibodies. (\*) Background product. In (B) and (C), the positions of prominent (•) and faint (○) cleavage products are indicated.

## IV.7 HCF-1<sub>PRO</sub>-repeat proteolysis does not interfere with Region II O-GlcNAcylation

It has been suggested that inhibition of HCF-1<sub>PRO</sub>-repeat proteolysis by E10A mutations in the HCF-1<sub>PRO</sub> repeat decreases general HCF-1 O-GlcNAcylation levels (Capotosti et al., 2011). To understand if there is a link between Region II O-GlcNAcylation and HCF-1<sub>PRO</sub>-repeat proteolysis, I tested whether O-GlcNAcylation levels in Region II are affected by the absence of HCF-1<sub>PRO</sub>-repeat proteolysis. For this purpose, I assessed the *in vivo* O-GlcNAcylation levels of a representative Region II peptide (peptide sequence underlined in Figure IV-8 A, HCF-1 sequence 901-933) by LC-MS/MS. The LC-MS/MS analysis of HCF-1rep1 proteins revealed that this Region II peptide can either contain 0, 1, 2 or 3 O-GlcNAcylation sites, showing that HCF-1 is modified sub-stoichiometrically. This pool of Region II peptides was either derived from wild-type (WT), or from cleavage inactive HCF-1rep1, and the proportions of their different O-GlcNAcylation forms was assessed. Cleavage inactive mutants containing the E10 mutations (E10A, E10D, E10Q and E10S), displayed decreased Region II O-GlcNAcylation levels when compared to peptides derived from WT uncleaved and cleaved proteins (Figure IV-8 B). These mass spectrometry results were confirmed in a second independent experiment and suggest that the inactivation of HCF-1<sub>PRO</sub>-repeat cleavage by an E10 mutation leads to decreased O-GlcNAcylation of Region II.

Additionally, I analyzed Region II peptides derived from the T17-22A mutant that inhibits cleavage through a defective HCF-1<sub>PRO</sub>-repeat–OGT TPR domain interaction (see Chapter II). Interestingly, Region II peptides derived from this mutant displayed nearly identical O-GlcNAcylation levels in comparison with peptides derived from WT HCF-1rep1. This suggests that OGT binding to the HCF-1<sub>PRO</sub> repeat is not required for O-GlcNAcylation of HCF-1 sequences. Furthermore, these results show that the decreased O-GlcNAcylation levels observed in the HCF-1rep1 E10 mutants are specific to mutations at the E10 cleavage site and not to the inhibition of cleavage *per se*. To support these results, I deleted the entire HCF-1<sub>PRO</sub> repeat in construct HCF-1rep1 to obtain a mutant called HCF-1rep1  $\Delta$ PRO and found nearly identical O-GlcNAcylation levels when compared to WT and T17-22A HCF-1rep1 (Figure IV-8 B).

I drew two conclusions from these results: First, the HCF-1<sub>PRO</sub> repeat is not necessary for O-GlcNAcylation of sequences lying outside of the HCF-1<sub>PRO</sub> repeat (LC-MS/MS analysis of other peptides derived from Region I gives similar results, data not shown). Second, mutations at the E10 cleavage enhance OGT–HCF-1<sub>PRO</sub>-repeat association (see Chapter II) and thereby prevent O-GlcNAcylation of flanking HCF-1 sequences. Thus, HCF-1<sub>PRO</sub>-repeat proteolysis *per se* does not interfere with O-GlcNAcylation of Region II and other flanking HCF-1 sequences.

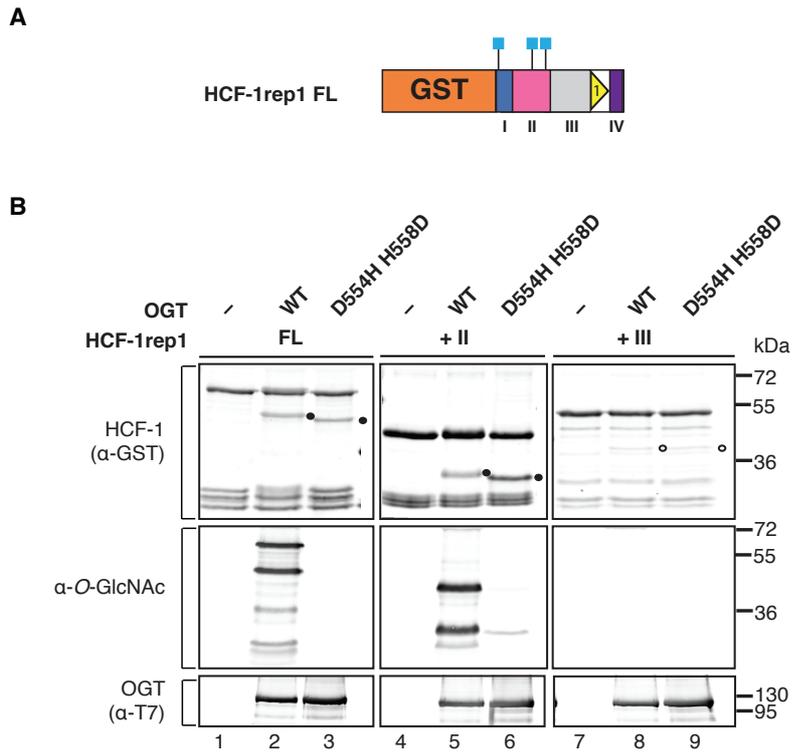


**Figure IV-8: The Region II O-GlcNAcylation status is influenced by HCF-1<sub>PRO</sub>-repeat E10 mutations but not by the absence of HCF-1<sub>PRO</sub>-repeat cleavage *per se*.**

(A) Top: Schematic representation of the full-length (FL) HCF-1rep1 precursor protein. *In vivo* O-GlcNAcylation sites are schematized as blue squares. Bottom: The Region II sequence (confident O-GlcNAcylation sites are colored in red, potential sites are colored in blue) and a representative peptide with an engineered trypsin cleavage site (901-933K) analyzed by LC-MS/MS for O-GlcNAcylation levels in (B). (B) Proportions of Region II peptides corresponding to HCF-1 residues 901-933K (see A) with 0, 1, 2, or 3 attached O-GlcNAc moieties. The samples were derived from HCF-1rep1 constructs synthesized in HEK 293 cells. HCF-1rep1 either contains wild-type (WT) or mutated (E10A, E10D, E10Q, E10S, T17-22A) HCF-1<sub>PRO</sub> repeats or a deletion of the entire HCF-1<sub>PRO</sub> repeat 1 (ΔPRO).

#### **IV.8 Region II enhances HCF-1<sub>PRO</sub>-repeat proteolysis independently of its O-GlcNAcylation status**

As the Region II O-GlcNAcylation status is not influenced by HCF-1<sub>PRO</sub>-repeat proteolysis *per se*, I asked whether HCF-1<sub>PRO</sub>-repeat cleavage could be influenced by Region II O-GlcNAcylation. To address this question, I tested HCF-1<sub>PRO</sub>-repeat proteolysis by an OGT swap mutant that retains wild-type cleavage activity, but is compromised for its O-GlcNAcylation activity (D554H H558D, kindly provided by Dr. Vaibhav Kapuria). The D554H H558D mutant displayed no difference in binding of Regions I, II or III as compared to wild-type (WT) OGT in an *in vitro* HCF-1–OGT binding assay (data not shown). I then compared cleavage and O-GlcNAcylation activities of WT and D554H H558D OGT on HCF-1rep1 FL or +II and +III substrates. Cleavage efficiencies of substrates cleaved by WT OGT or by D554H H558D OGT were nearly identical (Figure IV-9 B, upper panel, lanes 2, 3, 5, 6, 8 and 9). Importantly, whereas O-GlcNAcylation was detected in the HCF-1rep1 FL and +II construct when cleaved by WT OGT (lanes 2 and 5, middle panel), highly reduced O-GlcNAcylation was observed when cleaved with the D554H H558D mutant (lanes 3 and 6, middle panel), showing that HCF-1<sub>PRO</sub>-repeat cleavage can be observed although O-GlcNAcylation of HCF-1 sequences is compromised. (The lack of D554H H558D O-GlcNAcylation activity explains the difference in mobility of WT and D554H H558D OGT cleavage products during electrophoresis.) These results suggest that Region II enhances HCF-1<sub>PRO</sub>-repeat cleavage independently of its O-GlcNAcylation status. I thus conclude that HCF-1 O-GlcNAcylation and HCF-1<sub>PRO</sub>-repeat proteolysis do not influence each other.



**Figure IV-9: Region II O-GlcNAcylation is not fundamental for Region II activity on HCF-1<sub>PRO</sub>-repeat proteolysis.**

(A) Schematic representation of the HCF-1rep1 full-length (FL) construct. *In vivo* O-GlcNAcylation sites are schematized as blue squares. (B) *In vitro* cleavage and O-GlcNAcylation activities (1 h incubation time) of wild-type (WT) OGT and an O-GlcNAcylation compromised OGT mutant (D554H H558D) on selected HCF-1rep1 substrates. Cleavage and O-GlcNAcylation were analyzed by immunoblot using the indicated antibodies. The positions of prominent (•) and faint (◦) cleavage products are indicated.

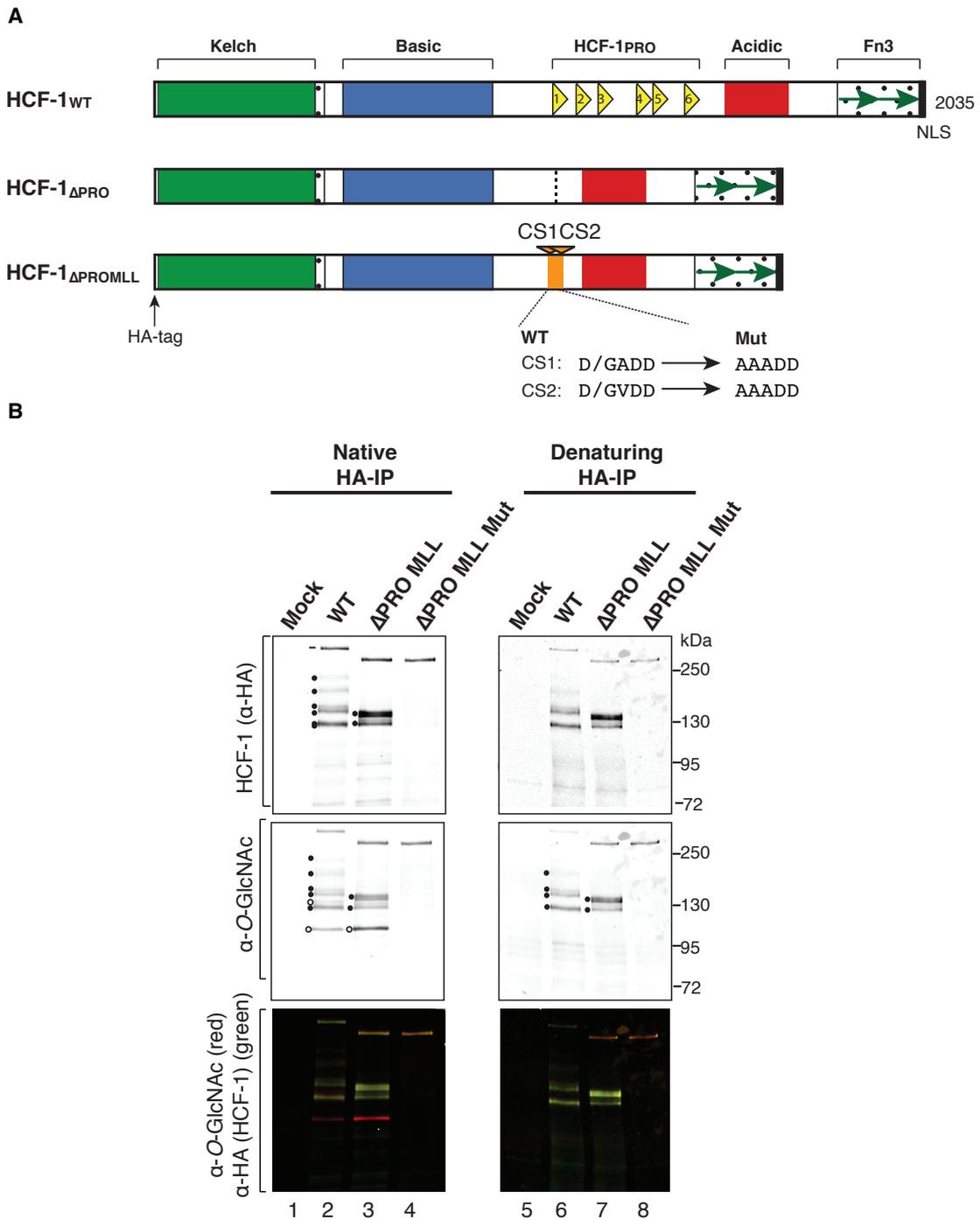
## IV.9 HCF-1 O-GlcNAcylation levels are independent of OGT-mediated HCF-1<sub>PRO</sub>-repeat processing

Capotosti et al. (2011) showed that OGT-mediated proteolysis is important for proper cell-cycle progression through the activation of the HCF-1<sub>C</sub> subunit. A full-length HCF-1 construct, in which the entire HCF-1<sub>PRO</sub>-repeat domain had been replaced by two Taspase1 cleavage sites, CS1 and CS2 of the MLL protein (see Figure I-4), was not able to rescue cell-division defects caused by HCF-1 knockdown in HeLa cells. As the HCF-1 construct containing the Taspase1 cleavage sites was efficiently cleaved, the authors concluded that OGT-mediated HCF-1 proteolysis, but not proteolysis *per se*, is important to activate HCF-1<sub>C</sub> functions. This led to the hypothesis that proper HCF-1<sub>N</sub> O-GlcNAcylation might be involved in the promotion of HCF-1 functionality. But the O-GlcNAcylation status of the Taspase1 cleaved HCF-1<sub>N</sub> and HCF-1<sub>C</sub> subunits was not assessed.

To determine whether the HCF-1 subunits produced by Taspase1 are still efficiently O-GlcNAcylated, I tested the O-GlcNAcylation status of the HCF-1 constructs full-length wild-type (WT), and HCF-1 full-length  $\Delta$ PRO with functional (MLL) or mutated (MLL Mut) MLL Taspase1 cleavage sites (Figure IV-10 A). I transiently transfected these HCF-1 recombinants in HEK 293 cells and subjected them to N-terminal immunoprecipitation (IP), followed by SDS-PAGE and immunoblot with  $\alpha$ -HA or  $\alpha$ -O-GlcNAc antibodies. Native IP of the ectopic HCF-1 full-length wild-type (WT) protein and subsequent immunoblotting using antibodies directed towards the N-terminus ( $\alpha$ -HA) revealed an uncleaved precursor band and N-terminal cleavage products (Figure IV-10 B, lane 2, upper panel).  $\Delta$ PRO MLL displayed an uncleaved precursor band and two N-terminal cleavage products (lane 3, upper panel), whereas the construct containing the mutated Taspase1 cleavage sites ( $\Delta$ PRO MLL Mut) displayed an uncleaved band and no cleavage products (lane 4, upper panel). HCF-1 WT,  $\Delta$ PRO MLL and  $\Delta$ PRO MLL Mut displayed prominent O-GlcNAcylation signals (lanes 2-4, middle panel). Comparing the HCF-1 WT with the HCF-1  $\Delta$ PRO MLL construct revealed no apparent difference in O-GlcNAcylation levels (compare lanes 2 and 3, middle panel). Interestingly, both constructs displayed a C-terminal O-GlcNAcylation cleavage product (lanes 2 and 3, middle and bottom panels), which disappeared upon boiling of the protein complexes before IP (denaturing IP, lanes 6 and 7, middle and bottom panels).

These results suggest that the HCF-1 construct containing Taspase1 cleavage sites ( $\Delta$ PRO MLL) is efficiently O-GlcNAcylated, despite the lack of OGT-mediated proteolysis and the lack OGT-HCF-1<sub>PRO</sub>-repeat association. It is important to point out that O-GlcNAcylation analysis by immunoblot cannot reveal subtle changes in O-GlcNAcylation levels or aberrant O-GlcNAcylation patterns, and my studies have shown that these can easily vary in unexpected ways (Figures IV-6 and IV-7). Nevertheless, these results are in concordance

with the O-GlcNAcylation analysis of a smaller HCF-1 fragment, the HCF-1rep1  $\Delta$ PRO mutant, which did not display different O-GlcNAcylation levels and patterns when compared to HCF-1rep1 WT (see Figure IV-8). This analysis also shows that the HCF-1<sub>C</sub> subunit is significantly O-GlcNAcylated and that it therefore also possesses sites for OGT recognition and O-GlcNAcylation. HCF-1 O-GlcNAcylation is therefore not restricted to just one of its subunits.



**Figure IV-10: HCF-1 O-GlcNAcylation levels are independent of OGT-mediated HCF-1<sup>PRO</sup>-repeat processing.**

(A) Schematic representation of different HCF-1 constructs. CS1 and CS2 indicate the two Taspase1 cleavage sites of the MLL protein in their wild-type (WT) or non-cleavable (Mut) form (Hsieh et al., 2003; constructs engineered by Sophie Guernier, former Herr laboratory member). (B) Native and denaturing N-terminal immunoprecipitation (IP): HEK 293 cells were transfected with transfection medium (mock) or with the indicated HCF-1 constructs and subjected to  $\alpha$ -HA IP using the HA-epitope tag. Precursor and N-terminal cleavage products were visualized by  $\alpha$ -HA and protein O-GlcNAcylation by  $\alpha$ -O-GlcNAc immunoblot. Precursor proteins (-), and the positions of N-terminal (•) and C-terminal O-GlcNAcylated (○) cleavage products are indicated.

## Discussion

### Region II represents a novel OGT-binding sequence

The Region II sequence, identified as a sequence-specific HCF-1<sub>PRO</sub>-repeat-cleavage enhancer in Chapter III, represents a novel OGT-binding sequence. Although a multitude of O-GlcNAcylated and OGT-interacting proteins have been identified in the past (see 1.5 above), only few OGT-binding sequences and their interactions with OGT have been characterized in detail: (i) A 14 amino acid sequence of casein kinase 2 (Lazarus et al., 2011; Lazarus et al., 2012), (ii) a 13 amino acid sequence derived from the innate immunity signaling protein TAB1 (Schimpl et al., 2012), and (iii) a 17 amino acid sequence derived from the HCF-1<sub>PRO</sub> repeat 2 (Lazarus et al., 2013). Region II represents another 58 amino acid OGT-binding sequence (Figure IV-2).

Region II, in addition to its potent OGT-association activity, also possesses strong cleavage-enhancement activity. Thus, I hypothesize that Region II recruits OGT close to the cleavage site at the first HCF-1<sub>PRO</sub> repeat to induce proteolysis. Consistent with this hypothesis, Region II cleavage-enhancement activity is particularly evident at the earliest time points of the cleavage reaction before reaching a plateau (Figure III-5). This indicates that OGT could first get recruited to Region II, and then promotes cleavage enhancement. The sequences covering Regions I, II, and III as well as the HCF-1<sub>PRO</sub> repeats are predicted to be highly unstructured and could form extensive loops, such that Region II could, via a loop, approach the first HCF-1<sub>PRO</sub> repeat (see model below). Region II might thus cause enrichment of OGT at the HCF-1<sub>PRO</sub> repeat 1 and thereby promote cleavage, perhaps because individual HCF-1<sub>PRO</sub> repeats do not represent effective OGT-binding sequences (see Chapter II and Figure IV-1).

Interestingly, OGT can bind to a scrambled Region II sequence (Figure IV-2), which does not enhance HCF-1<sub>PRO</sub>-repeat cleavage (Figure III-7). It is thus clear that OGT binding to Region II is not sufficient to promote cleavage. This could be explained by the following scenario: Although the scrambled Region II sequence is bound by OGT, the structural consequences (e.g. looping) can perhaps not be achieved, and thus the enhancement of cleavage would be impaired. In support of this hypothesis, the C-terminal Region II construct (+II\_C) is not cleaved (Figure III-8), but displays prominent OGT-binding activity (data not shown).

The mechanism by which OGT recognizes Region II is unknown. OGT does not display sequence specificity for Region II recognition, but the amino acid composition, and perhaps the enrichment in serines and threonines could play a role in Region II recognition (a serine and threonine density map of the HCF-1rep1 protein is shown in Figure Appendix-1).

Moreover, the OGT TPR domain might be involved in Region II recognition as N-terminal TPR deletions of OGT decreased Region II–OGT association (Figure IV-4) and the 5N-5A OGT mutant did not bind to Region II (Figure IV-2). Region II–OGT association is context dependent and can be enhanced or impaired by sequences adjacent to Region II and to the HCF-1<sub>PRO</sub> repeat 1. Region IV C-terminal of Region II enhances (Figures IV-3 and IV-4), whereas Region III impairs Region II–OGT association (Figure IV-1), correlating with the roles of these regions in HCF-1<sub>PRO</sub>-repeat cleavage enhancement (Chapter III). I therefore suspect, as already mentioned in Chapter III, that when OGT is bound to the HCF-1 protein, the regions surrounding Region II and the HCF-1<sub>PRO</sub> repeat might display structural properties supporting or impairing Region II–OGT association.

### **Region II is rich in O-GlcNAcylated serines and threonines**

Proteomic analysis of O-GlcNAcylation sites in the HCF-1rep1 precursor protein revealed that the respective activities of Regions I, II, and III on HCF-1<sub>PRO</sub>-repeat cleavage correlate with the number of O-GlcNAcylation sites identified in these regions (Figure IV-5), consistent with immunoblot results in Chapter III. Whereas identification of O-GlcNAcylated proteins by immunoblot revealed the big differences in O-GlcNAcylation levels, LC-MS/MS analysis revealed both O-GlcNAcylation patterns (the positions of O-GlcNAcylated residues) and levels (the proportions of O-GlcNAcylated residues in a sub-stoichiometrically O-GlcNAcylated peptide mixture). Region II contained the largest number of O-GlcNAcylated residues, correlating with Region II–OGT binding efficiency and HCF-1<sub>PRO</sub>-repeat cleavage-enhancement activity. It is therefore likely that strong O-GlcNAcylation in Region II is a result of strong OGT association to this sequence. Manipulation of Region II, for example by placing it C-terminal of the HCF-1<sub>PRO</sub> repeat or by splitting it into sub-regions, alters its O-GlcNAcylation pattern dramatically (Figures IV-6 and IV-7). Thus, O-GlcNAcylation of HCF-1 sequences appears to be highly variable because O-GlcNAcylation patterns vary with the sequence context.

Subdivision of Region II into two sub-regions and the subsequent isolation of these sub-regions from their surrounding sequences suggested, using an *in vivo* assay, that both of these sub-regions display weaker cleavage and O-GlcNAcylation activities with respect to the activity of the entire Region II sequence (Figure IV-7 C). Perhaps Regions II-a and II-b act together for HCF-1<sub>PRO</sub>-repeat cleavage enhancement and isolation of these two sub-regions from one another disrupts this cumulative or synergistic effect. Intriguingly, the HCF-1rep1 +II-b protein was unstable or insoluble during bacterial synthesis and purification. Region II-a could be required for stabilization or for correct folding of Region II-b. Nevertheless, the Region II-b construct was synthesized *in vivo* as well as the Region II-a construct, suggesting that other cellular proteins could interact with and stabilize the Region

II sequence. Interestingly, OGA, the enzyme that removes O-GlcNAc moieties from proteins, is a possible binding partner of HCF-1, and might therefore affect HCF-1 O-GlcNAcylation and maybe also cleavage *in vivo*.

### **Region II O-GlcNAcylation and HCF-1<sub>PRO</sub>-repeat proteolysis are independent**

It has been proposed that O-GlcNAcylation and proteolysis influence each other during HCF-1 protein maturation (Daou et al., 2011). According to the definition of cross-talk between post-translational modifications (see 1.2.2), a change in HCF-1<sub>PRO</sub>-repeat proteolysis efficiency should cause changes in HCF-1 O-GlcNAcylation patterns or levels and vice-versa. Unexpectedly, using proteomic and mutational analyses (Figures IV-8 and IV-9), HCF-1 O-GlcNAcylation and HCF-1<sub>PRO</sub>-repeat proteolysis appear to be two independent OGT activities. Upon replacement of the cleavage site E10 by alanine, glutamine, aspartate or serine in HCF-1<sub>rep1</sub> precursor constructs, O-GlcNAcylation levels of the HCF-1 proteins decreased, consistent with results obtained with HCF-1 E10A mutants (Capotosti et al., 2011). As I have discussed in Chapter II, amino acids different from glutamate at position 10 of the HCF-1<sub>PRO</sub> repeat promote OGT association to the HCF-1<sub>PRO</sub> repeat. Alanine, glutamine, aspartate or serine accommodate their side-chains more favorably in the OGT–UDP-GlcNAc–HCF-1<sub>PRO</sub>-repeat complex and they therefore bind better to OGT. Thus, it is very likely that the E10 mutations, which enhance HCF-1<sub>PRO</sub>-repeat–OGT binding, sequester available OGT and consequently reduce OGT binding to HCF-1 sequences outside of the HCF-1<sub>PRO</sub> repeat, resulting in reduced overall HCF-1 O-GlcNAcylation. I conclude that Region II O-GlcNAcylation is not influenced by HCF-1<sub>PRO</sub>-repeat proteolysis *per se* because HCF-1<sub>rep1</sub> mutants inactivated for cleavage using different mutational strategies (T17-22A and ΔPRO mutations) did not decrease Region II O-GlcNAcylation levels (Figure IV-8). The O-GlcNAcylation compromised OGT mutant (D554H H558D) cleaved HCF-1 substrates containing Region II as efficiently as wild-type OGT, suggesting that prominent O-GlcNAcylation of Region II is not fundamental for Region II activity on cleavage. I therefore propose that the cross-talk observed between HCF-1 O-GlcNAcylation and HCF-1<sub>PRO</sub>-repeat proteolysis (Daou et al., 2011) may have simply resulted from the fact that the same enzyme is responsible for both activities.

Why is Region II and, more generally, the HCF-1 N-terminal subunit highly O-GlcNAcylated? An HCF-1 construct cleaved via Taspase1-mediated proteolysis is defective for M-phase functions during the cell cycle (Capotosti et al., 2011), but displayed prominent O-GlcNAcylation in an *in vivo* assay (Figure IV-10), suggesting that HCF-1 O-GlcNAcylation levels are independent from OGT-mediated proteolysis. Yet, this experiment cannot rule out that the lack of OGT-mediated proteolysis changed the O-GlcNAcylation patterns (and not the levels) of the HCF-1 recombinant protein, as proteomic studies have not been performed.

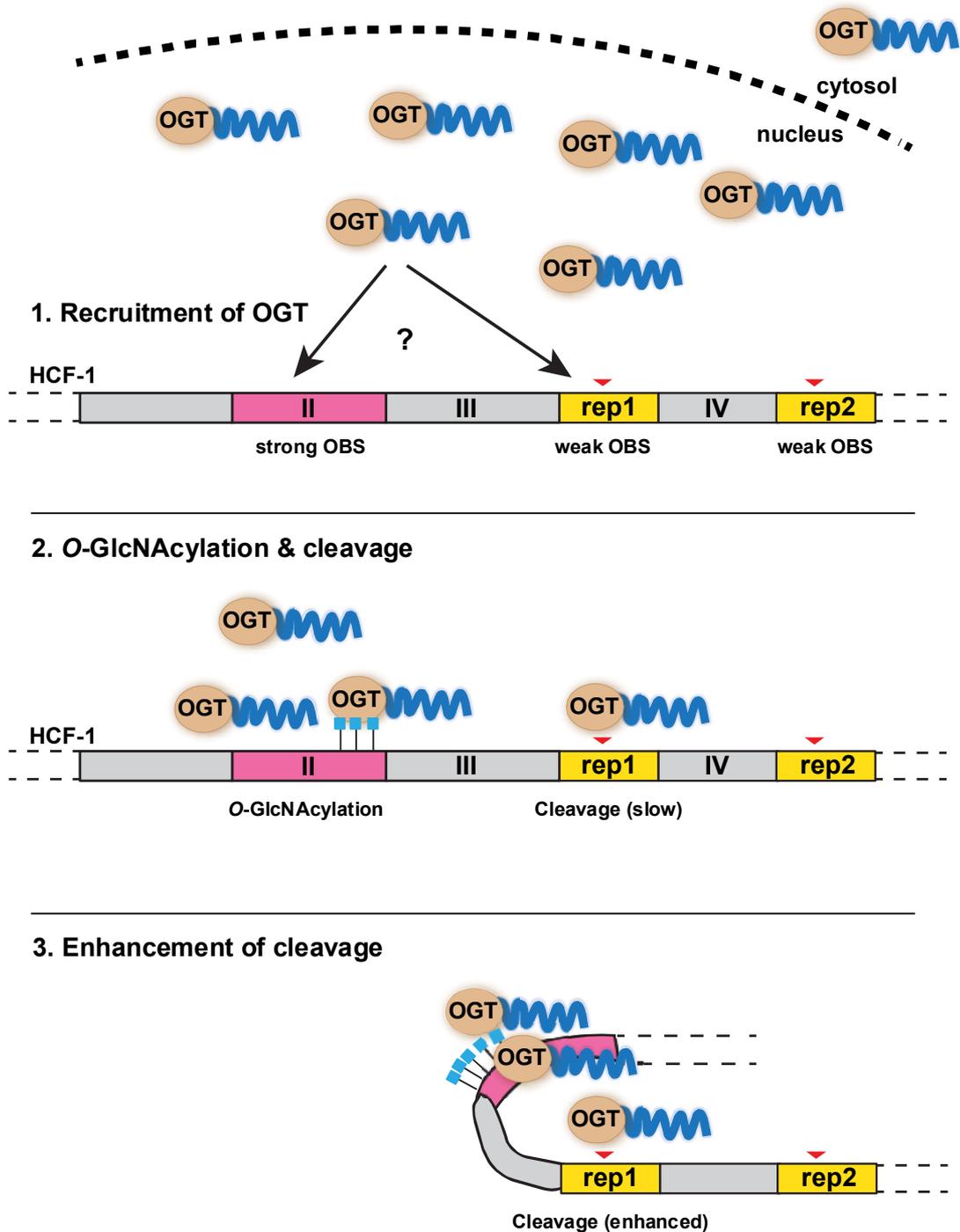
I have not identified any role of Region II O-GlcNAcylation for HCF-1 proteolysis. I rather propose that Region II O-GlcNAcylation is a result of enhanced OGT association to this sequence. Recently, two interesting studies shed light on more general roles of protein O-GlcNAcylation: The laboratory of Juerg Mueller showed that O-GlcNAcylation prevents aggregation of the polycomb protein Polyhomeotic in *Drosophila*, and interestingly, a similar role for *Drosophila* HCF has also been implicated in this study (Gambetta and Muller, 2014). Another study by the David Vocadlo laboratory proposed that O-GlcNAcylation promotes protein stability during and after protein translation of the transcription factor Sp1 (Zhu et al., 2015). It is thus possible that O-GlcNAcylation of HCF-1 promotes its stability and/or prevents its aggregation. These scenarios appear reasonable, given that HCF-1 is synthesized as a particularly large precursor protein of 2035 amino acids, which may require additional support for protein stability and integrity before stable HCF-1<sub>N</sub> and HCF-1<sub>C</sub> subunit association.

### **Proposed model for Region II activity on HCF-1<sub>PRO</sub>-repeat cleavage**

Summing up the results from Chapters III and IV, I propose three hypothetical steps for HCF-1<sub>PRO</sub>-repeat cleavage enhancement (Figure IV-11).

1. It has been shown that about 50% of OGT associates with HCF-1 in the nucleus (Daou et al., 2011). This is a large proportion of available nuclear OGT. I therefore suggest that several OGT molecules can associate with one HCF-1 molecule at distinct OGT-binding sequences (OBS), for example, the HCF-1<sub>PRO</sub> repeat and Region II. A single HCF-1<sub>PRO</sub> repeat represents an OGT-binding sequence with rather low affinity compared to the Region II–OGT-type association. Therefore, competition of these binding sequences for OGT recruitment might take place (Figure IV-11, step 1).
2. Region II represents a strong OGT-binding sequence and this sequence thus associates efficiently with OGT, which results in Region II O-GlcNAcylation. At the same time, HCF-1<sub>PRO</sub>-repeat proteolysis might be induced at slow rates (Figure IV-11, step 2).
3. Region II and its adjacent sequences might have structural properties, allowing to approach HCF-1<sub>PRO</sub> repeat 1. This results in more OGT molecules in proximity of the cleavage site. These processes may enhance HCF-1<sub>PRO</sub>-repeat proteolysis (Figure IV-11, step 3).

It is still unknown if Region II affects proteolysis of all six HCF-1<sub>PRO</sub> repeats in the HCF-1 full-length context. However, I suspect that HCF-1 contains more sequences, which may serve as binding sites for OGT..



**Figure IV-11: Proposed model for Region II cleavage enhancement of the first HCF-1<sub>PRO</sub> repeat.** OGT and the HCF-1 protein are schematically illustrated (not to scale). HCF-1 Regions II, III, and IV and the HCF-1<sub>PRO</sub> repeats 1 (rep1) and 2 (rep2) are illustrated. The cleavage sites within rep1 and rep2 are indicated by the red arrowhead; OBS, OGT-binding sequence. Prominent O-GlcNAcylation is indicated by blue squares. Step 1: HCF-1 and OGT are abundant in the nucleus. Region II (strong OBS) and rep1 (weak OBS) compete for OGT binding. Step2: Region II associates efficiently to OGT, whereas rep1 associates to OGT and gets cleaved at slow rates. Step 3: Region II (including bound OGT) bends towards rep1 and enhances rep1 cleavage.

# Chapter V :

## *Conclusions and Perspectives*

Two post-translational modifications (PTMs), reversible O-GlcNAcylation and irreversible site-specific proteolysis, catalyzed by a single enzyme — OGT — are the major focus in this thesis. OGT represents without doubt an unusual enzyme that executes the catalysis of two very different PTMs within a single active site (Lazarus et al., 2013). In this thesis, I characterized substrate requirements for O-GlcNAcylation and site-specific proteolysis of the transcriptional co-regulator HCF-1. In this chapter, I present the main conclusions elaborated here and I place them into a larger context. I also present ideas that could stimulate future studies on this topic.

### **V.1 The HCF-1<sub>PRO</sub> repeats are highly specific protease recognition sites containing three different elements important for cleavage**

Proteases must faithfully and specifically recognize their target substrates. This recognition depends largely on the complementarity between the protease active site and the sequence surrounding the scissile bond of the substrate. Whereas the five families of mammalian aspartyl, cysteine, metallo-, serine or threonine proteases display, on average, specificities between 4 and 6 amino acids surrounding the scissile bond, OGT displays a specificity of 21 residues within the HCF-1<sub>PRO</sub> repeat. The interactions between the HCF-1<sub>PRO</sub> repeat and OGT are highly specific being most prominent at the E10 cleavage site residue (P1' position) and within the threonine region (see II.2 and II.9; Lazarus et al., 2013). So far, homologous HCF-1<sub>PRO</sub>-repeat sequences in proteins other than HCF-1 proteins, have not been identified, it thus seems very unlikely that OGT could cleave substrates other than HCF-1 proteins.

#### **The HCF-1<sub>PRO</sub>-repeat threonine region**

The HCF-1<sub>PRO</sub>-repeat threonine region binds to the OGT TPR region through a large network of hydrogen bonds (Lazarus et al., 2013), and I showed using the example of a representative residue within the threonine region, T14, that this interaction is very important for binding and cleavage. Additionally, OGT–HCF-1rep1 cleavage assays using N-terminal and internal TPR deletion mutants engineered by Dr. Vaibhav Kapuria, revealed that the TPRs most C-terminal to the OGT catalytic domain are required for cleavage (Kapuria and Herr, unpublished results). The HCF-1<sub>PRO</sub>-repeat threonine region is not directly involved in the cleavage chemistry for HCF-1<sub>PRO</sub>-repeat proteolysis, as it lies far away from the OGT

catalytic domain. Nevertheless, it is fundamental for cleavage because its specific interactions with the OGT TPR domain probably allow for the correct accommodation of the HCF-1<sub>PRO</sub>-repeat cleavage region inside the OGT catalytic domain. The OGT TPRs were proposed to interact with other OGT O-GlcNAcylation substrates in a manner similar to how they bind to the threonine region (Lazarus et al., 2013). This hypothesis is questionable, given that the HCF-1<sub>PRO</sub>-repeat threonine region–TPR domain interactions display such high specificity. As the TPRs clearly support cleavage of the HCF-1<sub>PRO</sub> repeat, one might ask conversely: What is their role for O-GlcNAcylation or for substrate recognition of O-GlcNAcylation substrates? I have two hypotheses: (i) the OGT TPR domain could display no particular specificity for any substrate other than HCF-1<sub>PRO</sub> repeats. Perhaps the interactions between the HCF-1<sub>PRO</sub>-repeat cleavage region and the OGT catalytic domain require such specificity between the OGT TPRs and the threonine region. Other O-GlcNAcylation substrates probably do not contain a region similar to the HCF-1<sub>PRO</sub>-repeat cleavage region with its particular OGT-binding properties (see below). Thus, the OGT TPR domain might simply form favorable contacts with a variety of different substrates via a combination of factors, such as electrostatic interactions or hydrophobic pockets (as described in I.6.2). This suggests that OGT could bind to a multitude of O-GlcNAcylation substrates without great specificity. (ii) OGT could use HCF-1 as an adaptor protein that recruits OGT to HCF-1 binding partners that can then be O-GlcNAcyated. OGT could dissociate from cleaved HCF-1, implying that premature HCF-1 would be more effective to recruit OGT to its O-GlcNAcylation substrates. But OGT could also remain associated with Region II and therefore HCF-1 could bring OGT to its O-GlcNAcylation substrates, even after cleavage. The latter possibility could explain why OGT and HCF-1 are often found to co-occupy genomic regions (especially active promoters), as described in I.7.4. Further studies will be needed to clarify these scenarios.

### **The HCF-1<sub>PRO</sub>-repeat cleavage region**

The HCF-1<sub>PRO</sub>-repeat cleavage region binds in the OGT catalytic domain and forms a binding interface with the nucleotide sugar UDP-GlcNAc (Lazarus et al., 2013). I showed that the amino acid side-chain of glutamate at position 10 of the HCF-1<sub>PRO</sub> repeat (E10) surprisingly displays highly unfavorable interactions with the OGT-UDP-GlcNAc complex and is in fact the only residue in the cleavage region to display this unusual activity. The energy involved in binding E10 might be used to favor HCF-1<sub>PRO</sub>-repeat cleavage (see II.6 and II.7). Thus, the threonine region clamps the HCF-1<sub>PRO</sub>-repeat to OGT, whereas the cleavage region causes strains in the OGT-UDP-GlcNAc-HCF-1<sub>PRO</sub>-repeat complex. Enzymatic mechanisms involving substrate strains in catalysis have been previously described for bacterial carbon-carbon lyases (Phillips et al., 2014). These enzymes catalyze the hydrolytic cleavage of

either one of the amino acids tryptophan and tyrosine, using substrate strains to increase substrate specificity and catalytic efficiency. Interestingly, the substrate strains originating from the E10 residue in the HCF-1<sub>PRO</sub>-repeat cleavage region are dependent on the sugar moiety of UDP-GlcNAc, indicating that HCF-1 proteolysis is coupled to UDP-GlcNAc availability and maybe indirectly to the cellular nutrient status. The mechanisms by which substrate strains contribute to catalysis are not completely understood and could be a subject for future studies.

### **The HCF-1<sub>PRO</sub>-repeat Hinge region**

Although interactions between the OGT TPRs and the threonine region are quite rigid and favor binding of the cleavage region to the OGT catalytic domain, there is structural flexibility of residues between these two critical regions for proteolysis. These residues — the Hinge region — consist of three amino acid residues displaying flexibility in the OGT-UDP-GlcNAc-HCF-1<sub>PRO</sub>-repeat complex (see II.8). One of the major questions about OGT–HCF-1<sub>PRO</sub>-repeat binding is how OGT can fit a long polypeptide chain of 2035 amino acids inside its catalytic domain and inside the tunnel formed by the TPRs. One possibility is that OGT makes hinge-like motions between the TPRs and the catalytic domain (this region is called the OGT hinge; Lazarus et al., 2011). This movement could allow for the HCF-1 polypeptide to be accommodated inside OGT, potentially aided by unfolding and refolding of the TPR superhelical structure. Flexibility in the HCF-1<sub>PRO</sub>-repeat Hinge region, which is surrounded by the OGT hinge in the complex, could therefore be important for this process. Interestingly, the three residues in the Hinge region are highly conserved, arguing for an additional role beyond simply contributing to flexibility and this would still need to be explored.

## **V.2 A single invariant amino acid side-chain is essential for proteolysis**

The E10 side-chain displays particular properties within the HCF-1<sub>PRO</sub> repeat. Based on OGT–HCF-1<sub>PRO</sub>-repeat binding data and molecular dynamics analyses, I propose that E10 initiates the cleavage reaction by causing strains in the OGT-UDP-GlcNAc-HCF-1<sub>PRO</sub>-repeat complex. The products of cleavage of a single HCF-1<sub>PRO</sub> repeat are two HCF-1 polypeptide chains, one of which bears an N-terminal pyroglutamate formed by cyclization of E10. Pyroglutamate is a post-translational modification that can be found in humans (e.g., immunoglobulins are modified by pyroglutamate; Liu et al., 2011). The physiological roles of pyroglutamate are not completely understood. The modification has been proposed to contribute to protein stabilization *in vivo* but it can also be removed by pyroglutamate aminopeptidase (Kumar and Bachhawat, 2012). It is not known if the presence or absence of pyroglutamate at the HCF-1 cleaved polypeptide plays a role for HCF-1's cellular functions. The development of antibodies directed to N-terminal pyroglutamate of the HCF-1<sub>PRO</sub> repeat

would be a step forward toward understanding if and in which cellular compartment pyroglutamate modified HCF-1 occurs *in vivo*.

In Lazarus et al. (2013), we proposed a model for the initial step of the cleavage reaction, in which a deprotonated E10 side-chain leads a nucleophilic attack on the anomeric carbon atom of UDP-GlcNAc, which would result in transient O-GlcNAcylation of E10 (Figure II-14). This model was based on the observation that UDP-GlcNAc is required for cleavage and that a replacement of E10 by serine activates HCF-1<sub>PRO</sub>-repeat O-GlcNAcylation but blocks cleavage (Lazarus et al., 2013). Furthermore, I showed that the sugar moiety of UDP-GlcNAc is crucial for proper OGT–HCF-1<sub>PRO</sub>-repeat association (see II.5 and II.6). Nevertheless, the reactions following this initial step still remain to be elucidated. The cofactor analog UDP-5SGlcNAc could be an excellent tool to study the cleavage chemistry. I showed that this cofactor alters the OGT–HCF-1<sub>PRO</sub>-repeat binding mode dramatically (Figure II-7). UDP-5SGlcNAc might react with the HCF-1<sub>PRO</sub> repeat and an analysis of this reaction involving organic chemistry methods could lead to a better understanding of the cleavage reaction. Furthermore, to understand if UDP-GlcNAc indeed reacts during the cleavage reaction (if it is a cosubstrate), I propose to determine if there is free UDP after the *in vitro* HCF-1–OGT cleavage reaction. For this purpose, an HCF-1 substrate devoid of O-GlcNAcylation sites should be used, for example the HCF3R substrate. UDP in UDP-GlcNAc could be labeled with a fluorescent marker and an assay would need to be developed that can detect free fluorescent UDP.

A recent study proposed involvement of the  $\alpha$ -phosphate of UDP-GlcNAc in the O-GlcNAcylation mechanism (Schimpl et al., 2012). In fact, substitution of the  $\alpha$ -phosphate of UDP-GlcNAc by sulfur blocks O-GlcNAcylation of TAB1, a normally O-GlcNAcylated innate immunity signaling protein. It is not known if the proteolysis and the O-GlcNAcylation mechanisms are similar, thus studies addressing HCF-1<sub>PRO</sub>-repeat cleavage with UDP-GlcNAc  $\alpha$ -phosphate analogs would be highly valuable to extend studies of the HCF-1<sub>PRO</sub>-repeat cleavage mechanism. In summary, although 21 amino acids within the HCF-1<sub>PRO</sub> repeat are required for full cleavage activity, only a single amino acid side-chain within the repeat — E10 — is essential for HCF-1<sub>PRO</sub>-repeat proteolysis and thus for HCF-1<sub>C</sub> functions during the cell-division cycle.

### **V.3 How did OGT-induced cleavage of the HCF-1<sub>PRO</sub> repeats arise?**

OGT possesses dual enzymatic activity that, to my knowledge, is unique within the families of glycosyltransferases and proteases. The HCF-1<sub>PRO</sub> repeat adopts the conformation of an O-GlcNAcylation substrate in the OGT active site and replacement of E10 by serine transforms the cleavage substrate into an O-GlcNAcylation substrate. Interestingly, OGT has a preference for O-GlcNAcylation of substrates containing a proline residue close to the site

of O-GlcNAcylation, which enforces an extended structure of the substrate. The HCF-1<sub>PRO</sub> repeat also contains a highly conserved proline at the P2 position. Could it be that HCF-1 mimics an O-GlcNAcylation substrate to use OGT as its protease?

OGT-mediated HCF-1 cleavage can only be found in vertebrates. In insects, HCF proteins are cleaved by Taspase1 — the same protease that cleaves the MLL protein both in humans and flies. Why is the HCF-1 cleavage mechanism in vertebrates different from Taspase1-mediated HCF cleavage in insects? To answer this question, it might be useful to know the origin of the HCF-1<sub>PRO</sub> repeats. The hypothesis has been put forward that the HCF-1<sub>PRO</sub> repeats were inserted into the HCF-1 gene of an early vertebrate millions of years ago by a viral transposition event. Indeed, sequences homologous to the repeats are nowhere to be found other than being very highly conserved in HCF-1 proteins. Human HCF-1 contains eight repeats, two of which are inactive because they have diverged in sequence. These eight repeated sequences — and a large part of the Region III sequence — are encoded by a single extraordinarily large exon of 1476 base pairs, which supports the hypothesis of a viral transposition event. Sequence comparison of the HCF-1<sub>PRO</sub> repeats in vertebrates suggests that the HCF-1<sub>PRO</sub> repeats arose through serial duplication of progressively larger sets of repeats (Wilson et al., 1995a). Selective pressure might have removed the intervening sequences generating variable numbers of repeats in different vertebrate species.

Viruses utilize their host's replication, transcription and translation machineries to coordinate their infection cycle with the life cycle of the host. Moreover, viruses are known to often process generated polyproteins by proteases encoded by the viral genome. Perhaps, however, a virus, which infected a common ancestor of today's vertebrates, used a host protease to cleave its polyproteins. OGT could have been coopted as a protease for cleavage of a viral polyprotein containing the sequence of the HCF-1<sub>PRO</sub> repeat. The virus would thus have succeeded to couple the metabolic status of the host cell with the production and activation of its own proteins. Subsequent transposition of viral genomic sequences into the host HCF-1 gene locus might have led to an HCF-1 protein containing both Taspase1 and HCF-1<sub>PRO</sub>-repeat cleavage sites. HCF-1<sub>PRO</sub>-repeat cleavage by OGT has been apparently selected during evolution. This is not unreasonable given that this mechanism might allow the linkage of the cellular metabolic state to the control of cell-cycle progression. Intriguingly, human and other vertebrate HCF-1s indeed contain vestigial Taspase1 cleavage sites (see I.7.5) indicating that vertebrates did not just “lose” their Taspase1 cleavage sites, but that they were simply not retained intact when HCF-1<sub>PRO</sub>-repeat cleavage arose as the main proteolytic mechanism.

Bioinformatic evolutionary studies could help to shed light on the origins of the HCF-1<sub>PRO</sub> repeats by comparison of the genes encoding HCF-1 in different vertebrate species.

Analyses like these would lead to a better understanding of the evolutionary background of OGT-mediated proteolysis.

#### **V.4 Efficient HCF-1<sub>PRO</sub>-repeat cleavage requires HCF-1 sequences flanking the HCF-1<sub>PRO</sub> repeat**

HCF-1<sub>PRO</sub>-repeat cleavage is a slow and inefficient reaction compared to other proteolytic reactions (see Chapter III). A single HCF-1<sub>PRO</sub> repeat in isolation from its surrounding sequences represents a cleavage substrate with low activity (POUrep2 substrate, see III.1). As the HCF-1<sub>PRO</sub> repeat consists of 26 amino acids, additional requirements for efficient proteolysis were unexpected. My studies showed that HCF-1 sequences flanking the first HCF-1<sub>PRO</sub> repeat can enhance its cleavage efficiency. One of these flanking sequences, Region II, displays particularly strong, sequence-specific cleavage-enhancement activity correlating with strong O-GlcNAcylation and OGT-binding activities. In fact, I identified six clustered O-GlcNAcylation sites in the C-terminal half of Region II, but I could not find evidence for a role of Region II O-GlcNAcylation in HCF-1<sub>PRO</sub>-repeat proteolysis. Region II–OGT association can vary depending on the context of the surrounding HCF-1 sequences. It thus appears that the surrounding regions have properties that can interfere with Region II–OGT binding. Currently, there is no evidence for structured protein domains in the HCF-1 Basic region, in Regions I, II, and III or in the HCF-1<sub>PRO</sub> repeats. In fact, Regions I, II, and III are predicted to be unstructured. These properties have led me to propose a model in which Region II recruits OGT close to the cleavage site at the first HCF-1<sub>PRO</sub> repeat, followed by loop formation of Region II with the support of the other flanking regions (see discussion of Chapter IV). This might be a possible scenario to explain how HCF-1 sequences outside of the HCF-1<sub>PRO</sub> repeat can influence cleavage. Region II might cause enrichment of OGT molecules at the first HCF-1<sub>PRO</sub> repeat, perhaps because the HCF-1<sub>PRO</sub> repeat itself does not represent a strong OGT-binding sequence. For future studies, it would be useful to compare OGT-binding affinities to Region II and to the HCF-1<sub>PRO</sub> repeat to determine whether Region II is indeed a stronger OGT-binding site than the HCF-1<sub>PRO</sub> repeat. Affinity studies using Surface Plasmon Resonance techniques, for instance, could complement my *in vitro* HCF-1–OGT binding studies.

Interestingly, within the *HCFC1* gene, Regions I and II lie on exons and the HCF-1<sub>PRO</sub> repeats lie on a separate exon. This suggests that the sequences encoding Regions I and II were probably not inserted together with the HCF-1<sub>PRO</sub> repeats by the proposed viral transposition event described above. The high degree of conservation of these regions in vertebrates suggests, however, that these sequences have a conserved function (see III.1.1). In fact, co-evolution of OGT-mediated proteolysis and the cleavage-enhancing activities of

the flanking sequences might have led to conservation of both the flanking sequences and the HCF-1<sub>PRO</sub> repeats.

Because Region II is predicted to be highly unstructured and homologous structures of Region II could not be found in the PDB (Dr. Ute Roehrig, personal communication), it is unfeasible to predict how Region II could interact with OGT using molecular dynamics approaches. It would therefore be interesting to perform crystallization studies of Region II in complex with OGT to understand how Region II binds to OGT and if this binding mode resembles the one of the HCF-1<sub>PRO</sub> repeat. My studies using TPR deletion mutants indicate that the TPRs are involved in Region II recognition (see IV.3). A Region II–OGT crystal structure would not only shed light on the mechanisms of HCF-1<sub>PRO</sub>-repeat cleavage enhancement but it would also be a major advancement in the OGT field, which lacks an understanding of how OGT recognizes its O-GlcNAcylation substrates.

## V.5 The two faces of OGT

Site-specific proteolysis of HCF-1 is irreversible and promotes cell-cycle progression (Julien and Herr, 2003). O-GlcNAcylation, in contrast, is known to be a highly dynamic PTM that can respond to changes in the environment. It is not known if O-GlcNAcylation of HCF-1 is reversible. I identified extensive O-GlcNAcylation of the HCF-1<sub>N</sub> subunit, but I also detected O-GlcNAcylation of HCF-1<sub>C</sub>, consistent with available mass spectrometry data (Myers et al., 2013) (see IV.9). O-GlcNAcylation of HCF-1 does not appear to be regulated in a cell-cycle dependent manner (Capotosti, Michaud, Herr, unpublished results) and it is unknown if OGA associates with HCF-1. Studies aimed at identifying HCF-1-OGA complexes by immunoprecipitation and mass spectrometry, for example, could contribute to uncover if HCF-1 O-GlcNAcylation is reversible.

The role of O-GlcNAcylation for HCF-1 function remains enigmatic. Using a genetic approach to dissect O-GlcNAcylation from proteolysis, I did not find evidence for a role of HCF-1 O-GlcNAcylation in proteolysis (see IV.8). Furthermore, proteolysis at the HCF-1<sub>PRO</sub> repeats does not influence O-GlcNAcylation of sequences N-terminal (Regions I, II, and III) of the first HCF-1<sub>PRO</sub> repeat (see IV.7). Therefore, there is no evidence for cross-talk between O-GlcNAcylation and site-specific proteolysis of the HCF-1 protein. Nevertheless, I showed that an E10 mutation in the HCF-1<sub>PRO</sub> repeat decreases HCF-1 O-GlcNAcylation levels, consistent with data showing that E10 mutations cause an enhancement of OGT–HCF-1<sub>PRO</sub>-repeat association (see II.7). Because a mutated HCF-1<sub>PRO</sub> repeat traps OGT to this binding site, it is possible that there is competition for OGT binding between different OGT-binding sites including Region II and the HCF-1<sub>PRO</sub> repeat (see IV.7).

The proposed mechanism for cleavage in Chapter II (Figure II-14) included transient O-GlcNAcylation of the E10 residue. This seems contradictory with my proposal that O-

GlcNAcylation is not fundamental for HCF-1<sub>PRO</sub>-repeat proteolysis. I suggest that the potential O-GlcNAcylation of the E10 residue could be a non-conventional type of O-GlcNAcylation in a specific context — the HCF-1<sub>PRO</sub> repeat — that cannot be compared to classical O-GlcNAcylation of other substrates or of sequences lying outside of the HCF-1<sub>PRO</sub> repeat. Hence, even though transient O-GlcNAcylation of E10 might be required for proteolysis, O-GlcNAcylation outside of the HCF-1<sub>PRO</sub> repeat might not be fundamental for proteolysis.

Whereas site-specific proteolysis of HCF-1 is highly specific, I showed that O-GlcNAcylation of HCF-1 is highly variable and dependent on the sequence context (see IV.5 and IV.6). At least in HCF-1, there are no strong sequence motifs that promote HCF-1 O-GlcNAcylation. Interestingly, HCF-1 O-GlcNAcylation occurs mainly in regions that are predicted to be highly unstructured such as the Basic region and Region II. This is in concordance with other studies predicting that O-GlcNAcylated residues cluster in highly unstructured regions (Hunter, 2007). It was recently shown that O-GlcNAcylation can occur co-translationally and that the modification can stabilize highly unstructured protein domains during and/or after translation (Zhu et al., 2015). Moreover, O-GlcNAcylation was implicated in the prevention of protein aggregation (Gambetta and Muller, 2014). Intriguingly, HCF-1 is synthesized as a long precursor protein that might require stabilization of its unstructured regions. It would be interesting to conduct studies, addressing the role of O-GlcNAcylation for HCF-1 protein stability and integrity.

Importantly, such studies cannot be carried out using OGT inhibitors such as 5S-GlcNAc or RNA interference (RNAi) against OGT mRNA. These approaches would lead to a defect in HCF-1 proteolysis, and any effects on HCF-1 O-GlcNAcylation and functions would have also to be attributed to HCF-1<sub>C</sub> defects during cell-cycle progression and to other pleiotropic effects resulting from OGT inactivation. It is thus advisable to dissect O-GlcNAcylation from proteolysis carefully, for example by a genetic approach. OGT mutants that are defective for one PTM, but fully active for the other PTM are very useful for this purpose. The D554H H558D OGT mutant (Dr. Vaibhav Kapuria), for instance, is compromised for its O-GlcNAcylation activities but highly active for proteolysis (see IV.8). In the case of HCF-1, I would not recommend a mutational strategy to replace single O-GlcNAcylated serines or threonines by alanines for the following reasons: (i) there are too many O-GlcNAcylated serines and threonines in HCF-1<sub>N</sub>. Alanine mutations could also affect the functions of these HCF-1 sequences. (ii) My studies showed that HCF-1 O-GlcNAcylation is highly variable and dependent on the HCF-1 sequence context. Thus, an alanine substitution could prevent O-GlcNAcylation at one site but induce O-GlcNAcylation at another site that would normally not get O-GlcNAcylated.

In summary, OGT has two faces: One that O-GlcNAcylates HCF-1 heavily in regions predicted to be highly unstructured. This modification does not appear to be highly specific. OGT has another face that cleaves HCF-1, with high specificity, within the HCF-1<sub>PRO</sub> repeats. The cleavage efficiency at least at the first HCF-1<sub>PRO</sub> repeat can be enhanced by neighboring HCF-1 sequences that display strong OGT-association and O-GlcNAcylation activities.

## V.6 Final remarks

In the introduction to this thesis, I mentioned four methods that can contribute to the investigation of protein post-translational modifications: Mutational analyses, proteomics, structural biology, and molecular dynamics. In the present work, these approaches were combined to study the mechanisms of OGT-mediated HCF-1 maturation. Despite continuous improvement of the technologies for these approaches, the study of protein post-translational modifications (and their reciprocal influence) still represents a challenging endeavor.

Ancient types of post-translational modifications, such as acetylation and phosphorylation, were involved in energy sensing but have since then been coopted to other functions. The glycosyltransferase OGT is an example of an evolutionary old protein that has been coopted to perform a different function. Likely, OGT has been “hijacked” by HCF-1 as its protease, and therefore the mechanism by which OGT mediates HCF-1<sub>PRO</sub>-repeat cleavage is unusual. It seems that it is the HCF-1<sub>PRO</sub>-repeat substrate that dictates the type of the post-translational modification: OGT can either perform its normal function (O-GlcNAcylation) or it can engage in HCF-1<sub>PRO</sub>-repeat proteolysis. Within the HCF-1<sub>PRO</sub> repeat, despite multiple elements being important for cleavage, there is only a single amino acid side-chain that is decisive for the outcome of the OGT–HCF-1 interaction. A curiosity of nature! Additional mechanisms could have co-evolved to ensure efficient HCF-1<sub>PRO</sub>-repeat cleavage, for example through the HCF-1 sequences flanking the HCF-1<sub>PRO</sub> repeats.

In the light of how deeply fundamental cellular processes (e.g., circadian rhythm, metabolism and the cell division cycle) are interwoven with each other, it appears less unexpected that an important regulator of metabolism (OGT) and of cell-cycle progression (HCF-1) are connected with each other via this unusual maturation process.

# Materials and Methods

## Plasmid constructs and DNA template preparation

The bacterial expression plasmid pGEX-HCF-1rep1 (encoding HCF-1 amino acids 867-1071) containing the wild-type or E10A HCF-1<sub>PRO</sub> repeat 1 was described in Capotosti et al. (2011). Deletions of Region I (867-891), Region II (892-949), Region III (850-1009), and Region IV (1037-1071) were obtained by QuickChange site-directed mutagenesis (Agilent Technologies). The scrambled sequences of Regions II and III were generated by random permutation of the amino acids in Region II or Region III and the resulting oligonucleotide sequence with additional BamHI restriction sites at the 5' and 3' ends custom synthesized. The fragments were PCR amplified and inserted into the pGEX-HCF-1rep1 vector using the BamHI restriction sites. For the heterologous HdaA construct, a sequence encoding for amino acids 151-210 of the HdaA protein of *Caulobacter crescentus* was custom synthesized and inserted into the pGEX-HCF-1rep1 vector using the strategy specified above. The mutations E10D, E10Q, E10S, T17-22A (Thr 17, Thr 19, Thr 21, and Thr 22; see Capotosti et al., 2011), and  $\Delta$ PRO (deletion of HCF-1 residues 1010-1035) in the HCF-1<sub>PRO</sub> repeat were engineered by QuickChange site-directed mutagenesis. The HCF-1rep1 trypsin cleavage sites A933K and M951K for peptide generation in LC-MS/MS analysis were engineered by site-directed mutagenesis. The insertion of Region II C-terminal of the HCF-1<sub>PRO</sub> repeat (construct HCF-1rep1 +II\_C) was obtained by overlap extension PCR (Lee et al., 2004). The pCGNHCF-1rep123 construct was described in Capotosti et al. (2007 and 2011). Deletions of Region II and Regions I, II and III were obtained by QuickChange site-directed mutagenesis.

For *in vitro* transcription/translation, DNA templates encoding GST–HCF-1-fusion proteins were prepared by PCR amplification from the different pGEX-HCF-1rep1 plasmid constructs. PCR reactions were performed using a forward primer including the phage T7 promoter and human  $\beta$ -globin translation initiation codon and this template was directly used for *in vitro* transcription/translation using the wheat-germ extract based TNT T7 transcription/translation system as recommended (Promega).

For transient expression of the HCF-1rep1 deletion constructs in HEK 293 cells, a PCR fragment encoding the GST–HCF-1rep1 sequence was amplified and inserted into a pCGN vector, using XbaI and KpnI restriction sites. The sequence encoding the POUrep2 protein (Oct-1 sequence 280-439, in which the HCF-1<sub>PRO</sub> repeat 2 was embedded) was amplified from the pET11c vector and inserted as a GST-fusion protein into the pCGN vector,

using KpnI and BamHI restriction sites. Constructs containing Regions II or III upstream of the HCF-1<sub>PRO</sub> repeat were obtained by overlap extension PCR [42].

The bacterial expression plasmid pET24 containing N-terminal T7 and 8-His tags was described in Lazarus et al. (2013). Site-directed mutagenesis was used as described above to generate the catalytic domain D554H\_H558D OGT mutant, as well as the TPR domain mutants 5N-5A,  $\Delta$  1-6,  $\Delta$  1-8, and  $\Delta$  1-9.

### **Cell culture and plasmid transfections**

HEK 293 cells were grown on plates at 37°C in DMEM with 10% FBS. For plasmid transfection, 10<sup>5</sup> cells were seeded onto a 10 cm plate in 12 ml of DMEM with 10% FBS, and transfected one and a half days after seeding with 4.8  $\mu$ g of plasmid and 60  $\mu$ l of Lipofectamin in Opti-MEM medium as described (Invitrogen).

### **Bacterial protein expression**

#### *GST–HCF-1rep1 constructs*

The recombinant HCF-1rep1 protein encoding HCF-1 amino acids 867-1071 and its mutants were verified by sequencing. Proteins were synthesized in *E. coli* BL21 (DE3) as a fusion with N-terminal GST and C-terminal 6-Hist tag and purified using Nickel affinity chromatography according to the QIAexpressionist™ protocol (Qiagen) for native protein purification. Briefly, a starter culture of BL21 (DE3) was grown for 6 h at 37°C. 100 ml Luria Bertani (LB) medium containing 100  $\mu$ g/ml carbenicillin was inoculated with 1 ml of starter culture and grown overnight at 22°C. Protein synthesis was induced the next morning with 0.4 mM IPTG at 16°C for 6 h to reduce the synthesis of C-terminal truncation products of the GST–fusion protein. The bacterial pellet was resuspended in lysis buffer supplemented with Complete protease inhibitor cocktail (Roche) and 0.1 mg/ml lysozyme and incubated for 20 min on ice. The lysate was sonicated 12 times for 10 seconds, cleared at maximum speed for 30 min at 4°C, and incubated for 90 min at 4°C with Ni-NTA agarose superflow resin (Qiagen). The resin was washed three times with wash buffer and eluted in 4 ml elution buffer. The proteins were concentrated in Amicon concentration tubes (Milipore) and dialyzed against PBS supplemented with 1 mM DTT overnight at 4°C. The concentrated dialyzed protein was frozen in liquid nitrogen and stored at -70 °C.

#### *Human OGT*

Recombinant human OGT was synthesized as a His-tag fusion protein in *E. coli* BL21 (DE3) cells as described in Lazarus et al. (2013) with minor modifications: Bacterial cultures were grown in 200 ml volumes at 27°C after diluting an overnight culture 1 to 200 in fresh kanamycin supplemented LB media. Cells were grown to an OD<sub>600</sub> of 1.1 and the

temperature was lowered to 16°C for 30 min. Protein expression was induced with 0.4 mM IPTG at 16°C for 16 h. The bacterial pellet was resuspended in 20 ml lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1 % Triton X-100, 10 % glycerol) supplemented with Complete protease inhibitor cocktail (Roche) and 0.1 mg/ml lysozyme final concentration and incubated for 30 min on ice. The lysate was then sonicated 12 times for 10 sec to remove viscosity and ensure further lysis. The lysate was cleared by high-speed centrifugation for 45 min at 4°C and incubated for 90 min at 4°C with Ni-NTA agarose superflow resin (Qiagen). The flow-through was removed and the resin washed with 3 column volumes PBS supplemented with 40 mM imidazole. The protein was then eluted in PBS supplemented with 250 mM imidazole and the elute concentrated in Amicon concentration tubes (Milipore). The elute was dialyzed against PBS supplemented with 1 mM DTT overnight at 4°C. The concentrated dialyzed protein was supplemented with 1 mM DTT and stored at -70°C.

### ***In vitro* HCF-1–OGT binding assay**

GST–HCF-1rep1–OGT pull-down assays were performed as in Lazarus et al. (2013) with few modifications: Prior to the OGT pull-down, 20 µl of a slurry of α-T7 antibody-conjugated beads (polyclonal from goat, Abcam) were incubated with PBS containing 5% (w/v) bovine serum albumin (BSA) for 1 h at 4°C to decrease non-specific binding of GST–HCF-1rep1 to the agarose beads. Subsequently, the beads were washed extensively in PBS. For the OGT pull-down, 2.5 µg GST–HCF-1rep1 and 5 µg of OGT were pre-incubated in 0.5% NP40, 10 mM Tris-Cl (pH 8.0), 150 mM NaCl, 5 mM MgCl<sub>2</sub> (NP40 buffer; (Misaghi et al., 2009)) supplemented with 5 mM DTT, in the presence or absence of 1 mM UDP-GlcNAc, in a rotating incubator for 1 h at 20°C. After incubation, 9 % of the reaction was removed as an input control for the OGT pull-down. The washed α-T7 agarose beads were added to the reaction, the volume increased to 500 µl using NP40 buffer and the suspension incubated for 1 h at 20°C. The beads were washed in NP40 buffer at least three times for 5 min at room temperature and the OGT–GST–HCF-1rep1 complexes eluted by boiling for 5 min in 20 µl of 2-fold SDS sample buffer. HCF-1rep1–OGT binding was subsequently analyzed by SDS-PAGE followed by immunoblot.

### **HCF-1 cleavage and O-GlcNAcylation assays**

#### *In vitro* cleavage assay with wheat-germ extract synthesized substrates

DNA templates encoding GST–HCF-1rep1 or deletion constructs were prepared as specified above according to the manufacturer's protocol (Promega). For the *in vitro* cleavage assay, 1 µl of wheat-germ extract was incubated in 100 mM HEPES (pH 7.9), 5 mM MgCl<sub>2</sub>, 20 mM KCl, 5 mM DTT, and 10 % sucrose (cleavage buffer), 5 mM UDP-GlcNAc, and insect cell synthesized human OGT (Capotosti et al., 2011) at a final volume of 15 µl for 16 h at 37°C.

Resulting reaction products were resolved by SDS-PAGE and detected using autoradiography.

*In vitro cleavage assay with bacterially synthesized substrates and OGT*

*In vitro* HCF-1<sub>PRO</sub>-repeat cleavage and HCF-1 O-GlcNAcylation assays with bacterially synthesized substrates and OGT were performed in 100 mM HEPES (pH 7.9), 5 mM MgCl<sub>2</sub>, 20 mM KCl, 5 mM DTT, and 10 % sucrose at a final volume of 15  $\mu$ l at 37 °C for the indicated time. For a typical assay, bacterially purified GST–HCF-1rep1 (3  $\mu$ M; concentration measured by comparative analysis with a protein loading control on Coomassie gel) was incubated with bacterially purified OGT (0.6  $\mu$ M, concentration measured as described above) in the presence or absence of 1 mM UDP-GlcNAc and the reaction stopped by transfer of the tube to -20°C and boiling in SDS-sample buffer. HCF-1 cleavage and O-GlcNAcylation were examined by immunoblot.

*In vivo cleavage assay*

*In vivo* HCF-1 cleavage assays were performed as follows: HEK 293 cells were transfected in 10 cm dishes with pCGN vectors encoding the HCF-1rep1 or POUrep2 precursor proteins as described above. 48 hours after transfection, cells were washed in PBS and lysed in 0.5% NP40, 10 mM Tris-Cl (pH 8.0), 150 mM NaCl, 5 mM MgCl<sub>2</sub> (NP40 buffer; Misaghi et al., 2009) supplemented with Roche Complete protease inhibitors and 10  $\mu$ g/ml Pefabloc SC (AEBSF from Roche Life Science) final concentration for 20 min on ice. The lysate was cleared in a table centrifuge at maximum speed at 4°C for 20 min and incubated with 30  $\mu$ l mouse  $\alpha$ -HA-conjugated agarose beads (Sigma) overnight at 4°C. Beads were washed four times in 0.5% NP40 buffer, and immunoprecipitated material was eluted in 5 bead volumes of HA-peptide (200  $\mu$ g/ml), in order to avoid elution of the IgG chains from the antibody conjugate, which would interfere with cleavage band analysis by immunoblot. Samples were analyzed by immunoblot.

**Denaturing immunoprecipitation**

$\alpha$ -HA tag immunoprecipitation was performed as in Capotosti et al. (2011) with modifications: cells were lysed 24-48 h post-transfection in NP40 buffer supplemented with Roche Complete protease inhibitors and 10  $\mu$ g/ml Pefabloc SC (AEBSF from Roche Life Science). The lysates were adjusted to 1% SDS and incubated at 65°C for 10 min. The SDS concentration was adjusted to 0.1% by dilution with NP40 buffer, and lysates were incubated with  $\alpha$ -HA agarose beads (Sigma) for 1 h to overnight at 4°C. Beads were washed extensively in NP40 buffer, and immunoprecipitated material was eluted by boiling in one bead volume of 2-fold SDS sample buffer. Samples were analyzed by immunoblot.

### **Immunoblot and antibodies**

For immunoblot analysis, nitrocellulose membranes were incubated for 60 min with 5 ml of LI-COR blocking buffer, followed by incubation with the relevant antibodies in 50% LI-COR blocking buffer and 50% PBST (PBS containing 0.1% Tween 20), as described below, at 4°C over night. The membranes were washed four times and incubated with the appropriate secondary antibodies (dilution 1:10,000) in 50% LI-COR blocking buffer and 50% PBST at room temperature for 45 min. The membranes were extensively washed in PBS containing 0.5% Tween20 and scanned with an Odyssey infrared imager (LI-COR). Quantification of bands on immunoblots was performed using the LI-COR Image Quant quantification software.

Antibodies used to detect GST–HCF-1rep1, GST–POUrep2, O-GlcNAc and OGT were as follows: rabbit polyclonal  $\alpha$ -GST (1-109; Santa Cruz) used at 1:1000, mouse monoclonal  $\alpha$ -HA-epitope (12CA5; Field et al., 1988) used at 1:1000, mouse monoclonal (RL2)  $\alpha$ -O-GlcNAc (Abcam) used at 1:3000, mouse monoclonal  $\alpha$ -T7 (Novagen) used at 1:5000, and rabbit polyclonal  $\alpha$ -OGT (H-300; Santa Cruz) used at 1:1000.

Quantification of bands on immunoblots was performed using the LI-COR Image Quant quantification software and visualized using GraphPad Prism version 6.0e.

### **Sample preparation for O-GlcNAcylation analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS)**

Sample preparation for O-GlcNAcylation analysis by LC-MS/MS was performed as in Capotosti et al. (2011) with modifications. HA-tagged GST–HCF-1rep1 plasmid expression vectors with wild-type or with E10A, E10D, E10Q and T17-22A mutated HCF-1<sub>PRO</sub> repeats were transfected into eight 15 cm dishes of HEK 293 cells for 48 hours. Cells were lysed in 8 ml NP40 buffer supplemented with 1mM PMSF, 10  $\mu$ M PUGNAc and Roche Complete protease inhibitors. The lysates were adjusted to 1% SDS and incubated at 65°C for 10 min. The SDS concentration was adjusted to 0.1% by dilution with NP40 buffer, and lysates were incubated with  $\alpha$ -HA agarose beads overnight at 4°C. Immunoprecipitated material was eluted with HA-peptide as described above and the resulting elute concentrated using Amicon concentration tubes (Milipore). For proteomic analysis, see below.

### **O-GlcNAcylation analysis by LC-MS/MS**

For the O-GlcNAcylation analysis, purified proteins were resolved by SDS-PAGE and stained with Coomassie Blue. Gel bands were excised from SDS-PAGE gel and in-gel digested with sequencing-grade trypsin (Promega, Madison, WI, USA) as described (Shevchenko et al., 1996; Wilm et al., 1996). Extracted peptides were then cleaved with Glu-C endoproteinase

(Sigma-Aldrich, St. Louis, MO, USA). Data-dependent LC-MS/MS analyses were carried out on a hybrid linear trap LTQ-Orbitrap XL (Thermo Scientific) mass spectrometer interfaced to a nanocapillary HPLC equipped with a C18 reversed-phase column (Thermo Scientific), using CID (collision induced dissociation) mode for MS/MS fragmentation. MS/MS spectra were analyzed using Mascot 2.4 software (Matrix Science, London, UK). Mascot was set up to search a custom-built database containing the sequences of the HCF constructions and of contaminants (enzymes, keratins, etc.). Semi-specific cleavage at K, R (not before P), and at E, D were used as the enzyme definition. Mascot was searched with a fragment ion mass tolerance of 0.50 Da, a parent ion tolerance of 10 ppm, and allowing two missed cleavages. Iodoacetamide and propionamide derivatives of cysteine, deamidation of asparagine and glutamine, phosphorylation of serine and threonine, oxidation of methionine, and addition of acetylhexosamine (HexNAc) to serine and threonine were specified in Mascot as variable modifications. O-HexNAc modified and phosphorylated residues were considered confident if the ion score for the identified peptide was superior to 23 with a site localization score (Mascot Delta Score) higher than 70% and potential if the ion score was between 14 and 22 or localization score between 50 and 70 %.

### **Molecular dynamics simulations**

Molecular Dynamics (MD) simulations were performed using stochastic boundary conditions (SBC; Brooks et al., 1988), after solvating the systems with a 24 Å radius sphere of TIP3P water molecules, centered on the mutated residue. A 4 Å wide restrained buffer region, coupled to a heat bath using the Langevin equation of motion and a 250 ps<sup>-1</sup> friction constant (Brooks and Karplus, 1989), was used. Protein atoms outside the buffer region were held fixed. The water molecules were kept within the sphere by the use of a solvent boundary potential (Brooks and Karplus, 1983) and a friction constant of 62 ps<sup>-1</sup> was applied to the water oxygens (Brooks and Karplus, 1989). After energy minimization, the system was gradually heated and equilibrated at 300 K during 220 ps, while restraints initially applied on the heavy atoms were progressively removed. The MD production run, during which all atoms in the reaction region were unconstrained, was performed at 300 K for 2 ns. All molecular modeling calculations were performed starting from the high-quality crystal structure of the human OGT bound to the peptide from HCF-1<sub>PRO</sub> repeat 2 (1-26) E10Q and UDP-5SGlcNAc (PDB code 4N3B). This experimental structure was preferred to the lower-quality experimental structure with E10 and UDP-GlcNAc (PDB code 4N3C) to perform molecular modeling calculations. Before MD simulations, the 5S sulfur atom was replaced by an oxygen atom, while the Q10 residue of the HCF-1<sub>PRO</sub> repeat 2 was replaced by a E, D or A residue using the UCSF chimera program (Pettersen et al., 2004).

For the calculation of residue contributions to the binding-free energy, the role of each residue on the HCF-1–OGT binding free energy was estimated according to the MM-GBSA approach (Gohlke et al., 2003; Zoete et al., 2005; Zoete and Michielin, 2007), using the GB-MV2 implicit solvent model (Lee et al., 2003; Lee et al., 2002). This method allows the decomposition of the total binding free energy into residue contributions. The contribution of each residue of interest was calculated and averaged over 500 frames regularly extracted along a 2 ns SBC MD simulations centered on it.

Periodic-boundary MD simulations were carried out with GROMACS Version 4.6.5 (Bjellmar et al., 2010; Hess et al., 2008), using the all-atom CHARMM27 force field (MacKerell et al., 1998; Mackerell et al., 2004) and the TIP3P water model (Jorgensen et al., 1983). Electrostatic interactions were calculated with the Ewald particle-mesh method (Essmann et al., 1995) with a grid spacing of 1.2 Å and a spline interpolation of order 4. A cutoff of 10 Å was applied for the real-space direct sum part of the Ewald sum, and a cutoff of 14 Å for the van der Waals interactions. Dispersion corrections were applied to the energy. Bonds involving hydrogen atoms were constrained using the P-LINCS algorithm (Hess et al., 2008). The system was coupled to a barostat (Berendsen et al., 1984) with a relaxation time of 1 ps. The solute and the solvent were separately coupled to two thermostats (Bussi et al., 2007), each with a relaxation time of 0.2 ps. The time integration step was set to 2 fs, the temperature to 300 K, and the pressure to 1 bar. A cubic simulation cell with an edge length of 120 Å was used to prevent direct interactions between periodic images, resulting about 170,000 atoms per system. Sodium atoms were added to neutralize each system. Initial structures were optimized, heated from 0 to 300 K over a period of 0.1 ns, equilibrated for a further 0.9 ns restraining each solute non-hydrogen atom to its original position, and finally equilibrated for 0.5 ns without restraints before data collection. For each system, 8 simulations were carried out for a total of 40 ns of production time, saving coordinates every 0.05 ns.

OGT simulations in complex with HCF-1-derived peptides were based on the X-ray structure 4N3B (Lazarus et al., 2013), replacing UDP-5SGlcNAc with UDP-GlcNAc and glutamine 10 with glutamate. When replacing glutamine 10 by glutamate, it was observed that neutralizing its carboxylate group was necessary to obtain a stable OGT–HCF-1 complex. The added proton formed a stable hydrogen bond with the backbone oxygen of OGT glycine 654.

Molecular graphics were performed with the UCSF Chimera visualization software (Pettersen et al., 2004) or the PyMOL (Schrodinger, 2010) software v1.5.0.3 enhanced for Mac OS X.

## References

- Ajuh, P.M., Browne, G.J., Hawkes, N.A., Cohen, P.T., Roberts, S.G., and Lamond, A.I. (2000). Association of a protein phosphatase 1 activity with the human factor C1 (HCF) complex. *Nucleic Acids Res* **28**, 678-686.
- Alberts, B. (2008). *Molecular biology of the cell*, 5th edn (New York, Garland Science).
- Beltrao, P., Bork, P., Krogan, N.J., and van Noort, V. (2013). Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* **9**, 714.
- Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A., and Haak, J.R. (1984). Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* **81**, 3684-3690.
- Berg, J.M., Tymoczko, J.L., Stryer, L., and Stryer, L. (2002). *Biochemistry*, 5th edn (New York, W.H. Freeman).
- Bjelkmar, P., Larsson, P., Cuendet, M.A., Hess, B., and Lindahl, E. (2010). Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput* **6**, 459-466.
- Brooks, C.L., 3rd, and Karplus, M. (1989). Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. *J Mol Biol* **208**, 159-181.
- Brooks, C.L., and Gu, W. (2003). Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr Opin Cell Biol* **15**, 164-171.
- Brooks, C.L., and Karplus, M. (1983). Deformable Stochastic Boundaries in Molecular-Dynamics. *J Chem Phys* **79**, 6312-6325.
- Brooks, C.L., Karplus, M., and Pettitt, B.M. (1988). *Proteins : a theoretical perspective of dynamics, structure, and thermodynamics* (New York, J. Wiley).
- Brown, M.S., and Goldstein, J.L. (1997). The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331-340.
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J Chem Phys* **126**.
- Caballero, B., Allen, L., and Prentice, A. *Encyclopedia of human nutrition*, Third edition. edn.
- Capotosti, F., Guernier, S., Lammers, F., Waridel, P., Cai, Y., Jin, J., Conaway, J.W., Conaway, R.C., and Herr, W. (2011). O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144**, 376-388.
- Capotosti, F., Hsieh, J.J., and Herr, W. (2007). Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways. *Mol Cell Biol* **27**, 7063-7072.
- Carbone, M., Yang, H., Pass, H.I., Krausz, T., Testa, J.R., and Gaudino, G. (2013). BAP1 and cancer. *Nat Rev Cancer* **13**, 153-159.
- Cleland, W.W. (1967). The statistical analysis of enzyme kinetic data. *Adv Enzymol Relat Areas Mol Biol* **29**, 1-32.
- Comer, F.I., and Hart, G.W. (2001). Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40**, 7845-7852.

## REFERENCES

- D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem Sci* **28**, 655-662.
- Dalziel, M., Crispin, M., Scanlan, C.N., Zitzmann, N., and Dwek, R.A. (2014). Emerging principles for the therapeutic exploitation of glycosylation. *Science* **343**, 1235681.
- Daou, S., Mashtalir, N., Hammond-Martel, I., Pak, H., Yu, H., Sui, G., Vogel, J.L., Kristie, T.M., and Affar el, B. (2011). Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway. *Proc Natl Acad Sci U S A* **108**, 2747-2752.
- Davie, E.W., and Ratnoff, O.D. (1964). Waterfall Sequence for Intrinsic Blood Clotting. *Science* **145**, 1310-1312.
- Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* **105**, 10762-10767.
- Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., *et al.* (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO journal* **32**, 645-655.
- Dey, A., Seshasayee, D., Noubade, R., French, D.M., Liu, J., Chaurushiya, M.S., Kirkpatrick, D.S., Pham, V.C., Lill, J.R., Bakalarski, C.E., *et al.* (2012). Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541-1546.
- Drag, M., and Salvesen, G.S. (2010). Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* **9**, 690-701.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103**, 8577-8593.
- Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R.A., and Wigler, M. (1988). Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol Cell Biol* **8**, 2159-2165.
- Fong, J.J., Nguyen, B.L., Bridger, R., Medrano, E.E., Wells, L., Pan, S., and Sifers, R.N. (2012). beta-N-Acetylglucosamine (O-GlcNAc) is a novel regulator of mitosis-specific phosphorylations on histone H3. *J Biol Chem* **287**, 12195-12203.
- Fraser, D., and Powell, R.E. (1950). The kinetics of trypsin digestion. *J Biol Chem* **187**, 803-820.
- Fujiki, R., Hashiba, W., Sekine, H., Yokoyama, A., Chikanishi, T., Ito, S., Imai, Y., Kim, J., He, H.H., Igarashi, K., *et al.* (2011). GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* **480**, 557-560.
- Gambetta, M.C., and Muller, J. (2014). O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. *Dev Cell* **31**, 629-639.
- Gambetta, M.C., and Muller, J. (2015). A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma*.
- Gambetta, M.C., Oktaba, K., and Muller, J. (2009). Essential role of the glycosyltransferase *sxc/Ogt* in polycomb repression. *Science* **325**, 93-96.
- Gerster, T., and Roeder, R.G. (1988). A herpesvirus trans-activating protein interacts with transcription factor OTF-1 and other cellular proteins. *Proc Natl Acad Sci U S A* **85**, 6347-6351.
- Glinsky, G.V., Berezovska, O., and Glinskii, A.B. (2005). Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J Clin Invest* **115**, 1503-1521.

## REFERENCES

- Gloster, T.M., Zandberg, W.F., Heinonen, J.E., Shen, D.L., Deng, L., and Vocadlo, D.J. (2011). Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. *Nat Chem Biol* 7, 174-181.
- Gohlke, H., Kiel, C., and Case, D.A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J Mol Biol* 330, 891-913.
- Goto, H., Motomura, S., Wilson, A.C., Freiman, R.N., Nakabeppu, Y., Fukushima, K., Fujishima, M., Herr, W., and Nishimoto, T. (1997). A single-point mutation in HCF causes temperature-sensitive cell-cycle arrest and disrupts VP16 function. *Genes Dev* 11, 726-737.
- Goulet, B., and Nepveu, A. (2004). Complete and limited proteolysis in cell cycle progression. *Cell Cycle* 3, 986-989.
- Gross, B.J., Kraybill, B.C., and Walker, S. (2005). Discovery of O-GlcNAc transferase inhibitors. *J Am Chem Soc* 127, 14588-14589.
- Guinez, C., Filhoulaud, G., Rayah-Benhamed, F., Marmier, S., Dubuquoy, C., Dentin, R., Moldes, M., Burnol, A.F., Yang, X., Lefebvre, T., *et al.* (2011). O-GlcNAcylation increases ChREBP protein content and transcriptional activity in the liver. *Diabetes* 60, 1399-1413.
- Hailfinger, S., Nogai, H., Pelzer, C., Jaworski, M., Cabalzar, K., Charton, J.E., Guzzardi, M., Decaillet, C., Grau, M., Dorken, B., *et al.* (2011). Malt1-dependent RelB cleavage promotes canonical NF-kappaB activation in lymphocytes and lymphoma cell lines. *Proc Natl Acad Sci U S A* 108, 14596-14601.
- Hanover, J.A., Krause, M.W., and Love, D.C. (2012). Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nat Rev Mol Cell Biol* 13, 312-321.
- Hanover, J.A., Yu, S., Lubas, W.B., Shin, S.H., Ragano-Caracciola, M., Kochran, J., and Love, D.C. (2003). Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* 409, 287-297.
- Hart, G.W., Housley, M.P., and Slawson, C. (2007). Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446, 1017-1022.
- Hart, G.W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. (2011). Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem* 80, 825-858.
- Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G., *et al.* (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev* 2, 1513-1516.
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4, 435-447.
- Hsieh, J.J., Cheng, E.H., and Korsmeyer, S.J. (2003). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* 115, HEK 293-303.
- Huang, L., Jolly, L.A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubridge, C., Wiczorek, D., Tzschach, A., Cohen, M., *et al.* (2012). A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *Am J Hum Genet* 91, 694-702.
- Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell* 28, 730-738.

## REFERENCES

- Jackson, S.P., and Tjian, R. (1988). O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* 55, 125-133.
- Jinek, M., Rehwinkel, J., Lazarus, B.D., Izaurralde, E., Hanover, J.A., and Conti, E. (2004). The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* 11, 1001-1007.
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res* 36, W5-9.
- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292, 195-202.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 79, 926-935.
- Julien, E., and Herr, W. (2003). Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *The EMBO journal* 22, 2360-2369.
- Kaasik, K., Kivimae, S., Allen, J.J., Chalkley, R.J., Huang, Y., Baer, K., Kissel, H., Burlingame, A.L., Shokat, K.M., Ptacek, L.J., *et al.* (2013). Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. *Cell Metab* 17, 291-302.
- Kamemura, K., Hayes, B.K., Comer, F.I., and Hart, G.W. (2002). Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J Biol Chem* 277, 19229-19235.
- Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* 268, 10416-10424.
- Kelly, W.G., and Hart, G.W. (1989). Glycosylation of chromosomal proteins: localization of O-linked N-acetylglucosamine in *Drosophila* chromatin. *Cell* 57, 243-251.
- Khoury, G.A., Baliban, R.C., and Floudas, C.A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* 1.
- Kristie, T.M., Pomerantz, J.L., Twomey, T.C., Parent, S.A., and Sharp, P.A. (1995). The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. *J Biol Chem* 270, 4387-4394.
- Kumar, A., and Bachhawat, A.K. (2012). Pyroglutamic acid: throwing light on a lightly studied metabolite. *Curr Sci India* 102, 288-297.
- Lazarus, M.B., Jiang, J., Gloster, T.M., Zandberg, W.F., Whitworth, G.E., Vocadlo, D.J., and Walker, S. (2012). Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nat Chem Biol* 8, 966-968.
- Lazarus, M.B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W.F., Vocadlo, D.J., Herr, W., and Walker, S. (2013). HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* 342, 1235-1239.
- Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011). Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* 469, 564-567.
- Lee, J., Lee, H.J., Shin, M.K., and Ryu, W.S. (2004). Versatile PCR-mediated insertion or deletion mutagenesis. *Biotechniques* 36, 398-400.

## REFERENCES

- Lee, M.S., Feig, M., Salsbury, F.R., Jr., and Brooks, C.L., 3rd (2003). New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J Comput Chem* **24**, 1348-1356.
- Lee, M.S., Salsbury, F.R., and Brooks, C.L. (2002). Novel generalized Born methods. *J Chem Phys* **116**, 10606-10614.
- Liu, Y.D., Goetze, A.M., Bass, R.B., and Flynn, G.C. (2011). N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. *J Biol Chem* **286**, 11211-11217.
- Love, D.C., Kochan, J., Cathey, R.L., Shin, S.H., and Hanover, J.A. (2003). Mitochondrial and nucleocytoplasmic targeting of O-linked GlcNAc transferase. *J Cell Sci* **116**, 647-654.
- Lubas, W.A., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J Biol Chem* **272**, 9316-9324.
- Macfarlane, R.G. (1964). An Enzyme Cascade in the Blood Clotting Mechanism, and Its Function as a Biochemical Amplifier. *Nature* **202**, 498-499.
- Machida, Y.J., Machida, Y., Vashisht, A.A., Wohlschlegel, J.A., and Dutta, A. (2009). The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J Biol Chem* **284**, 34179-34188.
- MackKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616.
- Mackerell, A.D., Feig, M., and Brooks, C.L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **25**, 1400-1415.
- Marshall, S., Bacote, V., and Traxinger, R.R. (1991). Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *J Biol Chem* **266**, 4706-4712.
- Michaud, J., Praz, V., James Faresse, N., Jnbaptiste, C.K., Tyagi, S., Schutz, F., and Herr, W. (2013). HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. *Genome Res* **23**, 907-916.
- Minguez, P., Parca, L., Diella, F., Mende, D.R., Kumar, R., Helmer-Citterich, M., Gavin, A.C., van Noort, V., and Bork, P. (2012). Deciphering a global network of functionally associated post-translational modifications. *Mol Syst Biol* **8**, 599.
- Misaghi, S., Ottosen, S., Izrael-Tomasevic, A., Arnott, D., Lamkanfi, M., Lee, J., Liu, J., O'Rourke, K., Dixit, V.M., and Wilson, A.C. (2009). Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol Cell Biol* **29**, 2181-2192.
- Myers, S.A., Daou, S., Affar el, B., and Burlingame, A. (2013). 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**, 982-991.
- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., *et al.* (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3.
- Park, J., Lammers, F., Herr, W., and Song, J.J. (2012). HCF-1 self-association via an interdigitated Fn3 structure facilitates transcriptional regulatory complex formation. *Proc Natl Acad Sci U S A* **109**, 17430-17435.

## REFERENCES

- Parker, J.B., Palchaudhuri, S., Yin, H., Wei, J., and Chakravarti, D. (2012). A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function. *Mol Cell Biol* 32, 1654-1670.
- Patel, T., Gores, G.J., and Kaufmann, S.H. (1996). The role of proteases during apoptosis. *Faseb J* 10, 587-597.
- Pena-Llopis, S., Vega-Rubin-de-Celis, S., Liao, A., Leng, N., Pavia-Jimenez, A., Wang, S., Yamasaki, T., Zhrebker, L., Sivanand, S., Spence, P., *et al.* (2012). BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet* 44, 751-759.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.
- Phillips, R.S., Demidkina, T.V., and Faleev, N.G. (2014). The role of substrate strain in the mechanism of the carbon-carbon lyases. *Bioorg Chem* 57, 198-205.
- Plank, J.L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. *Mol Cell* 55, 5-14.
- Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* 4, 565-583.
- Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (2012). Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J Biol Chem* 287, 23549-23561.
- Roos, M.D., Su, K., Baker, J.R., and Kudlow, J.E. (1997). O glycosylation of an Sp1-derived peptide blocks known Sp1 protein interactions. *Mol Cell Biol* 17, 6472-6480.
- Ruan, H.B., Han, X., Li, M.D., Singh, J.P., Qian, K., Azarhoush, S., Zhao, L., Bennett, A.M., Samuel, V.T., Wu, J., *et al.* (2012). O-GlcNAc Transferase/Host Cell Factor C1 Complex Regulates Gluconeogenesis by Modulating PGC-1alpha Stability. *Cell Metab* 16, 226-237.
- Ruan, H.B., Nie, Y., and Yang, X. (2013). Regulation of protein degradation by O-GlcNAcylation: crosstalk with ubiquitination. *Mol Cell Proteomics* 12, 3489-3497.
- Sakabe, K., and Hart, G.W. (2010). O-GlcNAc transferase regulates mitotic chromatin dynamics. *J Biol Chem* 285, 34460-34468.
- Sakabe, K., Wang, Z., and Hart, G.W. (2010). Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc Natl Acad Sci U S A* 107, 19915-19920.
- Schimpl, M., Zheng, X., Borodkin, V.S., Blair, D.E., Ferenbach, A.T., Schuttelkopf, A.W., Navratilova, I., Aristotelous, T., Albarbarawi, O., Robinson, D.A., *et al.* (2012). O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nat Chem Biol* 8, 969-974.
- Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.
- Schroeter, E.H., Kisslinger, J.A., and Kopan, R. (1998). Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* 393, 382-386.
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68, 850-858.

## REFERENCES

- Slawson, C., and Hart, G.W. (2011). O-GlcNAc signalling: implications for cancer cell biology. *Nat Rev Cancer* *11*, 678-684.
- Slawson, C., Zachara, N.E., Vosseller, K., Cheung, W.D., Lane, M.D., and Hart, G.W. (2005). Perturbations in O-linked beta-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. *J Biol Chem* *280*, 32944-32956.
- Spiro, R.G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* *12*, 43R-56R.
- Takeda, S., Chen, D.Y., Westergard, T.D., Fisher, J.K., Rubens, J.A., Sasagawa, S., Kan, J.T., Korsmeyer, S.J., Cheng, E.H., and Hsieh, J.J. (2006). Proteolysis of MLL family proteins is essential for *taspase1*-orchestrated cell cycle progression. *Genes Dev* *20*, 2397-2409.
- Torres, C.R., and Hart, G.W. (1984). Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem* *259*, 3308-3317.
- Trinidad, J.C., Barkan, D.T., Gullede, B.F., Thalhammer, A., Sali, A., Schoepfer, R., and Burlingame, A.L. (2012). Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics* *11*, 215-229.
- Turk, B. (2006). Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* *5*, 785-799.
- Turk, B., Turk, D., and Turk, V. (2012). Protease signalling: the cutting edge. *The EMBO journal* *31*, 1630-1643.
- Tyagi, S., Chabes, A.L., Wysocka, J., and Herr, W. (2007). E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol Cell* *27*, 107-119.
- Tyagi, S., and Herr, W. (2009). E2F1 mediates DNA damage and apoptosis through HCF-1 and the MLL family of histone methyltransferases. *The EMBO journal* *28*, 3185-3195.
- UniProt Consortium (2015). Glycosylation (UniProt Consortium).
- Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., *et al.* (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* *49*, 645-656.
- Vosseller, K., Wells, L., Lane, M.D., and Hart, G.W. (2002). Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes. *Proc Natl Acad Sci U S A* *99*, 5313-5318.
- Wang, Z., Pandey, A., and Hart, G.W. (2007). Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics* *6*, 1365-1379.
- Wang, Z., Udeshi, N.D., Slawson, C., Compton, P.D., Sakabe, K., Cheung, W.D., Shabanowitz, J., Hunt, D.F., and Hart, G.W. (2010a). Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* *3*, ra2.
- Wang, Z.H., Udeshi, N.D., Slawson, C., Compton, P.D., Sakabe, K., Cheung, W.D., Shabanowitz, J., Hunt, D.F., and Hart, G.W. (2010b). Extensive Crosstalk Between O-GlcNAcylation and Phosphorylation Regulates Cytokinesis. *Sci Signal* *3*.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* *25*, 1189-1191.

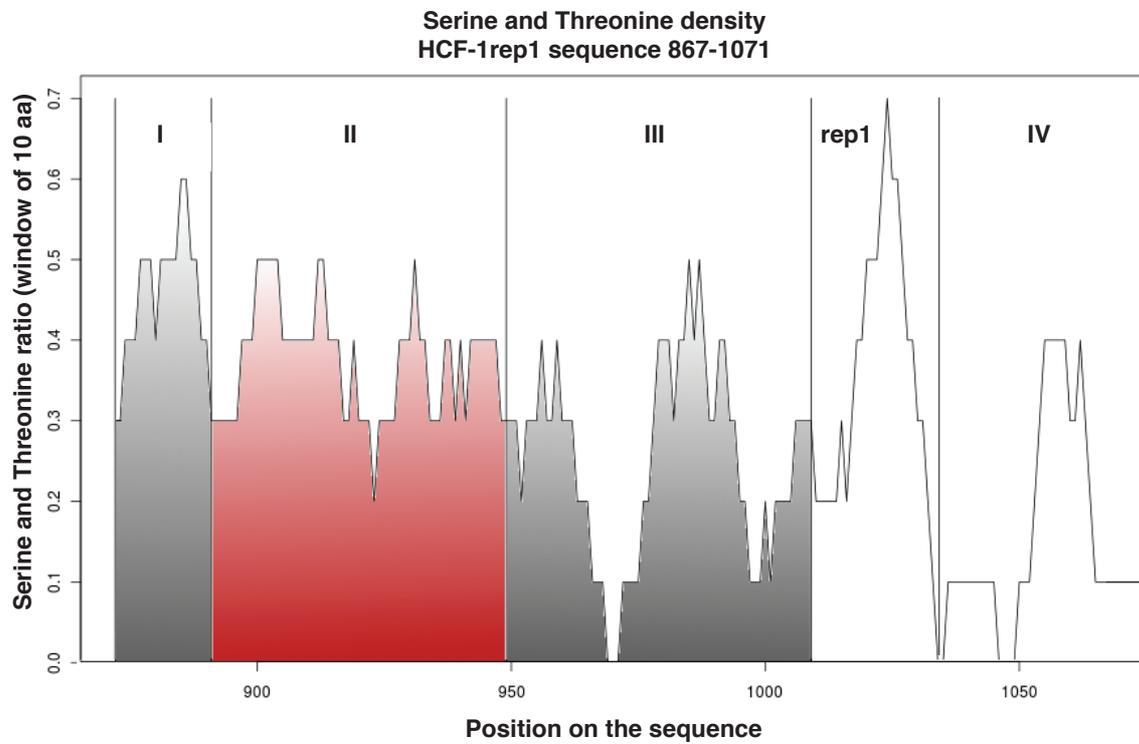
## REFERENCES

- Wells, L., and Hart, G.W. (2003). O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar. *FEBS Lett* **546**, 154-158.
- Wilm, M., Shevchenko, A., Houthaev, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469.
- Wilson, A.C., LaMarco, K., Peterson, M.G., and Herr, W. (1993). The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74**, 115-125.
- Wilson, A.C., Parrish, J.E., Massa, H.F., Nelson, D.L., Trask, B.J., and Herr, W. (1995a). The gene encoding the VP16-accessory protein HCF (HCFC1) resides in human Xq28 and is highly expressed in fetal tissues and the adult kidney. *Genomics* **25**, 462-468.
- Wilson, A.C., Peterson, M.G., and Herr, W. (1995b). The HCF repeat is an unusual proteolytic cleavage signal. *Genes Dev* **9**, 2445-2458.
- Wysocka, J., and Herr, W. (2003). The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends Biochem Sci* **28**, 294-304.
- Wysocka, J., Liu, Y., Kobayashi, R., and Herr, W. (2001a). Developmental and cell-cycle regulation of *Caenorhabditis elegans* HCF phosphorylation. *Biochemistry* **40**, 5786-5794.
- Wysocka, J., Myers, M.P., Laherty, C.D., Eisenman, R.N., and Herr, W. (2003). Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* **17**, 896-911.
- Wysocka, J., Reilly, P.T., and Herr, W. (2001b). Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol Cell Biol* **21**, 3820-3829.
- Yang, W.H., Kim, J.E., Nam, H.W., Ju, J.W., Kim, H.S., Kim, Y.S., and Cho, J.W. (2006). Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat Cell Biol* **8**, 1074-1083.
- Yang, X., Zhang, F., and Kudlow, J.E. (2002). Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110**, 69-80.
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* **24**, 5639-5649.
- Yoshida, A., Takauji, R., Inuzuka, M., Ueda, T., and Nakamura, T. (1996). Role of serine and ICE-like proteases in induction of apoptosis by etoposide in human leukemia HL-60 cells. *Leukemia* **10**, 821-824.
- Yu, H., Mashtalir, N., Daou, S., Hammond-Martel, I., Ross, J., Sui, G., Hart, G.W., Rauscher, F.J., 3rd, Drobetsky, E., Milot, E., *et al.* (2010). The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol* **30**, 5071-5085.
- Yu, H.C., Sloan, J.L., Scharer, G., Brebner, A., Quintana, A.M., Achilly, N.P., Manoli, I., Coughlin, C.R., 2nd, Geiger, E.A., Schneck, U., *et al.* (2013). An X-linked cobalamin disorder caused by mutations in transcriptional coregulator HCFC1. *Am J Hum Genet* **93**, 506-514.
- Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., *et al.* (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**, 971-975.
- Zachara, N.E., O'Donnell, N., Cheung, W.D., Mercer, J.J., Marth, J.D., and Hart, G.W. (2004). Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* **279**, 30133-30142.

## REFERENCES

- Zeytuni, N., Baran, D., Davidov, G., and Zarivach, R. (2012). Inter-phylum structural conservation of the magnetosome-associated TPR-containing protein, MamA. *J Struct Biol* *180*, 479-487.
- Zhang, K., Yin, R., and Yang, X. (2014). O-GlcNAc: A Bittersweet Switch in Liver. *Front Endocrinol (Lausanne)* *5*, 221.
- Zhu, Y., Liu, T.W., Cecioni, S., Eskandari, R., Zandberg, W.F., and Vocadlo, D.J. (2015). O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. *Nat Chem Biol*.
- Zoete, V., Meuwly, M., and Karplus, M. (2005). Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition. *Proteins-Structure Function and Bioinformatics* *61*, 79-93.
- Zoete, V., and Michielin, O. (2007). Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins* *67*, 1026-1047.

# Appendix



**Figure Appendix-1: Serine and threonine density in the HCF-1 amino acid sequence 867-1071.** The ratio of serines and threonines per position in a sliding window of 10 amino acids (aa) is shown. Analysis and figure by Dr. Viviane Praz (member of the Herr laboratory).

**Table Appendix-1: *In vitro* and *in vivo* activities of selected HCF-1 recombinant constructs.**

	constructs	short name	<i>in vitro</i>		<i>in vivo</i>		<i>in vitro</i> OGT binding		<i>in vitro</i> cleavage		<i>in vitro</i> OGT binding D554H / H558D		<i>in vitro</i> OGT binding TPR mutants											
			cleavage		O-GlcNAcylation		O-GlcNAcylation		UDP-GlcNAc		no cofactor		D554H / H558D		UDP-GlcNAc		no cofactor		TPR 5N-5A	TPR Δ1-6	TPR Δ1-8	TPR Δ1-9		
			+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-		
4 hours cleavage	<b>with PRO</b>	I. II. III. rep1. IV	FL	+++	+++	+++	+++	++	+++	+++	+++	+++	+++	+	(*)	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		II. III. rep1. IV	Δ I	+++	+++	+++	+++	-	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		III. rep1. IV	Δ II	++	(+)	++	(+)	-	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. rep1. IV	Δ III	++	+++	+++	+++	+	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		rep1. IV	Δ I.II.III	+	-	+	(+)	+	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. rep1. IV	+ I	++	+	++	(+)	++	+	++	++	++	++	++	++	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		II. rep1. IV	+ II	++++	++++	++++	+++	++++	++++	++++	++++	++++	++++	++++	++	++	N/D	N/D	N/D	N/D	N/D	N/D		
		III. rep1. IV	+ III	++	-	+	-	-	-	-	-	+	-	-	-	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1	Δ IV	+++	+++	+	(*)	+	(*)	-	+	N/D	+	+	+	N/D	+	+	N/D	N/D	N/D	N/D		
		Region II mutants	II. rep1.	II. rep1.	(+)	+++	(+)(*)	+++(*)	++	+++	N/D	+++	+	+	+	N/D	+	+	N/D	N/D	N/D	N/D	N/D	
			rep1. II. IV	+ II_C	-	++	(+)	++++	++	+++	N/D	+++	-	-	-	N/D	+	+	N/D	N/D	N/D	N/D	N/D	
			rep1. II. IV / E10A	+ II_C/E10A	-	++	-	+++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	+	+	N/D	N/D	N/D	N/D	N/D	
II. rep1. II. IV	+ II_N/C		-	+	-	(*)	++	+++	N/D	+++	-	-	-	N/D	+	+	N/D	N/D	N/D	N/D	N/D			
I. CKII. III. rep1. IV	CKII		N/D	N/D	+	(*)	+++(*)	N/D	N/D	N/D	N/D	N/D	N/D	N/D	+	+	N/D	N/D	N/D	N/D	N/D			
II. scrambled. rep1. IV	+ II_scramb.		(+)	+	(+)(*)	+++(*)	N/D	+++	N/D	+++	+	+	+	N/D	+	+	N/D	N/D	N/D	N/D	N/D			
III. rep1. II. IV	+ III_N/II_C		+	(*)	++++(*)	N/D	N/D	++(*)	N/D	++(*)	+	+	+	N/D	+	+	N/D	N/D	N/D	N/D	N/D			
II-a. rep1. IV	+ II-a		+++	+++	+	N/D	++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	+	+	N/D	N/D	N/D	N/D	N/D			
II-b. rep1. IV	+ II-b		+	?	+	+	++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	+	+	N/D	N/D	N/D	N/D	N/D			
Region III mutants	III. scrambled. rep1. IV		+ III_scramb.	++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
	HdaA. rep1. IV		+ HdaA	++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
no cleavage	<b>without PRO</b>		I. II. III. IV	Δ PRO	-	N/D	-	+++	N/D	N/D	-	-	-	-	N/D	-	-	N/D	N/D	N/D	N/D	N/D		
		II. IV	II_IV	-	N/D	-	N/D	+++	+++	-	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+		
		II	II_alone	-	N/D	-	N/D	++	++	-	(+)	+	+	+	(+)(*)	-	-	-	-	-	-	-		
		II. scrambled	+ II_scramb_alone	-	N/D	N/D	N/D	N/D	+	N/D	+	+	+	+	N/D	+	+	N/D	N/D	N/D	N/D	N/D		
POU domain mutants	POUs. rep2. POUh	POUrep2	(+)	-	(-)(*)	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D			
	POUs. II. rep2. POUh	+II_POUrep2	N/D	+	+++(*)	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D			
	POUs. III. rep2. POUh	+III_POUrep2	N/D	(+)	+++(*)	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D			
16 hours cleavage	<b>PRO mutants</b>	I. II. III. rep1. IV FL	FL	+++	+++	+++	+++	++	+++	+++	+++	+++	+++	+	+	N/D	N/D	N/D	N/D	N/D	N/D			
		I. II. III. rep1. IV E10A	E10A	-	+++	-	++	+++	+++	+++	+++	+++	+++	+++	+	+	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV E10D	E10D	-	+++	-	+++	+++	+++	+++	+++	+++	+++	+++	+	+	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV E10Q	E10Q	-	+++	-	+++	+++	+++	+++	+++	+++	+++	+++	+	+	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV E10S	E10S	-	+++	-	++	++++	+++	+++	+++	+++	+++	+++	+	+	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV P7A	P7A	+	+++	+++	+++	++	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV P8A	P8A	(+)	+++	(-)	++	++	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV C9A	C9A	+	+++	++	+++	++	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV T11A	T11A	+++	+++	+++	++	+++	+++	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV H12A	H12A	+++	+++	+++	++	+++	+++	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV H12G	H12G	++	+++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV H12P	H12P	+	+++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV E13A	E13A	++	+++	+++	+++	++	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV E13G	E13G	+	+++	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV E13P	E13P	+	+++	+++	+++	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV H12P/E13P	H12P/E13P	+	+++	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV H12G / E13G	H12G/E13G	+	+++	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV T14A	T14A	(-)	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV T14N	T14N	(-)	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV T14S	T14S	(-)	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV T17-22A	T17-22A	-	+++	-	+++	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. III. rep1. IV T17-22A / E10A	T17-22A/E10A	-	+++	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		II. III. rep1. IV / E10A	Δ I/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. III. rep1. IV / E10A	Δ II/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. II. rep1. IV / E10A	Δ III/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		I. rep1. IV / E10A	+ I/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		II. rep1. IV / E10A	+ II/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		III. rep1. IV / E10A	+ III/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		rep1. IV / E10A	Δ I.II.III/E10A	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		rep1. IV / E10S	Δ I.II.III/E10S	-	N/D	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		rep1. IV / T17-22A	Δ I.II.III/T17-22A	-	-	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		rep1. IV / E10A/T17-22A	Δ I.II.III/E10A/T17-22A	-	-	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		II. rep1. IV / T17-22A	+ II_T17-22A	-	+++	N/D	N/D	+	+	UDP: ++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D		
		<i>in vivo</i>	<b>HCF-1 Taspase mutants</b>	HCF-1 ΔPRO MLL	ΔPRO MLL	N/D	N/D	+++	+++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	
				HCF-1 ΔPRO MLL mut	ΔPRO MLL mut	N/D	N/D	-	+++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D
				HCF-1 wild-type	HCF-1 FL	N/D	N/D	+++	+++	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D	N/D

HCF-1 mutants were analyzed in multiple independent experiments with similar results unless marked by an asterisk. Cleavage activities were evaluated as enhanced (++++), moderate (+++), low (++), at detection level (+) or no cleavage detectable (-). (\*), the experiment was performed only once; N/D, not determined.

## HCF-1 is cleaved in the active site of O-GlcNAc transferase

The article by Lazarus et al. (2013) was first published in *Science* and is attached below or accessible in Pubmed (PMID: 24311690).

Lazarus, M.B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W.F., Vocadlo, D.J., Herr, W., and Walker, S. (2013). HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* 342, 1235-1239.

My individual contributions, as highlighted in Chapter II, are:

**Figure 1 C:** *In vitro* HCF-1rep1 cleavage assay. E10 is crucial of HCF-1<sub>PRO</sub>-repeat proteolysis: Substitutions of E10 by glutamine (E10Q), aspartate (E10D), or serine (E10S) block cleavage. The HCF-1<sub>PRO</sub>-repeat threonine region is important for cleavage because alanine substitutions of four consecutive threonines (HCF-1rep1 mutant T17-22A) inhibit cleavage.

**Figure 3 E:** *In vitro* HCF-1rep1–OGT binding assay. The HCF-1<sub>PRO</sub>-repeat threonine region is important for OGT association to the HCF-1<sub>PRO</sub>-repeat because the mutations in the threonine region (HCF-1rep1 mutant T17-22A) inhibit HCF-1<sub>PRO</sub>-repeat–OGT association.



## HCF-1 Is Cleaved in the Active Site of O-GlcNAc Transferase

Michael B. Lazarus *et al.*

*Science* **342**, 1235 (2013);

DOI: 10.1126/science.1243990

---

*This copy is for your personal, non-commercial use only.*

---

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

*The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of December 5, 2013):*

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/342/6163/1235.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2013/12/04/342.6163.1235.DC1.html>

This article **cites 37 articles**, 15 of which can be accessed free:

<http://www.sciencemag.org/content/342/6163/1235.full.html#ref-list-1>

20. H. Yang, P. J. Carney, R. O. Donis, J. Stevens, *J. Virol.* **86**, 8645–8652 (2012).
21. R. Xu, R. McBride, J. C. Paulson, C. F. Basler, I. A. Wilson, *J. Virol.* **84**, 1715–1721 (2010).
22. G. N. Rogers *et al.*, *J. Biol. Chem.* **260**, 7362–7367 (1985).
23. K. Tharakaraman *et al.*, *Cell* **153**, 1486–1493 (2013).
24. J. A. Belser *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7558–7563 (2008).
25. X. Xiong *et al.*, *Nature* **499**, 496–499 (2013).
26. J. Zhou *et al.*, *Nature* **499**, 500–503 (2013).
27. As a control, the HA from a representative human seasonal H1N1 strain (A/Kentucky/UR06-0258/2007, KY/07) displays characteristic strong binding to  $\alpha$ 2-6 SLNLN and negligible binding to  $\alpha$ 2-3 SLNLN (Fig. 2A).
28. I. Ramos *et al.*, *J. Gen. Virol.* **94**, 2417–2423 (2013).
29. Y. Shi *et al.*, *Science* **342**, 243–247 (2013); 10.1126/science.1242917.
30. H. Yang, P. J. Carney, J. C. Chang, J. M. Villanueva, J. Stevens, *J. Virol.* **87**, 12433–12446 (2013).
31. J. C. Paulson, R. P. de Vries, *Virus Res.* **178**, 99–113 (2013).
32. A. S. Gambaryan *et al.*, *Virol. J.* **5**, 85 (2008).
33. As a control, human HA KY07 binds broadly to  $\alpha$ 2-6-linked glycans on the array (Fig. 2B), and avian H7 HA recognizes a large number and variety of  $\alpha$ 2-3-sialylated glycans (fig. S2).
34. The L226Q mutant produced in insect cells showed robust binding to the array, revealing stronger avidity to most  $\alpha$ 2-3-glycans but still no binding to  $\alpha$ 2-6-sialosides.
35. R. Xu, I. A. Wilson, in *Structural Glycomics*, E. Yuriev, P. A. Ramsland, Eds. (CRC Press, Boca Raton, FL, 2012), pp. 235–257.
36. J. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17175–17180 (2009).
37. Xiong *et al.* used 3'-SLN as an avian-like receptor analog. Shi *et al.* used  $\alpha$ 2-3-SLNLN. In our study, we used L5Ta.
38. X. Xiong *et al.*, *Virus Res.* **178**, 12–14 (2013).
39. K. Srinivasan, R. Raman, A. Jayaraman, K. Viswanathan, R. Sasisekharan, *PLOS ONE* **8**, e49597 (2013).

**Acknowledgments:** This work was supported in part by NIH R56 AI099275 (I.A.W.) and the Skaggs Institute for Chemical Biology, the Scripps Microarray Core Facility, and a contract from the Centers for Disease Control (J.C.P.). R.P.d.V. is a recipient of a Rubicon grant from the Netherlands Organization for Scientific Research (NWO). Several glycans used for HA binding assays were provided by the Consortium for Functional Glycomics ([www.functionalglycomics.org/](http://www.functionalglycomics.org/)) funded by National Institute of General Medical Sciences (NIGMS) grant GM62116 (J.C.P.). X-ray diffraction data were collected at the Advanced Photon Source beamline 23ID-B (GM/CA CAT) and the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2. GM/CA CAT is funded in whole or in part with federal funds

from the National Cancer Institute (Y1-CO-1020) and the NIGMS (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy (DOE), Basic Energy Sciences, Office of Science, under contract no. DE-AC02-06CH11357. The SSRL is a Directorate of Stanford Linear Accelerator Center National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. DOE Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the NIH, NIGMS (including P41GM103393) and the National Center for Research Resources (NCRR, P41RR001209). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS, NCRR, or NIH. This is publication 24079 from the Scripps Research Institute. Coordinates and structure factors are deposited in the PDB (4N5J, 4N5K, 4N60, 4N61, 4N62, 4N63, and 4N64).

#### Supplementary Materials

[www.sciencemag.org/content/342/6163/1230/suppl/DC1](http://www.sciencemag.org/content/342/6163/1230/suppl/DC1)  
Materials and Methods  
Figs. S1 to S5  
Tables S1 to S3  
References (40–47)

25 July 2013; accepted 28 October 2013  
10.1126/science.1243761

## HCF-1 Is Cleaved in the Active Site of O-GlcNAc Transferase

Michael B. Lazarus,<sup>1,4\*</sup> Jiaoyang Jiang,<sup>1,\*†</sup> Vaibhav Kapuria,<sup>2</sup> Tanja Bhuiyan,<sup>2</sup> John Janetzko,<sup>4</sup> Wesley F. Zandberg,<sup>3</sup> David J. Vocadlo,<sup>3,5</sup> Winship Herr,<sup>2,5</sup> Suzanne Walker<sup>1,5</sup>

Host cell factor–1 (HCF-1), a transcriptional co-regulator of human cell-cycle progression, undergoes proteolytic maturation in which any of six repeated sequences is cleaved by the nutrient-responsive glycosyltransferase, O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT). We report that the tetratricopeptide-repeat domain of O-GlcNAc transferase binds the carboxyl-terminal portion of an HCF-1 proteolytic repeat such that the cleavage region lies in the glycosyltransferase active site above uridine diphosphate–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.

O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT) is a Ser/Thr (S/T) glycosyltransferase that O-GlcNAcylates nuclear and cytoplasmic proteins, thus influencing their activity, localization, and overall function (1–3). Because OGT activity is sensitive to uridine

diphosphate (UDP)–GlcNAc concentrations, OGT is proposed to regulate cellular responses to nutrient status (4–6). Human HCF-1 is a transcriptional coregulator involved in regulating cell-cycle progression (7, 8). In an unusual proteolytic maturation process (9–11), any of six centrally located 20 to 26 amino acid sequence repeats called HCF-1<sub>PRO</sub> repeats (Fig. 1A) are cleaved by OGT in the presence of UDP-GlcNAc (12), providing a link between cell-cycle progression and nutrient levels. The HCF-1<sub>PRO</sub> repeats contain two essential regions for proteolysis: a threonine-rich region proposed to be an OGT-binding site and the cleavage site, which contains a conserved Cys-Glu-Thr (CET) sequence (10, 11).

OGT can cleave a fragment of HCF-1, called HCF-1rep1, which contains the first HCF-1<sub>PRO</sub> repeat plus N-terminal HCF-1 sequences containing several O-GlcNAc sites (12) (Fig. 1A). To elucidate the cleavage process, we first analyzed the effect of amino acid substitutions in OGT (Fig. 1B and fig. S1) and the HCF-1<sub>PRO</sub> repeat

(Fig. 1C) on cleavage and glycosylation. Three OGT active site residues implicated in S/T 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; and H558, which contacts a backbone carbonyl of glycosylation substrates (13–17). Substitution of K842 with methionine prevented S/T glycosylation upstream of the proteolytic repeat as well as cleavage within the repeat region. Substitution of H498 or H558 with alanine decreased S/T glycosylation but had a negligible effect on the extent of cleavage after 16 hours. K842 is an essential residue for glycosylation (14, 15), and its importance in cleavage suggests that UDP-GlcNAc is involved in the cleavage mechanism.

Next, we tested substitutions in the proteolytic repeat of HCF-1rep1. We previously showed that alanine substitution of glutamate E10 leads to loss of cleavage (11, 12). To probe the role of E10 in more detail, we substituted it with glutamine (E10Q), aspartate (E10D), and serine (E10S). All three substitutions blocked cleavage (Fig. 1C), indicating that the chemical nature of the glutamate residue is critical for OGT-mediated HCF-1<sub>PRO</sub>-repeat cleavage. In contrast, the C9 position can tolerate alanine and serine substitution (12) (fig. S2).

Because S/T glycosylation upstream of the cleavage site in HCF-1rep1 complicates study of the cleavage requirements, we identified a cleavage substrate consisting of the first three proteolytic repeats (HCF3R, Fig. 1A), which did not undergo substantial glycosylation. No cleavage products were observed when HCF3R was incubated with OGT alone or in the presence of UDP, but several products were observed in reactions containing both OGT and UDP-GlcNAc (Fig. 2A). These products did not form if HCF3R was incubated with a K842A OGT mutant incapable of

<sup>1</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA. <sup>2</sup>Center for Integrative Genomics, University of Lausanne, Géopode, 1015 Lausanne, Switzerland. <sup>3</sup>Department of Chemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. <sup>4</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. <sup>5</sup>Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada.

\*These authors contributed equally to this work.

†Present address: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA.

‡Present address: School of Pharmacy, University of Wisconsin–Madison, Madison, WI 53705, USA.

§Corresponding author. E-mail: Suzanne\_Walker@hms.harvard.edu (S.W.); winship.herr@unil.ch (W.H.)

catalyzing glycosylation (13, 16) or when wild-type OGT was pretreated with 1.5 equivalents of a previously described inhibitor that covalently inactivates the enzyme by cross-linking the active site (18). Cleavage was also inhibited if UDP was added to reactions containing OGT and UDP-GlcNAc (fig. S3A) but accelerated by adding alkaline phosphatase, which destroys UDP. Because intact UDP-GlcNAc, but not UDP, promoted HCF3R cleavage, we tested cleavage in the presence of UDP-5SGlcNAc, an isostere of UDP-GlcNAc that adopts the same conformation within the active site, yet is resistant to glycosylation and hydrolysis (14, 19). Almost no peptide cleavage was observed (Fig. 2A and fig. S3B), implying that the UDP-GlcNAc does not simply fulfill a structural function but must react for HCF3R cleavage to occur. Consistent with this proposal, UDP-5SGlcNAc was found to inhibit cleavage of HCF-1rep1 when added to reactions containing UDP-GlcNAc (fig. S3C).

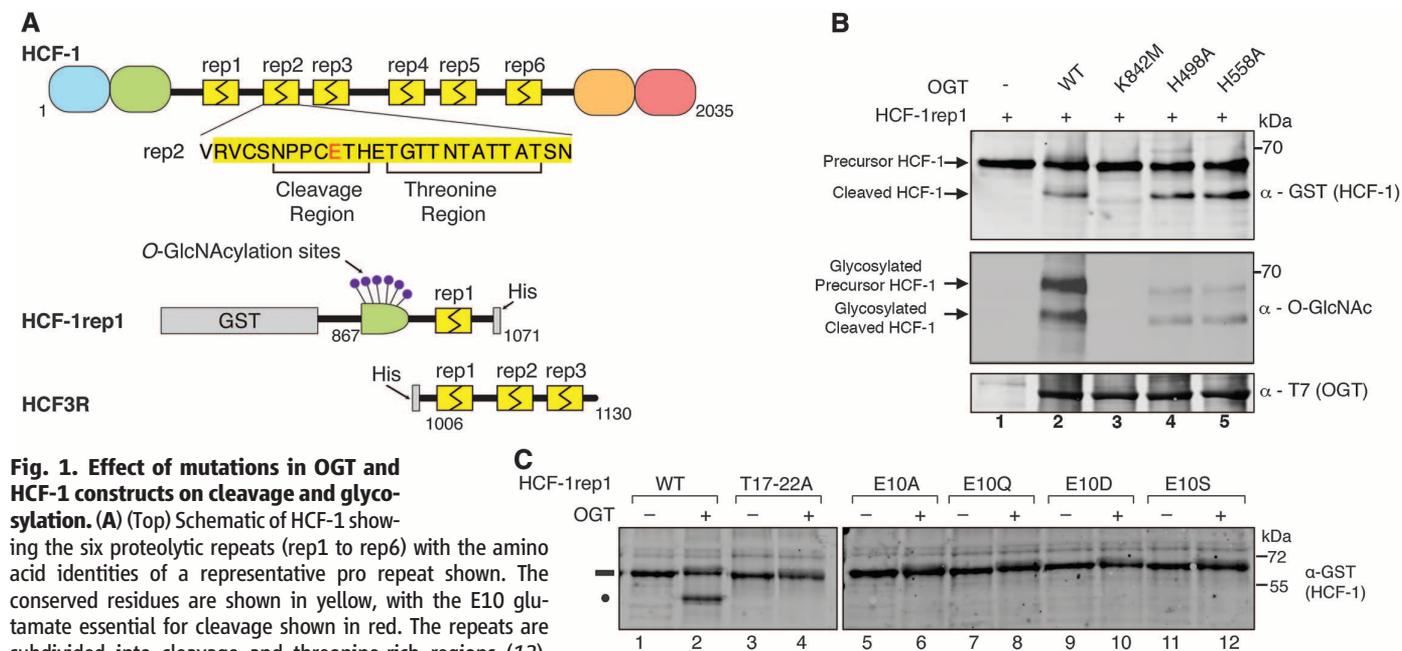
We next examined the cleavage products using liquid chromatography–mass spectrometry (LC-MS). Previous analysis of products isolated from cells concluded that cleavage occurs at the E-T peptide bond of the CET sequence (10, 11), but none of the four products we observed had masses consistent with E-T cleavage (Fig. 2B and fig. S4). Instead, they were consistent with cleavage at the preceding C-E bond except that all the C-terminal fragments were 18 daltons lower in

mass than expected. We hypothesized that the 18-dalton difference resulted from generation of an N-terminal pyroglutamate. Indeed, cleavage of a simplified HCF3R substrate possessing a single active HCF-1<sub>PRO</sub> repeat (HCF3R-EAA) led to the production of one dehydrated C-terminal product (Fig. 2C and fig. S5). Pyroglutamate aminopeptidase (PGAP) treatment reduced the mass of this product by 111 daltons, corresponding to loss of pyroglutamate (Fig. 2C and fig. S6). The earlier proposed cleavage site was probably misidentified because N-terminal pyroglutamates are resistant to the Edman sequencing method used in those studies, and cellular PGAPs likely processed some cleavage products (10, 11).

To gain additional insight into how OGT cleaves HCF-1, we solved crystal structures of OGT-HCF-1<sub>PRO</sub>-repeat complexes (tables S1 and S2). We solved a 1.8 Å structure of OGT containing UDP and a 16 amino acid peptide comprising the threonine-rich region of the HCF-1<sub>PRO</sub> repeat (Fig. 3, A and B). We also solved a 1.9 Å structure with UDP and a full HCF-1<sub>PRO</sub> repeat containing an E10A mutation, but density was only observed for the threonine-rich portion of the repeat (Fig. 3C and fig. S7). These two structures show the threonine-rich peptide bound in an extended conformation along the channel formed by the tetratricopeptide-repeat (TPR) domain (13, 20, 21) of OGT (Fig. 3B and fig. S7). 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 (Fig. 3C). A lysine side chain also contacts the peptide backbone. Four aspartates of OGT form hydrogen bonds to the threonine side chains of the HCF-1 repeat. The binding mode in the crystal structures is consistent with mutational data demonstrating the importance of the OGT asparagines (Fig. 3D and fig. S8) and the conserved HCF-1<sub>PRO</sub>-repeat threonines (12) (Figs. 3E and 1C) for both cleavage and binding.

Because intact UDP-GlcNAc is required for OGT-catalyzed cleavage of HCF-1<sub>PRO</sub>-repeats, we thought the UDP-5SGlcNAc analog might stabilize density for a full repeat. Indeed, we obtained a structure of OGT with UDP-5SGlcNAc and a 26 amino acid peptide corresponding to HCF-1<sub>PRO</sub> rep2, but with an E10Q substitution. 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 (Fig. 4A) and forms an extensive binding interface with UDP-5SGlcNAc (Fig. 4B). A structure containing UDP-GlcNAc and a wild-type repeat confirms the binding mode of the E10Q peptide (fig. S9). Remarkably, the cleavage region binds in a mode almost identical to that of a glycosylation-competent peptide substrate (14), and residue 10 aligns perfectly with the glycosyl acceptor amino acid (Fig. 4C). The structures suggested that an



**Fig. 1. Effect of mutations in OGT and HCF-1 constructs on cleavage and glycosylation.**

(A) (Top) Schematic of HCF-1 showing the six proteolytic repeats (rep1 to rep6) with the amino acid identities of a representative pro repeat shown. The conserved residues are shown in yellow, with the E10 glutamate essential for cleavage shown in red. The repeats are subdivided into cleavage and threonine-rich regions (12). Site-specific proteolysis by OGT leads to the formation of HCF-1<sub>N</sub> and HCF-1<sub>C</sub> subunits. (Bottom) Schematic of HCF-1 constructs used in this study. Glutathione S-transferase (GST)–HCF-1rep1 contains the first HCF-1<sub>PRO</sub> repeat and surrounding sequences fused to GST. Several S/T glycosylation sites are found in the HCF-1rep1 construct as schematized. The HCF3R construct contains only the first three HCF-1<sub>PRO</sub> repeats fused to an N-terminal His-tag. (B) Comparative cleavage and glycosylation activities of wild-type (WT) OGT and several catalytic domain mutants. GST–HCF-1rep1 was incubated in the absence (lane 1) or presence of WT OGT (lane 2) or the indicated mutants

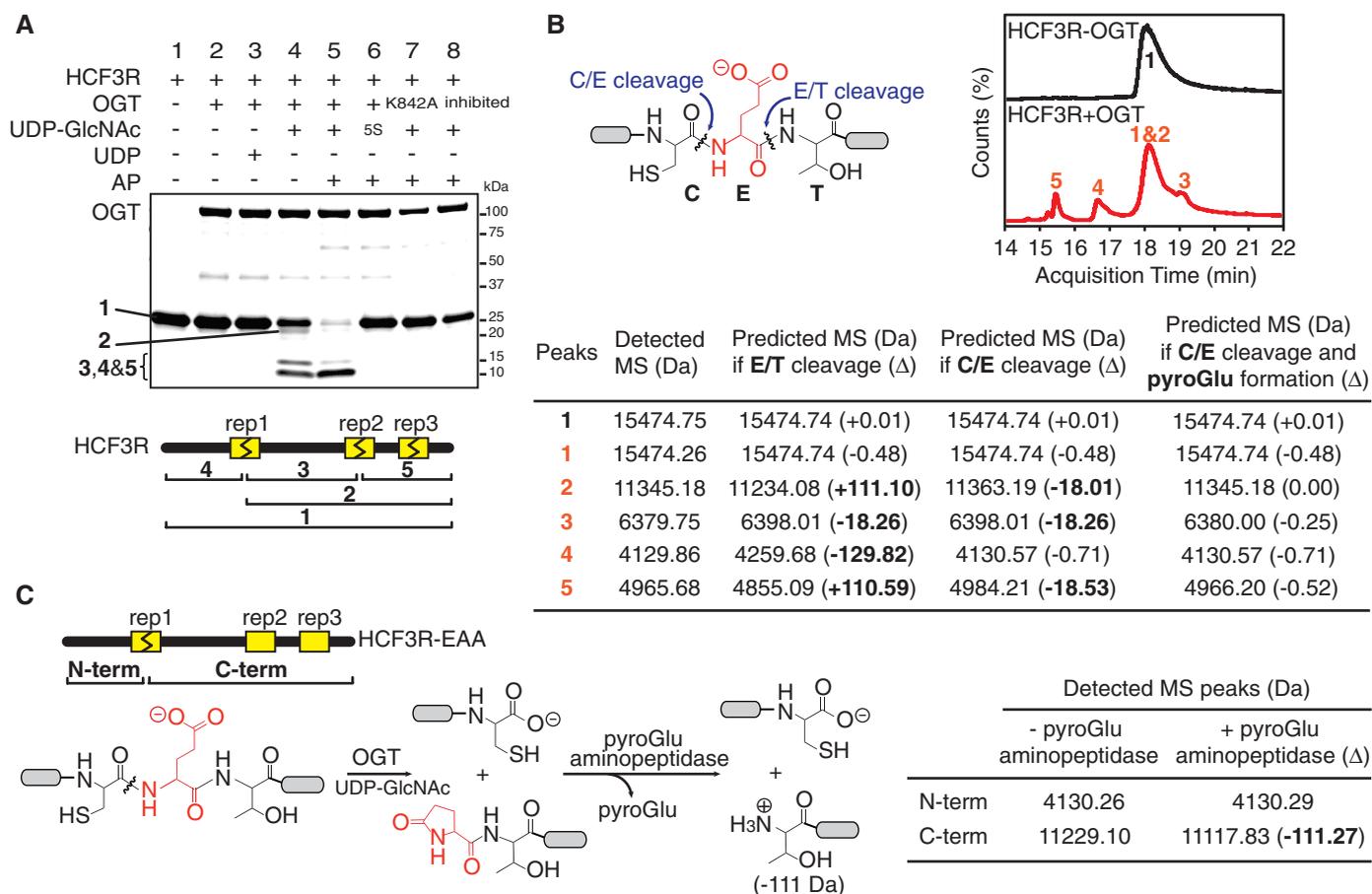
(lanes 3 to 5). HCF-1rep1 cleavage was detected by Western blot analysis with antibody to GST, and HCF-1rep1 glycosylation was detected with antibody to O-GlcNAc (RL2). (C) Cleavage activities of WT and mutant HCF-1rep1 constructs. WT GST–HCF-1rep1 (lanes 1 and 2) or a threonine-rich region mutant (T17-22A; lanes 3 and 4) or the indicated E10 cleavage site mutants (E10A, E10Q, E10D, and E10S; lanes 5 to 12) were incubated in the absence (–) or presence (+) of WT OGT as described in the supplementary materials. Cleavage was detected as in (B).

E10S mutation, which prevents cleavage (Fig. 1D), would be glycosylated at residue 10. Indeed, unlike HCF3R-EAA, the E10S analog (HCF3R-SAA) was glycosylated effectively (Fig. 4D and fig. S10). Thus, the identity of the amino acid at position 10 of an HCF-1<sub>PRO</sub> repeat—glutamate or serine—can dictate whether OGT cleaves or glycosylates the substrate.

Previous hypotheses suggested that OGT contains a dedicated protease active site or acts as a coprotease to template HCF-1 autocatalysis (12). Instead, OGT promotes cleavage of the HCF-1<sub>PRO</sub> repeat using the same catalytic region as for glycosylation. The threonine-rich region of the HCF-1<sub>PRO</sub> repeat binds in the channel formed by the TPR domain of OGT, stabilized by the contacts described above. The cleavage region threads into the active site and binds over UDP-GlcNAc in the same conformation that a glycosylation substrate would, with the glutamate side chain positioned near the anomeric carbon of the sugar. Because a pyroglutamate product is formed,

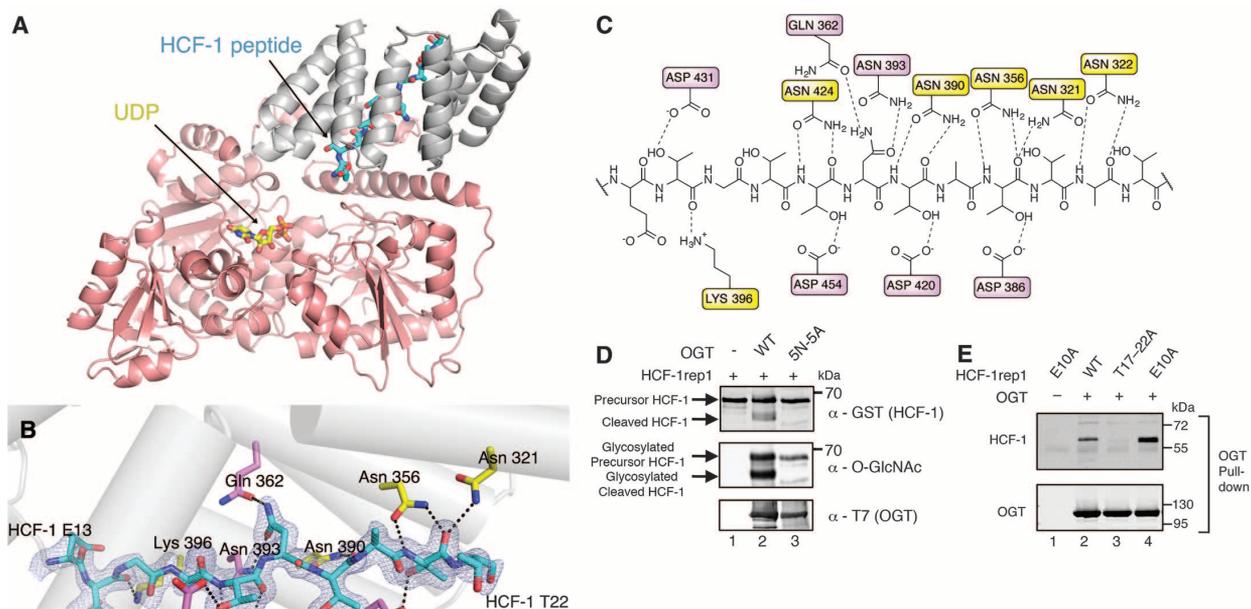
and spontaneous cyclization of N-terminal glutamates is kinetically very slow (22), the glutamate side chain is likely activated by formation of an ester species as part of the cleavage mechanism. We speculate that the glutamate side chain traps a transient oxocarbenium ion formed within the active site, producing a glutamyl ester that can undergo intramolecular attack, leading ultimately to formation of the N-terminal pyroglutamate. We note that pyroglutamates are proposed species in other biological phenomena (23). Possible mechanisms for cleavage proceeding from a glutamyl ester are suggested (fig. S11). Although direct physical evidence for a glutamyl-sugar intermediate has yet to be obtained, glycosylation of the glutamate side chain before 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<sub>PRO</sub> repeats requires UDP-GlcNAc and depends on an OGT residue, K842, which is essential for catalyzing glycosylation.

These studies provide insight into two important aspects of OGT function. First, they provide a view of a peptide bound to OGT's TPR domain, which is thought to play a central role in substrate selection (20, 21, 24, 25). The structures reported suggest that some glycosylation substrates may bind in a manner similar to the HCF-1<sub>PRO</sub> repeats, with the glycosylation site separated by several residues from a C-terminal recognition motif that binds in the channel formed by the TPR domain. Adaptor proteins that recruit glycosylation substrates to OGT might also contain threonine-rich recognition motifs. Second, they suggest an unprecedented mechanism of proteolysis in which OGT uses UDP-GlcNAc as a co-substrate in a cleavage reaction that takes place in the active site for glycosylation. Indeed, we show that two very different posttranslational protein modifications—proteolysis and addition of a sugar residue—can occur in the same active site, with the outcome determined by the identity of a single amino acid in the substrate.



**Fig. 2. HCF-1<sub>PRO</sub>-repeat cleavage results in formation of a pyroglutamate product.** (A) Cleavage of HCF3R requires UDP-GlcNAc. HCF3R was incubated with WT OGT (lanes 2 to 6), with K842A OGT (lane 7) or with OGT treated with a previously reported (18) covalent inhibitor BZX2 (lane 8), in the presence of UDP (lane 3), UDP-GlcNAc (lanes 4, 5, 7, and 8), or UDP-5S-GlcNAc (5S) (lane 6). Alkaline phosphatase (AP) was added to some reactions, as indicated. Cleavage products were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and stained with Coomassie Blue. (B) LC-MS analysis of

untreated HCF3R (black) and HCF3R cleavage products (red) after incubation with OGT and UDP-GlcNAc shows unexpected mass peaks. Detected and predicted MS peaks for different cleavage products are tabulated. (C) Mutation of E10 to alanine in the cleavage region of the second and third HCF-1<sub>PRO</sub> repeats produces a construct, HCF3R-EAA, containing only a single cleavable repeat. Pyroglutamate (pyroGlu) aminopeptidase removed a 111-dalton fragment from the HCF3R-EAA C-terminal cleavage product, confirming the formation of pyroglutamate in the cleavage reaction.

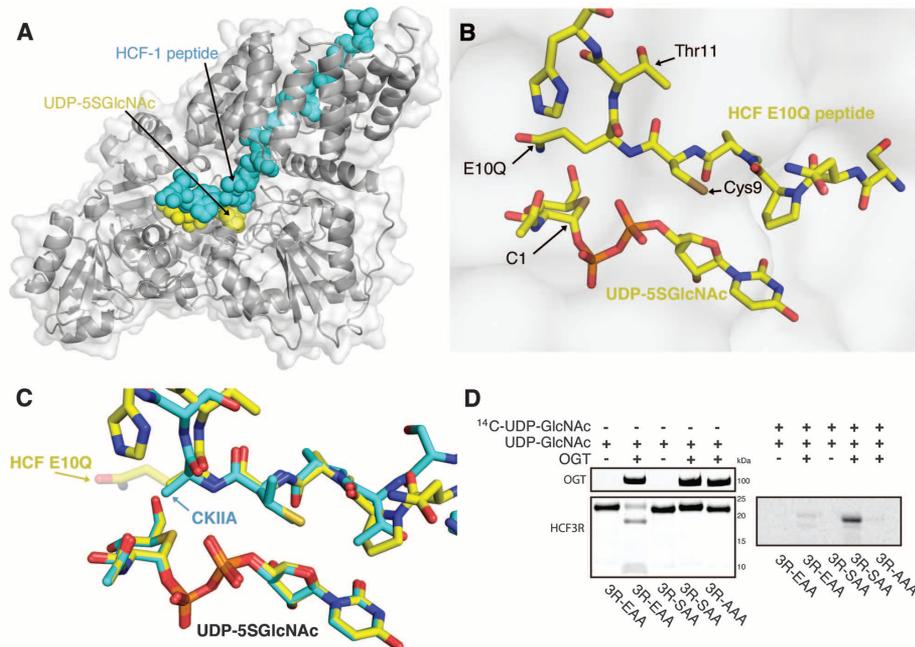


**Fig. 3. The threonine-rich region of the HCF-1<sub>PRO</sub> repeat binds in the channel formed by the TPR domain of OGT.** (A) Overall structure of the OGT:UDP:HCF-1<sub>11-26</sub> peptide complex. A 16-residue peptide comprising the threonine-rich region of

HCF-1<sub>PRO</sub> repeat 2 (THETGTTNTATTATSN) was cocrystallized with UDP and a previously described (13) N-terminally truncated OGT construct (hOGT<sub>4.5</sub>) and refined to 1.8 Å. The OGT catalytic domain (red) and TPR domain (gray), along with the HCF-1<sub>11-26</sub> peptide (cyan) and UDP (yellow), are shown. (B)

Close-up view of OGT-peptide interactions. The electron density around the visible portion of HCF-1<sub>11-26</sub> is shown as an  $F_0 - F_c$  difference map contoured at  $3\sigma$ . The peptide is shown in cyan. OGT side chains that contact the peptide backbone are shown in yellow, and OGT side chains that contact HCF-1 peptide side chains are shown in magenta. (C) Schematic of contacts between OGT side chains and the threonine-rich region of the HCF-1<sub>PRO</sub> repeat 2 from the complex of OGT:UDP:HCF-1-E10A<sub>1-26</sub>. OGT side chains are numbered and colored as in (B). (D) Mutations in the TPR domain of OGT (5N-5A) inhibit cleavage. Cleavage and glycosylation of GST-HCF-1rep1 were assayed, as in Fig. 1C, in the absence (lane 1) or presence (lane 2) of WT OGT or the 5N-5A TPR-domain mutant in which Asn residues 322, 356, 390, 424, and 458 are mutated to alanine (lane 3). (E) OGT does not bind effectively in vitro to an HCF-1<sub>PRO</sub> repeat mutant containing mutations in the threonine-rich region (T17–22A). WT (lane 2) and mutant (lanes 1, 3, and 4) GST-HCF-1rep1 substrates were tested for OGT binding in the presence of UDP-GlcNAc using an OGT-directed pull-down assay. Antibodies to GST and T7 were used to detect GST-HCF-1rep1 (top) and OGT (bottom), respectively, by Western blotting. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

**Fig. 4. HCF-1 cleavage takes place in the glycosyltransferase active site of OGT.** (A) Overall structure of the OGT:UDP-5SGlcNAc:HCF-1-E10Q<sub>1-26</sub> complex. The HCF-1 peptide is shown as spheres in cyan, with the UDP-5SGlcNAc in yellow. (B) Close-up view of the two substrate analogs shown in yellow in the OGT active site. The entire cleavage region can be seen, and the C-E10Q-T residues are annotated. The anomeric carbon of UDP-5SGlcNAc is indicated (C1). (C) Overlay of the substrate analogs from the OGT:UDP-5SGlcNAc:HCF-1 peptide complex (yellow) and the previously reported (14) OGT:UDP-5SGlcNAc:CKIIA complex (cyan). CKIIA is a well-characterized OGT glycosylation substrate. The E10Q side chain of the HCF-1 peptide is shown as transparent just after the  $\beta$  carbon. (D) Mutating E10 to S in an HCF-1<sub>PRO</sub> repeat converts a cleavage substrate (HCF3R-EAA) into a glycosylation substrate (HCF3R-SAA), which is defective in cleavage. (Left) Cleavage products of HCF3R-EAA and HCF3R-SAA were separated by SDS-PAGE and stained with Coomassie Blue. (Right) Glycosylation of WT and mutant HCF3R substrates was carried out with <sup>14</sup>C-UDP-GlcNAc and analyzed by PAGE. Full gels are shown in fig. S10.



## References and Notes

1. L. Wells, K. Vosseller, G. W. Hart, *Science* **291**, 2376–2378 (2001).
2. L. K. Kreppel, M. A. Blomberg, G. W. Hart, *J. Biol. Chem.* **272**, 9308–9315 (1997).
3. J. A. Hanover, M. W. Krause, D. C. Love, *Nat. Rev. Mol. Cell Biol.* **13**, 312–321 (2012).
4. G. W. Hart, M. P. Housley, C. Slawson, *Nature* **446**, 1017–1022 (2007).
5. W. Yi *et al.*, *Science* **337**, 975–980 (2012).
6. R. Dentin, S. Hedrick, J. Xie, J. Yates 3rd, M. Montminy, *Science* **319**, 1402–1405 (2008).
7. J. Wsocka, W. Herr, *Trends Biochem. Sci.* **28**, 294–304 (2003).
8. Z. Zargar, S. Tyagi, *Transcription* **3**, 187–192 (2012).
9. S. Daou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2747–2752 (2011).
10. T. M. Kristie, J. L. Pomerantz, T. C. Twomey, S. A. Parent, P. A. Sharp, *J. Biol. Chem.* **270**, 4387–4394 (1995).
11. A. C. Wilson, M. G. Peterson, W. Herr, *Genes Dev.* **9**, 2445–2458 (1995).
12. F. Capotosti *et al.*, *Cell* **144**, 376–388 (2011).
13. M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz, S. Walker, *Nature* **469**, 564–567 (2011).
14. M. B. Lazarus *et al.*, *Nat. Chem. Biol.* **8**, 966–968 (2012).
15. M. Schimpl *et al.*, *Nat. Chem. Biol.* **8**, 969–974 (2012).
16. C. Martinez-Fleites *et al.*, *Nat. Struct. Mol. Biol.* **15**, 764–765 (2008).
17. A. J. Clarke *et al.*, *EMBO J.* **27**, 2780–2788 (2008).
18. J. Jiang, M. B. Lazarus, L. Pasquina, P. Sliz, S. Walker, *Nat. Chem. Biol.* **8**, 72–77 (2011).
19. T. M. Gloster *et al.*, *Nat. Chem. Biol.* **7**, 174–181 (2011).
20. M. Jínek *et al.*, *Nat. Struct. Mol. Biol.* **11**, 1001–1007 (2004).
21. W. A. Lubas, J. A. Hanover, *J. Biol. Chem.* **275**, 10983–10988 (2000).
22. D. R. Twardzik, A. Peterkofsky, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 274–277 (1972).
23. S. A. Khan, B. W. Erickson, *J. Biol. Chem.* **257**, 11864–11867 (1982).
24. L. K. Kreppel, G. W. Hart, *J. Biol. Chem.* **274**, 32015–32022 (1999).
25. S. P. Iyer, G. W. Hart, *J. Biol. Chem.* **278**, 24608–24616 (2003).

**Acknowledgments:** This work was supported by National Institutes of Health (NIH) grant R01GM094263 to S.W., Swiss National Science Foundation (SNSF) grant 31003A\_130829 and the University of Lausanne to W.H., and the Natural

Sciences and Engineering Research Council of Canada (NSERC) and Simon Fraser University to D.J.V. M.B.L. is a Merck Fellow of the Helen Hay Whitney Foundation. V.K. is an European Molecular Biology Organization (EMBO) long-term fellowship recipient. J.J. is an NSERC PGS-M and D3 fellowship recipient. D.J.V. is a recipient of an E. W. R. Steacie Memorial Fellowship and holds a Canada Research Chair in Chemical Glycobiology. We thank A. Heroux (B.N.L.) for assistance with x-ray data collection. We thank V. Zoete for helpful discussions and M. Bogyo for reading the manuscript. Coordinates and structure factors have been deposited in the Protein Data Bank with accession nos. 4N39, 4N3A, 4N3B, and 4N3C.

## Supplementary Materials

www.sciencemag.org/content/342/6163/1235/suppl/DC1  
Materials and Methods  
Figs. S1 to S11  
Tables S1 and S2  
References (26–39)

30 July 2013; accepted 28 October 2013  
10.1126/science.1243990

# Crosstalk Between Microtubule Attachment Complexes Ensures Accurate Chromosome Segregation

Dhanya K. Cheerambathur, Reto Gassmann,\* Brian Cook, Karen Oegema, Arshad Desai†

The microtubule-based mitotic spindle segregates chromosomes during cell division. During chromosome segregation, the centromeric regions of chromosomes build kinetochores that establish end-coupled attachments to spindle microtubules. Here, we used the *Caenorhabditis elegans* embryo as a model system to examine the crosstalk between two kinetochore protein complexes implicated in temporally distinct stages of attachment formation. The kinetochore dynein module, which mediates initial lateral microtubule capture, inhibited microtubule binding by the Ndc80 complex, which ultimately forms the end-coupled attachments that segregate chromosomes. The kinetochore dynein module directly regulated Ndc80, independently of phosphorylation by Aurora B kinase, and this regulation was required for accurate segregation. Thus, the conversion from initial dynein-mediated, lateral attachments to correctly oriented, Ndc80-mediated end-coupled attachments is actively controlled.

The four-subunit Ndc80 complex, whose Ndc80 subunit harbors direct microtubule-binding activity, is the central component of the microtubule end-coupled attachments that segregate chromosomes on mitotic spindles (1, 2). In metazoans, initial lateral capture of microtubules by dynein motors localized to kinetochores kinetically accelerates the formation of end-coupled attachments and ensures their correct orientation (3–7). How kinetochores transition from an initial laterally bound state to the final end-coupled state is unclear.

The kinetochore dynein module is composed of the three-subunit RZZ (Rod-Zw10-Zwilch) complex, which recruits dynein to kinetochores via Spindly (Fig. 1A) (7–9). Formation of end-coupled microtubule attachments was assessed during the first division of the *Caenorhabditis elegans* embryo by visualizing chromosome dynamics (Fig. 1B) and by quantifying the kinetics of spindle pole separation (Fig. 1C) (10, 11). Removal of Spindly (SPDL-1 in *C. elegans*) was nearly equivalent to removal of NDC-80 (Fig. 1, B and C). As expected (7, 12), the failure to establish end-coupled attachments resulting from SPDL-1 depletion was suppressed by codepletion of RZZ (Fig. 1B); the double inhibition exhibited only the mild delay in end-coupled attachment formation expected for loss of kinetochore dynein. Thus, RZZ inhibits the formation of NDC-80-mediated microtubule attachments, and relief of this inhibition requires SPDL-1.

Aurora B kinase inhibits microtubule binding of Ndc80 by phosphorylating its basic tail (13–15). To determine whether RZZ inhibits NDC-80 by promoting Aurora B-mediated phosphorylation of NDC-80, we created an RNA interference (RNAi)-resistant transgenic system (fig. S1, A and B) to replace endogenous NDC-80 with transgene-encoded NDC-80<sup>WT</sup> or a phosphorylation-resistant NDC-80<sup>4A</sup> mutant (fig. S1, C and D) (13). NDC-80<sup>WT</sup> and NDC-80<sup>4A</sup> mutants both rescued the severe chromosome segregation defect and embryonic lethality of NDC-80 depletion (Fig. 1D and fig. S1, E and F). Furthermore, NDC-80<sup>WT</sup> and NDC-80<sup>4A</sup> were equally sensitive to inhibition by RZZ—after SPDL-1 depletion, both exhibited severe chromosome segregation (Fig. 1D) and pole separation (fig. S1G) defects indicative of a failure to form end-coupled attachments. Thus, Aurora B-mediated phosphorylation of the NDC-80 tail is not required for RZZ to inhibit NDC-80-mediated end-coupled attachments.

We next tested if the RZZ complex directly interacts with NDC-80 and inhibits its microtubule-binding activity. ROD-1 has an N-terminal  $\beta$ -propeller domain that binds to Zwilch<sup>ZWL-1</sup> followed by an extended  $\alpha$  solenoid that binds Zw10<sup>CZW-1</sup> (Fig. 2A) (16). The N-terminal  $\beta$ -propeller domain of ROD-1 and the N-terminal microtubule-binding region of NDC-80 interacted in a yeast two-hybrid assay (Fig. 2B and fig. S2A). Deletion of the basic tail of NDC-80 abolished its interaction with ROD-1 without affecting binding to its Nuf2<sup>HIM-10</sup> partner; by contrast, mimicking an Aurora B-phosphorylated NDC-80 tail by mutation of four target sites to aspartic acid (4D) did not affect the ROD-1 interaction (Fig. 2B). Binding assays with a partially reconstituted RZZ complex composed of the N terminus of ROD-1 and Zwilch<sup>ZWL-1</sup> (termed R<sup>N</sup>Z) confirmed a direct tail-dependent interaction between R<sup>N</sup>Z and NDC-80 (Fig. 2B). To test if the ROD-1–NDC-80 interaction regulates NDC-80 microtubule binding, we used reconstituted *C. elegans* NDC-80 complex (fig. S2B)

Ludwig Institute for Cancer Research, Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093, USA.

\*Present address: Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal.

†Corresponding author. E-mail: abdesai@ucsd.edu



## Supplementary Materials for

### **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,\* Suzanne Walker\*

\*Corresponding author.

E-mail: Suzanne\_Walker@hms.harvard.edu (S.W.); winship.herr@unil.ch (W.H.)

Published 6 December 2013, *Science* **342**, 1235 (2013)

DOI: 10.1126/science.1243990

#### **This PDF file includes:**

Materials and Methods

Figs. S1 to S11

Tables S1 and S2

References

## Materials and Methods

### *Bacterial protein expression*

The GST–HCF-1rep1 fusion construct encoding HCF-1 amino acids 867-1071 was purified as described(12). This construct was mutated using QuickChange site-directed mutagenesis (Stratagene) to obtain GST–HCF-1rep1 with mutated HCF-1rep1<sub>PRO</sub>-repeats: E10A, E10D, E10Q, E10S and T17-22A. These constructs were subsequently synthesized and purified from *E. coli* BL21(DE3) using IMAC. Recombinant human ncOGT was likewise synthesized as a fusion protein in BL21(DE3) cells with N-terminal T7 and 8-His tags as described(26) with minor modifications: Starter cultures were grown at 37°C after diluting an overnight culture 1 to 250 in fresh kanamycin supplemented LB media. Cells were grown to an  $A_{600}$  of 1.1, after which they were transferred to 16°C for 30 min. They were then induced with 0.2 mM IPTG and grown overnight at 16°C for 16 h. The bacteria were pelleted and re-suspended in lysis buffer (20 mM Tris, pH 7.4, 250 mM NaCl, 0.5% NP40 supplemented with 1 mM PMSF and 0.1 mg/ml lysozyme) for 30 min on ice. The lysate was then sonicated 6 times for 10 s and centrifuged at 14,000 RPM in an SS-34 rotor for 1 h at 4°C. Imidazole was added to the supernatant to a final concentration of 40 mM before the lysate was incubated with Ni-NTA agarose superflow resin (Qiagen) overnight at 4°C, which was prewashed with TBS + 40 mM imidazole for nickel affinity purification. The flow through was removed and the resin washed with 3 column volumes of lysis buffer supplemented with 50 mM imidazole. The protein was then eluted with PBS supplemented with 250 mM imidazole. To concentrate the eluate, Amicon concentration tubes (Millipore) were used. The concentrated protein was dialyzed against PBS supplemented with 1 mM DTT overnight at 4°C. The concentrated dialyzed protein was supplemented with 1 mM DTT and stored at -80°C until needed. To generate mutants of OGT, we used site-directed mutagenesis (Agilent technologies) to create catalytic domain (K842M, H498A, H558A) and the multi-asparagine (N322A, N356A, N390A, N424A, N458A) mutants. The mutants were verified by sequencing.

### *HCF-1rep1 cleavage and glycosylation assay*

For a typical *in vitro* cleavage/glycosylation assay, bacterially purified GST–HCF-1rep1 (2 µg) was incubated with purified OGT (1 µg) in the presence or absence of UDP-GlcNAc (1 mM, unless otherwise noted) as described earlier(9) at 37°C overnight. HCF-1rep1 cleavage and glycosylation were examined by performing SDS-PAGE followed by western blotting to detect HCF-1rep1 cleavage (anti-GST) and HCF-1 glycosylation (anti-O-GlcNAc).

### *CKII glycosylation assay*

CKII glycosylation assay was performed by incubating CKII (NEB, 2500 U) with 1 µg of recombinant human ncOGT (WT or mutant) in the presence of 1 mM UDP-GlcNAc. The reaction was incubated at 37°C for 90 min, and the extent of CKII glycosylation was assessed by western blotting using anti-O-GlcNAc antibody.

## ***Antibodies***

Antibodies used to detect GST–HCF-1rep1, O-GlcNAc and OGT were as follows: rabbit polyclonal anti-GST (1-109) (Santa Cruz), mouse monoclonal (RL2) anti-O-GlcNAc (Abcam) and mouse monoclonal anti-T7 (Novagen). Casein kinase II $\alpha$  antibody (1AD9) was obtained from Santa Cruz biotech.

## ***In vitro HCF-1–OGT binding assay***

20  $\mu$ l of a slurry of anti-T7 antibody-conjugated beads (polyclonal from goat, Abcam) were incubated with PBS containing 5% (w/v) bovine serum albumin (BSA) for 1 h at 4°C to decrease non-specific binding of GST–HCF-1rep1 to the agarose beads. Subsequently, the beads were washed extensively in PBS followed by 0.5 % NP40 buffer as described(27). For the OGT-pulldown, 2.5  $\mu$ g GST–HCF-1rep1 and 5  $\mu$ g OGT were pre-incubated in the presence of 5 mM DTT and 1 mM UDP-GlcNAc in a rotating incubator for 1 h at 20°C. After incubation, 9% of the reaction was taken as an input-control for the OGT-pulldown. The washed anti-T7 agarose beads were added to the reaction, and the volume was increased to 500  $\mu$ l using NP40 buffer and the suspension incubated for 1 h at 20°C. The beads were washed in NP40 buffer three times for 5 min at room temperature, and the OGT/GST–HCF-1rep1 complexes were eluted by boiling for 5 min in 20  $\mu$ l of 2x SDS sample buffer.

## ***Purification and crystallization***

Human OGT<sub>4.5</sub> was prepared as previously described(13). All peptides for crystallization were purchased (Biomatik,  $\geq$ 95% purity, HPLC). A summary of the peptides and their sequences is listed in table S1 below. All complexes were prepared by incubating OGT<sub>4.5</sub> at 7 mg/ml with nucleotide-sugar (1 mM) and peptide (3 mM). UDP-5SGlcNAc was synthesized as previously described(14, 19). All crystals were grown with the hanging drop method by combining 2  $\mu$ l protein complex with 1  $\mu$ l reservoir solution. Crystals for the C-terminal fragment of HCF-1 were grown with reservoir consisting of 1.5 M potassium sodium tartrate and 0.1 M Tris pH 8.5. The OGT:UDP:HCF-1-E10A complex crystals were obtained with reservoir consisting of 0.88 M potassium phosphate dibasic and 0.72 M potassium phosphate monobasic. Crystals of the complex containing OGT and UDP-5SGlcNAc and the E10Q 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 of the complex of OGT with UDP-GlcNAc and the wild-type peptide were grown in 0.86 M potassium phosphate dibasic and 0.86 M sodium phosphate monobasic. Crystals were grown at room temperature and cryoprotected in solutions consisting of the reservoir solution plus 28% glycerol or xylitol and flash frozen in liquid nitrogen.

## ***Data collection, structure determination, and refinement***

All data were collected at NSLS beamline X29 or X25 at Brookhaven National Laboratory. Data was processed using iMosflm(28) and scaled with Scala(29). The previously determined OGT-UDP-CKII ternary complex (3PE4) was used as a search model for molecular replacement in this work, using Phaser(30). The models were subsequently refined using Phenix(31), with 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(32, 33)), with interspersed manual adjustments using Coot(34). Geometric restraints for UDP-GlcNAc or UDP-5SGlcNAc were generated using Phenix Elbow(35), and these restraints were used throughout refinement. All structural figures were made with Pymol(36). Crystallization software installation support was provided by SBGrid(37).

### ***HCF3R and OGT constructs***

The HCF3R construct (residues 1006-1130) was amplified from a cDNA library (Promega) and cloned into a modified N-terminal 8-His pET47b vector (EMD). Full length OGT constructs were purified as previously described(13). To generate mutants of HCF3R, an HCF3R-AAA construct (containing alanines in all three E10 sites) was synthesized by overlap extension PCR to get around the problems of repeating nucleotide sequences, using primers designed with DNAworks(38). The HCF3R-AAA insert was also cloned into an 8-His pET47b vector. Mutations were then made in this vector using QuickChange mutagenesis. The constructs were subsequently expressed and purified from *E. coli* using IMAC.

The DNA sequence of the HCF3R-AAA insert is:

```
CCGGGTACCGTTACCCTGGTTTTGCTCTAACCCGCCGTGCGCTACCCACGAAACCGGTACCACCAA  
CACCGCTACCACCACCGTTGTTGCTAACCTGGGTGGTCACCCGCAGCCGACCCAGGTTTCAGTTTCGTTTTGCG  
ACCGTCAGGAAGCTGCTGCTTCTCTGGTTACCTCTACCGTTGGTCAGCAGAATGGTAGCGTGGTCCGAGTC  
TGTTTCGAATCCTCCCTGTGCAACACACGAGACGGGCACGACGAATACGGCCACAACGGCCACCTCCAACAT  
GGCTGGACAACATGGATGTTCTAATCCTCCTTGTGCTACTCATGAAACTGGAACTACTAATACTGCTACTA  
CTGCTATGTCTTCTGTTGGAGCTAATTAA
```

### ***HCF3R cleavage assays***

Purified His-HCF3R constructs were incubated in 20  $\mu$ l reactions with PBS pH 7.4 buffer and 1 mM THP for 5 h at 37°C in the presence or absence of OGT constructs and different reagents as specified in the figures. For reactions described in Figure 2, the concentrations were: 200  $\mu$ M HCF3R; 2.5 $\mu$ M wild-type or K842A ncOGT; 100 U/ml alkaline phosphatase (NEB) supplemented with 20 mM MgCl<sub>2</sub>; 1 mM UDP-GlcNAc, UDP-5SGlcNAc, and UDP; 3.75  $\mu$ M BZX inhibitor (Compound 2) (18). For Figure 4 and for the cysteine mutants, reagent concentrations were: 100  $\mu$ M HCF3R, 2.67  $\mu$ M ncOGT and 0.1 U (100 U/mL) calf-alkaline phosphatase (NEB) in 10  $\mu$ l TBS pH 7.4 with 20 mM MgCl<sub>2</sub> and 1 mM UDP-GlcNAc except for the <sup>14</sup>C cleavage, which used a 1 mM mixture of UDP-GlcNAc with <sup>14</sup>C/<sup>12</sup>C (1:3), as specified. After incubation at 37°C, the reaction were quenched by 1:1 addition of 2x SDS loading buffer and the products were separated on a 4-20% Tris-Glycine SDS-PAGE gel and detected either by Coomassie staining or autoradiography.

### ***LC-MS detection of cleavage products***

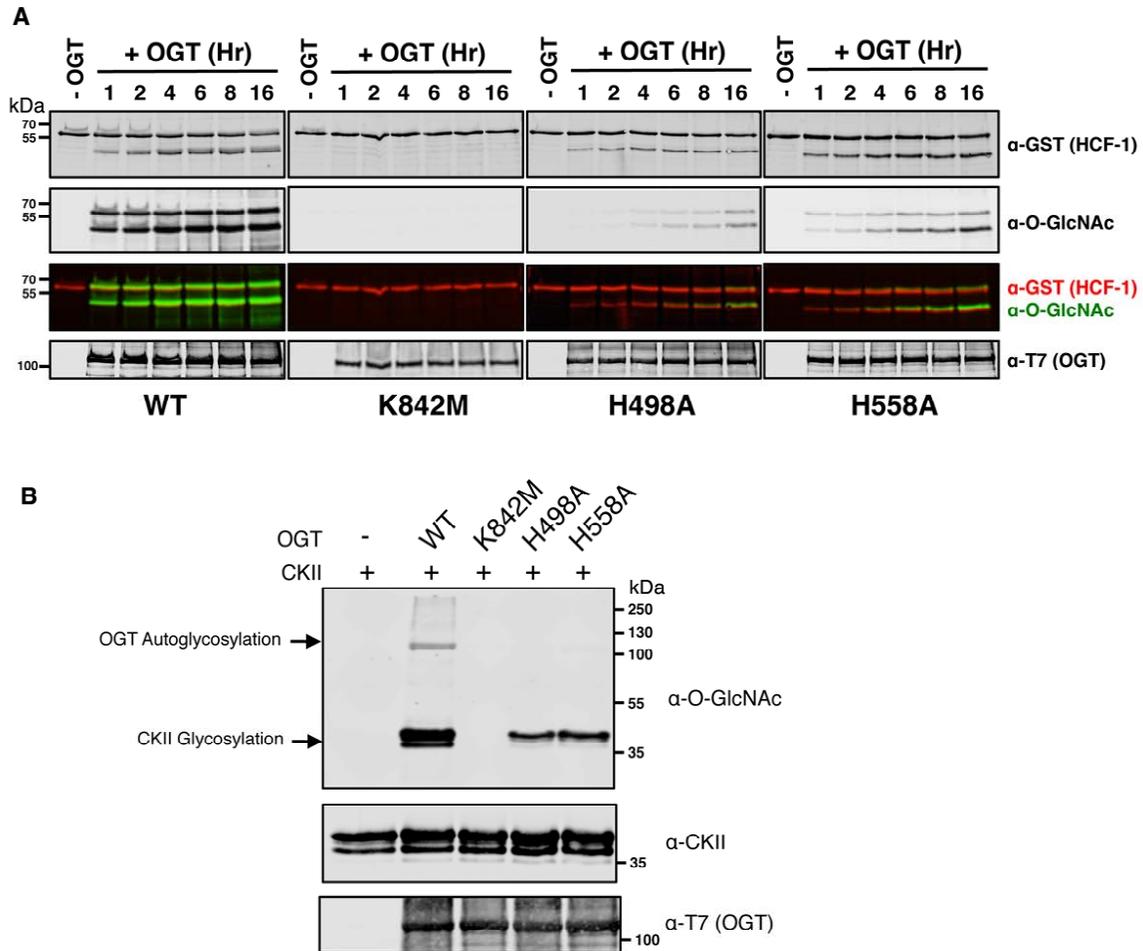
Purified His-HCF3R (100  $\mu$ M) was incubated with 2.5  $\mu$ M ncOGT and 1 mM UDP-GlcNAc using the same reaction conditions specified above, except that incubation time was only 1 h. The reaction products were separated using an Agilent 6520 LC-QTOF system equipped with a Phenomenex Gemini-NX C18 column (5  $\mu$ m, 110 Å, 50 x

2.00 mm), pre-equilibrated with 6% acetonitrile and 0.1% aqueous formic acid at a flow rate of 0.4 ml/min with a linear gradient of 6–60% buffer B (90% acetonitrile and 0.1% aqueous formic acid) over 20 min after desalting for 10 min with 6% buffer B. The data was analyzed using Agilent MassHunter software.

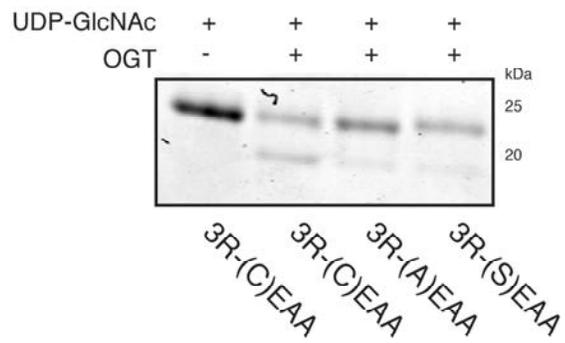
***Pyroglutamate aminopeptidase treatment of HCF-1 cleavage products***

Purified HCF3R-EAA (100  $\mu$ M) was incubated in reaction buffer with alkaline phosphatase at 37°C overnight as described above to generate cleaved products. Pfu pyroglutamate aminopeptidase (Takara Bio) was then added to the reaction at 12.5 U/L and the incubation was continued at 37°C for another 5 h to cleave the pyroglutamate. The reaction products were tested on Q-TOF LC-MS as described above.

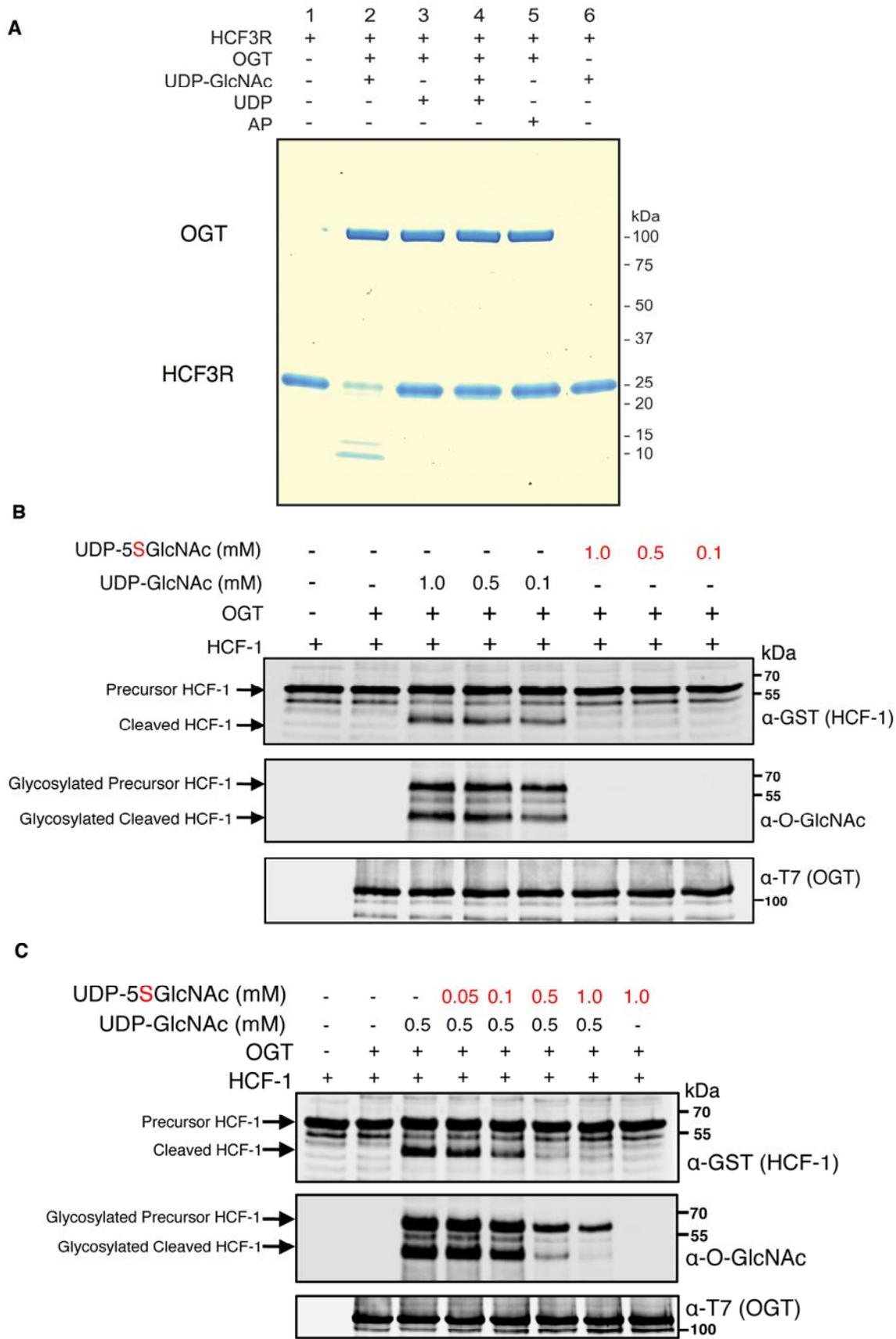
## Supplementary Figures



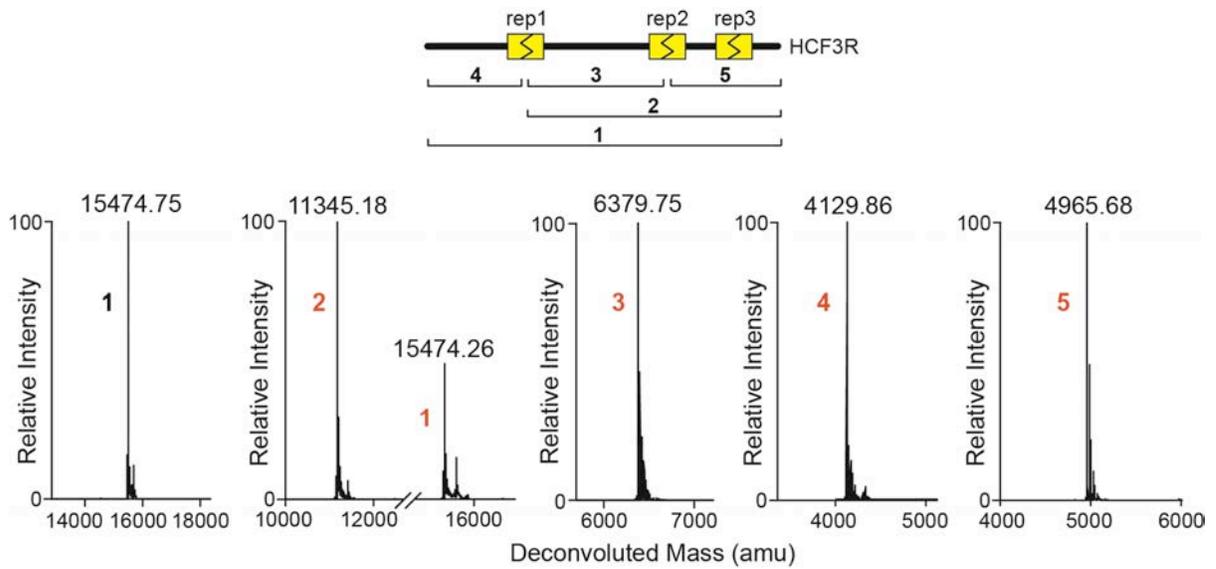
**Fig. S1. Cleavage vs. Glycosylation by OGT.** (A) GST-HCF-1rep1 was incubated with WT human OGT or its mutants (K842M, H498A, H558A) in the presence of UDP-GlcNAc for indicated time periods. HCF-1rep1 cleavage and glycosylation were examined by western blotting as described in Materials and Methods. (B) CKII was incubated in the presence of WT OGT or the specified mutants and UDP-GlcNAc, as described in Materials and Methods. The glycosylation of CKII was examined by western blotting using O-GlcNAc antibody (RL2).



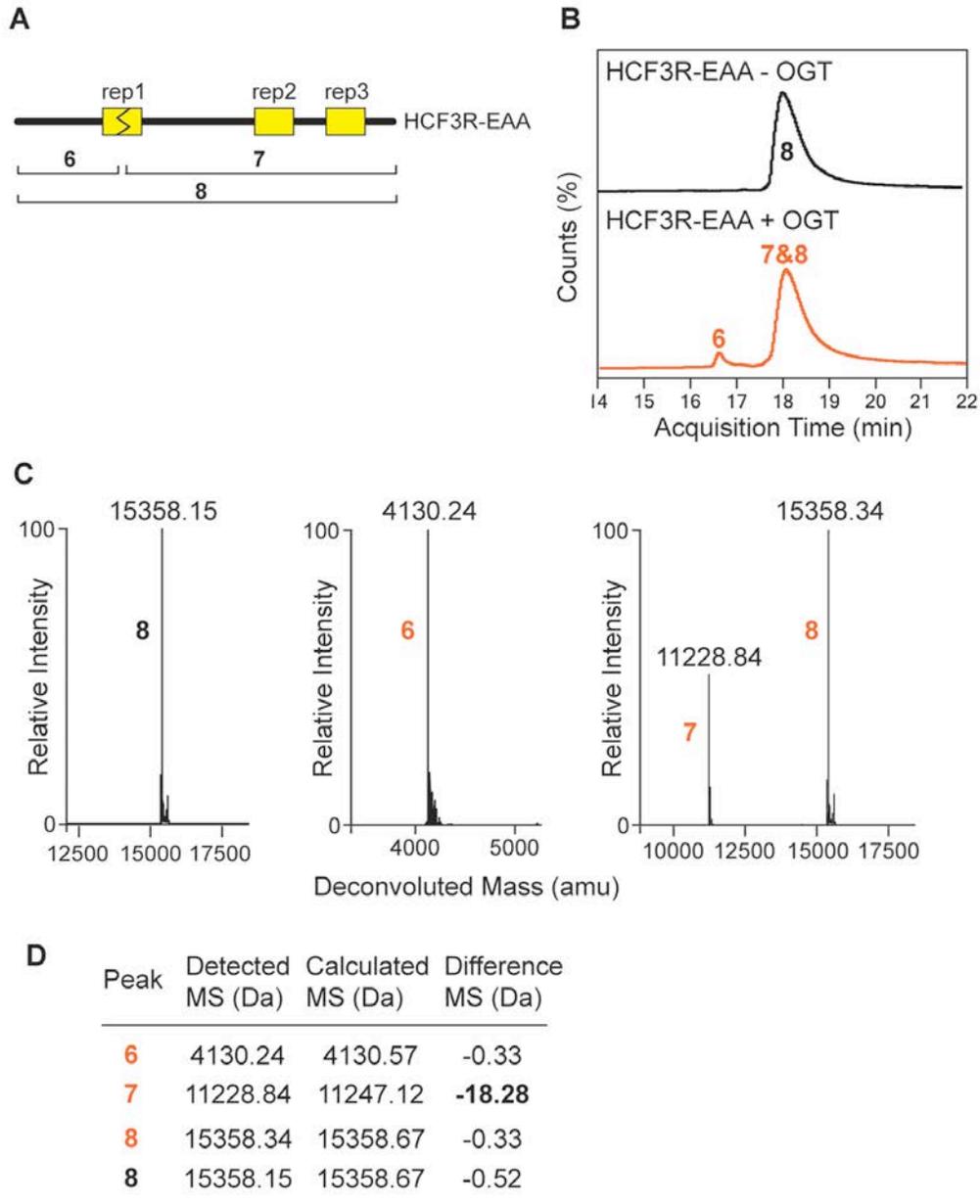
**Fig. S2. The C9 cysteine is not essential for HCF-1 pro-repeat cleavage.** We tested serine and alanine substitutions of the cysteine at the cleavage site in the HCF3R-EAA construct. Both substitutions weakened but did not abolish the cleavage, consistent with previous work on the cysteine to alanine substitution(12).



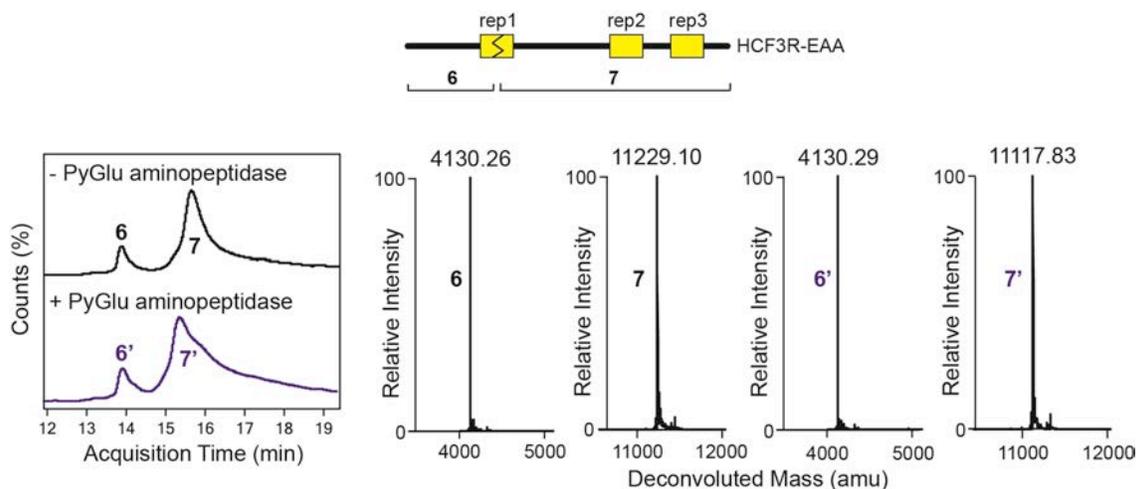
**Fig. S3. HCF-1rep1 cleavage requires UDP-GlcNAc.** (A) HCF3R was incubated with OGT (lanes 2-5) in the presence of UDP-GlcNAc (lane 2), UDP (lane 3), both (lane 4), or neither of them (lane 5). Alkaline phosphatase (AP) was added to reaction 5, as indicated. In reaction 6, HCF3R was incubated with UDP-GlcNAc alone. Cleavage products were separated by SDS-PAGE and stained with Coomassie Blue. (B) GST-HCF-1rep1 (precursor HCF-1) was incubated with WT human OGT in the absence or presence of increasing concentrations of UDP-GlcNAc or its inhibitory analogue UDP-5SGlcNAc. The reaction was incubated for 16 h at 37°C and HCF-1rep1 cleavage and glycosylation were examined by western blotting as described in Materials and Methods. HCF-1rep1 cleavage and glycosylation were observed only when UDP-GlcNAc was present in the reaction. (C) A competition assay between UDP-GlcNAc and its analogue UDP-5SGlcNAc was performed by incubating GST-HCF-1rep1 (precursor HCF-1) with WT human OGT in the presence of UDP-GlcNAc alone (0.5 mM) or in combination with varying concentrations of UDP-5SGlcNAc (0.05-1.0 mM). The reaction was incubated for 16 h at 37°C and HCF-1 cleavage and glycosylation were examined by western blotting as described in Materials and Methods. The presence of UDP-5SGlcNAc inhibited HCF-1 proteolysis.



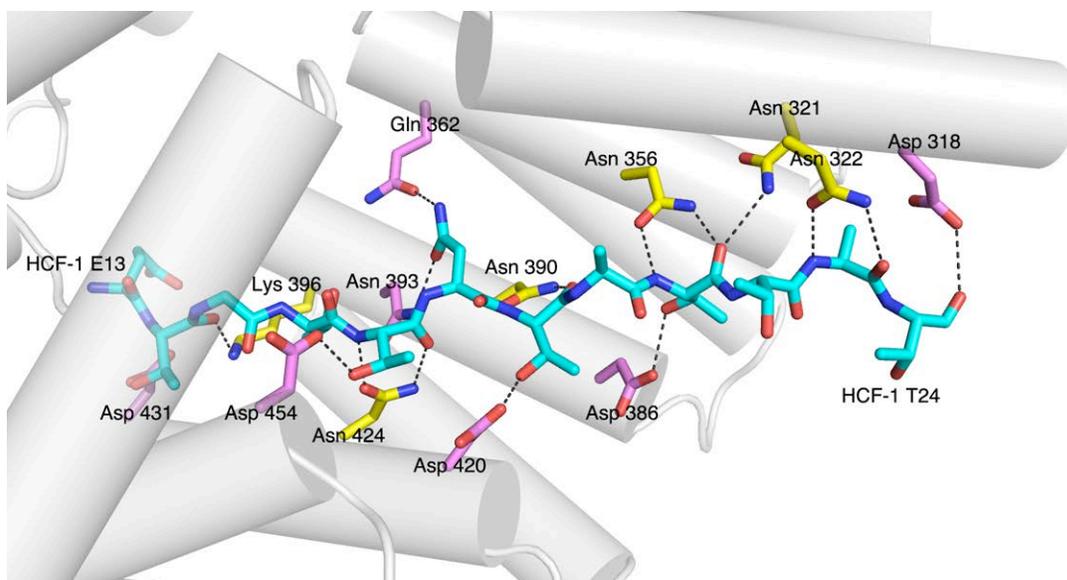
**Fig. S4. LC-MS detection of HCF3R cleavage products.** OGT was incubated with either HCF3R and UDP-GlcNAc (red) or with HCF3R alone as a control (black). The detected MS peaks are summarized in the table in Figure 2B.



**Fig. S5. LC-MS analysis of HCF3R-EAA cleavage.** (A) HCF3R-EAA construct and the numbering of detected mass peaks. (B) Total Ion Chromatogram (TIC) of the cleavage products from the incubation of HCF3R-EAA, OGT, and UDP-GlcNAc (red), HCF3R-EAA alone as a control (black). (C) The detected MS peaks for HCF3R-EAA cleavage. (D) Table summarizing the detected mass peaks from HCF3R-EAA cleavage.

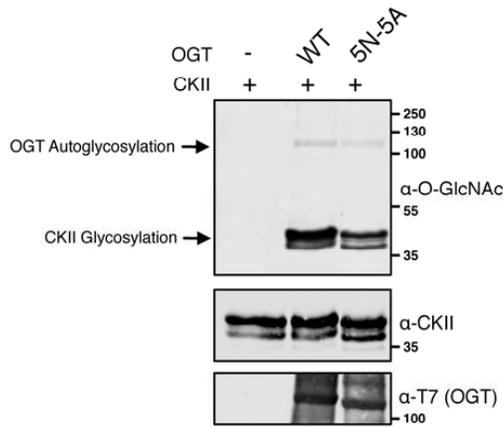


**Fig. S6. Treatment of HCF3R-EAA cleavage products with PyGlu aminopeptidase removes a pyroglutamate from the C-terminal fragment.** The HCF3R-EAA cleavage products in the absence of PyGlu aminopeptidase are shown in black (6 and 7) and those in the presence of PyGlu aminopeptidase are shown in purple (6' and 7'). The MS data are summarized in the table in Figure 2C. Loss of 111 Da from the C-terminal fragment (7) to produce 7' confirms formation of an N-terminal PyGlu during cleavage.

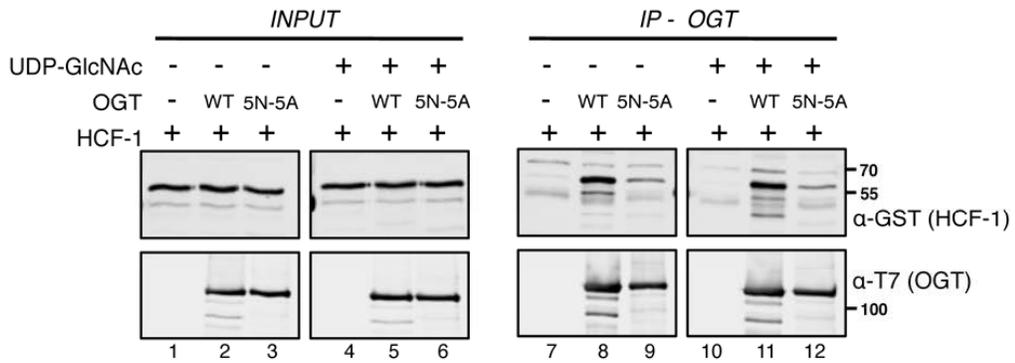


**Fig. S7. Interactions between the TPR domain of OGT with the HCF-1<sub>PRO</sub> repeat E10A peptide.** The 26-residue full HCF-1 repeat containing the E10A mutation was crystallized with hOGT<sub>4.5</sub> and UDP. Compared with the structure shown in Figures 3A and 3B, two additional residues (23 and 24) are now visible, both of which make contacts with the TPRs of OGT. A schematic of the interactions in this structure is shown in Figure 3C. The HCF-1 peptide is shown in cyan, with the N-terminus of the peptide labeled as residue E13 and the C-terminus as residue T24. The OGT TPRs are shown as gray cylindrical helices. OGT side chains that make contacts to the HCF-1 peptide backbone are colored in yellow; and OGT side chains that contact HCF-1 side chains are colored in violet.

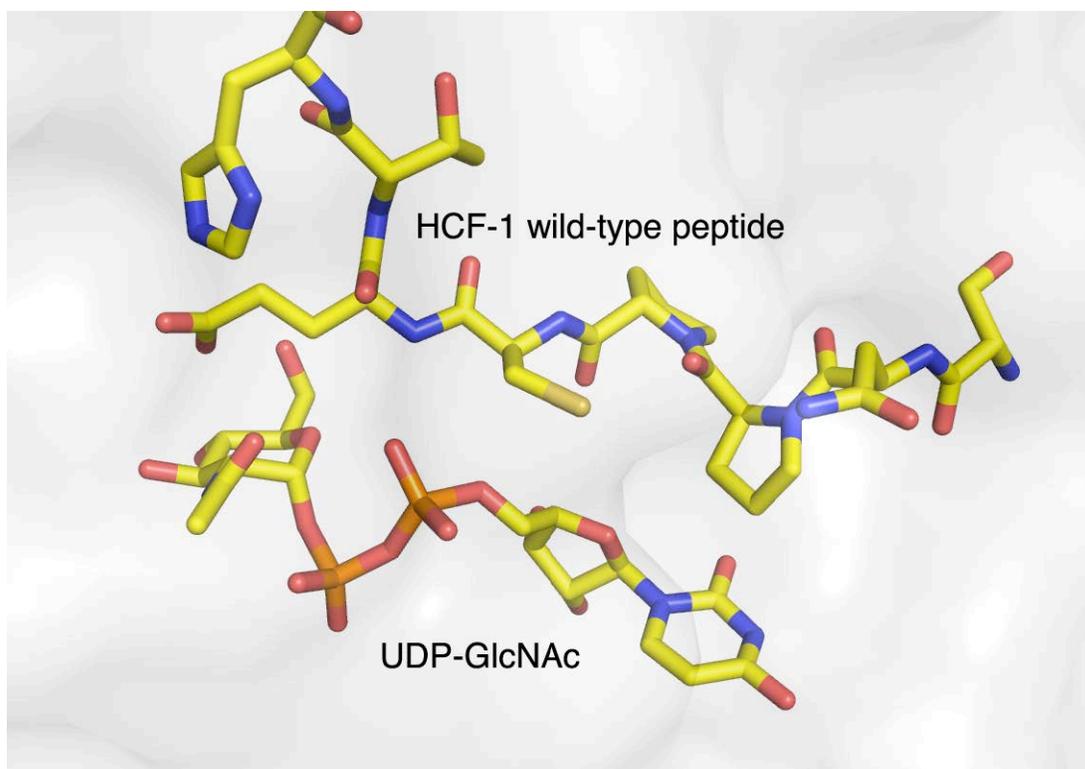
**A**



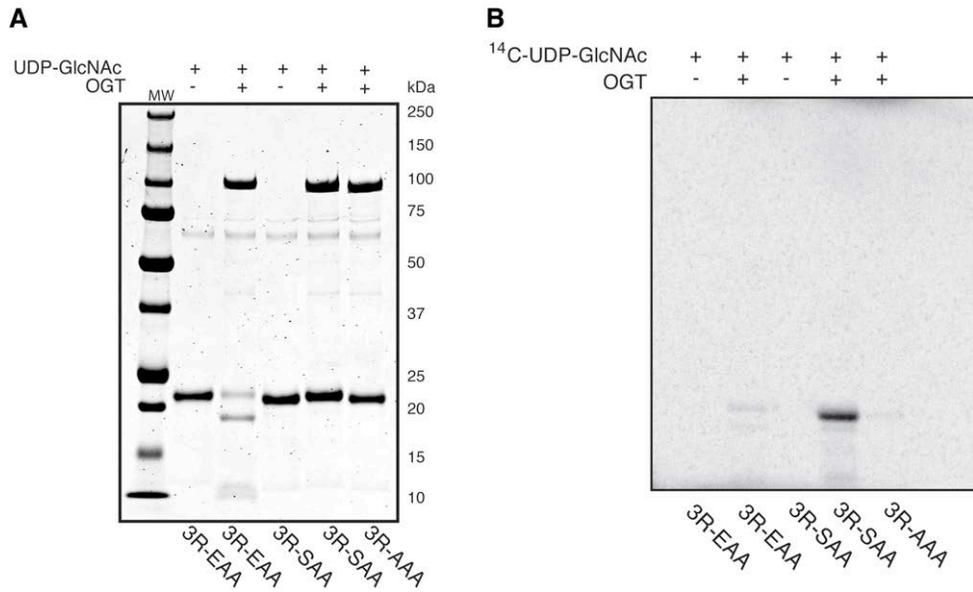
**B**



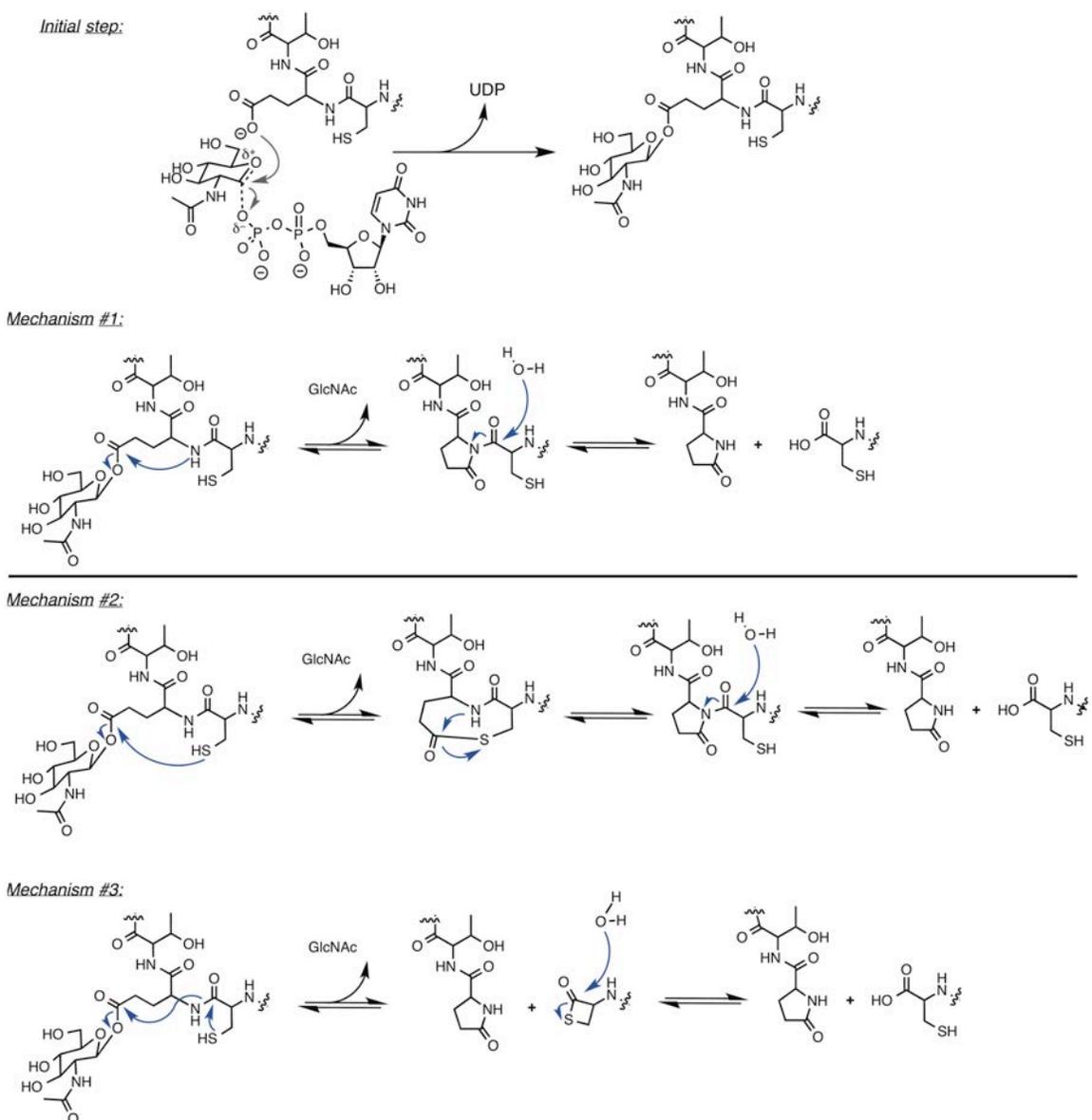
**Fig. S8. The 5N-5A OGT mutant is defective in binding to HCF-1rep1. (A)** Glycosylation of CKII is minimally affected by the 5N-5A mutations in the TPR domain of OGT. CKII (2500 U) was incubated in the presence of WT human OGT or 5N-5A OGT and CKII glycosylation was examined as described above. **(B)** *In vitro* binding assay to study GST–HCF-1rep1 interaction with WT human OGT and 5N-5A OGT was performed as described in Materials and Methods.



**Fig. S9. Crystal structure of the wild-type HCF-1<sub>PRO</sub> repeat peptide bound to OGT with UDP-GlcNAc.** The structure for the native complex is nearly identical to the ternary complex of OGT with UDP-5SGlcNAc and the HCF-1<sub>PRO</sub> repeat E10Q peptide (0.185 Å rmsd). We note that the density for the GlcNAc is poorer in this complex, which may be related to the reactivity of UDP-GlcNAc.



**Fig. S10. A single amino acid substitution converts HCF-1 from a cleavage substrate into a glycosylation substrate of OGT.** Full-size versions of the gels shown in Figure 4D. (A) Coomassie stained SDS-PAGE gel of HCF3R constructs. Only the EAA with UDP-GlcNAc is cleaved. (B) Autoradiogram of HCF3R constructs incubated with OGT and <sup>14</sup>C-UDP-GlcNAc. Only HCF3R-SAA is glycosylated by OGT.



**Fig. S11. Possible mechanisms for OGT cleavage of HCF-1.** We speculate that cleavage occurs after formation of a glutamyl-sugar intermediate. After this initial step there are several ways this could proceed to cleavage. We outline 3 possible mechanisms here. Mechanisms 2 and 3 involve the C9 cysteine of the HCF-1<sub>PRO</sub> repeat through the formation of thioester intermediates. Other biological mechanisms with peptide backbone cleavage involve cysteines or pyroglutamate(23, 39). However, the fact that the C9 cysteine residue is not essential for cleavage of HCF-1 (fig. S2 and ref. 12) supports the first mechanism, although other mechanisms are possible.

## Supplementary Tables

**Table S1.** Peptides used for crystallization in this study.

<b>Peptide 1</b>	
Name	HCF-1 <sub>11-26</sub>
Sequence	THETGTTNTATTATSN

<b>Peptide 2</b>	
Name	HCF-1-E10A <sub>1-26</sub>
Sequence	VRVCSNPPCATHETGTTNTATTATSN

<b>Peptide 3</b>	
Name	HCF-1-E10Q <sub>1-26</sub>
Sequence	VRVCSNPPCQTHETGTTNTATTATSN

<b>Peptide 4</b>	
Name	HCF-1 <sub>1-26</sub>
Sequence	VRVCSNPPCETHETGTTNTATTATSN

**Table S2.** X-ray Data collection and refinement statistics for crystal structures.

	OGT:UDP: HCF-1 <sub>11-26</sub> Complex	OGT:UDP: HCF-1-E10A <sub>1-26</sub> Complex	OGT:UDP- 5SGlcNAc: HCF-1- E10Q <sub>1-26</sub> Complex	OGT:UDP-GlcNAc: HCF-1 <sub>1-26</sub> Complex
<b>Data collection</b>				
Space group	P3221	P6122	P6122	P6122
Cell dimensions				
<i>a, b, c</i> (Å)	101.17 101.17 131.73	98.93 98.93 367.02	98.91 98.91 364.93	98.88 98.88 365.93
<i>a, b, g</i> (°)	90.00 90.00 120.00	90.00 90.00 120.00	90.00 90.00 120.00	90.00 90.00 120.00
Resolution (Å)	35.00 - 1.76 (1.86 - 1.76)	47.76 - 1.88 (1.98 - 1.88)	55.55 - 2.17 (2.29 - 2.17)	77.56 - 2.55 (2.69 - 2.55)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.093 (0.628)	0.087 (0.599)	0.123 (0.762)	0.151 (0.807)
<i>I</i> / <i>sI</i>	9.1 (2.1)	7.3 (2.1)	11.5 (2.0)	6.0 (2.1)
Completeness (%)	99.6 (100.0)	95.0 (96.1)	99.9 (99.5)	96.1 (92.7)
Redundancy	5.1 (5.2)	3.8 (3.8)	7.6 (4.7)	3.0 (3.1)
<b>Refinement</b>				
Resolution (Å)	33.12 - 1.76	45.86 - 1.88	49.59 - 2.17	62.51-2.55
No. reflections	77299	82481	56713	33941
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1872 / 0.2158	0.2158 / 0.2239	0.1922 / 0.2228	0.1837 / 0.2261
No. atoms				
Protein	5598	5610	5670	5663
Ligand/ion	25	25	39	39
Water	551	148	124	102
<i>B</i> -factors				
Protein	25.89	38.85	50.67	50.04
Ligand/ion	17.23	23.22	38.27	58.34
Water	35.53	33.51	46.78	45.21
R.m.s. deviations				
Bond lengths (Å)	0.005	0.002	0.003	0.003
Bond angles (°)	0.899	0.72	0.73	0.749

\*Values in parentheses are for highest-resolution shell.

## References

1. L. Wells, K. Vosseller, G. W. Hart, Glycosylation of nucleocytoplasmic proteins: Signal transduction and O-GlcNAc. *Science* **291**, 2376–2378 (2001). [doi:10.1126/science.1058714](https://doi.org/10.1126/science.1058714) [Medline](#)
2. L. K. Kreppel, M. A. Blomberg, G. W. Hart, Dynamic glycosylation of nuclear and cytosolic proteins: Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9308–9315 (1997). [doi:10.1074/jbc.272.14.9308](https://doi.org/10.1074/jbc.272.14.9308) [Medline](#)
3. J. A. Hanover, M. W. Krause, D. C. Love, Bittersweet memories: Linking metabolism to epigenetics through O-GlcNAcylation. *Nat. Rev. Mol. Cell Biol.* **13**, 312–321 (2012). [doi:10.1038/nrm3334](https://doi.org/10.1038/nrm3334) [Medline](#)
4. G. W. Hart, M. P. Housley, C. Slawson, Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017–1022 (2007). [doi:10.1038/nature05815](https://doi.org/10.1038/nature05815) [Medline](#)
5. W. Yi, P. M. Clark, D. E. Mason, M. C. Keenan, C. Hill, W. A. Goddard, 3rd, E. C. Peters, E. M. Driggers, L. C. Hsieh-Wilson, Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science* **337**, 975–980 (2012). [doi:10.1126/science.1222278](https://doi.org/10.1126/science.1222278) [Medline](#)
6. R. Dentin, S. Hedrick, J. Xie, J. Yates, 3rd, M. Montminy, Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* **319**, 1402–1405 (2008). [doi:10.1126/science.1151363](https://doi.org/10.1126/science.1151363) [Medline](#)
7. J. Wysocka, W. Herr, The herpes simplex virus VP16-induced complex: The makings of a regulatory switch. *Trends Biochem. Sci.* **28**, 294–304 (2003). [doi:10.1016/S0968-0004\(03\)00088-4](https://doi.org/10.1016/S0968-0004(03)00088-4) [Medline](#)
8. Z. Zargar, S. Tyagi, Role of host cell factor-1 in cell cycle regulation. *Transcription* **3**, 187–192 (2012). [doi:10.4161/trns.20711](https://doi.org/10.4161/trns.20711) [Medline](#)
9. S. Daou, N. Mashtalir, I. Hammond-Martel, H. Pak, H. Yu, G. Sui, J. L. Vogel, T. M. Kristie, B. Affar, Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2747–2752 (2011). [doi:10.1073/pnas.1013822108](https://doi.org/10.1073/pnas.1013822108) [Medline](#)
10. T. M. Kristie, J. L. Pomerantz, T. C. Twomey, S. A. Parent, P. A. Sharp, The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. *J. Biol. Chem.* **270**, 4387–4394 (1995). [doi:10.1074/jbc.270.9.4387](https://doi.org/10.1074/jbc.270.9.4387) [Medline](#)
11. A. C. Wilson, M. G. Peterson, W. Herr, The HCF repeat is an unusual proteolytic cleavage signal. *Genes Dev.* **9**, 2445–2458 (1995). [doi:10.1101/gad.9.20.2445](https://doi.org/10.1101/gad.9.20.2445) [Medline](#)
12. F. Capotosti, S. Guernier, F. Lammers, P. Waridel, Y. Cai, J. Jin, J. W. Conaway, R. C. Conaway, W. Herr, O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144**, 376–388 (2011). [doi:10.1016/j.cell.2010.12.030](https://doi.org/10.1016/j.cell.2010.12.030) [Medline](#)
13. M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz, S. Walker, Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–567 (2011). [doi:10.1038/nature09638](https://doi.org/10.1038/nature09638) [Medline](#)

14. M. B. Lazarus, J. Jiang, T. M. Gloster, W. F. Zandberg, G. E. Whitworth, D. J. Vocadlo, S. Walker, Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nat. Chem. Biol.* **8**, 966–968 (2012). [doi:10.1038/nchembio.1109](https://doi.org/10.1038/nchembio.1109) [Medline](#)
15. 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, D. M. van Aalten, O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nat. Chem. Biol.* **8**, 969–974 (2012). [doi:10.1038/nchembio.1108](https://doi.org/10.1038/nchembio.1108) [Medline](#)
16. C. Martinez-Fleites, M. S. Macauley, Y. He, D. L. Shen, D. J. Vocadlo, G. J. Davies, Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. *Nat. Struct. Mol. Biol.* **15**, 764–765 (2008). [doi:10.1038/nsmb.1443](https://doi.org/10.1038/nsmb.1443) [Medline](#)
17. A. J. Clarke, R. Hurtado-Guerrero, S. Pathak, A. W. Schüttelkopf, V. Borodkin, S. M. Shepherd, A. F. Ibrahim, D. M. van Aalten, Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J.* **27**, 2780–2788 (2008). [doi:10.1038/emboj.2008.186](https://doi.org/10.1038/emboj.2008.186) [Medline](#)
18. J. Jiang, M. B. Lazarus, L. Pasquina, P. Sliz, S. Walker, A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. *Nat. Chem. Biol.* **8**, 72–77 (2011). [doi:10.1038/nchembio.711](https://doi.org/10.1038/nchembio.711) [Medline](#)
19. T. M. Gloster, W. F. Zandberg, J. E. Heinonen, D. L. Shen, L. Deng, D. J. Vocadlo, Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. *Nat. Chem. Biol.* **7**, 174–181 (2011). [doi:10.1038/nchembio.520](https://doi.org/10.1038/nchembio.520) [Medline](#)
20. M. Jínek, J. Rehwinkel, B. D. Lazarus, E. Izaurralde, J. A. Hanover, E. Conti, The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat. Struct. Mol. Biol.* **11**, 1001–1007 (2004). [doi:10.1038/nsmb833](https://doi.org/10.1038/nsmb833) [Medline](#)
21. W. A. Lubas, J. A. Hanover, Functional expression of O-linked GlcNAc transferase: Domain structure and substrate specificity. *J. Biol. Chem.* **275**, 10983–10988 (2000). [doi:10.1074/jbc.275.15.10983](https://doi.org/10.1074/jbc.275.15.10983) [Medline](#)
22. D. R. Twardzik, A. Peterkofsky, Glutamic acid as a precursor to N-terminal pyroglutamic acid in mouse plasmacytoma protein (protein synthesis-initiation-immunoglobulins-pyrrolidone carboxylic acid). *Proc. Natl. Acad. Sci. U.S.A.* **69**, 274–277 (1972). [doi:10.1073/pnas.69.1.274](https://doi.org/10.1073/pnas.69.1.274) [Medline](#)
23. S. A. Khan, B. W. Erickson, An equilibrium model of the metastable binding sites of alpha 2-macroglobulin and complement proteins C3 and C4. *J. Biol. Chem.* **257**, 11864–11867 (1982). [Medline](#)
24. L. K. Kreppel, G. W. Hart, Regulation of a cytosolic and nuclear O-GlcNAc transferase: Role of the tetratricopeptide repeats. *J. Biol. Chem.* **274**, 32015–32022 (1999). [doi:10.1074/jbc.274.45.32015](https://doi.org/10.1074/jbc.274.45.32015) [Medline](#)
25. S. P. Iyer, G. W. Hart, Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J. Biol. Chem.* **278**, 24608–24616 (2003). [doi:10.1074/jbc.M300036200](https://doi.org/10.1074/jbc.M300036200) [Medline](#)
26. B. J. Gross, B. C. Kraybill, S. Walker, Discovery of O-GlcNAc transferase inhibitors. *J. Am. Chem. Soc.* **127**, 14588–14589 (2005). [doi:10.1021/ja0555217](https://doi.org/10.1021/ja0555217) [Medline](#)

27. S. Misaghi, S. Ottosen, A. Izrael-Tomasevic, D. Arnott, M. Lamkanfi, J. Lee, J. Liu, K. O'Rourke, V. M. Dixit, A. C. Wilson, Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol. Cell. Biol.* **29**, 2181–2192 (2009). [doi:10.1128/MCB.01517-08](https://doi.org/10.1128/MCB.01517-08) [Medline](#)
28. A. G. W. Leslie, Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26** (1992).
29. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006). [doi:10.1107/S0907444905036693](https://doi.org/10.1107/S0907444905036693) [Medline](#)
30. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). [doi:10.1107/S0021889807021206](https://doi.org/10.1107/S0021889807021206) [Medline](#)
31. 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, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). [doi:10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925) [Medline](#)
32. J. Painter, E. A. Merritt, Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 439–450 (2006). [doi:10.1107/S0907444906005270](https://doi.org/10.1107/S0907444906005270) [Medline](#)
33. J. Painter, E. A. Merritt, TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109–111 (2006). [doi:10.1107/S0021889805038987](https://doi.org/10.1107/S0021889805038987)
34. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). [doi:10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) [Medline](#)
35. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, Electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009). [doi:10.1107/S0907444909029436](https://doi.org/10.1107/S0907444909029436) [Medline](#)
36. The PyMol Molecular Graphics System, Schrödinger, LLC; [www.pymol.org](http://www.pymol.org).
37. A. Morin, B. Eisenbraun, J. Key, P. C. Sanschagrin, M. A. Timony, M. Ottaviano, P. Sliz, Collaboration gets the most out of software. *eLife* **2**, e01456 (2013); [doi:10.7554/eLife.01456](https://doi.org/10.7554/eLife.01456).
38. D. M. Hoover, J. Lubkowski, DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002). [doi:10.1093/nar/30.10.e43](https://doi.org/10.1093/nar/30.10.e43) [Medline](#)
39. J. P. Gogarten, A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, E. Hilario, Inteins: Structure, function, and evolution. *Annu. Rev. Microbiol.* **56**, 263–287 (2002). [doi:10.1146/annurev.micro.56.012302.160741](https://doi.org/10.1146/annurev.micro.56.012302.160741) [Medline](#)