DOCTOR OF PHILOSOPHY

Developing O-GlcNAc transferase inhibitors - insights from substrate recognition

Rafie, Karim

Award date:
2017

Developing O-GlcNAc transferase inhibitors – insights from substrate recognition

Karim Rafie

A thesis submitted for the degree of Doctor of Philosophy.

Center for Gene Regulation and Expression

School of Life Sciences

University of Dundee

September 2017

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Å	Angstrom (1 Å = 10 nm)
aa	Amino acid
Acetyl-CoA	Acetyl Coenzyme A
Ahx	6-aminohexanoic acid
B-factor	Temperature / Debye-Waller factor
BADGP	Benzyl 2-acetamido-2-deoxy-α-D-galactopyranoside
Bis-Tris propane	1,3-bis[tris(hydroxymethyl)methylamino]propane
BSA	Bovine serum albumin
BZX	4-methoxyphenyl 6-acetyl-2oxobenzo[d]oxazole-3(2H)-carboxylate
CAT	Catalytic domain
CBSI	cyclic bisubstrate inhibitor
C. elegans	*Caenorhabditis elegans*
C. perfringens	*Clostridium perfringens*
CK2	Casein-kinase 2
*Cp*OGA	*C. perfringens* O-GlcNAc hydrolase
CPP	Cell-penetrating peptide
CRMP2	Collapsin response mediator protein 2
C-terminus	Carboxy-terminus
Da	Dalton
DAPI	4,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethylsulfoxide

D. melanogaster	*Drosophila melanogaster*
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetra-acetate
EGTA	Ethyleneglycol bis-(2-aminoethylether)-N'N'-tetra-acetate
FBS	Fetal bovine serum
FITC	Fluorescein isothiocyanate
FL	Full length
Floc	5(6)-fluorescein carboxamide
FP	Fluorescence polarimetry
Fruc	Fructose
g	gram
GABA	γ-butyric acid
GH	Glycoside hydrolase
GFAT	Glutamine fructose-6-phsophate amidotransferase
Glc	Glucose
GlcNAc	N-acetylglucosamine
GST	Glutathione-S-transferase
GT	Glycosyl transferase
h	Hour
HAT	Histone-acetyl transferase
HCF1	Host-cell factor 1
HBP	Hexosamine biosynthetic pathway
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
HEK293	Human embryonic kidney 293 (cells)

HeLa	Henrietta Lacks (cell line)
HIV	Human immunodeficiency virus
hOGA	Human O-GlcNAc hydrolase
hOGT	Human O-GlcNAc transferase
HSP	Heat shock protein
HSV	Herpes simplex virus
IC_{50}	Half-maximal inhibitory concentration
INTD	Intervening domain
IR	Infrared
ITC	Isothermal titration calorimetry
k	Kilo
K	Kelvin
Kd	Equilibrium dissociation constant
Ki	Inhibitory constant
LDS	Lithium dodecyl sulphate
m	Milli
M	Molar
µ	Micro
min	Minute
mgea5	meningioma-expressed antigen 5
n	Nano
N.A.	Not available
N.D.	Not determinable
ndx	N-(6-(2,5-dioxopyrrolidine-1-yl)hexyl)
N-terminus	Amino-terminus
OD	Optical density

O/N	Over-night
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PUGNAc	N-Acetylglucosaminono-1,5-lactone O-(phenylcarbamoyl)oxime
PPD	Proteolytic processing domain
RMSD	Root mean square deviation
s	Second(s)
SD	Standard deviation
SDS	Sodium dodecyl sulphate
S.E.M.	Standard error of the mean
*Sm*NodC	*Sinorhizobium meliloti* NodC
T. adherence	*Trichoplex adherence*
TAB1	TAK1-binding protein 1
TAK1	Transforming growth factor beta-activated kinase
TCEP	Tris(2-carboxyethyl)phosphine
Tris	Tris(hydroxymethyl)aminomethane
TBS	Tris buffered saline
TBST	Tris buffered saline Tween-20
tod	4-(2-(2-(methylamino)ethoxy)ethoxy)butane-2-one
TPR	Tetratricot peptide repeat
TRAK 1	Trafficking-kinesin binding protein 1
UDP	Uridine diphosphate
UDP-GlcNAc	Uridine 5’-diphosphate-N-acetylglucosamine
UMBP	Uridine methylene bisphosphonate
UMP	Uridine monophosphate

UTP	Uridine triphosphate
V	Volt
v/v	Volume per volume
vdW	van-der-Waals
w/v	Weight per volume
WT	Wild type
X. campestris	*Xanthomonas campestris*

List of Amino acid code

Amino acid	*Three letter code*	*One letter code*
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Random amino acid	Xaa	X

Acknowledgements

The work presented in this thesis would not exist without the essential contribution, guidance and advice provided by many people in the School of Life Sciences.

First and foremost, I would like to thank Daan for the extraordinary support he has given over the last 5 years and the trust he has placed in me. His guidance, mentoring and encouragement made my PhD an exciting and instructive time. All the long, often heated but always constructive discussions we had about scientific research have made me the researcher I am today. Thanks for taking a chance on me 5 years ago and sorry for the frustrations I have caused with my unique writing style.

Secondly, I would like to thank all the present and past lab members who always made for exciting days in the lab. I would especially like to thank Vladimir Borodkin for the fantastic, fruitful and interactive collaboration we had on the OGT bisubstrate project and any other odd project we ended up working together. I would also like to thank Riccardo Trapannone for his help in the early stages of the cell culture studies needed for the bisubstrate project and Andrii Gorelik for taking over after Riccardo left. Special thanks to Olawale Raimi for the fruitful collaboration on the OGT linear fusion project. Thanks to Andrew Ferenbach for indulging me on my cloning wishes, Villo Muha for showing me the ropes in Tissue Culture and the fly room and Sergio Galan Bertual for the informative discussions about crystallography. I would also like to thank Marianne Schimpl for her guidance during my Master’s project and the early days of my PhD.

Thanks to my thesis committee Ron Hay, Thimo Kurz, Sara Ten Have and Adrian Saurin wo gave me valuable input.

Thanks to the media kitchen for supplying buffers, media and plates, thanks to the ESRF and Diamond Light Source for beam time and the local contacts for support. My work would not have been possible without their support.

Last but not least, I want to thank all my friends who made my time in Dundee unforgettable. Special thanks to Riccardo in whom I found a friend in arms and for life. Special thanks also to Franzi, with whom I shared many a whisky and many a conversation. Thanks to Sofia, Helge, Sergio, Mat, Wale, Mehmet, Paul, and everyone else I have forgotten.

I would like to thank my family for their unwavering support. Mom thanks for everything you have given me, without you I would be a fraction of the man I am today. Fernando and Sandra thanks for being there for me, whenever I need you. Thanks to my three beautiful nieces Chiana, Ai-Vy and Mayla who have made my life so much more worth living. Thanks to Nona, Dahlia, Teymour and Naila who are always here with me. Thanks to Paul and Mai for being part of our growing family and everything you've given us. Thanks to Margret for everything, you are a pillar of strength in our family. We don't know what'd have happened without you. Thanks to Sascha, Yvonne and Noah, you are and always will be family. Thanks to Alex Fuchs and Simon Herrman for their friendship.
Finally, thanks to my late grandfather Salah, my role model and the person I turn to when I need moral support, I dedicate this work to you.

Declaration

I hereby declare that the following thesis is based on the results of investigations conducted by myself and that this thesis is of my own composition. Work other than my own is clearly indicated in the text by reference to the relevant researchers or of their publications. This dissertation has not in whole, nor in part, been previously submitted for a higher degree.

Karim Rafie

I certify that Karim Rafie has spent an equivalent of at least nine terms in research work at the School of Life Sciences, University of Dundee, and that he has fulfilled the conditions of the Ordinance No. 14 of the University of Dundee and is qualified to submit the accompanying thesis in application for the degree of the Doctor of Philosophy.

Daan van Aalten

List of Publications

Work done during my PhD has been published in the following peer reviewed journals:

Aneika C. Leney, **Karim Rafie**, Daan M.F. van Aalten, Albert J. R. Heck, "Direct Monitoring of protein O-GlcNAcylation by high-resolution native mass spectrometry", *ACS Chemical Biology,* 2017

Karim Rafie*, Olawale Raimi*, Andrew T. Ferenbach, Vaibhav Kapuria, Daan M.F. van Aalten, "Recogntiion of a glycosylation substrate by the O-GlcNAc transferase TPR repeats", *Open Biology*, 2017

Riccardo Trapannone*, **Karim Rafie***, Daan M.F. van Aalten, "*O-GlcNAc transferase inhibitors: current tools and future challenges*.", *Biochemical Society Transaction*, 2016

Shalini Pathak, Jana Alonso, Marianne Schimpl, **Karim Rafie**, David E. Blair, Vladimir S. Borodkin, Alexander W. Schüttelkopf, Osama Albarbarawi, Daan M.F. van Aalten, "The active site of O-GlcNAc transferase imposes constraints on substrate sequence.", *Nature Structure & Molecular Biology*, 2015

Vladimir S. Borodkin, Marianne Schimpl, Mehmet Gundogdu, **Karim Rafie**, Helge C. Dorfmueller, David A. Robinson, Daan M.F. van Aalten, **"**Bisubstrate UDP-peptide conjugates as human O-GlcNAc transferase inhibitors.", *Biochemical Journal, 2014*

Summary

The decoration of serine and threonine residues of > 1000 different nucleocytoplasmic proteins with a single N-acetylglucosamine (O-GlcNAcylation), is a highly abundant and dynamic protein post-translational modification. Remarkably, this glycosylation is only regulated by two enzymes, the O-GlcNAc transferase (OGT) that transfers the sugar onto proteins and the O-GlcNAc hydrolase (OGA), which in turn removes it. However, the precise mechanism by which OGT and OGA recognise their plethora of substrates is still elusive. O-GlcNAcylation of nuclear and intracellular proteins is a highly conserved mechanism that can be found in all metazoa and is linked to a broad range of cellular processes and was found to be essential for the proper development of organisms. Genetic knock-outs of either OGA or OGT results in embryonic lethality, hampering the use of animal models to study its biological role. More recently, mutations in the *ogt* gene were linked to X-linked intellectual disability, supporting an essential role of OGT in proper development. An alternative route to study the biological role of O-GlcNAcylation, would be to utilise chemical tools to modulate O-GlcNAc cycling. In recent years multiple, highly potent and selective inhibitors of OGA were reported and are widely used to study the importance of the removal of the GlcNAc residue. Unfortunately, the lack of potent and selective inhibitors of OGT has been a bottle neck in studying its function in cell culture or on the organism level. The aim of work presented in the following thesis, is to investigate the molecular mechanism by which OGT recognises its plethora of substrates and to utilise the information gained to develop mechanism-inspired inhibitors that can ultimately be used in cellular studies. In the first results chapter (Chapter 2) of the thesis, the role of the

tetratricopeptide repeat (TPR) domain of OGT in the recognition and binding of a glycosylation substrate is investigated. O-GlcNAcylation of the TAK1-binding protein 1 (TAB1) modulates TAK1-mediated cytokine release, implicating O-GlcNAc in innate immune signalling. A linear fusion approach was utilised to capture a substrate complex of OGT and an 18-mer TAB1 peptide resulting in the identification of key interactions mediated by the TPR domain. Using *in vitro* systems and structurally guided mutagenesis, it was shown that five key asparagine residues contribute significantly to recognition and binding of an 18-mer TAB1 peptide. This interaction was disrupted in a quintuple mutant of OGT that lacked these five key asparagine residues.

Bisubstrate inhibitors consist of two conjugated substrates/substrate analogues that bind to a bisubstrate enzyme, possibly yielding highly specific and potent inhibitors. The second results chapter (Chapter 3) of this thesis describes the work performed to improve on the recently reported class of Goblin bisubstrate hOGT inhibitors. Modification of the peptide, linker or UDP moiety of the bisubstrate scaffolds yielded the bisubstrate inhibitor ThioGoblin 1 exhibiting low μM inhibition potency, a 10-fold increase over its recently reported progenitor Goblin1. High-resolution macromolecular x-ray crystallography allowed the identification of the exact binding conformation of bisubstrate inhibitors to hOGT. This revealed differences in the linker geometry between serine and cysteine linked bisubstrate scaffolds, which may explain the increase in potency from Goblin 1 to ThioGoblin 1. Finally, a novel high-throughput amenable fluorescence polarimetry assay system for the identification of hOGT binders, using an N-terminal fluorescein modified ThioGoblin 1 scaffold, is described.

In the third and final chapter of the present work the decoration of ThioGoblin 1 with cell penetrating peptide (CPP) sequences to achieve cellular uptake is

described. Attachment of linear CPPs culminated in bisubstrate scaffolds with hOGT inhibition potencies similar to their parent compounds ThioGoblin 1, however showed no inhibition of hOGT activity *in cellulo*, due to limited endosomal escape. Recent studies reported the use of cyclic CPPs to promote cellular uptake and early endosomal escape with a up to 12-fold more efficient delivery of biomolecular cargo into the nucleocytoplasm. Incorporation of a short cell penetrating hexapeptide sequence (RRRRΦF) into the ThioGoblin 1 scaffold and subsequent cyclisation culminated in a small set of cyclic bisubstrate conjugates that exhibited varying degrees of hOGT inhibition *in vitro*. These compounds were capable of translocating across the cell membrane and showed *in cellulo* activity in preliminary experiments. Structural characterisation of hOGT in complex with the cyclic bisubstrate inhibitor dvB_1399 revealed a binding mode of dvB_1399 similar to that of ThioGoblin 1, with potential steric clashes with the TPR domain of hOGT and allowed the design of the more potent cyclic bisubstrate inhibitor dvB_1523. Binding affinity studies of dvB_1523 reveal a potential sub-μM affinity to hOGT, making it the most potent inhibitor of hOGT known to date and a potential tool for future studies of OGT in cell biology.

1. Introduction

1.1. Glycans in biology

Post-translational modification (PTM) is a process, where modifying groups are covalently linked to one or more amino acid side chains, changing the properties of proteins. PTMs range from the attachment of small charged groups, such as phosphorylation [1], to complex structures such as poly-ubiquitin chains [2] and large glycan structures [3]. They realise a multitude of roles within the living organism [4,5] from transcriptional and translational regulation [6], signalling [1,2], protein localisation [7–9], protein folding and stability [10–12], enzyme activity [1,13], protein-protein interactions [14] to protein degradation [2].

The modification of polypeptides with carbohydrate structures, varying in size and complexity, has long been appreciated as being one of the most diverse and complex post-translational modifications found in nature [15]. Indeed, glycosylation events can be found throughout the kingdom of life from archaea to eubacteria to metazoa [16,17] and are regulated by a large number of enzymes [18]. The high diversity of carbohydrate structures can be attributed to two major factors. Firstly, the multitude of unique monosaccharide units that can be used to generated large glycan structures creates a large platform for diversity. Secondly, multiple hydroxyl groups on each monosaccharide unit allow for a large variation in glycosidic linkages. The combination of both allows for the generation of extensive and highly branched structures. The first glycopeptide bond, discovered in 1961, was the β-GlcNAc-Asn linkage found on ovalbumin [19]. Since then, a large array of glycopeptide linkages were identified, including a significant number of functional groups found on peptides, decorated with the most common monosaccharides to polysaccharides (Fig. 1.1) [3].

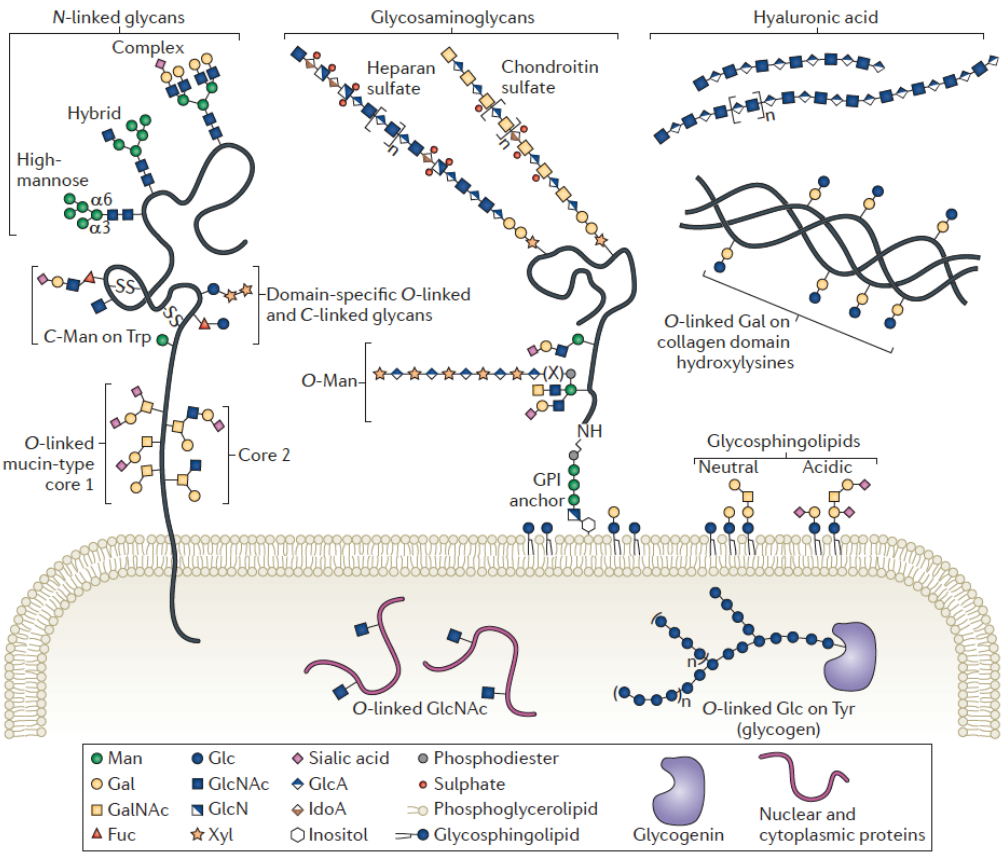

Figure 1.1: Schematic representation of glycans found in vertebrates. Figure taken from [3].

The two most common types of glycopeptide linkages are N- and O-linkages found on asparagine and serine / threonine / tyrosine residues, respectively (Fig. 1.2). Glycan structures fulfil a multitude of biological and physiological tasks, from interactions with the extracellular environment, cell-cell signalling and pathogen invasion (reviewed in [20,21]), protein function and signalling [22] as well as protein folding, quality control and secretion [23,24]. Indeed, ≥ 7000 different glycan structures generate the spectrum of mammalian glycans, regulated by ~700 proteins [25–27], commonly referred to as carbohydrate active enzymes. Recent advances in sequencing technologies allow the identification of novel protein sequences with an ever-increasing rate, broadening the gap between the number of known and experimentally characterised proteins. In order to classify novel protein sequences the Carbohydrate-Active Enzyme (CAZy) database was

introduced [28]. The CAZy-database aims to classify novel putative carbohydrate active enzymes into two major families; (i) the glycoside hydrolase (GH) families, which catalyse the fission of glycosidic bonds and (ii) the glycoside transferase (GT) families, which transfer mono- or polysaccharide structures onto proteins, lipids or other carbohydrates, also known as glycoconjugates. The classification of enzymes is based on their amino acid sequence and their similarity with previously either biochemically or structurally characterised enzymes. Currently > 800.000 protein sequences have been assigned into 103 GT and 145 GH families. Each family is made up of enzymes that can be grouped together, based on their substrate specificity and well conserved structural motifs.

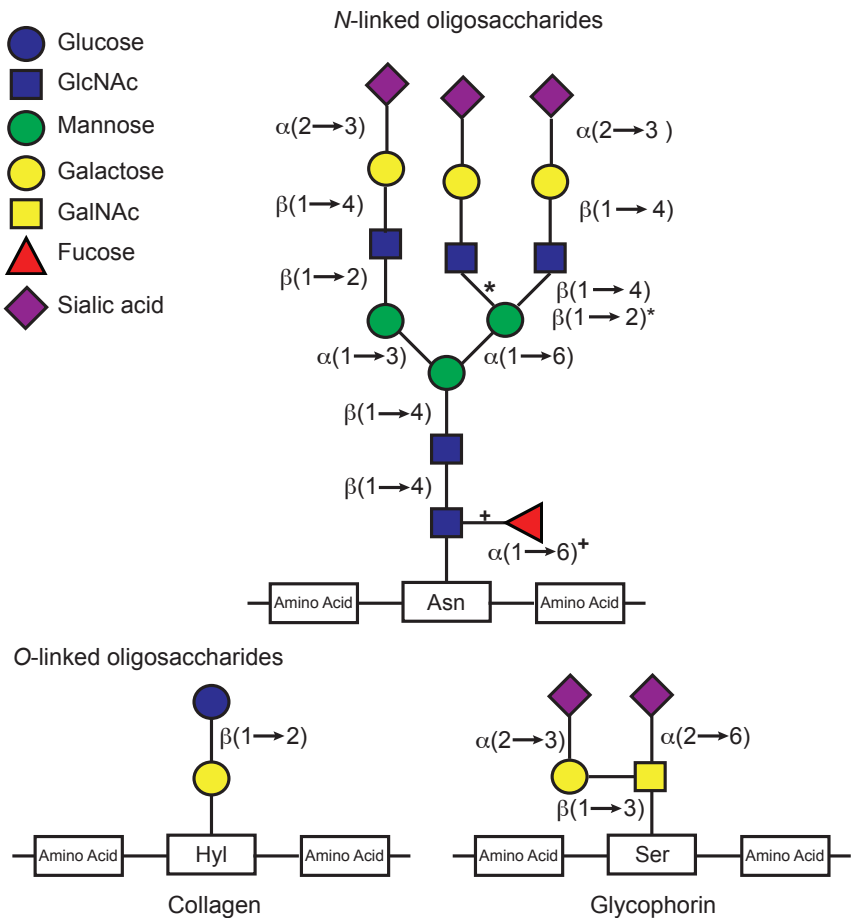

Figure 1.2: Simplified structures of the most common N- and O-linked glycosylation patterns. Hyl = hydroxyl-lysine. Figure adapted from [18].

Glycans can be attached to various proteins and other biomolecules, such as phosphatidylinositol [29], through various linkages. Protein glycosylation can occur on the δ-nitrogen of an asparagine side chain (*N*-linkage), through the hydroxyl-oxygen found on serine, threonine, tyrosine and hydroxylysine/proline side chains (*O*-linkage), as well as the C2 of tryptophan side chains (*C*-linkage). The most studied form of glycosylation is the N-linked glycosylation. In eukaryotic organisms, asparagine residues of nascent polypeptide chains are first glycosylated in the endoplasmatic reticulum (ER), regulating proper folding. After the initial glycosylation inside the ER, glycoproteins are transferred to the Golgi apparatus where they undergo glycan maturation, a process regulated by a multitude of nucleotide sugar donors in concert with GHs and GTs that are found

in the different Golgi compartments and are subsequently secreted or presented on the extracellular membrane. The diversity and amount of various N-glycans is determined by cell lineage and organism, controlling intercellular communication during development and cell migration [30]. The majority of O-linked glycans found in eukaryotes are synthesised similarly to N-glycans and can be found in the extracellular matrix [3]. To date seven different types of O-glycans have been identified in humans, classified by the initial monosaccharide attached to the acceptor residue (Fig. 1.1). More recently however, the modification of serine and/or threonine residues of nucleocytoplasmic proteins with β-*N*-acetylglucosamine has been described, identifying the first intracellular glycosylation of proteins [31].

1.2. The post-translational modification of nucleocytoplasmic proteins with O-linked β-N-acetylglucosamine

The post-translational modification of nucleocytoplasmic proteins with O-linked β-N-acetylglucosamine (O-GlcNAc), found in metazoans, was first described in 1984 [31]. Remarkably, this novel glycosylation rewrote the dogma of protein glycosylation, formerly thought to be exclusively localised to the endoplasmic reticulum, the Golgi apparatus and the extracellular environment. Subsequent experiments on subcellular fractions, derived from rat liver, showed a large abundance of O-GlcNAcylated proteins in cells, with the majority localised to the nuclear envelope and nucleus [32]. Since its discovery more than three decades ago, a large number (> 1000) of nucleocytoplasmic proteins have been found be O-GlcNAc modified [33–43], in various eukaryotic organisms, from the most basal metazoan *Trichoplax adhaerence* [44], to *Caenorhabditis elegans* [45,46], *Drosophila melanogaster* [47,48], mice [49–51] and humans [52]. Interestingly, prokaryotic organisms such as *Listeria monocytogenes* utilise O-GlcNAcylation to modify their extracellular flagella, although the precise role is still elusive [53,54]. The eukaryotic O-GlcNAc proteome (O-GlcNAcome) includes proteins involved in a plethora of cellular processes such as transcription and translation [55–57], trafficking and localisation [58,59] and cell cycle progression [60–63] (Fig. 1.3). Further O-GlcNAc can be seen as a sensor of the metabolic state of a cell, as global O-GlcNAc levels are directly tied to the cellular concentration of UDP-GlcNAc [64–66]. UDP-GlcNAc is the final product of the hexosamine biosynthetic pathway (HBP). The HBP utilises 2-5% of total available glucose and metabolites from amino acid (glutamine), fatty acid (Acetyl CoA) and nucleotide (UTP) metabolism [67] (Fig. 1.3). First glucose is converted to gucose-6-phosphate (Glc-6-P) by the Hexokinase (HK), followed by conversion into fructose-6-phosphate (Fruc-6-P) by the Glc-6-

P-isomerase (GPI). Next Fruc-6-P is converted to glucosamine-6-phosphate (GlcN-6-P) by Glutamine-fructose 6-P aminotransferase (GFAT) before being acetylated to N-acetylglucosamine-6-phsophate (GlcNAc-6-P) by the Glucosamine 6-P N-acetyltransferase (GNPNAT), using acetyl CoA. GlcNAc-6-P is converted to N-acetylglucosamine-1-phosphate (GlcNAc-1-P) by the Phosphoacetylglucosamine mutase (PGM), before finally being converted UDP-N-acetylglucosamine (UDP-GlcNAc) by the UDP-N-acetylglucosamine pyrophosphorylase (UAP). While some of the UDP-GlcNAc pool is directed towards the ER and Golgi apparatus, a substantial amount remains in the cytoplasm serving as donor substrate for protein O-GlcNAcylation [67]. Not surprisingly, deprivation or abundance of metabolites feeding into the HBP such as glucose and glutamine regulate the availability of UDP-GlcNAc, linking its availability tightly to the metabolic status of the cell, which was shown to affect cellualr O-GlcNAc levels [68–71]. Furthermore, OGT activity was found to be modulated by UDP-GlcNAc availability [72].

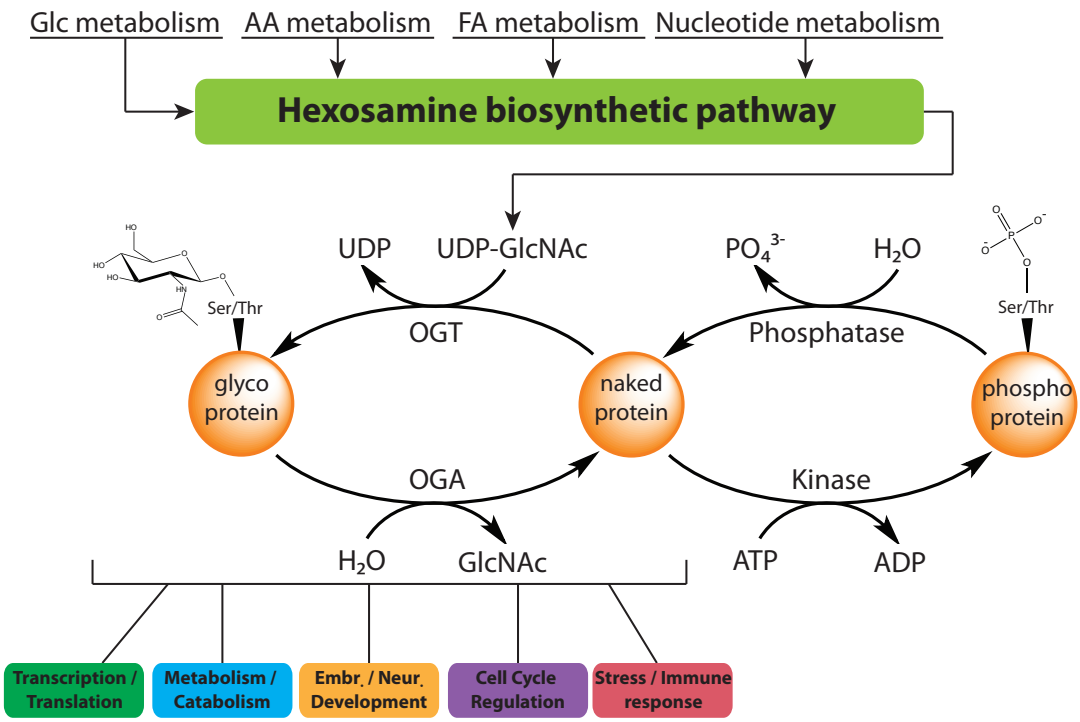

Figure 1.3: Schematic representation of the link between the hexosamine biosynthetic pathway (HBP) and O-GlcNAc cycling. Roughly 2-5% of the total glucose (Glc) is fed into the HBP, among products from other metabolic pathways such as glutamine (amino acid metabolism, AA), acetyl CoA (fatty acid metabolism, FA) and UTP (nucleotide metabolism). The final product, UDP-GlcNAc, is then used by OGT to modify nucleocytoplasmic proteins with N-acetylglucosamine, which is reciprocally removed by OGA. Together OGT and OGA regulate the O-GlcNAcome. Figure adapted from [73].

Strikingly, only two enzymes regulate this ubiquitous, reversible and dynamic modification, the O-GlcNAc transferase (OGT), which adds N-acetylglucosamine to Ser/Thr residues using the sugar nucleotide UDP-GlcNAc and the O-GlcNAc hydrolase (OGA), which removes it. Both enzymes are encoded by single genes and are essential for proper cellular development and viability [74,75]. Furthermore, a close relationship between O-GlcNAc and O-phospho of Ser/Thr residues was reported [76–78]. This “ying-yang” relationship between both modifications is thought to co-regulate each other, as both can be found on the same, proximal and distal

residues [76,77]. For instance, the AMP-activated protein kinase (AMPK) is a crucial regulator of multiple intracellular signalling processes and a vital nutrient sensor [79], which is O-GlcNAcylated on its α- and γ-subunits by OGT [80] and phosphorylated on Thr^{172} by the Ca^{2+}/calmoduline-activated protein kinase kinase (CaMKK2) or the liver kinase B1 (LKB1) [81]. Further, Bullen *et al.* describe a potential link between elevated O-GlcNAc levels in differentiated C2C12 skeletal muscle cells and reduced AMPK activity. However, a direct correlation between AMPK O-GlcNAcylation and its kinase activity has yet to be identified. Interestingly, AMPK itself can phosphorylate OGT on Thr^{444} altering its cellular localisation and substrate specificity [80]. Other cross-talk between O-GlcNAcylation and phosphorylation can be found throughout cells. Inhibition of GSK3β results in changes of O-GlcNAcylation levels on multiple proteins [82], whereas an overexpression of OGT causes polyploidy in tissue culture cell lines by significantly reducing proline-directed phosphorylation [63] as well as reducing phosphorylation of cyclin-dependent protein kinase 1 (CDK1) substrates [83]. Furthermore, cross-talk between O-GlcNAc and O-phospho can happen reciprocally by being modified at the same site under different conditions, such as on Thr^{58} of the oncogene protein c-Myc [84] or the Ser^{16} of the estrogen receptor β (ER-β) [85]. Finally, both protein substrates can be modified in a competitive fashion, where proximal or distal sites affect each other. Vimentin, an intermediate filament protein, is a substrate for both OGT and the Aurora B kinase and is heavily modified by both during M phase [86]. Other reciprocally modified proteins are the Calcium/calmodulin-dependent kinas IV (CAMKIV) [87], p53 [88] and FOXO1 [64,89]. Given its ubiquitous presence and involvement in multiple cellular processes, dysregulation of O-GlcNAc can result in severe phenotypes [73,77,90]. Strikingly, catalytically inactive mutants of either OGT/OGA or a knock-out of

either of these two genes, results in embryonic lethality [49,50,75,91]. Changes in the catalytic activity of either enzyme, which lead to an imbalance in global O-GlcNAcylation are linked to severe diseases such as diabetes [92–95], Parkinson's [37,96,97], Alzheimer's [59,98] and cancer [52,92,99–102]. Therefore, O-GlcNAc might play a crucial role in many biological processes, making it of paramount importance to understand its biological function.

1.3. The biological role of O-GlcNAc

Compared to other post-translational modifications such as phosphorylation, ubiquitination and acetylation, the precise biological role of O-GlcNAc is still unknown. Studies have confirmed that O-GlcNAc is an essential modification and the absence or presence of catalytically inactive mutants of either OGT or OGA were shown to have severe effects on development. Furthermore, the lack of specific tools, such as specific and potent inhibitors, antibodies and other reagents hampered advances in understanding the role of O-GlcNAc. For instance, the isolation and identification of O-GlcNAc sites is difficult due to the labile nature of the glycosydic bond creating difficulties during mass spectrometric analysis. However, it is clear that it plays a significant role in various cellular processes.

1.3.1. O-GlcNAcylation of the nuclear envelope and gene regulation

O-GlcNAc is highly abundant on the nuclear envelope and the nucleus, with the nuclear pore proteins being amongst the most O-GlcNAcylated proteins [103–107]. Various sites were mapped to nucleoporin (NUP) 62/69 and thought to be linked to their stability [108], nuclear pore permeability and maintenance as a selectivity filter [109]. Furthermore, various chromatin-binding proteins, such as transcription

factors and histones, are known to be O-GlcNAc modified [110]. Interestingly, various studies described the O-GlcNAcylation of histones. For instance, histone H2B is supposedly O-GlcNAcylated on Ser[112] and thought to promote monoubiquitination of histone H2B Lys[120], resulting in transcriptional activation [111]. More recently, it was shown that the ten-eleven translocation (TET) proteins recruit OGT to chromatin [112,113], a process potentially important for DNA methylation. The C-terminal domain of the RNA-Polymerase II is heavily O-GlcNAc modified and modulates, in concert with phosphorylation, gene expression at different stages of the transcription cycle [76], although further investigation is needed. The host-cell factor 1 (HCF1) is a ubiquitously expressed chromatin-bound transcriptional regulator and is involved in multiple cellular processes such as cell-cycle progression and was shown to be heavily O-GlcNAc modified shown to be important for the proteolytic maturation of HCF1 [57]. Another transcription factor that is O-GlcNAc modified is p53 [88]. O-GlcNAcylation of p53 on Ser[149] supposedly increases its stability by reducing of phosphorylation of Thr[155], blocking ubiquitin-directed proteasomal degradation [88].

1.3.2. O-GlcNAc and cell signalling

Investigation of the O-GlcNAc proteome by high-throughput proteomics identified numerous O-GlcNAcylated proteins involved in various signalling pathways [34,41,114,115], For instance, proper signalling of cytokine-mediated immune response via the TGFβ-activated kinase 1 (TAK1) [116], via O-GlcNAcylation of its binding partner, the TAK1-binding protein 1 (TAB1) [117]. Here Pathak *et al.* investigated the effect of O-GlcNAcylation of TAB1 Ser[395], in response to interleukin-1 (IL-1) induced immune response and osmotic stress. Modification of Ser[395] results in an increased activation of TAK1, which leads to the consequent activation of

downstream effectors. Other studies reported the O-GlcNAcylation of several proteins involved in insulin signalling, such as the insulin receptor (IR), the insulin receptor substrates 1 and 2 (IRS 1/2), the phosphoinositide 3-kinase (PI3K), the protein kinase B (Akt) and the phosphoinositide-dependent kinase 1 (PDK1) [118]. Indeed, it was shown that increased O-GlcNAc levels leads to the inhibition of insulin-stimulated phosphorylation of Akt but not PDK1, in 3T3-L1 adipocytes [94,118]. Further studies reported the importance of O-GlcNAc in insulin signalling, showing that increased O-GlcNAc impairs the localisation of the glucose transporter GLUT4 on the plasma membrane, via the modification of the key regulator of membrane translocation and fusion Munc18c [119], which may alter glucose uptake. Remarkably, O-GlcNAcylation of Ser^{529} of phosphofructokinase 1 (PFK1) in response to oxidative stress, inhibited its activity in cancer cell lines and redirecting the glucose flux through the pentose phosphate pathway [78]. Even though evidence exists for the involvement of O-GlcNAc in signalling, the majority of O-GlcNAc proteins identified in proteomic studies have yet to be investigated in the context of cellular signalling.

1.3.3. O-GlcNAc increases protein stability

Recent studies linked increased stability of proteins to their modification by O-GlcNAc. For instance, O-GlcNAcylation of p53 prevents the phosphorylation on a proximal site, which in turn blocks ubiquitin-driven proteasomal degradation [88]. Furthermore, the ubiquitously expressed specificity protein 1 (Sp1) transcription factor was shown to be more stable upon O-GlcNAcylation, when studied in rat kidney (NRK) cells treated with forskolin [120,121]. More recently, a study described the co-translation modification of nascent polypeptide chains with O-GlcNAc, preventing their premature degradation. Using cell-free rabbit reticulocyte

expression systems Zhu *et al.* showed that OGT dependent O-GlcNAcylation of nascent protein chains prevents their ubiquitination and therefore promotes proper folding and maturation, which can be prevented by OGT inhibition [10].

1.3.4. The role of O-GlcNAc in cell stress and apoptosis

Evidence emerged that O-GlcNAc plays a vital role in cell stress and apoptosis. O-GlcNAc levels increase upon stimulation of cells under various stress conditions such as oxidative stress, UV light irradiation and heat shock [122,123]. For instance, treatment of rat hearts with an O-GlcNAcase inhibitor was shown to be cardioprotective in an O-GlcNAc dependent manner [124,125]. Other studies found a direct correlation between the increase of O-GlcNAc levels post-heat shock [123]. Interestingly, increased transcription of heat shock proteins (HSP) 40 and 70 were associated with increased global O-GlcNAc levels, the latter being directly O-GlcNAcylated itself [33,66]. Furthermore, elevated O-GlcNAc levels, via either glucosamine, OGT overexpression or O-GlcNAcase inhibition, protects hypoxia induced cell death in cardiomyocytes [126]. A glucosamine-induced increase in global O-GlcNAc levels also affect the phosphorylation levels of the p38 mitogen activated protein kinase (MAPK), decrease ischemic contracture and reduce the number of reperfusion-induced arrhythmias [127]. Further, O-GlcNAc was linked to synaptosomal-associated protein 29 (SNAP-29) directed autophagy, upon nutrient starvation or energy depravation [128]. SNAP29 is necessary for the fusion of autophagosomes and endosomes/lysosomes and promotes the autophagic flux [129]. SNAP29 was shown to be O-GlcNAcylated on multiple different sites and that mutation of these sites or the siRNA induced knock-down of OGT lead to an increase of autophagy [128]. Additionally, O-GlcNAc was found to regulate

autophagy through modification of the transcription factor forkhead box protein (dFOXO), in *Drosophila melanogaster* [130].

1.3.5. O-GlcNAc in development

In recent years the importance of O-GlcNAc in the proper development of organisms moved into focus [46,48,60,75,131–133]. Interestingly, the first evidence of O-GlcNAc linked developmental defects were reported in 1984, identifying the polycomb gene *sxc* from *D. melanogaster* as an important contributor of proper development [48]. Only later was *sxc* identified as the gene encoding for OGT, identifying a role of OGT in the polycomb transcriptional repression by glycosylation of Polyhomeotic, a transcriptional repressor that is needed for proper segmentation during development [47]. Indeed, O-GlcNAcylation of Polyhomeotic is thought to prevent its aggregation [133]. Later studies revealed the importance of O-GlcNAc in various model organisms, such as *C. elegans* [45,131,134,135], zebrafish [75], *Xenopus laevis* [136] and mouse [60]. Strikingly however, genetic knock-outs of *ogt* alleles in *C. elegans* are not embryonically lethal but lead to insulin-like phenotypes [131,135]. Knockdown of *ogt* in zebrafish and *X. laevis* lead to severe growth defects, shortened body axis and retarded nervous system [75,136]. A brain specific conditional knock-out of *oga* in cells results in severe metabolic deregulation and semi-penetrant perinatal lethality [137], whereas a brain-specific knock-out also results in neuronal, anatomical and behavioural phenotypes [138]. Recently, O-GlcNAc was reported to be important for the proper development of T-cells by metabolically controlled levels of O-GlcNAc [71]. Swamy *et al.* showed that loss of O-GlcNAc, by using a conditional knock-out of OGT, leads to a blocking of T cell progenitor renewal, malignant transformation and the expansion of peripheral T cell clones [71]. The importance of proper O-GlcNAc

cycling during embryonic development was shown previously [47,60,132]. Neuron and thymocyte specific conditional knock-outs of the *ogt* gene allowed the study of the role of OGT in development [60], as embryonic knock-outs of *ogt* are not viable in mice [50]. Interestingly, *ogt-null* cells underwent increased apoptosis and therefore failed to contribute to the peripheral T-lymphocyte repertoire, whereas the generation of neuron specific *ogt-null* cells resulted in neuronal dysfunction, expressed in severely deficient locomotor activity and postnatal death [60]. In *Drosophila melanogaster*, mutants of OGT lacking catalytic activity or genetic knock-out of the *sxc* gene result in deferred development and do not progress beyond the pharate adult stage [47,48,56]. Strikingly, a recent study investigates whether OGT catalytic activity is necessary for post-pupal development in *D. melanogaster* [132]. Here, Mariappa *et al.* rescued the developmental phenotype of *ogt-null* flies by driving the transgenic expression of WT *D. melanogaster* OGT or one of its point mutants described in the study. As expected no F1 offspring could be seen for flies lacking any DmOGT transgene, whereas WT DmOGT was fully capable of rescuing proper development. Interestingly however, hypomorphic mutants of DmOGT were also able to rescue pupal lethality but the F1 progeny was not able to produce viable offspring [132]. These results indicate a dual functionality of OGT, one catalytic and one non-catalytic activity and would also suggest an essential role of OGT in the formation of important mulitprotein complexes, necessary for proper development. Indeed, few studies reported the formation of various OGT-protein complexes with Host-cell factor 1 [83,139,140] (described in detail in chapter 1.6.3), the motor-adapter trafficking kinesin binding protein-1 (TRAK1) [141], the ten-eleven translocation proteins [112,113,142–144], histones [110,145,146] and ataxin 10 [147].

All studies on the developmental role of O-GlcNAc have so far been achieved using genetic alterations of the genome or knock-down of either *ogt* or *oga*. However, this does not allow the dissection of the potential roles of OGT. Furthermore, it is not yet fully understood what the spatiotemporal effect of protein O-GlcNAcylation is in the context of disease and development. The lack of potent and selective OGT inhibitors so far halted these investigations in model organism. Any advances in the design of these tools would greatly facilitate our understanding of the role of OGT in development.

1.4. O-GlcNAc in health and disease

The involvement of O-GlcNAc in virtually all cellular processes makes it one of the most abundant post-translational modifications. It was shown to be crucial for the proper development of organisms and genetic manipulation of either OGA or OGT leads to severe phenotypes. However, the precise role of O-GlcNAc in the development of certain metabolic and neurological diseases is still elusive. Additionally, the lack of potent and selective inhibitors of OGT has severely hampered the investigation of its role in these pathologies.

1.4.1. Neurological pathologies and X-linked intellectual disabilities

In recent years, a number of studies identified O-GlcNAc to be highly abundant in neurons and that levels of OGT are high in neurons, especially in the nucleus, dendrites, synaptic vesicles and the presynaptic terminal [37,148,149]. Interestingly, aberrant O-GlcNAcylation has been linked to Alzheimer's (AD) [150,151] and Parkinsons [67,152] disease indicating at an important role of O-GlcNAc in these neuropathologies. The microtubule associated protein Tau has long been appreciated as a key player in the development of AD [153–155]. Tau is post-translationally modified at various positions [156] and hyper-phosphorylation of Tau

is linked to the formation of toxic Tau species [153]. A genetic approach to investigate the role of O-GlcNAc in Tau linked AD was carried out using *C. elegans* models of human neurodegenerative disease [152]. Here, Wang *et al.* proposed a link between proteotoxicity and increased levels of O-GlcNAc, where toxic effects of tau aggregation were alleviated in *ogt-null* mutants and increased in *oga-null* mutants [152]. Interestingly however, the opposite can be seen when investigated in human and mouse [51,59,97,157]. Strikingly, experiments done with human neuroblastoma cells reveal a tight balance between levels of O-GlcNAcylation and O-phosphorylation and altering of this balance results in the nuclear localisation of tau [59]. In an animal model of starved mice, decrease of glucose uptake resulted in a decreased O-GlcNAcylation but increased phosphorylation of Tau, mimicking an AD phenotype [97]. Remarkably, pharmacological inhibition of OGA prevents the hyper-phosphorylation of Tau and its resulting aggregation [51,157]. Long treatment of an AD rTg4510 tau transgenic mouse model with an OGA inhibitor strongly increased Tau O-GlcNAcylation, reduced the number of dystrophic neurons and protected against the formation of pathological species, without altering the phosphorylation levels of non-toxic Tau species [51,157],

Intellectual disability (ID) is defined as an early-onset of cognitive function and significantly decreased/limited adaptive behaviour, affecting up to 1% of the total global population [158]. Intellectual disability is categorised by two factors, (i) the severity of cognitive impairment into mild (IQ 50-70) and profound (< 30) and (ii) the presence (syndromic) and absence (non-syndromic) of dysmorphic features [158]. *Ogt* is located on the long arm of the X chromosome in region q13 (Xq13) [50] and recent reports of point mutations in the *ogt* gene link OGT to pathogenic X-linked ID (XLID) [159–162]. So far, four mutations in *ogt* were reported, of which three

lead to an amino acid change (L254F [160], A319T [159], R284P [161]) in the TPR domain and one that leads to a splicing defect in the same domain [161]. Interestingly, global cellular O-GlcNAc levels are not significantly affected by either the L254F nor the R248P mutant. However, variations in the protein levels of OGT seem to be balanced out by a reduction in OGA levels [161] to maintain cellular O-GlcNAc homeostasis [163]. Although *ogt* was identified as a potential cause for XLID, the precise mechanism by which they cause the ID phenotypes has yet to be elucidated.

1.4.2. O-GlcNAc and metabolic diseases

Given its involvement in multiple cellular processes and metabolic state of cells, O-GlcNAc was identified as a post-translational modification dynamically modulated in metabolic diseases, such as cancer [52,164] and diabetes [92,122,165]. Many components of the insulin signalling pathway are dynamically O-GlcNAcylated, tightly linked to nutrient levels [118]. Cancer is a multifactorial disease characterised by the unchecked proliferation of cells in a variety of tissues. Cancer cells have an altered glucose metabolism, leading to a production of ATP primarily through glycolysis. The resulting increase in glucose uptake, termed the “Warburg effect” [166], also leads to an increase of O-GlcNAc in cancer cell lines [164]. Indeed, multiple transcription factors such as p53 [88], cMyc [84] and NFκB [99] are O-GlcNAcylated and known to be elevated in cancer cells.

1.5. The O-GlcNAc hydrolase

The removal of O-linked β-N-acetylglucosamine from nucleocytoplasmic proteins is regulated by the evolutionary conserved [167] O-GlcNAc hydrolase (OGA), first isolated from rats [168]. OGA was initially described as the meningioma-expressed antigen 5 (*mgea5*) [169], found in a screen of a meningioma expression library [169], located on chromosome 10q24.1-q24.3 [169,170]. Initially classified as a hyaluronidase, due to its sequence similarity to homologues found *Caenorhabditis elegans* (*C. elegans*) [169], it was quickly shown that the protein encoded by the *mgea5* gene is, the O-GlcNAc hydrolase. Interestingly, further characterisation of OGA [171] showed that this enzyme most likely represents the Hexosaminidase C, which was identified together with the lysosomal Hexoaminidase A and B (HexA/B) [172–174]. A single isoform of OGA was reported for *Drosophila melanogaster* [165] and *C. elegans* [131] and two predicted isoforms, generated by alternative splicing [175], in humans [176]. The larger of the two isoforms has a predicted molecular weight of 103 kDa (FL-OGA), consisting of a hydrolase domain belonging to the GH84 Family [28] and a predicted histone-acetyl transferase-like domain (HAT) with a putative acetyl transferase activity (Fig. 1.4). Initial studies described interactions between histones and HAT activity *in vitro* [177,178]. However, these results could not be reproduced by others [179–181], leaving the function of the HAT-like domain still highly debated. The smaller isoform, with a predicted molecular weight of 76 kDa, lacks the HAT-like domain (Fig. 1.4).

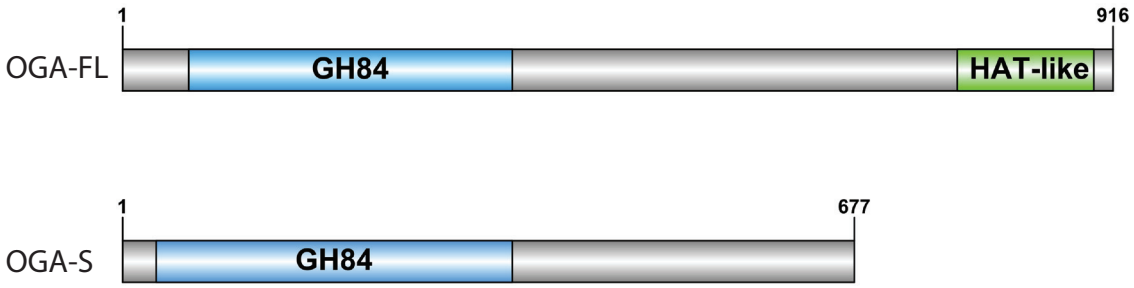

Figure 1.4: Domain structures of full length OGA (OGA-FL) and its shorter isoform (OGA-S). The GH84 domain is highlighted in blue and the HAT-like domain in green. The sizes of the domains are scaled to the total size of the protein.

Interestingly, OGA-FL is thought to be localised mainly to the cytoplasm [167], whereas the shorter isoform resides mostly in the nucleus [175] and putatively on the surface of lipid droplets [8]. The precise role of each O-GlcNAcase isoform is still unknown, however their regulation is thought to be modulated by various factors such as co-regulation by OGT forming an O-GlcNAczyme [182,183], post-translational phosphorylation of Ser^{364} (human) [184] and Ser^{405} (mice) [38], or proteolytic cleavage after Asp^{413} by caspase-3 [179]. Initial insights into the structure and mechanism of the O-GlcNAcase was realised by the reported structure of a bacterial O-GlcNAcase from *Clostridium perfringens* (*Cp*NagJ / *Cp*OGA) [185]. The structure revealed a three-domain fold, with the N-terminal domain consisting of an α/β-fold of seven β-sheets, with flanking α-helices (Fig. 1.5). The middle domain forms a β/α-barrel and the C-terminus consists of a stalk domain made up of an α-helix bundle (Fig. 1.5 A). Interestingly, OGA binds its substrates in a conserved groove in a similar conformation and orientation, characterised by interactions of the backbone of peptide substrates with OGA, in the -4 through the +3 position, surrounding the GlcNAc site [186,187]. Strikingly, when bound to OGA substrates adopt a “V”-shaped conformation with the residue side-chains pointing away from the active site, which explains how a single enzyme can recognise > 1000 different substrates [186]. An additional, important interaction is

formed by hydrophobic interactions of -1 and -2 peptide bonds and a surface exposed tyrosine (Tyr[69] in human OGA). Indeed, this interaction significantly contributes to the binding affinity of OGA substrates [186].

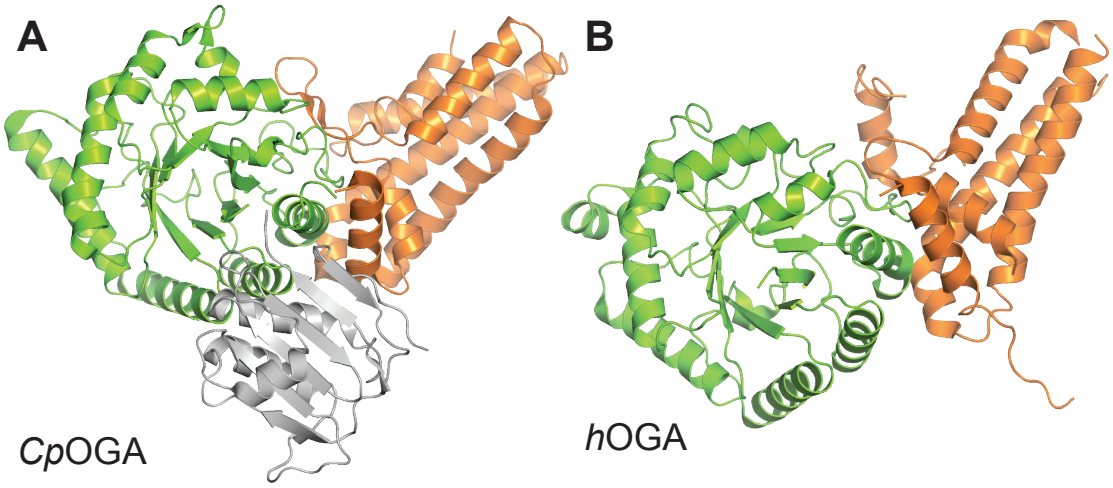

Figure 1.5: Structures of the bacterial O-GlcNAcase from *C. perfringens* and the human O-GlcNAcase. (A) The structure of *Cp*OGA reveals a three domain fold with the N-terminal domain consisting of a α/β-fold, consisting of seven β-sheets with flanking α-helices (grey), the catalytic domain formed by a β/α-barrel fold (green) and the stalk bundle domain (orange) (PDBID 2YDQ [186]). **(B)** Structure of the human O-GlcNAcase showing a β/α-barrel fold (green) for the catalytic domain and a stalk bundle domain consisting of six α-helices (PDBID 5UN8 [188]). **(A+B)** The protein is shown in cartoon representation.

More recently however, the structure of the human O-GlcNAcase has been solved [188–190], revealing significant differences in the stalk-domain compared to *Cp*OGA and other bacterial O-GlcNAcases [185,187,191,192]. Most notably, the last α-helix of the stalk domain in the human O-GlcNAcase structure interacts with its counterpart from the sister monomer allowing them to form a homo-dimer, which is not present in the bacterial structures (Fig. 1.5). Furthermore, the structure revealed a significant interaction of the hydrophobic patch of the stalk domain with substrates [188]. Here, the dimerization of hOGA forms a long hydrophobic cleft that allows the binding of peptide substrates. More remarkably however, is

the fact that the residues involved in the formation of this hydrophobic pocket are highly conserved in most eukaryotes, indicating its importance in evolution [188]. Interestingly, the GlcNAc residues forms extensive polar interactions with the enzyme and are crucial for substrate recognition (Fig. 1.6).

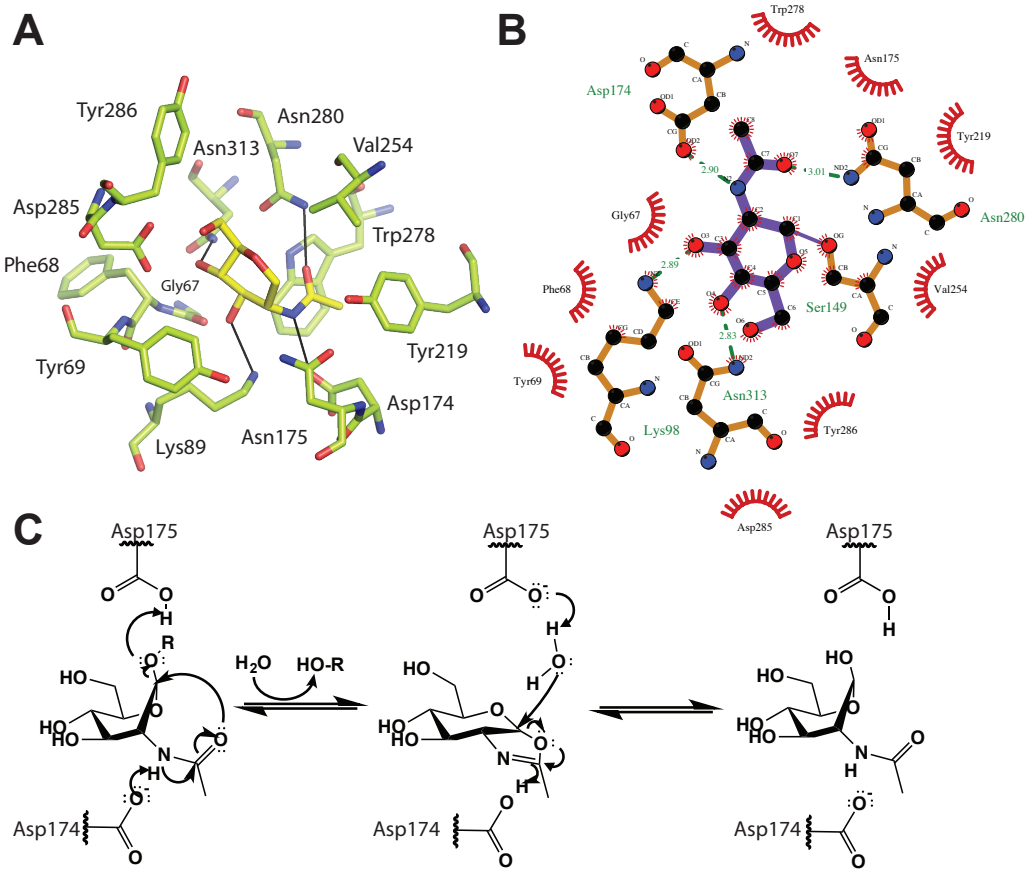

Figure 1.6: OGA recognises the GlcNAc residue via multiple polar interactions and cleaves the glycosydic bond via substrate-assisted catalysis. (A) Structural representation of GlcNAc bound to hOGA D175N (PDBID 5UN8 [188]). Residues interacting with GlcNAc are shown as green sticks, the GlcNAc residue is shown as yellow sticks. Hydrogen bonds are shown as black dashes. **(B)** Schematic representation of the ligand environment. The GlcNAc ligand is shown in purple. Residues forming hydrogen bonds are shown as orange sticks. The drawing was generated using PDB 5UN8 [188] and LigPlot+ [193]. **(C)** Proposed catalytic mechanism of O-GlcNAc hydrolysis catalysed by OGA. Adapted from [194].

OGA cleaves the glycosidic bond, using a double-displacement substrate assisted catalytic mechanism, retaining the configuration of the anomeric carbon

[185,187,192,195]. Here, the two aspartic acids Asp[174/175], forming part of the catalytic DDX motif, and the 2-acetamido group, which acts as the nucleophile, form a catalytic triad to cleave the glycosidic bond via a bicyclic oxazoline intermediate, similar to hydrolases found in the GH20 family [28,196]. In the proposed mechanism, the carboxylate of Asp[174] orients and polarises the 2-acetamido group increasing its nucleophilicity and priming it for an attack on the anomeric centre to form an oxazoline intermediate. Simultaneously, the protonated Asp[175] acts as a general acid to facilitate the fission of the glycosydic bond (Fig. 1.6). Subsequently, Asp[175] deprotonates an incoming water molecule, allowing it to act as a nucleophile, attacking the anomeric centre and promote the disassembly of the oxazoline intermediate, which is supported by the polarising effect of Asp[174].

Initial studies of the role of OGA in biological systems were studied using the potent inhibitors PUGNAc [197] and Streptozotocin [198]. PUGNAc is a competitive inhibitor, first described in 1991 [197], capable of inhibiting all known human hexosaminidases (Hex A/B and OGA). PUGNAc is a (phenylcarbamoyl)oxime (Fig. 1.7), which mimics the transition state and binds to hOGA with low nM affinity [168,195] and was extensively used in biological studies [199–206]. However, PUGNAc targets Hex A/B equally well as it inhibits human OGA [168,207,208] resulting in adverse side effects [195,206]. For example, Mehdy *et al.* reported the existence of unusual free oligosaccharides, containing 2-5 HexNAc residues, in Chinese Ovarian Hamster (CHO) cells, treated with PUGNAc [206]. Another study found that desensitisation of cells to insulin treatment was not a result of PUGNAc treatment, as the more specific OGA inhibitor NButGT (discussed below) did not recapitulate the effect [195]. Streptazotocin (STZ, Fig. 1.7), a GlcNAc mimic first described as an antibiotic in 1960 [198], is routinely used to induce diabetes in rodents by killing insulin-producing pancreatic β-cells [209]. STZ is a naturally occurring molecule and

readily taken up by the GLUT2 glucose transporter selectively expressed in β-cells [210], thus being one of the first systems probed for effects of protein O-GlcNAcylation [211]. Liu *et al.* reported a STZ linked elevation of O-GlcNAc levels of treated β-cell islets, accompanied by increased cell death [211]. Based on these findings it was concluded that insulin resistance in β-cells is tightly linked to O-GlcNAc levels and causes decreased cell viability [211]. Shortly after, Konrad *et al.* showed that significantly higher levels of STZ are needed to inhibit OGA in homogenised rat brain, linking it to the reduced expression levels of GLUT2 transporters compared to those in pancreatic β-cells [212]. In the same study Konrad *et al.* proposed an irreversible inhibition of OGA, without providing mechanistic studies however. Another study proposed a competitive inhibition of OGA by STZ by partial catalysis into a more stable reaction intermediate, than the oxazoline intermediate [213]. However, studies comparing the effects of STZ and PUGNAc on cell toxicity showed clearly that STZ induces cell death independent of OGA inhibition [204,214]. Furthermore, STZ induced cell toxicity was linked to increased DNA fragmentation [215] and induction of apoptosis [214,216]. Due to their adverse side effects both, PUGNAc and STZ should not be used as tools to study OGA in biological systems.

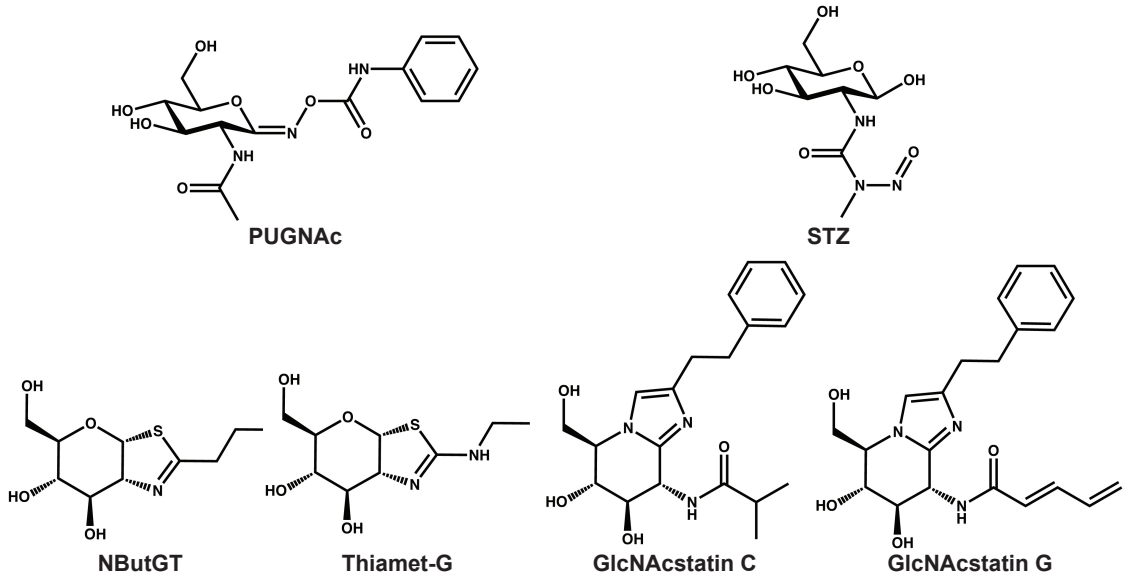

Figure 1.7: Chemical structures of OGA inhibitors. PUGNAc [197], Streptazotocin (STZ) [198], NButGT [195], Thiamet-G [217], GlcNAcstatin C [207] and GlcNAcstatin G [218].

Elucidation of the catalytic mechanism [195] and the report of bacterial O-GlcNAcase structures [185,192] greatly accelerated the design and synthesis of highly potent and selective OGA inhibitors such as NButGT [195], Thiamet-G [217] and the family of GlcNAcstatins [207,218,219]. An important step in the mechanism of OGA-catalysed fission of the O-GlcNAc bond is the formation of the oxazoline intermediate. Based on this Macauly *et al.* synthesised a set of NAG-thiazoline derivatives capable of mimicking the reaction intermediate. Remarkably, NButGT (Fig. 17) reportedly inhibited OGA with low μM affinity and was found to be active in COS-7 cells at 50 μM [195]. Furthermore, it was shown to not cause insulin resistance [94,205] as seen with PUGNAc and STZ as well as having a 1500-fold selectivity over Hex A/B [195].

Additional work on NAG-thiazolines culminated in the development of Thiamet-G (TMG, Fig. 1.7) that showed 31000-fold selectivity against OGA over Hex A/B [217]. Initial cell culture studies showed global O-GlcNAc level elevation in neuronal-like PC-12 cells with maximum inhibition of OGA seen at concentrations > 200 nM [217]. In the same study, Yuzwa *et al.* investigated whether TMG is a tool to

study the link between O-GlcNAc and AD in a living organism, given its water-solubility. Another important factor, addressed in this experiment, is the capability of TMG to cross the blood-brain barrier. Remarkably, administration of TMG resulted in 10-fold increase of global O-GlcNAc levels after 24 h and a significant decrease in the phosphorylation of Tau [217]. Since then, TMG has been used in various studies [220,221], showing a link between O-GlcNAcylation and the phosphorylation of the p38 mitogen-activated kinase (MAPK) in response to high-glucose concentrations, indirectly activating the expression of mesangial matrix protein coding genes [220]. Mi *et al.* reported that artificially elevated levels of O-GlcNAc in TMG treated A549, H1299 and HT92 lung cancer cells, result in increased colony formation capability [221].

In parallel to the development of the NAG-thiazolines a set of transition-state mimicking glucoimidazoles, called GlcNAcstatins, were reported [207,218,219]. The GlcNAcstatins are capable of inhibiting eukaryotic OGAs from the μM [44] down to the pM range *in vitro* [218,219], have a high selectivity towards OGA over Hex A/B [207,218,219] and to be cell-penetrant [218]. Most likely the best characterised is GlcNAcstatin C (Fig. 1.7) showing moderate selectivity over Hex A/B and inhibitory activity in HEK293, SH-SY5Y and HeLa cells [218,219]. Furthermore, GlcNAcstatin C was used to inhibit OGA in mouse embryonic stem cells (mESC) linking increased levels of O-GlcNAc to the impairment of cell differentiation and upregulation of genes normally epigenetically silenced in mESCs [145]. The most specific OGA inhibitor to date, is GlcNAcstatin G (Fig. 1.7) with a 90000-fold selectivity over HexA/B [218] and capable of increasing global O-GlcNAc levels in HEK293 cells at 1 μM [117]. In 2012 Pathak *et al.* used GlcNAcstatin G to investigate the role of TAB1 O-GlcNAcylation in innate immunity [117]. Here, they showed that the Ser^{395} O-GlcNAc on TAB1 is essential for the proper activation

of TAK1 and therefore its downstream effector NFκB [117]. Later studies used GlcNAcstatin G to artificially enrich the fraction O-GlcNAcylated proteins in a large scale proteomic analysis of the HEK293 proteome [34] and to characterise a bacterial OGA isolated from the thermophilic prokaryote *Thermobaculum terrenum* [191,222]. However, the biggest drawback of the GlcNAcstatins is their low solubility in aqueous solutions and their complicated chemical synthesis [207,223]. Given their high potency and selectivity the family of GlcNAcstatins and the NAG-thiazolines, especially TMG, have contributed greatly to our understanding of the role of OGA and are crucial tools in the further investigation of the biological role of O-GlcNAc.

1.6. The O-GlcNAc transferase

1.6.1. Identification and characterisation of the O-GlcNAc transferase

O-GlcNAc transferase activity was first described by Haltiwanger *et al.* in 1990 by exposing synthetic peptide substrates, designed based on three O-GlcNAcylated proteins discovered previously [224], to rabbit reticulocyte extract and rat liver [225]. Only two years later OGT was purified from rat liver using $(NH_4)_2SO_4$ precipitation, hydrophobic interaction and anion exchange chromatography, followed by a sequence of affinity chromatographies and a final size-exclusion chromatography [226]. Initially identified as a 340 kDa heterotrimeric complex of two 110 kDa units with glycosyl transferase activity and one 78 kDa unit with no detectable activity [226]. It was later shown that these subunits were the result of alternative splice variants of OGT, differing at their N-terminus and known as nucleocytoplasmic OGT (ncOGT) and short OGT (sOGT) (Fig. 1.8 A) [91]. Interestingly, a third isoform harbouring a mitochondrial targeting sequence (MTS) known as mitochondrial OGT, was reported in the same study [91]. The first insight into the domain structure of OGT was reported by Kreppel *et al.* in 1997, where for the first time they have cloned the *ogt* gene from a rat liver cDNA library [227]. Strikingly, the authors were able to assign part of the *ogt* sequence to a diverse group of proteins, of which all contained a similar motif designated the tetratricopeptide repeat domain (TPR) [227]. Using Southern blot analysis, the authors were also able to identify orthologues of *ogt* in *C. elegans*, mouse, dog, cow, rabbit and humans and were also able to show that it is a unique in the genome and does not belong to a multigene family [227]. Western blot analysis allowed the identification of a single OGT at ~ 110 kDa in several rat tissues, whilst the smaller sOGT was only detected on occasion in some of the tissues [227]. Later, sequencing results showed that OGT contained 13.5 TPRs [228] and two

distinct domains bearing similarity to GalNAc transferases and designating them as conserved domain (CD) I and II, interrupted by an uncharacterised intervening sequence [229].

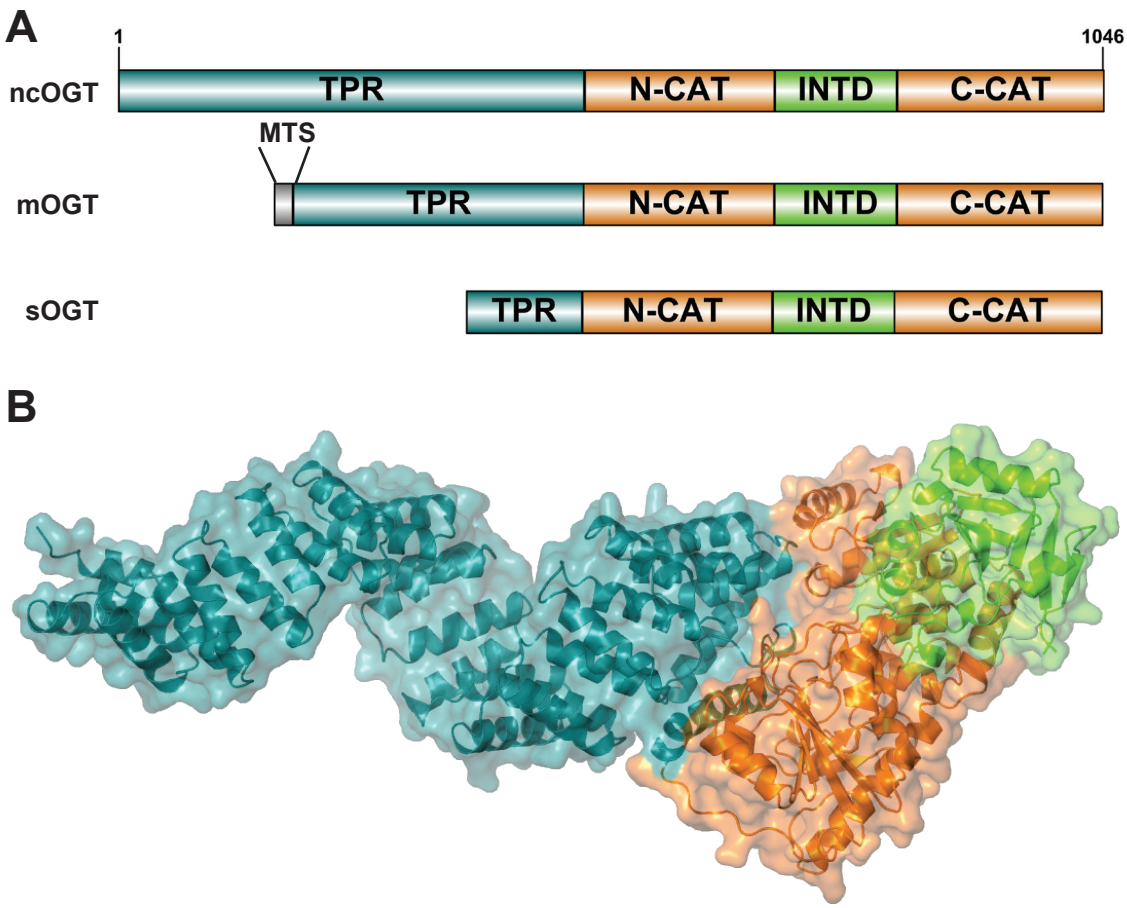

Figure 1.8: Modelled structure of the human O-GlcNAc transferase. (A) The three isoforms of OGT are comprised of two major domains, the tetratricopeptide repeat domain (teal) and the catalytic domain (orange). The latter is further comprised of two catalytic lobes (N-CAT / C-CAT, orange) that form the active site and an intervening domain (INTD, green), of unknown function. The mitochondrial isoform mOGT carries an additional N-terminal mitochondrial targeting sequence (MTS). **(B)** Full length structure of the human O-GlcNAc transferase, modelled using the published structures of the TPR domain (PDBID 1W3B [230]) and the catalytic domain (PDBID 5C1D [231]).

The domains CDI and CDII were initially assigned catalytic activity, conferring nucleotide sugar specificity [229]. In parallel Shafi *et al.* showed that *ogt* resides on the X chromosome (Xq13) and is essential for the proper development for mESCs [50].

1.6.2. Substrate recognition and catalytic mechanism of OGT

Given the large number (> 1000) of O-GlcNAcylated proteins found in in the cell, the question arises how a single enzyme is capable of recognising and modifying all of them. Four, not necessarily mutually exclusive, possible modes of recognition were postulated (i) the globular domain or disordered tails of protein substrates interact with the TPR domain, (ii) the TPR domain allows the binding of adapter proteins that modulate activity and specificity (iii) OGT recognises its substrates via a preferred sequence and finally (iv) substrates are recognised by the catalytic domain. Initial experiments with elegant truncations of the TPR domain at various points and competition experiments have demonstrated that the domain confers recognition of protein substrates such as Nup62, Casein kinase II (CKII), GSK3β, RNA Polymerase II and OIP106 (now known as Trafficking kinesin-binding protein 1 – TRAK1) but is not essential for catalytic activity on peptide substrates [72,76,141,232]. The first structure of the TPR domain was solved in 2004 [230] and revealed a right-handed, elongated superhelix, consisting of α-helical bundles and exhibits structural similarities to armadillo repeat (ARM)-containing proteins such as β-catenin [233] and importin α [234]. The TPR superhelix measures a length of ~ 100 Å, a width of 35 Å and a pitch of 55 Å, with seven TPR repeats forming a complete turn. The concave surface of the superhelix is lined with, highly conserved repeating asparagine residues, similar to importin α [234] and are suggested to be involved in substrate recognition [230].

These asparagine residues form a ladder on the inner surface of the TPR repeats [230] similar to the asparagine array found in the ARM-repeat proteins importin α [235] and β-catenin [236], which form hydrogen-bonds with the peptide backbone of substrates. Indeed, a recent crystallographic study has shown that OGT recognises host-cell factor 1 (HCF1) using this recognition mode [237] (discussed in more detail in chapter 1.6.3).

Initial structural insights into the overall fold of OGT came from a bacterial homologue produced by *Xanthomonas campestris* [238,239], revealing that the active site is made up of two catalytic lobes, both adopting a glycosyl transferase B (GT-B) fold tightly fused to a superhelical TPR domain [238]. Shortly after, the structure of the human O-GlcNAc transferase was reported, revealing a similar fold for the catalytic and TPR domains [240]. The catalytic domain harbours an intervening domain (Fig. 1.8), of as of yet unknown function, separating the two GT-B fold lobes forming the active site. Although initially it was thought that OGT binds its substrates in a random bi-bi mechanism [72], the determination of a complex of OGT, UDP and a peptide suggests that OGT binds its substrates in an ordered bi-bi mechanism [240]. The structural characterisation of a pseudo-Michaelis complex of the donor sugar analogue UDP-5S-GlcNAc (discussed in greater detail in chapter 1.7.1) and a peptide derived from TAB1 gave further insights into how OGT binds its substrates. UDP-GlcNAc binds first in a deep pocket, adopting a “back-bent” conformation, forming extensive interactions with OGT followed by the binding of the acceptor substrate, which forms mostly van der Waals interactions (Fig. 1.9) [240,241]. Interestingly, the nucleotide forms the majority of interactions with OGT (Fig. 1.9 D) [240–242]. The uracil ring forms hydrogen bonds with the backbone carbonyl of Ala^{906} and Arg^{904}, while the 2’- and 3’-hydroxy groups of the ribose form hydrogen bonds with Asp^{935} and Lys^{908},

respectively. Remarkably, the largest number of interactions is formed by the pyrophosphate moiety of UDP. Interestingly, the α-phosphate forms hydrogen bonds with Gln^{849} and the amide group of the acceptor residue backbone. Notably, the above mentioned “back-bent” conformation brings one of the non-bonding oxygens of the α-phosphate into close proximity (~ 2.8 Å, Fig. 1.9 A) of the acceptor hydroxyl group. The *β*-phosphate alone forms a large number of polar interactions with OGT. Aside from the interaction with Lys^{842}, which is crucial for tethering UDP-GlcNAc, the β-phosphate forms an extensive array of interactions with His^{920}, Thr^{921} and Thr^{922}. Additionally, the latter three residues form part of an *α*-helix creating an induced dipole moment, which introduces yet another interaction with the *β*-phosphate [241,243]. Interestingly, the sugar is tethered into place by interactions with the three residues Thr^{570}, Leu^{663} and His^{930}. Furthermore, it has been shown that the sugar moiety barely contributes to binding [241].

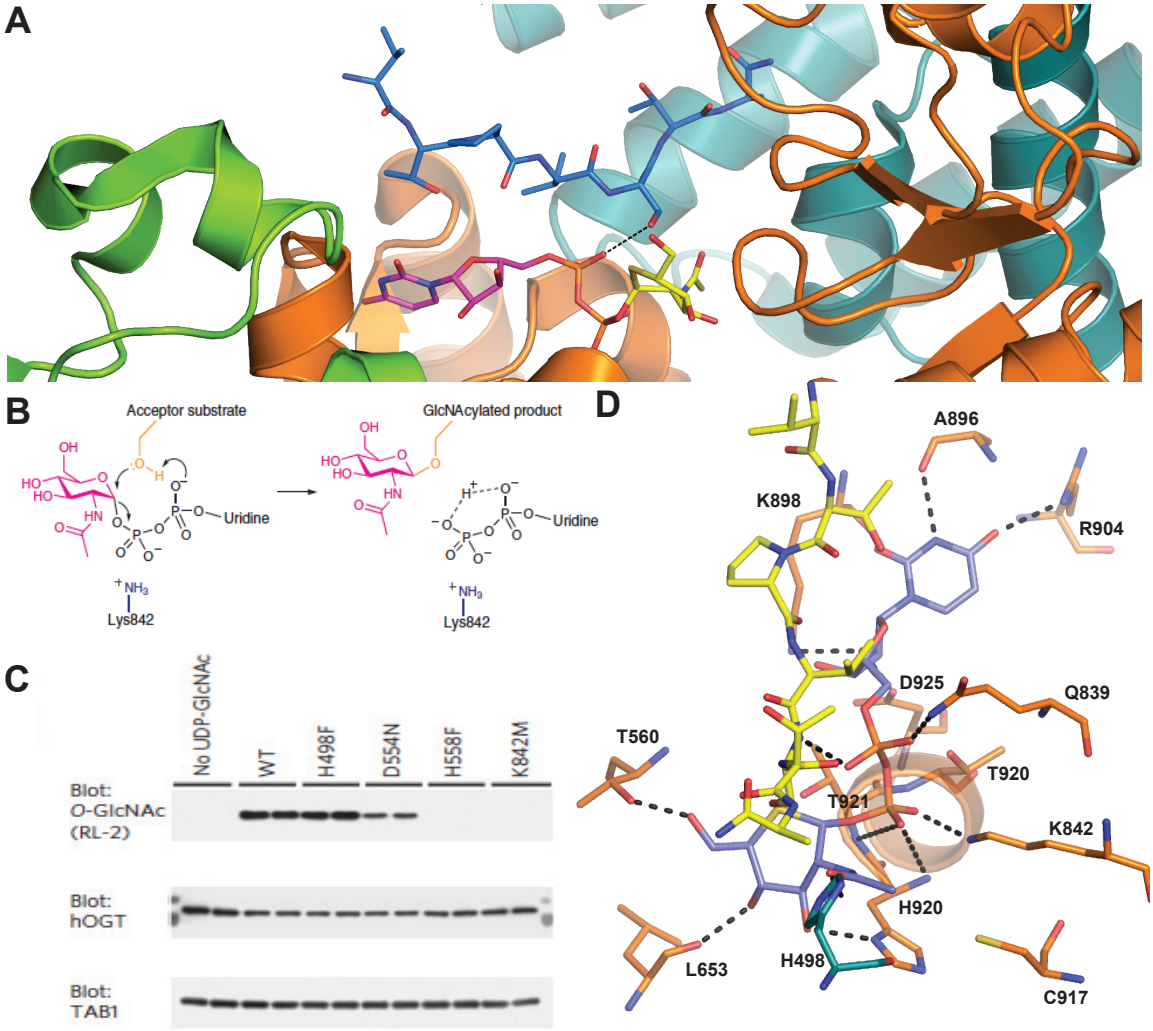

Figure 1.9: OGT binds its substrates in a ordered bi-bi mechanism and forms an intricate network of interactions. (A) The donor substrate analogue UDP-5S-GlcNAc binds first in a deep pocket followed by the binding of the acceptor substrate. The dashed-line highlights the close distance between the α-phosphate of UDP-5S-GlcNAc and the acceptor hydroxyl-group of the acceptor substrate. OGT is shown in cartoon and surface representation, with the catalytic lobes in orange, the intervening domain in green and the TPR domain in deep teal. The surface is shown in white. UDP-5S-GlcNAc, a non-hydrolysable UDP-GlcNAc analogue, is shown as magenta (UDP) and yellow (GlcNAc) sticks and the acceptor peptide VTPVSTA as blue sticks. **(B)** Schematic drawing of the proposed catalytic mechanism by which OGT glycosylates its acceptor substrates, via a substrate-assisted catalysis. (Figure taken from [241]). **(C)** Western blot analysis showing the activity of point-mutants of OGT and their activity on the TAK1-binding protein (TAB1) (Figure taken from [241]). **(D)** Stick representation of a pseudo-

Michaelis complex of OGT and its substrate analogue UDP-5S-GlcNAc and a peptide derived from the Retinoblastoma-like 2 protein (RB2L). The acceptor peptide substrate is shown as yellow sticks, UDP-5S-GlcNAc is shown as blue sticks and the hOGT residues as orange (catalytic domain) and teal (TPR domain) sticks. The α-helix creating the electric dipole, which interacts with the β-phosphate of the nucleotide is shown in cartoon representation. Polar interactions are shown as black dashed lines (PDBID 5C1D [231], Figure adapted from [241]).

The determination of the human OGT structure in 2011 [240], allowed the probing of the active site and identification of the catalytic base and elucidation of the catalytic mechanism. Shortly after, two different catalytic mechanisms of O-GlcNAc transfer were independently proposed. First, Lazarus *et al.* proposed a Grotthus mechanism, where a chain of water molecules, acting as a proton relay, links the proposed general base Asp^{554} to the anomeric carbon [242]. Although it is possible for the water molecules to act in this way, there were no experiments provided to investigate whether a mutation of Asp^{554} leads to a loss in catalytic activity. The second proposed mechanism claims that OGT invokes the α-phosphate of UDP-GlcNAc to act as a catalytic base (Fig. 1.9 B) [241]. Schimpl *et al.* argue that due to the unique conformation of the bound UDP-GlcNAc, the α-phosphate is in an optimal position to polarise the hydroxyl of the acceptor residue, thereby raising the nucleophilicity of the hydroxyl-oxygen [241]. Remarkably, when comparing the torsion angles of all other structurally characterised GT-B family enzymes, both retaining and inverting, the angles are clustered closely together contrary to the angle of UDP-GlcNAc bound to the human O-GlcNAc transferase [241]. Lys^{842}, even though not directly involved in the catalytic mechanism, is crucial for the proper placement and binding of the nucleotide sugar and stabilises the negative charge on the leaving group.

Mutagenesis studies showed that a loss of Lys[842] results in a catalytically inactive enzyme (Fig. 1.9 C) [241]. Additional experiments, using phosphorothioate UDP-GlcNAc derivatives, which do not allow the abstraction of the acceptor hydroxyl proton, further supported the mechanism proposed.

Recent advances in high-throughput proteomics and enrichment tools have greatly extended not only the list of the O-GlcNAcome but also allowed the identification of specific O-GlcNAc sites [33–39,149,244]. Upon compiling the identified O-GlcNAc sites certain sequence motifs surrounding the O-GlcNAc modification emerged [43,231,245–251]. Three main approaches were taken in order to investigate a common motif for O-GlcNAcylation (i) sites identified by pull-down experiments form cell or tissue lysates, followed up by mass spectrometry (MS) analysis [43,248,250], (ii) measuring OGT activity on a library of peptide substrates followed up by MS analysis of identified substrates [231,247,252,253] and (iii) bioinformatic prediction based on all known O-GlcNAc sites [245,246,249,251]. Arguably the most unbiased approach in identifying a common sequence motif (sequon) for OGT substrates is the use of large pull-down studies of the O-GlcNAcome using (i) antibodies recognising the O-GlcNAc epitope [33,254,255], (ii) affinity enrichments with lectins like wheat germ agglutinin [36,39,43,256] or a recently reported affinity enrichment using a catalytically inactive mutant of a bacterial OGA [244] or (iii) the chemo-enzymatic labelling of the GlcNAc moiety [34,38,83,257,258]. Although a lot of knowledge on the stoichiometry, abundance and sites of O-GlcNAcylation is gained by performing proteomic analysis, it is still difficult to elucidate a common sequence motif for O-GlcNAc substrates. Most studies report a serine/threonine rich sequences flanking the glycosylation site [36,39,43]. This however is not necessarily surprising, as there is evidence for the processive modification of long stretches of serine/threonine residues by OGT, such as the C-terminal domain

(CTD) of RNA Polymerase II [259,260] and various NUPs [108]. It is very likely that peptides, containing a large number of O-GlcNAc modifications prevent the detection of peptides with lower occupancy. However, one striking feature emerged from the proteomic studies, which is that a large population of identified O-GlcNAc peptides contained small hydrophobic residues such as alanine and valine around the modification site and a proline residue in the -2 position.

A more biased experimental approach taken to identify common features of O-GlcNAc substrates, is measuring OGT activity on libraries containing a diverse set of peptides [231,247,252,253]. In their study Leavy and Bertozzi, used a small library of α-crystalline derived peptides by substituting residues flanking the O-GlcNAc site and visualised OGT activity using an azido-ELISA [247,261,262]. Furthermore, they showed that the peptide can be improved, suggesting an advantage of having a proline in the +2 position but not in the -2 position as found in proteomic studies [34,36,38,43,83,257]. Liu *et al.*, used information from previous proteomic data to design a peptide library based on the a common sequence PPVS/TATT they identify from frequency analysis [253]. They reported a series of preferred amino acids found in the -2, -1 and +2 position surrounding the O-GlcNAc site, with the most preferred residues being proline, alanine and alanine, respectively [253]. More recently, Pathak *et al.* identified a degenerate sequon of OGT peptide substrates ([TS][PT][VT][S/T][RLV][ASY]), by screening a library consisting of 720 physiologically relevant peptides against the human O-GlcNAc transferase [231]. Furthermore, crystallographic studies of four of the peptides, identified as being substrates for OGT, bound in an extended conformation with similar conformation of the residues flanking the acceptor residue in the -3 to +2 positions [231]. Finally, Shi *et al.* developed a peptide-microarray screen, based on kinase substrates and nuclear receptor binding co-regulators, to detect novel OGT substrates [252].

The authors identified a peptide derived from the retinoblastoma-like 2 (RBL2) protein (410-422, CGKENSPCVTPSTA) as the best OGT substrates [252], confirming previous results [231]. They further tried to identify the O-GlcNAc site using an alanine-scanning approach, where by substituting individual sites with alanine residues lead to a loss in glycosylation [252]. They were able to confirm the previously identified Ser^{420} site as being one of the possible O-GlcNAc sites [252]. Interestingly, mass spectrometric analysis of an *in vitro* O-GlcNAcylated, RB2L peptide spanning residues 411-422 only identified a single O-GlcNAc site on residue Ser^{420} [231]. Furthermore, Pathak *et al.* could show that the position of the acceptor site in the active site is good agreement with the MS results, despite the presence of two additional potential O-GlcNAc sites [231].

A more theoretical approach is the prediction of O-GlcNAc sites based on common features, such as sequence and structure [246,249,251,263]. Bioinformatic software make use of publically available proteomics data [264,265] and structure databases [266] to identify common features. Although having potential, computational methods to predict O-GlcNAc sites have only been used with limited success [245,246,267–269]. A recent study investigated the contribution of structural motifs of OGT substrates [251]. Interestingly, the authors found that although some O-GlcNAc sites can be found in ordered regions of protein substrates, the vast majority lies in disordered regions [251]. The authors further claim that OGT may induce a partial unfolding of protein substrates, thereby introducing disorder making it easier to glycosylate [251]. This however is rather unlikely as the unfolding of a protein under physiological conditions, requires external force [270]. It is more likely that O-GlcNAc sites found on ordered regions of proteins are remnants of the co-translational addition of O-GlcNAc to nascent polypeptide chains [10].

Structural and biochemical characterisation of the TPR domain and the catalytic domain of OGT have greatly accelerated our understanding of this unique glycosyltransferase. Recent studies proposed various mechanisms on how OGT recognises and glycosylates its substrates on a peptide and protein level, which may or may not be mutually exclusive. It was demonstrated that the TPR domain is important for substrate recognition and specificity and more recently in the binding of long disordered peptides, via the asparagine array on its concave surface. Structural studies showed that OGT binds its substrates in an ordered bi-bi mechanism, with the nucleotide sugar binding ahead of the acceptor residue. Finally, high-throughput proteomic experiments and peptide library screens revealed a degenerate sequence motif (sequon), by which OGT binds substrates. However, all these experiments addressed one particular part of substrate binding/recognition in isolation and the full picture is still elusive. Furthermore, some studies suggested the possibility of multiple binding modes working in concert. Finally, no study has addressed the role of the intervening domain, located between the two catalytic lobes, which could be involved in the recognition of the globular domain of protein substrates.

1.6.3. Mechanism of OGT catalysed proteolytic maturation of HCF-1

Host-cell factor 1 (HCF-1) was initially characterised as one of two host factors present in the VP-16 induced complex by the Herpes Simplex Virus (HSV) to initiate its early gene expression [271] but was soon found to be a ubiquitously expressed transcriptional co-regulator involved in a broad range of cellular processes such as cell-cycle progression (reviewed in [272]). HCF1 is initially expressed as a 210 kDa protein, encoded by a single gene and consists of 4 distinct domains (i) an N-terminal chromatin-associating Kelch-domain followed by fibronectin-like (FN3) repeats, (ii) a disordered region with a high proportion of basic residues (basic region), (iii) a C-terminal region containing a high proportion acidic residues (acidic region) and (iv) two additional FN3 repeats followed by a nuclear localisation sequence [271,273,274]. To achieve full functionality HCF-1 has to undergo proteolytic maturation by limited proteolysis inside a stretch of 20 aa ($HCF1_{PRO}$) repeats within the central proteolytic processing domain (PPD), located between the basic and acidic region [275,276]. Proteolytic cleavage occurs at one or more $HCF1_{PRO}$ repeats, generating multiple species of N- and C-terminal subunits, capable of forming heterodimers via the FN3 repeats [277]. Early studies suggested that proteolytic maturation of HCF-1 is required to coordinate two major functions in cell cycle progression, where the N-subunit is needed for proper transition from G1 to S-phase and the C-subunit for progression through mitosis [278]. However, the precise role of the precursor and the mature form is still elusive. Interestingly, in 2003 it was reported that OGT glycosylates and associates tightly with HCF-1, even after modification took place [83,139]. Shortly after Daou *et al.* reported evidence for a symbiotic interaction between OGT and HCF-1 [57]. The authors found that OGT is essential for HCF-1 cleavage and is in turn stabilised by it. Moreover, they show that OGT glycosylates HCF-1 in the

PPD domain, which is however not essential for proteolysis [57]. In parallel Capotosti *et al.* identified OGT as the protease catalysing the proteolytic maturation of HCF-1, suggesting a catalytic dual-functionality of OGT [279]. First insights into the mechanism of $HCF1_{PRO}$ binding and proteolytic cleavage were reported recently [237]. Lazarus *et al.* found that a short $HCF1_{PRO}$ peptide (VRVCSNPPCETHETGTTNTATTATSN) bound in an extended conformation inside the TPR domain [237] (discussed in greater detail in chapter 2).

A mechanistic study revealed that OGT cleaves HCF1 via step-wise, substrate assisted catalysis, invoking UDP-GlcNAc and a glutamate present in the $HCF1_{PRO}$ repeat sequence (VRVCSNPPCETHETGTTNTATTATSN) [280]. After initial O-GlcNAcylation of the glutamate a nucleophilic attack on the carbonyl carbon via the amide nitrogen takes place (Fig. 1.8, intermediate 1) forming a pyroglutamate (Fig. 1.8, intermediate 2), which ultimately leads to the fission of the amide bond between the adjacent cysteine and the pyroglutamate via a final hydrolysis by a water molecule (Fig. 1.8) [280].

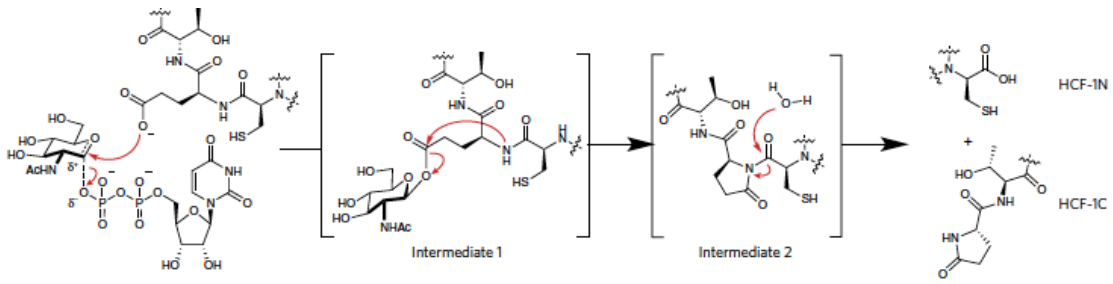

Figure 1.10: Proposed molecular mechanism of the proteolytic maturation of HCF1 by OGT. OGT induces cleavage of the HCF1 protein by O-GlcNAcylation of a glutamate, priming the formation of a pyroglutamate, which ultimately allows the fission of the peptide bond by a nucleophilic attack of a water molecule. Figure taken from [280].

1.7. O-GlcNAc transferase inhibitors

Elucidating the biological role of O-GlcNAc has been a slow process, due to various bottlenecks. As mentioned above, both OGA and OGT are essential proteins, necessary for proper development [49,50,75,91] and any genetic approaches to knock-out either of these two genes have resulted in embryonic lethality. However, in recent years inducible and localised knock-downs of OGT and OGA have been achieved allowing the study of their functionality in certain tissues [138,281,282]. However, the genetic manipulation of *oga* or *ogt* has additional drawbacks (i) total gene knock-outs result in a complete loss of OGA and OGT protein not allowing the investigation of potential non-catalytic functions [283], (ii) genetic manipulations of organisms is labour intensive and finally, (iii) they do not allow the spatiotemporal investigation of OGA/OGT activity. As described previously, potent and selective inhibitors were reported for OGA [207,217–219,284] and already extensively used in cell culture experiments and in some cases organisms [34,117,145,217,220,221].

However, there remains a need for potent and selective inhibitors of OGT (Fig. 1.13) as inhibitors reported to date lack selectivity and/or potency and results of cell culture experiments remain questionable. The most commonly used OGT inhibitors are discussed in the following chapter.

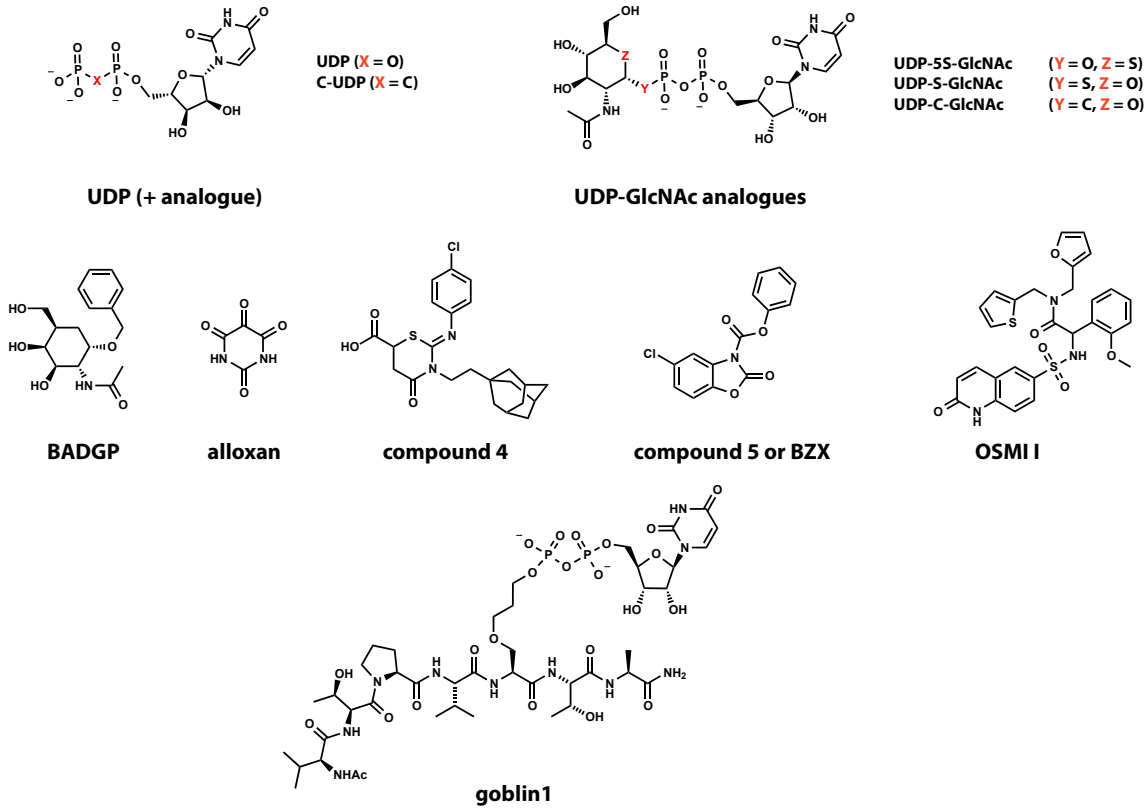

Figure 1.11: Chemical structure of the various OGT inhibitors reported. Substrate and product analogues: UDP (+analogue), UDP-GlcNAc (+analogues), BADGP, alloxan; Small molecule inhibitors: compound 4, compound 5 or BZX, OSMI1; Bisubstrate inhibitor: goblin 1. Figure taken from [243].

1.7.1. Substrate and product analogue inhibitors

Initially several compounds were found to inhibit OGT, among them 2,4,5,6-tetraoxypyrimidine more commonly known as alloxan and benzyl 2-acetamido-2-deoxy-α-D-galactopyranoside, also known as BADGP (Fig. 1.13). Alloxan is a uracil analogue, capable of inhibiting OGT in the mid µM range (IC_{50} = 100 µM) [285,286] and a glucose mimetic, which is readily taken up GLUT2 transporters [287]. It has been widely used in in cell culture and animal models [199,200,288–292]. However, the resulting insights are questionable due to a large number of off-target effects including OGA [293], general cellular toxicity, short half-life [294] and the generation of reactive oxygen species (ROS) [295]. For instance, comparative study of alloxan and compound 5 (Fig. 1.13, discussed later) proving the requirement of O-

GlcNAc for entry into M-phase [296], previously not seen [289]. BADGP is able to decrease global O-GlcNAc levels by inhibiting the biosynthesis of UDP-GlcNAc [67]. Unsurprisingly, BADGP affected the glycosylation of and synthesis of mucin oligosaccharide chains , highlighting its unspecific nature [297]. Interestingly, the most potent inhibitor of OGT to date is the reaction product UDP (IC_{50} = 1.8 μM) [241,285], although it is unsuitable for cell biology studies as it is not cell permeable and highly unspecific, likely targeting other UDP binding proteins. Nevertheless, the high potency of UDP inspired the synthesis of various UDP and UDP-GlcNAc analogues, in the hope of utilising the high binding affinity of UDP towards OGT [241]. Dorfmüller *et al.*, reported three substrate/product analogues that inhibited OGT with low μM affinity; UDP-S-GlcNAc (IC_{50} = 93 μM), UDP-C-GlcNAc (IC_{50} = 41 μM) and C-UDP (IC_{50} = 9 μM) [285]. However, they are also not usable in cell biology studies due to their polar nature and low cell permeability. An elegant solution to the problem of cell permeability was recently reported by Gloster *et al.* [298]. The authors reported the development the GlcNAc analogue 5S-GlcNAc that could be used to “hijack” the hexosamine biosynthetic pathway (HBP) and be metabolised into the substrate analogue UDP-5S-GlcNAc, capable of inhibiting OGT [298]. The sugar analogue was administered to COS-7 cells in its, more soluble peracetylated form Ac_4-5S-GlcNAc and significantly reduced global O-GlcNAc levels at low μM concentrations *in cellulo* (EC_{50} = 5 μM). Since then Ac_4-5S-GlcNAc has been used in multiple studies [299–302]. For example, Ac_4-5S-GlcNAc was used to investigate the role of O-GlcNAc mediated Sp1 modulation [299]. The authors found that decreased O-GlcNAcylation of the transcription factor Sp1 leads to a down-regulation of the pro-angiogenic factor VEGF-A (vascular endothelial growth factor – A), which is involved in retinal vascularisation [299]. Another study investigated the effect of O-GlcNAcylation in pancreatic cancer

cells, finding a suggestive correlation between hyper-O-GlcNAcylation and cell-death (apoptosis) via modulation of NF-κB anti-apoptotic transcriptional activity [301]. However, the major drawback of Ac_4-5S-GlcNAc is that, by hijacking the HBP, it reduces the intracellular UDP-GlcNAc, thereby affecting other UDP-GlcNAc utilising enzymes. Consequentially, N-glycosylation of proteins and extracellular glycan synthesis might be affected [303].

1.7.2. Small molecule inhibitors of OGT

A different approach for identifying inhibitors of enzymes, utilises high-throughput screens of compound libraries containing a large number of small, drug-like molecules. Using this approach, Gross *et al.* have identified a handful of molecules that were capable of inhibiting human OGT *in vitro* [296]. Among these were 3-(2-adamantanaylethyl)-2-[(4-chlorophenyl)azamethylene]-4-oxo-1,3-thiazapehydroine-6-carboxylic acid, also known as compound 4 (Fig 1.11) and 4-methoxyphenyl 6-acetyl-2oxobenzo[d]oxazole-3(2H)-carboxylate, referred to as BZX or compound 5 (Fig. 1.11) [296]. Both compounds showed low μM inhibition of full length human OGT [296] and showed activity *in cellulo* [100,202,304–307]. However, the molecular mechanism by which compound 4 inhibits OGT is still unknown, whereas it was shown that BZX is a covalent inhibitor [308]. Crystallographic and mechanistic studies revealed that BZX covalently links Lys^{842} and Cys^{917} together, via a double-displacement mechanism, rendering the enzyme inactive [308]. Even though BZX acts as a potent inhibitor of OGT, one cannot exclude potential off-target [308]. More recently, another small molecule inhibitor, OSMI1, of OGT was reported [309]. This non-competitive inhibitor was active in cellular studies significantly reducing global O-GlcNAc levels [309] and inhibiting the replication cycle of the herpes simplex virus (HSV) [310]. One major drawback however, is the

off-target effects of OSMI1, affecting surface glycosylation through an unknown mechanism [309].

1.7.3. Bisubstrate inhibitors targeting OGT

The advances we have made in the understanding of the binding mode of substrates [240] and the catalytic mechanism of OGT [241] has facilitated the development of bisubstrate conjugates that inhibit OGT with low μM affinities [311]. The OGT bisubstrate-linked inhibitor Goblin 1 was created by tethering a short peptide (VTPVSTA) to UDP via a short three-carbon linker (Fig. 1.11). Structural studies revealed that Goblin 1 binds in the active site of human OGT to form a pseudo-Michaelis complex, mimicking substrate binding, and inhibiting both human [311] (IC_{50} (human) =18 μM) and *Trichoplax adhaerence* OGT [44] *in vitro*, (IC_{50} (*T. adherence*) = 27 μM). The major drawback of this approach is the limited cell permeability exhibited, due to the high polarity and hydrophilicity of this conjugate, barring it from being used in cell culture. Furthermore, the specificity of Goblin 1 has not been exhaustively tested, apart from testing it against the bacterial GlcNAc transferase *Sm*NodC [311]. Future developments need to address the lack of cell permeability to make this a viable tool to study O-GlcNAc transferase biology.

1.7.4. Common features exhibited by OGT inhibitors

Examining the binding mode of OGT inhibitors, based on structural data, allows the elucidation of common binding features. Not surprisingly, product and substrate analogues UDP [240] and UDP-5S-GlcNAc [241], as well as the Michaelis-complex mimicking inhibitor Goblin 1 [311] possess similar binding features. All form extensive polar interactions between OGT and the UDP moiety, especially with

the pyrophosphate. As described earlier, the β-phosphate forms the majority of interactions with OGT, including Lys^{842} and an induced dipole moment of an active site α-helix formed by His^{920}, Thr^{921} and Thr^{922}. Strikingly, the GlcNAc residue forms few interactions with the enzyme and has previously been shown to not significantly contribute to binding [241]. Mutagenesis studies of active site residues have shown the importance of Lys^{842} in the catalytic activity [241], which is further supported by the binding mode of BZX [308], making it an interesting target of inhibitor design. Therefore, future inhibitors should be designed to specifically target this crucial residue.

1.8. Aim of the thesis

The biological role of the O-GlcNAc transferase and the mechanism by which it recognises its substrates is still poorly understood. Recent work has indicated that OGT binds its substrates using different recognition modes simultaneously. In my first aim I investigate the role of the asparagine residues, found in regular spacing throughout the concave surface of the TPR domain, contribute to binding of glycosylation substrates of OGT. Using biochemical analysis and steady-state kinetics I investigate, whether the binding mode of the a short TAB1 18-mer tightly fused to OGT via a three-glycine linker represents a physiologically relevant binding mode. Analysis of this linear fusion construct could pave the way to trap OGT in complex with a protein substrate allowing us to structurally characterise interactions between the globular domains of a substrate protein and the catalytic and TPR domain of OGT. Furthermore, this would be the first insight into the cooperative effect of the different binding modes.

The lack of potent and specific OGT inhibitors hampers our understanding of the spatiotemporal role of OGT catalytic activity during various stages in development and disease progression. In my second aim I design and develop novel potent OGT inhibitors suitable for cell culture and *in vivo* studies. Using high-resolution macromolecular X-ray crystallography, biochemical and biophysical methods I examine a panel of bisubstrate OGT inhibitors based on the published Goblin 1 structure (Fig. 1.13). Apart from increasing the potency of these bisubstrate conjugates I try to achieve *in cellulo* activity by attaching cell-penetrating peptide sequences to the inhibitor scaffolds.

2. The human O-GlcNAc transferase recognises its substrates inside the superhelical fold of its tetratricopeptide repeat domain

The work from this chapter has been published in the following article:

Rafie & Raimi *et al.*, “Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR Repeats.”, Open Biology, 2017

2.1. Introduction

OGT is a multi-domain protein with a catalytic core at the C-terminus and 13 tetratricopeptide (TPR) repeats at the N-terminus, making up about half of the enzyme. Early experiments suggested that the TPR domain is involved in substrate recognition and/or protein-protein interactions [72,76,141,147,227,230,232,237]. The structure of the isolated OGT TPR domain revealed topological similarity to other helical repeat proteins, such as importin α (Fig. 2.1), folding in a superhelical fold comprised of α-helices and led to speculation that this domain might bind substrates in an extended conformation [230].

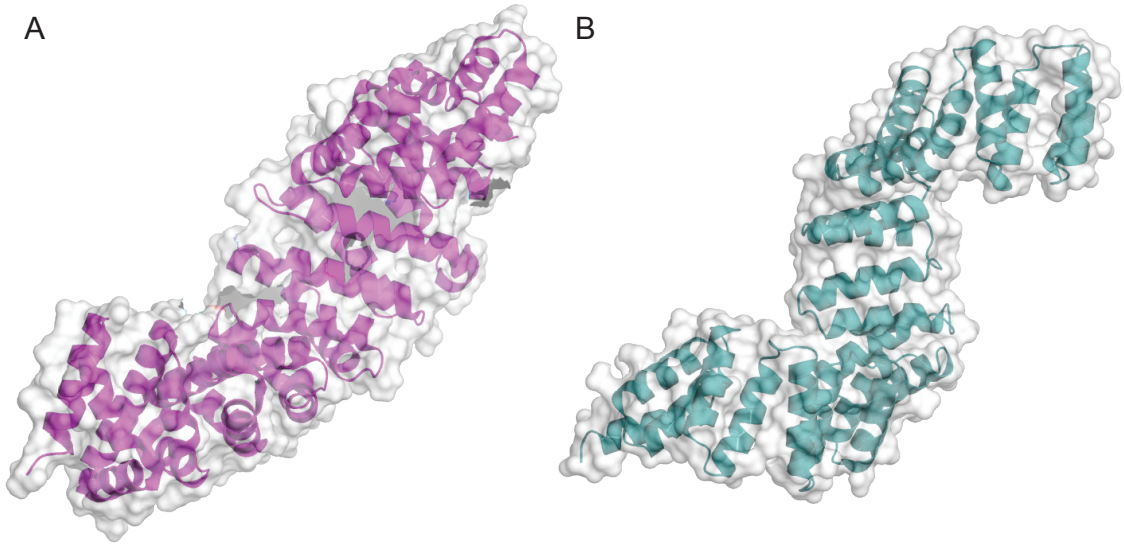

Figure 2.1: The superhelical fold of the human O-GlcNAc transferase TPR domain is reminiscent of the Importin α structure. (A) Structure of Importin α (PDBID 5K9S) The protein is shown as a composition of a cartoon and surface representation in magenta and white, respectively. **(B)** The hOGT TPR domain (16-400, PDBID 1W3B) folds in a superhelical structure comprised of pairs of α-helices. The TPR domain is shown as a composition of a cartoon and surface representation in teal and white, respectively.

The first structural insights into the OGT catalytic domain came from an OGT orthologue from the bacterium *Xanthomonas campestris* (*Xc*OGT) [238,239]. This structure revealed that the sugar donor binding site is made up of the two lobes of the GT-B fold, tightly fused to the superhelical TPR domain [238]. The subsequent structure of human OGT [240] revealed a very similar fold with the addition of an intervening domain of unknown function, which is flanked by the two catalytic lobes [240] (Fig. 2.2).

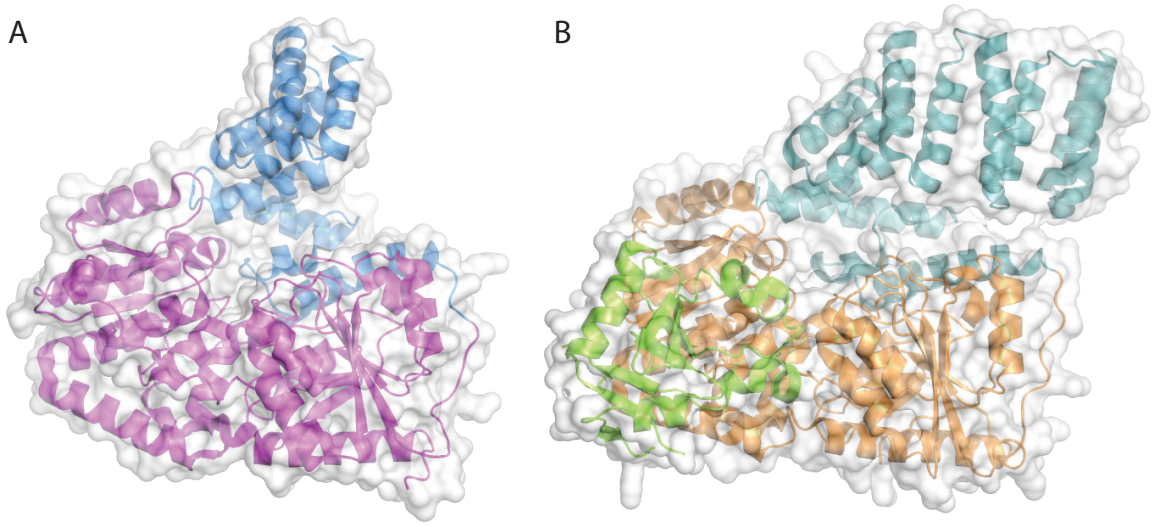

Figure 2.2: The human OGT structure reveals similarities to its *Xc*OGT orthologue. **(A)** The structure of the O-GlcNAc transferase from *Xanthomonoas campestris* (*Xc*OGT, PDBID 2XGM [238]) reveals a catalytic domain (magenta), comprised of two catalytic lobes tightly fused to its TPR domain (blue). The protein is shown in cartoon and surface (white) representation. **(B)** The structure of the human O-GlcNAc transferase (312-1031) is comprised of two distinct domains. The catalytic domain, which is comprised of the two catalytic lobes (orange) separated by an intervening domain of unknown function (green), and its TPR domain (teal). The protein is shown in cartoon and surface (white) representation.

The structure suggested an ordered bi-bi mechanism of substrate binding, in which UDP-GlcNAc binds before the acceptor substrate [240]. Initial structural studies exploring Michaelis/substrate complexes with short acceptor peptides

have revealed limited substrate interactions with the enzyme [241,242]. More recently Pathak *et al.* investigated the common binding modes of acceptor peptides to OGT [231]. Starting from a peptide library they identified preference for certain acceptor peptide sequences, leading to definition of a degenerate sequon of OGT peptide substrates ([TS][PT][VT][S/T][RLV][ASY]). Crystal structures of complexes of OGT with some of these peptides revealed that OGT binds all the acceptor peptides studied so far in an extended conformation with similar conformation of the residues in the -3 to +2 position around the acceptor serine/threonine. Although the C-termini of these peptides point towards the TPR domain these structural data do not explain how OGT recognises larger protein substrates for glycosylation. The short sequence patterns alone are not sufficient to accurately predict the O-GlcNAc proteome, suggesting other mechanisms contribute to substrate recognition.

A clue to how this might work came from the unusual OGT substrate host cell factor 1 (HCF1). HCF1 is a ubiquitously expressed chromatin-associated protein and a major transcriptional co-regulator involved in numerous cellular processes such as cell cycle progression (reviewed in [272]), which has also been shown to be heavily O-GlcNAcylated [279]. HCF1 is initially expressed as a ~210 kDa protein that is activated by limited proteolysis (protease maturation) within the proteolytic processing domain (PPD), consisting of multiple 20-residue repeats [276,312]. Strikingly, in 2011 it was discovered that OGT not only glycosylates HCF1 but is also needed for its proteolytic maturation [57,279]. A depletion of OGT leads to an accumulation of full length HCF1 protein and the PPD is proteolytically cleaved by OGT via an unusual glycosylated glutamate intermediate [237,279,280]. A recent structural study of a short PPD ($HCF1_{PRO}$) repeat in complex with OGT revealed that part of the substrate bound in extended conformation in the TPR repeats [237].

OGT was shown to form an extensive array of polar interactions with the backbone of the $HCF1_{PRO}$ repeat peptide, as well as specific side chain interactions that were demonstrated to be essential for $HCF1_{PRO}$ binding [237]. However, it is as yet not clear if this binding mode also extends to OGT glycosylation substrates.

A well characterised OGT glycosylation substrate is the TGFβ activated kinase 1 (TAK1) binding protein 1 (TAB1), a pseudophosphatase involved in the TGFβ-mediated inflammatory signalling pathway and found to be an essential activator of TAK1 [313,314]. The structure of the TAB1 N-terminal pseudophosphatase domain has been reported and revealed similarity to the PPM family of protein Ser/Thr protein phosphatases [315]. Previous studies have shown that phosphorylation at a C-terminal region of TAB1 regulates TAK1 activity [316–318]. We have recently discovered that TAB1 is dynamically O-GlcNAcylated at Ser^{395} in the C-terminal domain [117]. This glycosylation appears to be required for full activity of TAK1 and activation of downstream transcription and secretion of pro-inflammatory cytokines. Here, a novel approach to covalently trap OGT-substrate complexes is used to explore how OGT recognises glycosylation substrates through its TPR domain. The structure of hOGT in complex with the TAB1 C-terminal domain combined with mutagenesis studies reveals that OGT recognises the TAB1 substrate, and by extension a group of glycosylation substrates with similar disordered regions, through extensive essential interactions with the TPR repeats.

2.2. Materials & Methods

Molecular biology was performed by Dr. Andrew T. Ferenbach, peptide synthesis was performed by Dr. Vladimir S. Borodkin and the fusion construct expression, purification, crystallisation and crystallographic data analysis was performed by Dr. Olawale Raimi.

2.2.1. In vitro TAB1 glycosylation assay

For glycosylation assays, TAB1 and hOGT (WT and mutants) were expressed and purified as described previously [231,240,241,315]. 10 μM of TAB1 was incubated with 50 nM hOGT (WT or mutants) in TBS reaction buffer (0.1 M Tris-HCl pH 7.4, 150 mM NaCl) supplemented with 0.5 mM TCEP and 0.1 mg/mL BSA. The reaction was started by adding UDP-GlcNAc to a final concentration of 100 μM and incubating the reaction mixtures at 25 °C. 10 μL of sample mixtures was taken at indicated times and mixed with 4x LDS sample loading buffer to a final volume of 50 μL and boiled at 95 °C for 5 minutes. Proteins were resolved using precast SDS-PAGE gels (NuPAGE 4-12% Bis-Tris gels, Invitrogen) and blotted onto nitrocellulose membranes (GE Healthcare). The primary antibodies were used at the following concentrations: Anti-TAB1-O-GlcNAc (1:1000 [117]), anti-TAB1 (1:1000, Division of Signal Transduction and Translation, University of Dundee) and anti-OGT (1:2000, DM17, Sigma-Aldrich, Cat#: O6264). Li-Cor secondary antibodies (IRDye 680 Donkey anti-rabbit and IRDye 800 Donkey anti-rabbit, anti-sheep) were used at dilutions of 1:10,000. Blots were imaged using the Li-Cor Odyssey infrared imaging system (Li-Cor, Lincoln, NE). Quantification of the O-GlcNAc specific signal (gTAB1) was performed using imageStudioLite (Li-Core) and normalised to total OGT (tOGT) and total TAB1 (tTAB1) signal. Data was plotted using GraphPad Prism 7.

2.2.2. Steady-state kinetics

hOGT activity was determined in reactions containing 50 nM either WT or 5N5A His_6-hOGT (312-1031), 50 mM Tris-HCl pH 7.4, 0.1 mg/mL BSA, 10 μM sodium dithionate and varying concentrations of the TAB1 peptide KKPVSVPYSSAQSTSKTSVTLSL or at a fixed concentration of 10 μM of the TAB1 peptide KKPVSVPYSSAQSTS, in a total volume of 100 μL. Reaction mixtures were preincubated for 15 minutes before initiating the reaction by adding UDP-GlcNAc to a final concentration of 50 μM. Reactions were incubated for 30 minutes at 21 °C before addition of 200 μL of 75 μM pyrocatechol violet / 15 μM fluorophore, a UDP-sensitive xanthene-based Zn(II) compound [231,311,319], in 25 mM HEPES pH 7.4, 10 mM NaCl, 50% (v/v) MeOH. UDP formation was detected on a Gemini EM fluorescent Microplate reader (Molecular Devices) using excitation and emission wavelengths of 485 nm and 530 nm, respectively. Turnover did not exceed 10% for either substrate. Data are presented as average of three measurements, with error bars showing s.e.m. Data were analysed using GraphPad Prism 7.

2.2.3. *Western blot analysis of purified TAB1:OGT*

Samples of purified TAB1:OGT fusion protein were incubated for 30 minutes at 37 °C in the presence and absence of ~10 μg/mL *Cp*OGA, a promiscuous bacterial O-GlcNAc hydrolase [185]. Samples were supplemented with 4x LDS-loading buffer and boiled for 5 minutes at 95 °C. A total of 0.5 μg of each untreated and treated TAB1:OGT fusion protein were subjected to SDS-PAGE analysis and transferred onto a nitrocellulose membrane (GE Healthcare), using a wet-transfer system (Invitrogen). The membrane was blocked in 5% BSA for 30 minutes at 21 °C before incubating with anti-O-GlcNAc AB (RL2, 1:1000, Abcam, Cat#: ab2739)

and anti-OGT AB (1:2000, Abcam, Cat#: 177941). Li-Cor secondary antibodies IRDye 680 Donkey anti-mouse (anti-O-GlcNAc) and IRDye 800 Donkey anti-rabbit (anti-OGT) were used at dilutions of 1:10000. Blots were imaged using the Li-Cor Odyssey infrared imaging system (Li-Cor, Lincoln, NE).

2.3. Results and Discussion

2.3.1. *The TAB1 O-GlcNAc site resides in a disordered region with similarity to other OGT targets*

The O-GlcNAcylation sites on the OGT substrates TAB1 [117], collapsin response mediator 2 protein (CRMP2) [38] and casein kinase 2 (CK2) [320] are located in disordered regions close to the C-terminus (Fig. 2.3 A). Although short peptides derived from these sites can be co-crystallised with OGT [231,241,242], we have been unsuccessful in using this approach with longer sequences/intact proteins to explore the role of the OGT TPR domain in substrate recognition. Aligning the sequences around the O-GlcNAc sites reveals similarities near the site of modification (Fig. 2.3 B). Remarkably, this is also similar to the proteolytic cleavage site of a $HCF1_{PRO}$ repeat with the major difference being a glutamate at the acceptor position (Fig. 2.3 B). Mutating this glutamate to a serine is sufficient to change the peptide from a proteolytic to a glycosylation substrate [237]. In the OGT:$HCF1_{PRO}$ structure [237] the peptide substrate spans the whole length of the TPR domain (Fig. 2.4 B). The peptide interacts with the TPRs through some of its side chains, but intriguingly five regularly spaced asparagine side chains in OGT form hydrogen bonds with the $HCF1_{PRO}$ peptide backbone in a sequence-independent manner (Fig. 2.4 B). We noted the fortuitous proximity of the $HCF1_{PRO}$ C-terminus to the OGT N-terminus ($C\alpha_{HCF1\text{-}1340}$-$C\alpha_{OGT313}$ ~12 Å, Fig. 2.4 B), and wondered whether this would enable the direct tethering of substrates to OGT via a fusion linker to allow us to explore OGT-glycosylation substrate complexes.

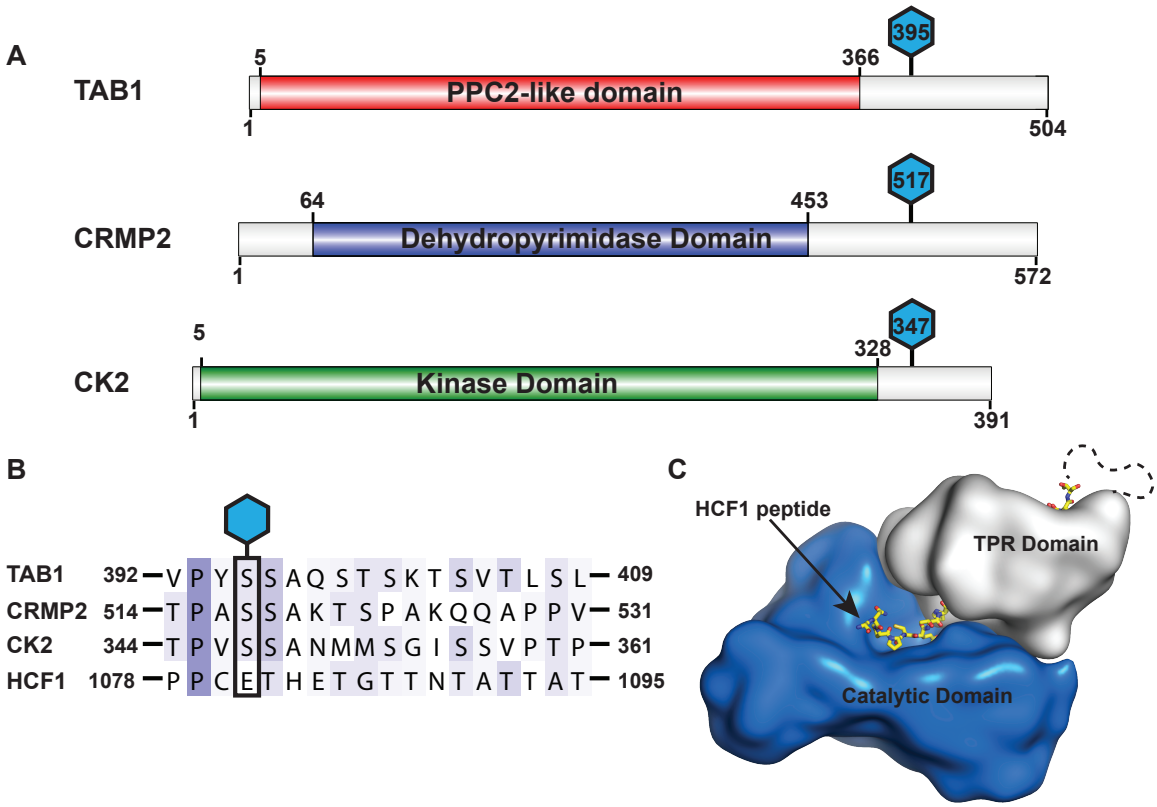

Figure 2.3: O-GlcNAc sites for a subset of proteins are located in a C-terminal disordered domain. (A) Cartoon depicting the domain structure and location of the glycosylation sites of the three OGT substrates TAB1, CRMP2, and CK2. The O-GlcNAc sites are depicted as blue hexagons. **(B)** Sequence alignment of the O-GlcNAc sites of TAB1, CRMP2 and CK2 as well as an $HCF1_{PRO}$ repeat. The O-GlcNAc sites for TAB1, CRMP2 and CK2, as well as the corresponding glutamate for the $HCF1_{PRO}$ repeat are highlighted with a black box. The sequence alignment shows the similarity between TAB1, CK2, CRMP2 and HCF1 surrounding the O-GlcNAc sites. **(C)** Schematic representation of OGT with a bound HCF1 peptide. The catalytic domain is shown in blue, the tetratricot-peptide repeat domain is shown in grey and the HCF1 peptide as yellow sticks (PDBID 4N3B [237]).

2.3.2. A linear fusion of OGT and $HCF1_{PRO}$ reproduces the $HCF1_{PRO}$ binding mode

To explore whether a fusion of the C-terminus of a peptide substrate to the N-terminus of a truncated OGT (312-1031) would generate physiological OGT-substrate complexes, we explored this approach first with $HCF1_{PRO}$. A construct was designed where an 18-mer $HCF1_{PRO}$ repeat peptide (PPCETHETGTTNTATTAT) was fused to the N-terminal Thr^{315} of OGT via a three glycine (3xGly) linker (Fig. 2.4 A). The fusion construct was overexpressed as a His_6-tagged protein in *E. coli*, purified and crystallised. Well-diffracting protein crystals were obtained and synchrotron data were collected to 1.9 Å. Molecular replacement and subsequent refinement revealed continuous unbiased $|F_o-F_c|$ density for both the $HCF1_{PRO}$ peptide and the 3xGly linker (Fig. 2.4 B). Encouragingly, the conformation of the $HCF1_{PRO}$ peptide in the fusion protein was nearly identical to that observed in the previously published OGT-HCF1 peptide complex [237] (RMSD on $C\alpha$ atoms = 0.2 Å). Thus, just as with the free peptide, the tethered $HCF1_{PRO}$ peptide backbone binds the OGT TPR domain in an extended conformation, interacting with residues lining the concave surface of the TPR superhelix (Fig. 2.4 B). Therefore, a linear fusion of OGT and $HCF1_{PRO}$ reproduces the $HCF1_{PRO}$ binding mode.

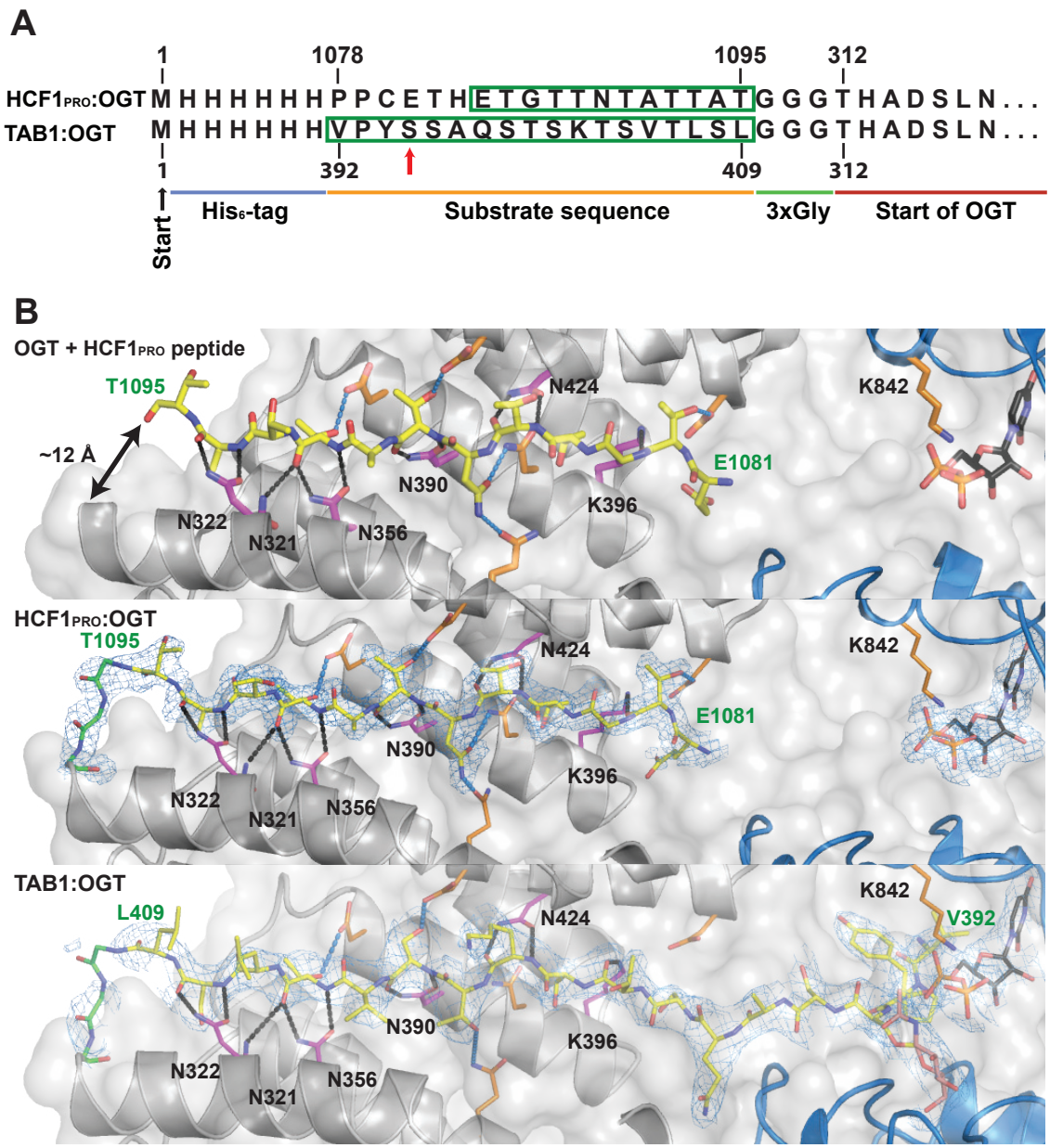

Figure 2.4: Design and structures of OGT:substrate fusion proteins. (A) Partial sequence of the fusion proteins showing the His_6-tag, the sequence of the substrate peptide, the 3xGly linker and the start of the OGT protein. Construct boundaries are indicated. The reported O-GlcNAc site Ser^{395} on TAB1 [117] is highlighted with a red arrow. Visible residues in the respective fusion construct structures are highlighted by a green box. **(B)** Three panels showing the structures of the free $HCF1_{PRO}$ repeat (PDBID 4N39) bound to OGT (top), the fusion protein $HCF1_{PRO}$:OGT (middle) and the fusion protein TAB1:OGT (bottom). OGT is shown in cartoon representation with the TPR and catalytic domains in grey and blue, respectively. The substrate peptides, UDP and the GlcNAc residues are shown as yellow, black and pink sticks, respectively. The 3xGly linker in the

fusion constructs is shown as green sticks. OGT residues interacting with the backbone of the substrate peptide are shown as magenta sticks and residues interacting with side chains are shown as orange sticks. The catalytic important K^{842} residue is shown as orange sticks. The green labels highlight the start and end residues of the substrate part of the fusion constructs. The distance from the N-terminus of OGT and the C-terminus of the free $HCF1_{PRO}$ peptide (top) is shown as a black double-headed arrow. Polar interactions between OGT and the backbone and side chains of substrate peptides are shown as black and blue dashed lines, respectively. The $|F_o\text{-}F_c|$ map for the fusion constructs $HCF1_{PRO}$:OGT and TAB1:OGT are shown as light-blue mesh contoured to 2.5 σ.

2.3.3. *A linear TAB1:OGT fusion suggests that TAB1 makes extensive interactions with the OGT TPRs*

We next explored the OGT-substrate fusion approach as a means of trapping complexes of OGT with TAB1. We generated a TAB1:OGT fusion construct matching the $HCF1_{PRO}$:OGT fusion, using an 18-mer TAB1 peptide derived from the Ser^{395} glycosylation site (VPY**S**SAQSTSKTSVTLSL) (Fig. 2.4 A). The chimeric protein was overexpressed as a His_6-fusion construct in *E. coli* and purified as described for the $HCF1_{PRO}$:OGT fusion protein (Fig. 2.4 A). We were able to generate crystals of the TAB1:OGT fusion protein, solve the structure by molecular replacement and refine the complex against 2.5 Å synchrotron diffraction data to R_{work}/R_{free} = 0.22/0.25. The unbiased $|F_o\text{-}F_c|$ density allowed the unambiguous building of the linker and peptide (Fig. 2.4 B). The first eight amino acids of the TAB1 peptide (VPY**S**SAQS), covering the glycosylation site, were found in a similar conformation in the active site as the free TAB1 peptide in complex with OGT reported previously [241] (Fig. 2.4 B, RMSD on Cαs = 1.4 Å). The electron density revealed Ser395 to be glycosylated as a result of self-

glycosylation during expression in *E. coli*, which was confirmed by Western blot analysis (Fig. 2.6 A). The sugar occupies the same position as observed in a complex with a short synthetic TAB1 glycopeptide [241] (Maximum atomic shift = 0.1 Å). Intriguingly there appears to be some extra electron density near Ser[396] and Ser[399] suggestive of additional glycosylation sites (Fig. 2.5) that could be an artefact of the very high (local) concentrations of the fused substrate peptide, or glycosylation occurring in *trans* as a result of the high protein concentrations (~10 mg/mL) used in the crystallisation experiments.

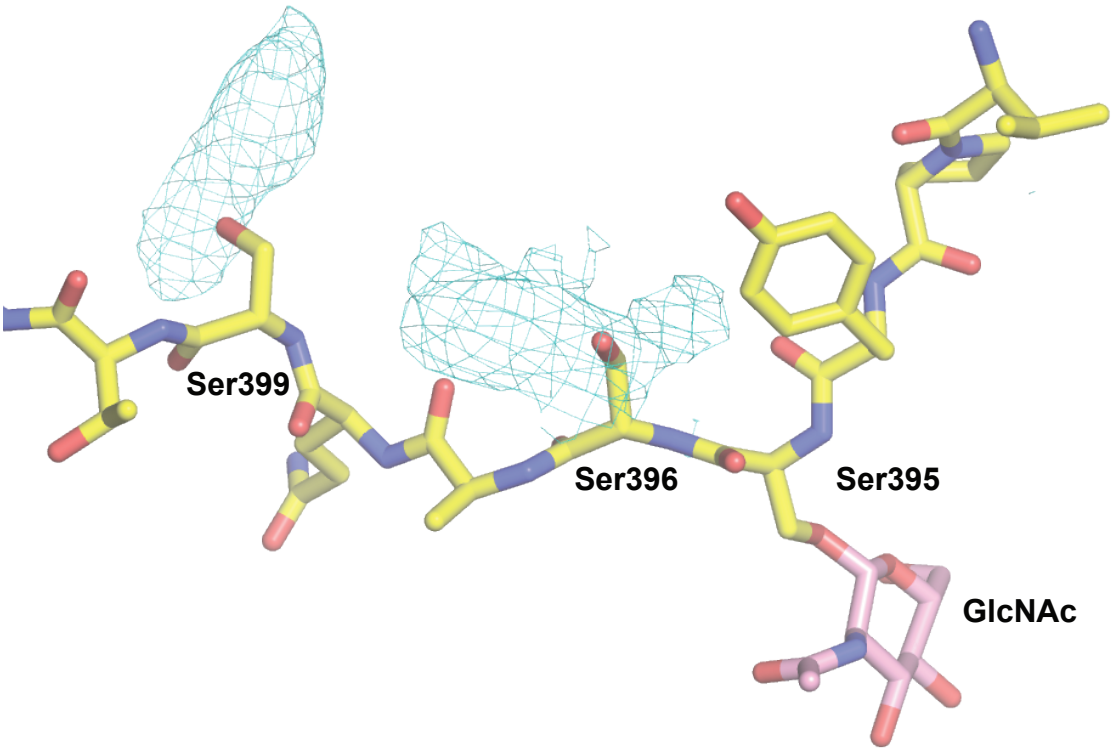

Figure 2.5: Close-up view of the isolated TAB1 peptide of the TAB1:OGT fusion construct. The substrate peptide of the fusion construct is shown as yellow sticks. The GlcNAc residue decorating Ser[395] is shown as pink sticks. The |Fo-Fc| map is shown as light-blue mesh contoured to 3.5 σ.

In the TAB1:OGT fusion structure, the TAB1 peptide forms two side chain mediated interactions (Ser^{404}/Thr^{406}) with the TPR domain of OGT (Asp386/Asp420) (Fig. 2.4 B). These are remarkably similar to the interactions between the same OGT residues and Thr^{1090} and Thr^{1092} of the $HCF1_{PRO}$ repeat (Fig. 2.4 B). Similarly, the interactions between the TAB1/HCF1 peptide backbones and the five regularly spaced OGT TPR asparagines are conserved (Fig. 2.4 B). Furthermore, the overall conformations of the TAB1 and HCF1 peptides in the respective fusion constructs is similar (RMSD on Cαs = 1.3 Å). Thus, a linear TAB1:OGT fusion suggests that the TAB1 OGT substrate makes extensive interactions with the OGT TPRs.

2.3.4. *Interactions with the OGT TPRs contribute to TAB1 O-GlcNAcylation*

Although the similarity to the HCF1 peptide binding mode and the presence of glycosylation on Ser^{395} suggests we have trapped a physiologically relevant TAB1:OGT complex, I further tested this model by structure guided site-directed mutagenesis in the context of truncated OGT (312-1031) and TAB1 (7-409) as separate proteins. Two types of OGT mutants were generated, a single point mutant in the active site (K842M), known to abolish catalytic activity [241] and a quintuple mutant where the five key asparagine residues that form the bulk of interactions in the TPR domain (Asn^{322}, Asn^{325}, Asn^{356}, Asn^{390} and Asn^{424}, Fig. 2.4 B) were all mutated to alanines (from here on referred to as the 5N5A mutant). Based on the TAB1:OGT fusion protein complex, the 5N5A mutations would be expected to disrupt the binding of the C-terminal region of TAB1 to the TPR domain**.** Using Western blot analysis, I probed OGT activity on TAB1 and blotted for O-GlcNAcylation using an O-GlcNAc Ser^{395} specific antibody [117].

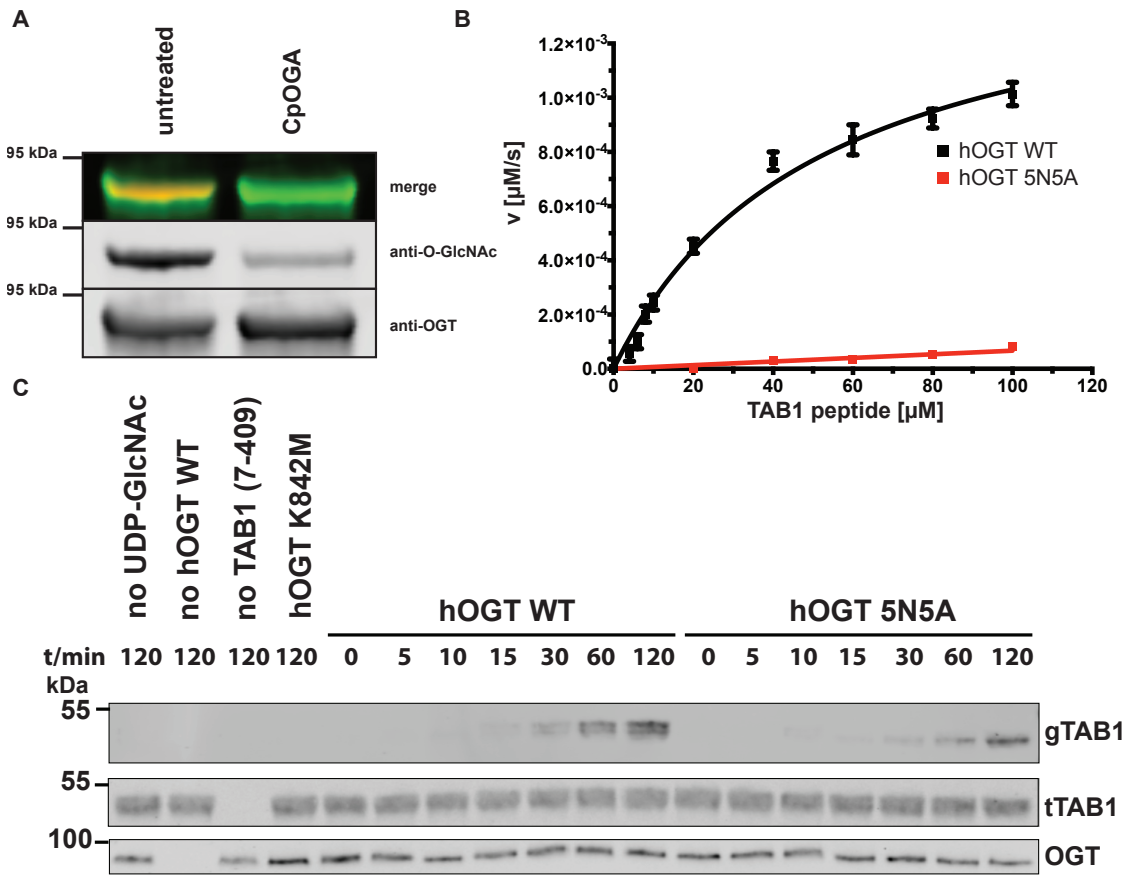

Figure 2.6: Activity of wild type and mutant OGT on peptide and protein substrates. **(A)** Western blot analysis of CpOGA treated purified TAB1:OGT fusion protein. Briefly, 0.5 µg of TAB1:OGT was incubated in the presence and absence of ~ 10 µg/mL CpOGA for 30 minutes at 37 °C. Reactions were stopped by addition of LDS-loading buffer and boiling at 95 °C. **(B)** Graph showing the steady-state kinetics to determine the K_M of the TAB1 peptide (KKPVSVPYSSAQSTSKTSVTLSL). Briefly, 50 nM hOGT(WT/5N5A) was pre-incubated with varying concentrations of TAB1 peptide before starting the reaction by adding UDP-GlcNAc to a final concentration of 50 µM. The reaction was stopped before 10% of substrate was converted by addition of detection reagent in 50% MeOH. K_M (WT) = 42 ± 7 µM, K_M (5N5A) = not detectable **(C)** Western blot analysis of *in vitro* glycosylation reactions of TAB1 with OGT (WT/mutants). Briefly, 10 µM TAB1 protein was incubated with 50 nM hOGT in 100 µL 0.1 M Tris-HCl pH 7.4, 0.15 M NaCl, 0.5 mM TCEP buffer. Reactions were started by the addition of UDP-GlcNAc to a final concentration of 100 µM and incubated at 25 °C. Samples were taken at indicated time points and reactions stopped by boiling for 5 minutes at 95 °C in LDS-loading buffer.

As demonstrated previously, TAB1 is readily O-GlcNAcylated by WT OGT whereas no glycosylation is observed with the catalytically inactive K842M mutant [241] (Fig. 2.6 C). The 5N5A mutant shows significantly reduced activity on a free TAB1 peptide (KKPVSVPYSSAQSTSKTSVTLSL) matching the peptide used in the fusion construct (Fig. 2.6 B), in agreement with the interactions formed by the key asparagines in the TPR domain of OGT observed in the structure (Fig. 2.4 B). However, when using a shorter synthetic TAB1 peptide (KKPVSVPYSSAQSTS, ending just before the start of the TPR repeats), the 5N5A mutant shows the same activity levels as WT OGT (Fig. 2.7).

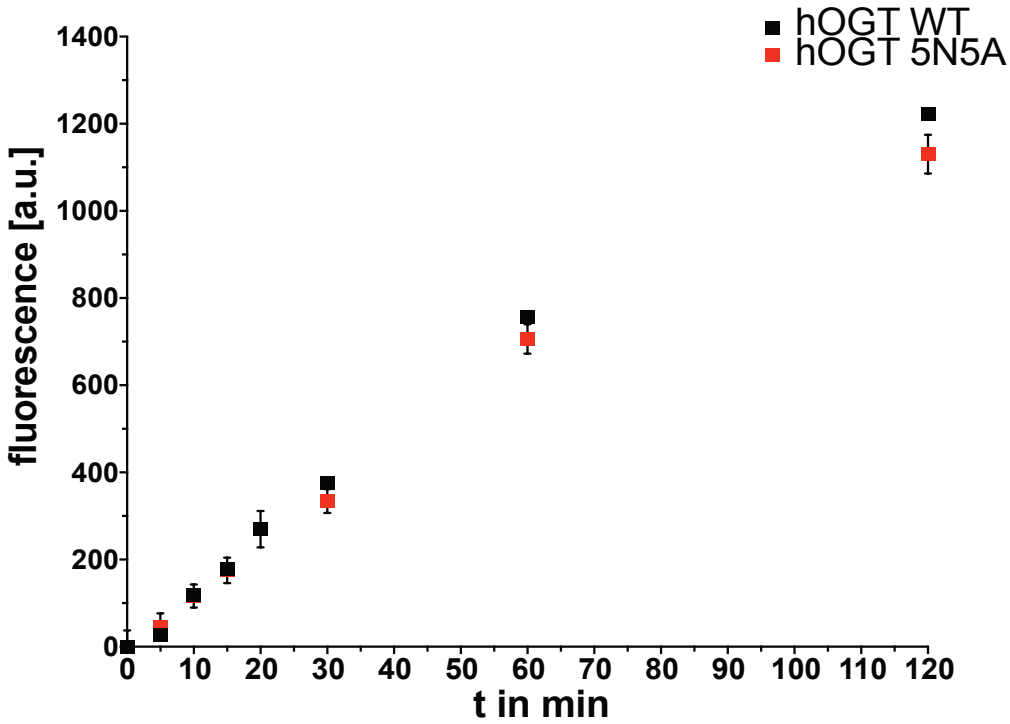

Figure 2.7: Graph showing the time-dependent glycosylation of a TAB1 peptide (KKPVSVPYSSAQSTS). Briefly, 50 nM of WT or 5N5A hOGT (312-1031) was pre-incubated with 10 µM of TAB1 peptide before starting the reaction by adding UDP-GlcNAc to a final concentration.

Intriguingly, the 5N5A mutant appears to show a more modest reduction of glycosyltransferase activity (~50%), as calculated by quantifying fluorescent

signal from the fluorophore conjugated secondary antibodies used in the Western blot analysis (Fig. 2.8), when using the TAB1 (7-409) protein.

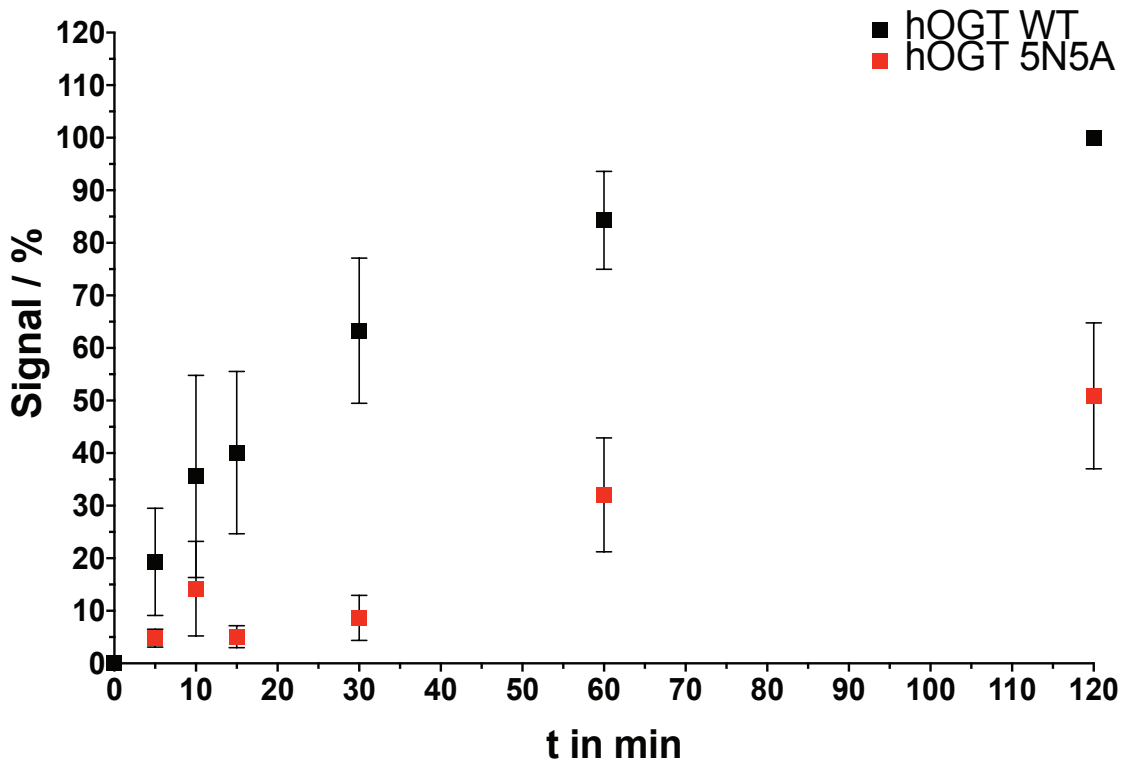

Figure 2.8: Graph showing the time-dependent glycosylation of a TAB1 protein (7-409). Briefly, 50 nM of WT or 5N5A hOGT (312-1031) was pre-incubated with 10 µM of TAB1 protein (7-409) before starting the reaction by adding UDP-GlcNAc to a final concentration of 100 µM.

Suggesting that while interactions of the TAB1 C-terminus with the OGT TPRs are important, further interactions with the globular pseudophosphatase domain of TAB1 may exist. Nevertheless, interactions with the OGT TPRs contribute to TAB1 O-GlcNAcylation.

2.4. Concluding remarks

The human O-GlcNAc transferase is a multi-domain protein and is essential in metazoa [50,56,74]. However, it is still unclear how a single OGT enzyme recognises its multitude of substrates. Previous work has proposed sequence specificity targeting -2 to +3 relative to the acceptor residue [231,253]. Previous work has also suggested the involvement of the TPR domain in substrate recognition by incrementally removing repeats from the TPR domain resulting in a loss of activity on substrates even on a peptide level, although the molecular basis of this was as yet unclear [72,76,141,147,227,230,232]. Using the proteolytic OGT substrate HCF1, Lazarus *et al.* revealed the involvement of multiple OGT residues on the concave surface of the TPR domain in binding side chains and backbone of the $HCF1_{PRO}$ repeat proteolytic substrate [237]. Here we used a fusion approach to trap OGT-substrate complexes to investigate the role of the TPRs in recognition of glycosylation substrates. We first demonstrated that this fusion approach recapitulates the published $HCF1_{PRO}$ peptide binding mode and then used that to reveal how the C-terminus of the OGT glycosylation substrate TAB1 is recognised by the enzyme. The TAB1 C-terminus binds in an extended conformation in the TPR domain, making extensive contacts with the concave surface through regularly spaced asparagines in OGT. An OGT mutant lacking these asparagines was deficient in glycosylation of TAB1. Interestingly, the data show a complete loss of O-GlcNAcylation of a free TAB1 C-terminal peptide, whereas activity on a TAB1 protein is more modestly reduced. These findings, coupled with recently published work on an OGT substrate sequence preference [231] suggest that OGT may bind its substrates through a combination of mechanisms. It is interesting to note that many other OGT substrates (e.g. Casein kinase II and CRMP2, Fig. 2.3 A) also possess similarly disordered regions C-

terminal of the O-GlcNAcylation site, suggesting that this may be a general mode of OGT substrate recognition. However, O-GlcNAc sites have also been reported to reside in/close to secondary structure motifs, as is the case for Histone H2B [321], p53 [88], the glucose-6-phosphate dehydrogenase G6PD [322] and SNAP-29 [128]. It is possible that a subset of substrates is O-GlcNAcylated in a co-translational fashion as proposed by recent work [10]. In the present work we have shown, using crystallography and site-directed mutagenesis, that the OGT substrate TAB1 binds the enzyme in the same way as the proteolytic substrate HCF1 [237] and that the five asparagine residues found on the concave surface of the TPR domain (Asn^{321}, Asn^{322}, Asn^{356}, Asn^{390} and Asn^{424}) are important for binding. Future studies could be directed at dissecting which other parts of OGT and/or substrate proteins contribute to substrate binding.

3. Development of low μM bisubstrate conjugate inhibitors targeting the human O-GlcNAc transferase

Parts of this work have been published or are currently prepared as a manuscript:

(1)

Borodkin *et al.*, “Bisubstrate UDP–peptide conjugates as human O-GlcNAc transferase inhibitors”, Biochemical Journal, 2014

3.1. Introduction

In order to dissect the catalytic and non-catalytic roles of OGT and understand its biological role better, potent and selective inhibitors are needed. The most frequently used inhibitor of OGT is the non-hydrolysable substrate analogue UDP-5S-GlcNAc [298–301]. However, UDP-5S-GlcNAc might target other UDP-GlcNAc utilising enzymes, resulting in the emergence of O-GlcNAc independent phenotypes [303]. Other inhibitors such as the small-molecules BZX [296,308] and OSMI-1 [309] show promising results, however their mode of action and specificity have not been exhaustively investigated and may have unknown off-target effects. Based on OGTs unique reaction mechanism, it would be interesting to pursue mechanism-inspired inhibitors that make use of the high affinity of UDP (IC_{50} = 1.8 μM [285]), to date the strongest OGT binder, and the selectivity of OGT substrate peptides. This culminated in the development of the bisubstrate inhibitor Goblin 1, exhibiting low μM inhibition (IC_{50} = 18 μM [311]). Goblin 1 combines UDP and an acceptor peptide, derived from the retinoblastoma-like protein 2 protein (RBL2) [231], via a three-carbon linker joining the β-phosphate of UDP to the oxygen of the acceptor serine residue. Structural characterisation showed that Goblin 1 binds OGT by forming a pseudo-Michaelis complex, in which the three carbon linker spans the space created by omitting the GlcNAc residue [311]. Both UDP and the peptide bind in the same way as found in a complex formed by a free RBL2 peptide and the substrate analogue UDP-5S-GlcNAc [231] ($RMSD_{C\alpha(RBL2peptide)}/RMSD_{C\alpha(Goblin1peptide)}$ = 0.45 Å, $RMSD_{UDP(5S\text{-}GlcNAc)}/RMSD_{UDP(Goblin1)}$ = 0.21 Å). However, the low inhibitory efficacy and lack of cell permeability makes Goblin 1 unsuitable for *in cellulo* studies. Fortunately, the Goblin 1 scaffold allows for further alterations by optimising the linker length/type, peptide sequence and nucleotide, potentially increasing potency. Cell

permeability of theses scaffolds could be addressed by attaching cell-penetrating peptide/peptoid motifs [323,324] and will be discussed in chapter 4.

Here, I am reporting a panel of novel thiol-linked bisubstrate conjugates, which exhibit a higher inhibitory potency towards the human O-GlcNAc transferase (hOGT) compared to their O-linked progenitors, reaching binding affinities close to that of UDP. Using high-resolution macromolecular X-ray crystallography, I could show that, by using a thioether- instead of an ether-linkage, the linker adopts a more “relaxed” conformation, which could explain the increased inhibitory potency. Finally, I have used this novel bisubstrate conjugate to design a high throughput fluorescence anisotropy assay system to rapidly screen for novel binders of OGT and determine their binding affinities.

3.2. *Materials & Methods*

Chemical synthesis was performed by Dr. Vladimir S. Borodkin and HeLa cell lysate treatment was performed by Andrii Gorelik. If not stated otherwise, all chemical structures have been supplied by Dr. Vladimir S. Borodkin.

3.2.1. *Crystallography*

Human OGT (312-1031) was recombinantly expressed as a cleavable GST-fusion protein and purified as described previously [241]. A solution of 8 mg/mL hOGT and 1 mM bisubstrate conjugate in 25 mM Tris-HCl pH 8.0, 20 mM NaCl, 0.5 mM TCEP was incubated on ice for 15 minutes. Hanging-drop crystallisation experiments were performed by mixing drops containing 1.2 µL of protein-ligand solution and 0.6 µL reservoir solution [1.45 M K_2HPO_4, 8 mM EDTA, 1% xylitol] supplemented with 0.5 M $(NH_4)_2SO_4$ for the ThioGoblin 1:OGT / dvB_1173:OGT complex and [1.3 M DL-Malic acid, 0.1 M Bis-Tris propane pH 6.4] supplemented with hOGT crystal seeds, grown in the same condition as for the Goblin1:OGT / dvB_1505:OGT complex. Crystals grew O/N and were cryo-protected by short immersion in 2.5 M sodium malonate pH 7.0, supplemented with 1 mM of respective bisubstrate conjugate and flash-frozen in liquid nitrogen. Data were collected at the European Synchrotron Radiation Facility (ESRF) on beamline ID30A-1 (Goblin1), ID29 (ThioGoblin 1), ID23-2 (dvB_1505) and the Diamond Light Source on ID04-1 (dvB_1173). Data were processed with XDS [325] and scaled to 1.68 Å (Goblin1), 1.85 Å (ThioGoblin 1), 1.97 Å (dvB_1505) and 1.85 Å (dvB_1173) using aimless [326]. 5% of total reflections were set aside as an R_{free} test set. Crystals belonged to space group F222 with one molecule per asymmetric unit, a solvent content of 63% and a Matthews coefficient of 3.33 for both crystals. The structures were solved with MOLREP [327], using chain A of PDB

3PE4 [240] as a search model. The structure was fully refined using iterative cycles of Refmac5 [328] and COOT [329]. Ligand topologies were generated with PRODRG [330].

3.2.2. *hOGT activity measurement*

Human OGT (312-1031) was recombinantly expressed as a His_6-fusion protein and was purified as described previously [231]. OGT activity was determined by setting up reactions containing 5 or 50 nM His_6-hOGT in 50 mM Tris-HCl pH 7.5, 0.1 mg/mL BSA, 10 µM sodium dithionate and 10 µM of peptide (KKENSPAVTPVSTA) in a total volume of 100 µL. Reaction mixtures were pre-incubated with varying concentrations of ligands for 15 min and initiated by addition of UDP-GlcNAc to a final concentration of 3.2 µM. The reaction was stopped after 30 minutes (50 nM hOGT) or 2 h (5 nM hOGT) at 21 °C by addition of 200 µL of 25 mM HEPES pH 7.4, 10 mM NaCl, 75 µM pyrocatechol violet, 50% (v/v) MeOH and 15 µM fluorophore [231,319]. UDP formation was detected fluorimetrically on a Gemini EM / SpectraMax i3x multimode-plate reader (Molecular Devices) at excitation and emission wavelengths of 485 nm and 530 nm, respectively. Data analysis was performed with GraphPad Prism 7. Turnover did not exceed 10% for either substrate.

3.2.3. *Fluorescence polarimetry measurements*

The binding affinity of F-ThioGoblin 1 and *L*-F-ThioGoblin 1 were determined by titrating hOGT at a fixed concentration of probe (125 nM) in 0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5 mM TCEP and 5% DMSO in a total volume of 25 µL and incubated in the dark for 30 min before read-out.

Binding affinities for ligands were determined by F-ThioGoblin 1 displacement. Reactions contained 0.8 μM hOGT and 0.75 μM F-ThioGoblin 1 in 0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5 mM TCEP and 5% DMSO in a total volume of 25 μL. Reaction mixtures were incubated with varying concentrations of inhibitor for 0.5 h in the dark. Fluorescence polarimetry was measured on a PHERAStar plate reader (BMG LABTECH) at excitation and emission wavelengths of 485 nm and 520 nm, respectively. Data analysis was performed in GraphPad Prism and binding constants for the labelled F-ThioGoblin 1 and *L*-F-ThioGoblin 1 were determined by fitting a one-site – total binding curve. Binding constants for ligands in displacement experiments were determined using GraphPad Prism 7 (4-parameter non-linear regression curve fit) and the equation reported by Nikolovska-Coleska *et al.*[331].

3.3. Results and Discussion

3.3.1. *An S-linked bisubstrate conjugate inhibits hOGT activity stronger than its O-linked predecessor*

In the original synthesis of Goblin 1, a serine carrying a phosphorylated three-carbon linker on the hydroxyl-oxygen was used as a building block during solid phase peptide synthesis (SPPS) of the Goblin 1 peptide [311]. The final bisubstrate scaffold was assembled by formation of the pyrophosphate bond using an activated UMP derivative (2',3'-*O*-diacetyl-UMP imidazolide) [311]. However, the formation of the pyrophosphate bond is a slow process, limiting the use of this synthetic route for the generation of large bisubstrate inhibitor libraries. This prompted the search for an alternative synthetic strategy that allows the efficient assembly of bisubstrate conjugates, independent of peptide sequence, linker length/type and nucleotide modifications. Substitution of the "acceptor" serine with cysteine allowed for carbon-sulphur bond formation using a photo-initiated thiol-ene conjugation (TEC). A modular synthetic approach was developed based on TEC directed addition of allyl-UDP to cysteine containing peptides (Rafie et al., Manuscript in preparation). Based on the structural characterisation of Goblin 1 we aimed to modulate activity by modifying three different positions (i) the peptide part, in length and sequence, (ii) linker modifications and finally (iii) the UDP-moiety (Fig. 3.1).

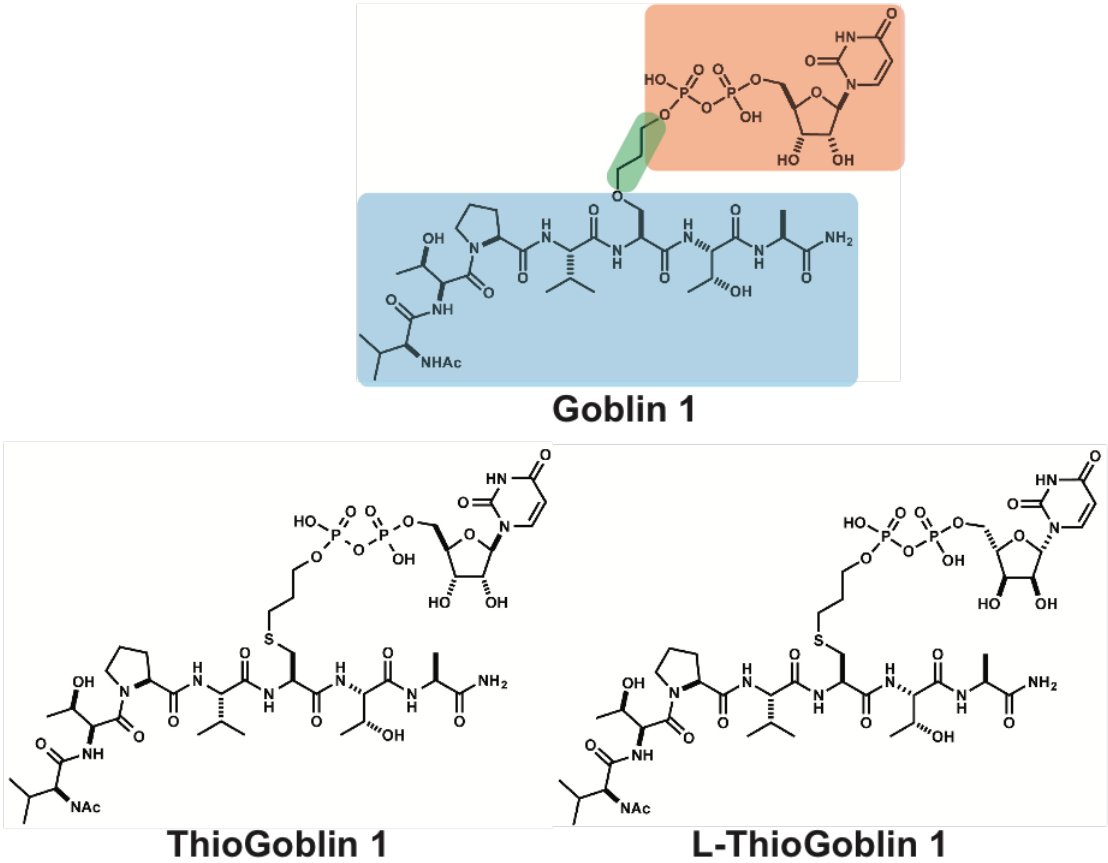

Figure 3.1: The Goblin 1 scaffold allows for further optimisation. Chemical structure of Goblin 1, ThioGoblin and L-ThioGoblin 1. The colouring corresponds to the modifiable parts of the construct. (Blue) Peptide sequence / length, (green) linker, (orange) UDP-moiety.

To investigate whether the introduction of a thiol-linkage into the scaffold (ThioGoblin 1, Fig. 3.1) would affect potency of the bisubstrate inhibitor, the activity of hOGT was measured in the presence of increasing concentrations of ThioGoblin 1. Remarkably, the substitution of the linker serine to cysteine resulted in a 10-fold increase in potency (IC_{50} (ThioGoblin 1) = 2 µM) compared to Goblin 1 (IC_{50} = 18 µM [311]) (Fig. 3.2).

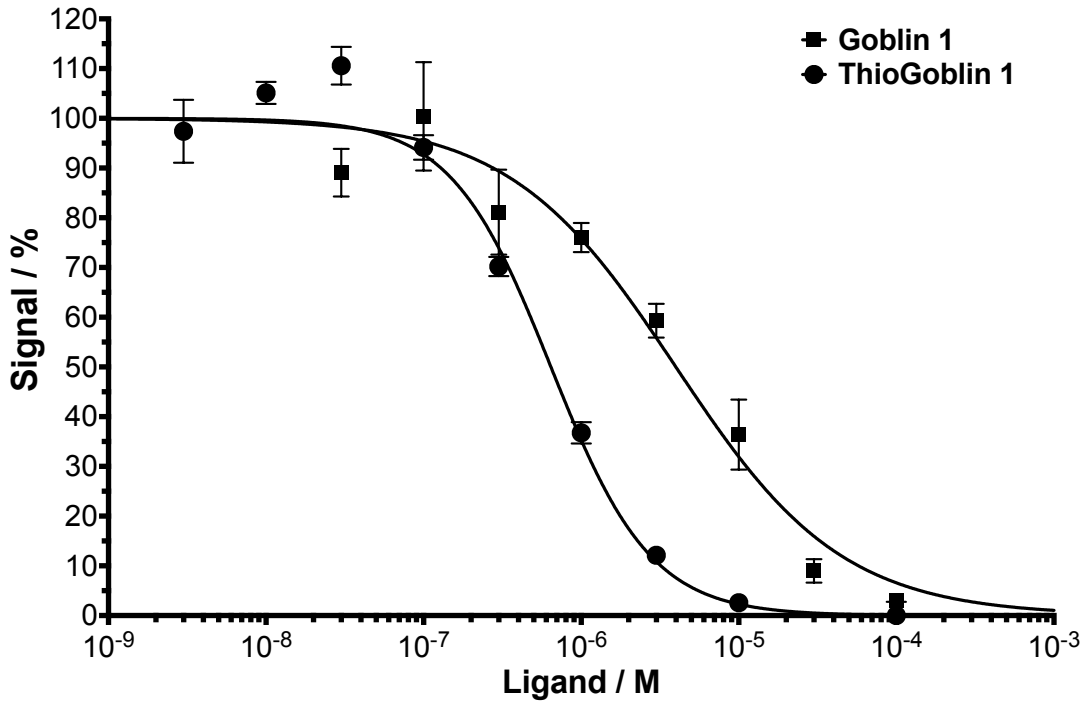

Figure 3.2: ThioGoblin 1 inhibits OGT with low µM affinity. Dose-dependent inhibition of hOGT in the presence of varying concentrations of ligands. In short, 5 nM (ThioGoblin 1/L-ThioGoblin 1) or 50 nM (Goblin 1) hOGT was incubated with varying concentration of ligands in 50 mM Tris-HCl pH 7.5, 0.1 mg/mL BSA, 10 µM $Na_2S_2O_4$ and 10 µM of the acceptor peptide KKENSPAVTPVSTA. The reaction was *initiate*d by the addition of UDP-GlcNAc to a final concentration of 3.2 µM. The reaction was stopped before more than 10% of substrate were turned over. Data for Goblin 1 were previously reported in [311].

The increased potency may stem from a change in the binding mode of the bisubstrate scaffold likely associated with the change from an ether to a thioether function. To investigate geometrical difference, structural studies of the human O-GlcNAc transferase (hOGT) in complex with ThioGoblin 1 were performed. However, previous crystallisation conditions of hOGT yielded crystals with a diffraction limit of ~ 3 Å, making them unsuitable for the dissection of small changes in the binding of ligands. The improvement of a previously reported crystallisation condition [240] by addition of 0.5 M $(NH_4)_2SO_4$ and the discovery of a

new crystallisation condition [231] yielded well diffracting hOGT crystals, suitable for mechanistic studies. The previously reported structure of hOGT in complex with Goblin 1 was solved to a resolution of 3.15 Å [311], making it necessary to collect diffraction data using the improved conditions. High-resolution synchrotron diffraction data of hOGT in complex with Goblin 1 (1.68 Å, R_{work}/R_{free} = 0.19/0.22) and ThioGoblin 1 (1.85 Å, R_{work}/R_{free} = 0.22/0.24) (Table 3.1) were collected to allow a comparison of the binding modes of each and dissect any small changes that might explain the effects seen.

Table 3.1: X-ray diffraction data collection and structure refinement statistics for hOGT:Goblin 1 and hOGT:ThioGoblin1. Values for the highest resolution shell in parenthesis.

	hOGT + Goblin 1	**hOGT + ThioGoblin 1**	**hOGT + dvB_1505**
Data collection			
Beamline, wavelength	ID30-A1, 0.98 Å	ID29, 0.976 Å	ID23-2, 0.873 Å
Space group	*F222*	*F222*	*F222*
Cell dimensions (Å)	*a*=137.87, *b*=151.33, *c*=200.48	*a*=138.04, *b*=150.95, *c*=200.54	*a*=138.3, *b*=150.9, *c*=200.57
Resolution (Å)	46.10-1.68 (1.71-1.68)	46.01-1.85 (1.88-1.85)	46.02-1.97 (2.01-1.97)
R_{merge}	0.05 (1.285)	0.06 (0.871)	0.126 (1.225)
$I/\sigma I$	13 (1.0)	9.7 (1.3)	6.9 (1.0)
$CC_{1/2}$	0.99 (0.45)	0.99 (0.65)	0.99 (0.35)
R_{meas}	0.06 (1.624)	0.078 (1.202)	0.153 (1.507)
R_{pim}	0.04 (0.979)	0.048 (0.711)	0.09 (0.87)
Completeness (%)	99.3 (99.9)	99.5 (99.8)	98.2 (99.3)
Redundancy	4.7 (4.8)	4.0 (4.1)	4.3 (4.3)
Refinement			
Resolution (Å)	46.01-1.68	46.01—1.85Å	50-1.97
No. total reflections	547394	349113	312676
No. unique reflections	117257	88158	72168
R_{work}, R_{free}	0.193 / 0.222	0.215 / 0.247	0.195 / 0.23
No. atoms			
Protein	5508	5478	5506
Ligand	78	78	82
Water	478	362	307
<*B*-factor> (Å^2)			
Protein	36.8	46.78	37.3
Ligand	32.5	38.65	34.9
R.m.s.d.			
Bond lengths(Å)	0.0077	0.0115	0.0098
Bond angles (°)	1.248	1.5	1.46

Structure solution by molecular replacement followed by iterative cycles of refinement and model building revealed continuous $|F_o\text{-}F_c|$ electron density for both ligands allowing the unambiguous placement of Goblin 1 and ThioGoblin 1 (Fig. 3.3). The fully refined models revealed both bisubstrate inhibitors bind to the OGT active site in a conformation closely resembling a previously reported pseudo-Michaelis complex of hOGT with UDP-5S-GlcNAc and an acceptor peptide [231].

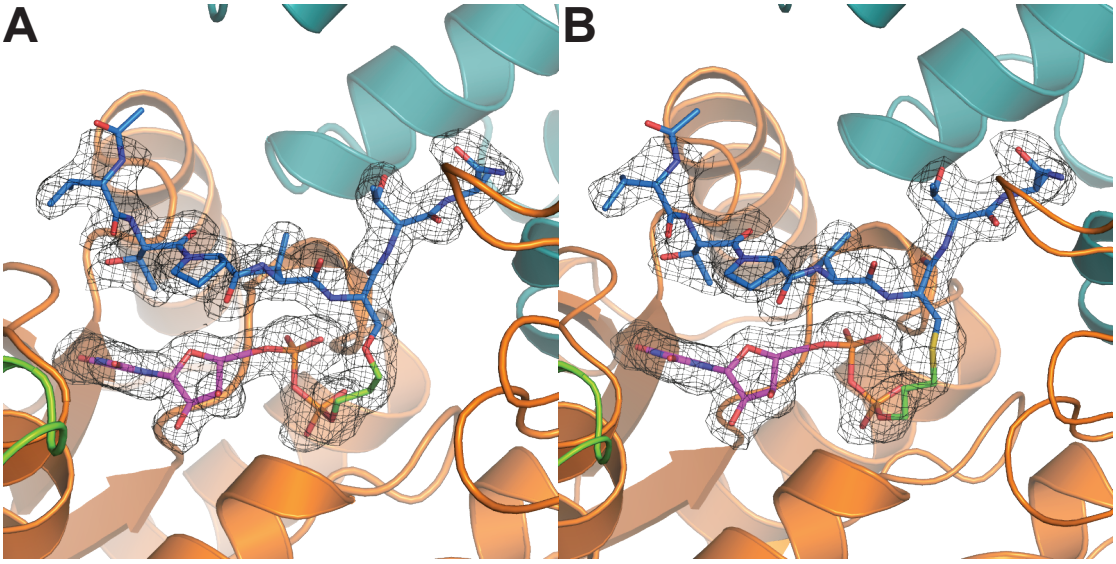

Figure 3.3: Goblin 1 and ThioGoblin 1 bind in the active site of hOGT recapitulating a pseudo-Michaelis-complex. Goblin 1 **(A)** and ThioGoblin 1 **(B)** bound to the human O-GlcNAc transferase. hOGT is shown in cartoon representation in teal (TPR domain), orange (catalytic lobes) and green (intervening domain). The bisubstrate inhibitors Goblin 1 and ThioGoblin1 are both in stick representation with the peptide part shown in blue, the linker in green and the UDP moiety in magenta. The unbiased |Fo-Fc| maps for both ligands are shown as a black mesh, contoured to 2.25 σ.

The maximum atomic shift between the UDP moieties of either bisubstrate conjugate to the corresponding substrates/substrate analogue is than 0.8 and 0.7 Å for Goblin 1 and ThioGoblin 1, respectively. The largest atomic shift between the two bisubstrate conjugates was measured between the two C_3 atoms (Fig.

3.4). A superposition of the unbiased $|F_o\text{-}F_c|$ electron density maps shows a clear difference in the conformation of the linker between the two ligands (Fig. 3.4 C).

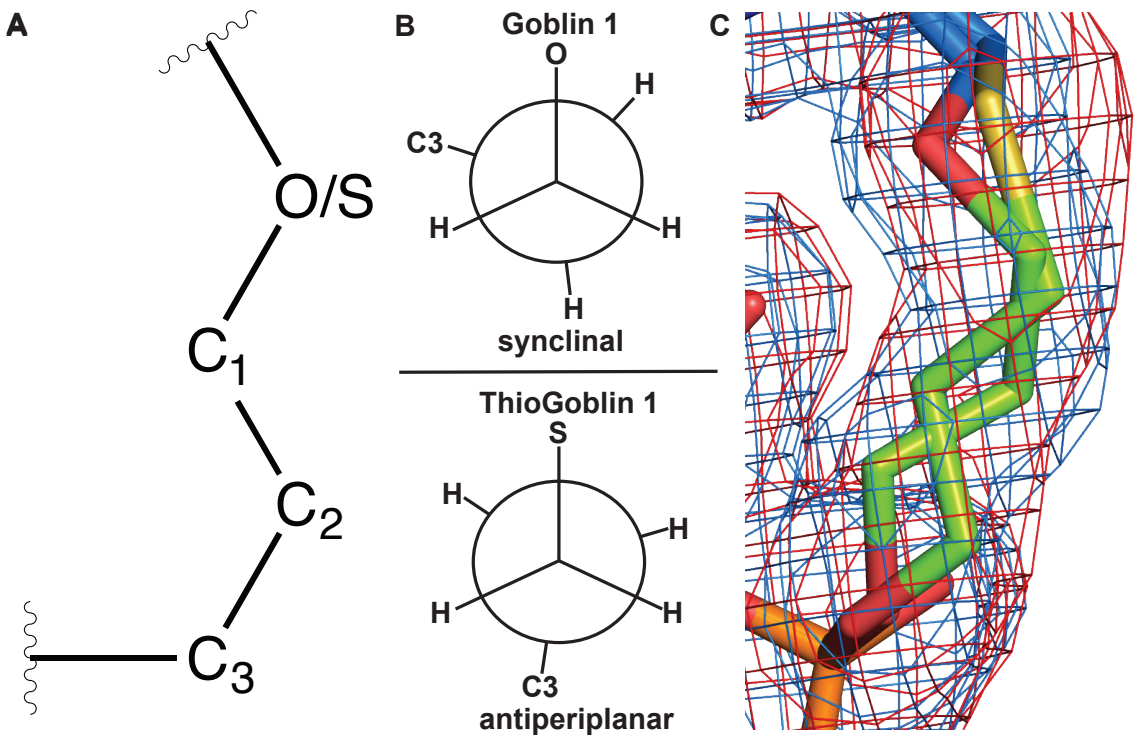

Figure 3.4: The linkage type defines the linker geometry. (A) Schematic drawing of the linker moiety of the bisubstrate conjugates. **(B)** Newman projection of the dihedral angles of Goblin 1 (dihedral angle = 72°) and ThioGoblin 1 (dihedral angle = 171.4°). **(C)** Superposition of Goblin 1 and ThioGoblin1 with their respective |Fo-Fc| maps in blue and red, respectively. The mesh is contoured to 1 σ. The superimposed electron density maps were calculated using the PHENIX Software Superpose suit [332].

In the Goblin 1 complex the linker adopts a synclinal conformation (dihedral angle $O\text{-}C_1\text{-}C_2\text{-}C_3$ = 72°, Fig. 3.4 B), which was not discernible in the lower resolution structure published previously [333]. In contrast, in the S-linked analogue ThioGoblin 1 the linker adopts an antiperiplanar conformation (dihedral angle $S\text{-}C_1\text{-}C_2\text{-}C_3$ = 171.4°, Fig. 3.4 B). This difference may contribute to the increased potency of ThioGoblin 1, as the antiperiplanar conformation is energetically more favourable than the synclinal conformation of Goblin 1. Additionally, the difference in geometry of the linkers in the inhibitors, may introduce additional

interactions. Taking into account the positioning of the sulphur of the acceptor cysteine and its larger van der Waals (vdW) radius (vdW (oxygen) = 152 pm, vdW (sulphur) = 180 pm), additional vdW-interactions with hOGT may be introduced, which are not present with Goblin 1. VdW interactions are weak attractions that occur between atoms in close proximity (3-6 Å). The positioning of the linker oxygen (Goblin 1) and sulphur (ThioGoblin 1) allows them both to interact with the His^{558}, Thr^{560} and Thr^{633} side chains, as well as the carbonyl oxygen of Gly^{654}. However, the 0.8 Å difference in positioning between the linker atoms brings the sulphur into closer proximity of His^{558}, Thr^{560}, Thr^{633} and Gly^{654}. In combination with the larger vdW radius of sulphur (ΔvdW(S-O) = 18 pm) may promote stronger interactions between ThioGoblin 1 and OGT, possibly contributing to the increased inhibitory efficacy.

3.3.2. *Using a fluorescein modified ThioGoblin 1 as a fluorescent probe to develop a high-throughput fluorescence polarimetry assay system*

The structures of the bisubstrate inhibitors in complex with hOGT show that the N-terminus of the bisubstrate peptide part is solvent exposed (Fig. 3.3). This should allow the introduction of various modification, such as a fluorophore or a biotin linker without affecting inhibitory potency. Common methods for measuring the activity of hOGT are via radiolabelling of substrates using tritiated UDP-[3H]-GlcNAc [231,285,296], a fluorometric read-out of UDP [231,311,334], a fluorescence polarimetry assay [296], a protease protection assay [335], and various other ELISA-based assay systems reviewed here [336]. Although there is a wide array of available assay systems to measure hOGT activity most of them are labour intensive, require radioactive materials to work and/or need a large amount of starting material. To address the lack of high throughput assay systems allowing the identification hOGT binders, we decided to label our bisubstrate conjugate with a fluorescein molecule (F-ThioGoblin 1), to explore its use as a fluorescent probe in a fluorescence polarimetry (FP) assay system. FP is the phenomenon in which emitted photons from a fluorophore have different intensities along varying axes of polarisation. This difference is quantifiable and is dependent on the rotation speed of the probe in solution. The degree of polarisation is calculated by measuring the fluorescence intensities parallel and perpendicular with respect to the plane of linearly polarised excitation light. The faster a probe rotates, the smaller the difference between the different planes of light . Vice versa, reducing the rotational velocity increases the difference in polarisation between the two planes of light. Measuring FP allows the determination of binding affinities of fluorescent probes to biological macromolecules and the binding affinity of unlabelled ligands. The latter is measured by a dose-dependent

displacement of the fluorescent probe by the ligand of interest. The binding affinity of F-ThioGoblin 1 to hOGT was measured by titrating protein concentrations at a fixed probe concentration in a total volume of 25 µL and measuring the FP (Fig. 3.5 B), allowing the calculation of the dissociation constant of F-ThioGoblin 1 (K_d = 1.3 µM), which is in good agreement with the IC_{50} measured for ThioGoblin 1 in the OGT activity assay (Fig. 3.2).

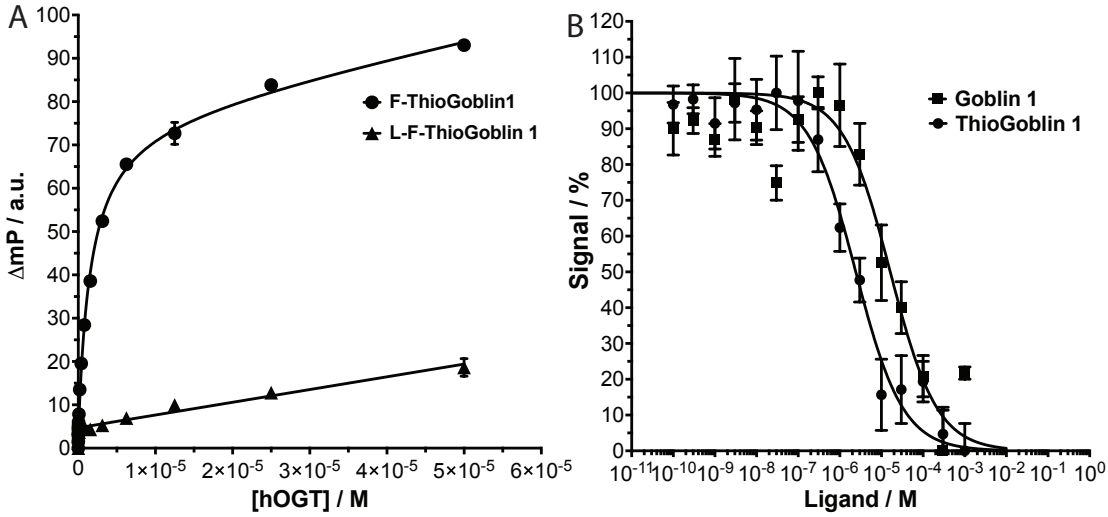

Figure 3.5: A fluorescein labelled ThioGoblin 1 conjugate binds to hOGT with low µM affinity. (A) K_d determination of F-ThioGoblin 1 (black) and its stereoisomer L-F-ThioGoblin 1 (red). Data points are s.e.m. of three replicates. **(B)** K_i determination of Goblin 1 and ThioGoblin 1. Data points are s.e.m. of three replicates.

Characterisation of various hOGT structures [240,241] revealed an intricate network of polar interactions between the protein and nucleotide. Interestingly, the interactions formed by hOGT residues Lys^{898} and Asp^{925} and the 2'- and 3'-hydroxy groups of the ribose are stereo-specific in respect to the *D*-configuration of the pentose sugar. The degree by which the binding affinity for the *L*-UDP stereoisomer of ThioGoblin 1 (L-ThioGoblin 1) (Fig. 3.1) towards hOGT is reduced, could give insights into the contribution of the interactions formed by the nucleotide and if hOGT has evolved to accommodate *L*-configured UDP. Moreover, a complete loss of binding or significant decrease in affinity would

make *L*-ThioGoblin 1 a suitable negative control in cell culture experiments to observe any OGT independent phenotypes.

To investigate whether the inversion of the stereocenters on the ribose moiety would affect binding to hOGT, the stereoisomer *L*-ThioGoblin 1 and the corresponding fluorescein-tagged analogue F-*L*-ThioGoblin 1 were synthesised. As hypoyhesised, binding of F-*L*-ThioGoblin 1 to hOGT was significantly reduced, preventing the determination of a dissociation constant (Fig. 3.5 A). Furthermore, *L*-ThioGoblin 1 did not inhibit hOGT activity (Table 3.2, dvB_1655). The low μM affinity of F-ThioGoblin 1 towards hOGT makes it suitable to be used as a tool to develop a high-throughput system to rapidly identify ligands that bind in the active site of hOGT and measure their binding affinities. Performing a dose-dependent displacement assay with an unlabelled ligand yields an 'apparent' IC_{50} value, which can then be used to calculate a corresponding inhibition constant K_i. In classical inhibition experiments the Cheng-Prusoff equation [337] has been used to calculate K_i values directly from IC_{50} values. This is acceptable in cases where the total concentration of receptor protein is far below the total concentration of an inhibitor, as is the case in classical inhibition experiments where the enzyme concentration is significantly lower than the substrate concentration [E] << [S]. However, FP is measured under non-receptor depleting conditions ([E] > [Fluorescent probe]), which would result in an overestimation of the K_i, if Cheng-Prusoff is used [331]. It is therefore recommended to use the equation described recently by Nikolovska-Coleska *et al.* [331], which corrects for this overestimation. Goblin1 and ThioGoblin 1 were able to displace F-ThioGoblin 1 in a dose-dependent manner, validating this assay system as a platform to identify ligands of hOGT and measure their binding affinity (Fig. 3.5 B). The derived values of K_i (Goblin 1) = 9.7 ± 4.4 μM and K_i (ThioGoblin 1) = 1.3 ± 0.8 μM are in good

agreement with previously reported binding affinities for Goblin 1 (Borodkin et al., 2013) and the dissociation constant measured for F-ThioGoblin 1 (1.3 μM). Using this assay system, in combination with hOGT activity measurements, the potency of 36 different bisubstrate inhibitors were determined (Table 3.2).

Table 3.2: Summary of bisubstrate conjugates discussed in the present chapter. DvB codes are inhouse ID numbers, the acceptor residue in each peptide is highlighted in red. Resolutions in the final column represent the highest resolution shell. Unless stated otherwise, all IC_{50} values were obtained using the fluorometric steady-state kinetics assay system and all K_i values determined using the fluorescence polarimetry assay. N.a. = not available, n.d. = not determinable, * = value amibgous, Floc = 5(6) fluorescein carboxamide, Ahx = 6-aminohexanoic acid, toad = 4-(2-(2-(methylamino)ethoxy)ethoxy)butan-2-one, ndx = N-(6-(2,5-dioxopyrrolidin-1-yl)hexyl), ITC = Isothermal titration calorimetry, UMBP = Uridine methylene bisphosphonate, Φ = *D*-Naphtilalanine.

dvB_Code	Peptide sequence	Linker length	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure [Y/N]
958	VPYSSAQ	C4	Ser	UDP	11	n.a.	N
1051	KVTPVSTA	C3	Ser	UDP	29	n.a.	N
1059	KKKKKKVTPVSTA	C3	Ser	UDP	19	n.a.	N

dvB_Code	Peptide sequence	Linker length	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure [Y/N]
1060	VPYSSAQ	C3	Ser	UDP	146	n.a.	N
1063	VTPVSTA	C4	Ser	UDP	68	n.a.	N
1085 (Goblin 1)	VTPVSTA	C3	Ser	UDP	18	10	Y / 1.68 Å
1090	YSPTSPS	C4	Thr	UDP	63	n.a.	N
1099	VTPVSTA	C3	Ser	UDP	7	n.a.	N
1161	---	---	---	UMP-phosphonoacetate	67	n.a.	Y / 2.05 Å
1173	VTPVSTA	C2	Bishomo-Ser	UMP-phosphonoacetate	250	n.a.	Y / 1.85 Å

dvB_Code	Peptide sequence	Linker length	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure [Y/N]
1177 (ThioGoblin 1)	VTPVCTA	C3	Cys	UDP	2	1.3	Y / 1.85 Å
1223	VTPVCTA	C3	Cys	UMP-phosphonoacetate	360	n.a.	N
1245	VTPV(*D*-C)TA	C3	*D*-Cys	UDP	18	n.a.	N
1246	KENSPAVTPVCTA	C3	Cys	UDP	3	n.a.	N
1248	PKTVTPACSAKTSPAKQQAPPVRNLH	C3	Cys	UDP	90*	n.a.	N
1256	VTPVCTA (all *D*)	C3	*D*-Cys	UDP	50	n.a.	N

dvB_Code	Peptide sequence	Linker length	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure[y/n]
1267	VTPVCTA	C3	Cys	2'amino-2'deoxy-UDP	81	n.a.	N
1282	Floc-VTPVCTA	C3	Cys	UDP	n.a.	1.6	N
1294	VTPVCTA (all *D*)	C3	*D*-Cys	UMBP	>500	250*	N
1297	VTPVCTA	C3	Cys	UMBP	n.a.	180	N
1323	VTPVCTA	C3	Cys	5-fluoro-UDP	12	13	N
1372	VTPVCRA	C3	Cys	UDP	2	n.a.	N

dvB_Code	Peptide sequence	Linker length	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure[y/n]
1395	SPTCPSYSPTSPSYSPTS	C3	Cys	UDP	n.a.	70	N
1396	TPACSAKTSPAKQQPPT	C3	Cys	UDP	n.a.	260	N
1457	Floc-*Ahx*-VTPVCTA	C3	Cys	*L*-UDP	n.a.	n.d.	N
1505	VTPVCTA	2-acetamido C3	Cys	UDP	n.a.	1.4	Y / 1.97 Å
1624	SVPYCSA	C3	Cys	UDP	n.a.	2.3	N
1625	PVFTCRS	C3	Cys	UDP	n.a.	12.5	N
1626	STPVCSA	C3	Cys	UDP	n.a.	1.2	N

dvB_Code	Peptide sequence	Linker length	Acceptor residue	Nucleotide	IC_{50} / µM	Ki / µM	Crystal structure [y/n]
1630	SVPYCSAQS	C3	Cys	UDP	n.a.	15	N
1631	VTPVCTATH	C3	Cys	UDP	n.a.	12.7	N
1632	PVFTCRSAA	C3	Cys	UDP	n.a.	56	N
1635	PVCTATHSLSRLH	C3	Cys	UDP	n.a.	80	N
1636	Floc-STPVCSA	C3	Cys	UDP	n.a.	3.7	N
1648	VTPVCTA	Dimethyl-C3	Cys	UDP	n.a.	4	N
1655	VTPVCTA	C3	Cys	*L*-UDP	n.d.	n.a.	N

3.4. Modifications of the peptide sequence, linker nature and UDP moiety have an impact on potency

The bisubstrate conjugates presented in this study consist of three main parts (i) a short peptide, (ii) a short three- or four-carbon linker and (iii) UDP or an analogue. To investigate the effect of modifying either of these three parts has on the inhibitory potency of bisubstrate inhibitors, a set of 36 bisubstrate conjugates with different combinations of peptide sequence/length, linker length/style and UDP/UDP-analogue was characterised.

The majority of bisubstrate inhibitors tested were comprised of a short 7-mer peptide, a three-carbon linker and a UDP moiety (Table 3.2) and showed little to no changes in potency when the sequence was varied. However, this is not surprising as they closely resemble the common sequence motif (sequon) [TS][PT][VT][S/T][RLV][ASY] reported recently [43,231,247,248,250,252]. In agreement with the sequon, bisubstrate conjugates containing a proline in the -2 position, appear to be better binders (Table 3.2). This is further supported, as a bisubstrate conjugate carrying a phenylalanine in the -2 position (dvB_1625) binds to hOGT with a 10-fold lower affinity. Antagonistic effects on the binding affinity of the bisubstrate inhibitors were seen when the length of the peptide was varied. Addition of two residues to the C-terminus of the peptide generally resulted in a 10-fold drop in binding affinity, whereas addition of residues to the N-terminus do not affect potency (Table 3.2). A C-terminal extension of the peptide part could lead to steric clashes with the TPR domain of OGT, which would make it unfavourable for the inhibitors to bind. Changing the stereo configuration of the acceptor residue, or even the whole peptide, from an *L*- to a *D*-configuration only leads to a minor decrease in binding affinity (Table 3.2). This is in good agreement with previous studies reporting that acceptor peptides form mostly

weak van der Waals interactions with hOGT, which do not contribute to binding to the same extent that the interactions with UDP do [231,240,241].

A greater effect on inhibitory potency could be seen between bisubstrate inhibitors that differ in linker length (Table 3.2). In general, the longer four-carbon linker conjugates seem to have a reduced affinity towards hOGT, when compared to the three-carbon linker conjugates (Table 3.2). Examination of the Goblin 1 / ThioGoblin 1 structures allowed the determination of the distance between the linker-O/S and the β-phosphate that the three-carbon linker occupies, revealing a distance of 4.4 Å (Goblin 1) and 5.0 Å (ThioGoblin 1), closely mimicking the distance between the acceptor serine hydroxyl-group and the β-phosphate oxygen of UDP-5S-GlcNAc (4.6 Å), found in a previously reported structure (PDBID 5C1D [231]). Addition of a fourth carbon to the linker would likely make it more challenging for the bisubstrate conjugate to bind to hOGT.

Closer inspection of the ThioGoblin 1:OGT structure revealed that the C_2 carbon of the bisubstrate linker occupies a similar position to the anomeric carbon of the GlcNAc residue in the hOGT-substrate complex (Fig. 3.6), making it suitable for modifications that could recapitulate interactions formed by the GlcNAc sugar. The addition of an acetamido group to the linker appeared to have little effect on binding affinity binding affinity (K_d (dvB_1505) = 1.4 µM, Table 3.2). To investigate whether the addition of the acetamido groups had effects on the binding mode of the bisubstrate conjugate, high resolution diffraction data was collected of hOGT in complex with dvB_1505 (1.97 Å, R_{work}/R_{free} = 0.19/0.23) (Table 3.1). Structure solution by molecular replacement and subsequent refinement revealed continuous $|F_o\text{-}F_c|$ electron density, allowing for the unambiguous placement of the ligand (dvB_1505). The fully refined model revealed a well order linker and

acetamido group. The acetamido group occupies a similar space as the N-acetyl group of the GlcNAc residue (Fig. 3.6).

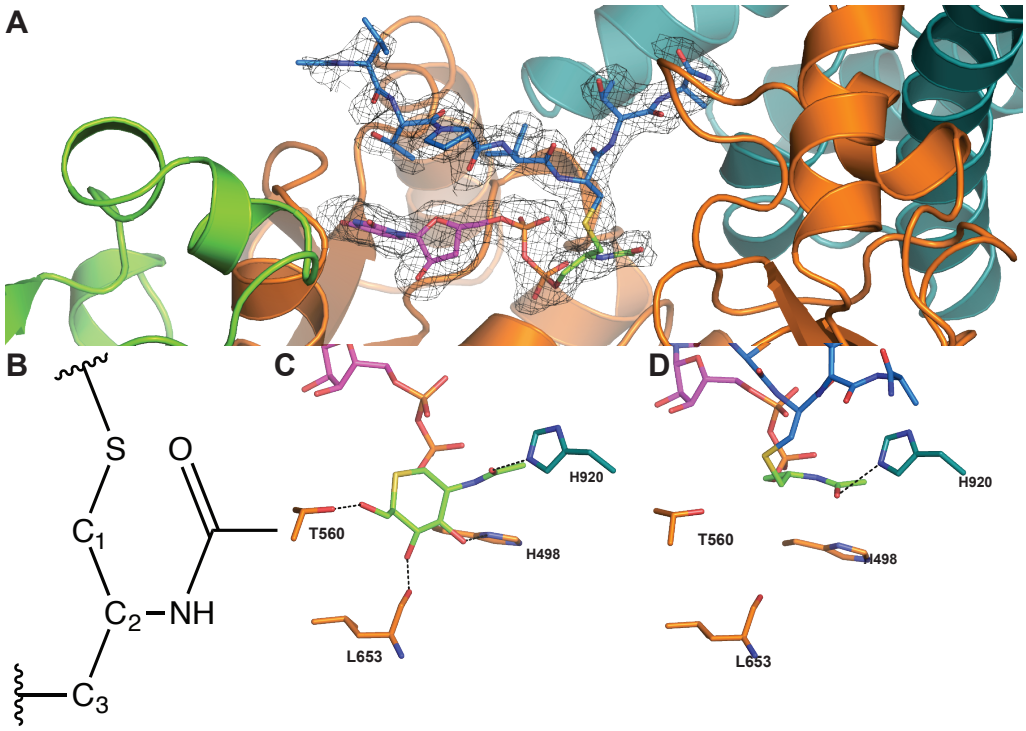

Figure 3.6: An acetamido modified linker adopts a similar conformation to the N-acetyl group of UDP-GlcNAc. (A) Structure of hOGT in complex with dvB_1505, a ThioGoblin 1 derivative harbouring an N-acetyl mimicking group on the C_2-carbon of the linker. hOGT is shown in cartoon representation with the TPR domain shown in teal, the catalytic lobes in orange and the intervening domain in green. The unbiased |Fo-Fc| maps for dvB_1505 is shown as a black mesh contoured to 2.25 σ. **(B)** Schematic drawing of the 2-acetamido modified linker. **(C**) Sticks representation of hOGT residues (orange & teal) forming hydrogen bonds with the GlcNAc residue (green) of UDP-5S-GlcNAc The UDP moiety is shown as magenta sticks (PDBID 5C1D [231]). (D) Stick representation of hOGT residues (orange & teal) usually forming hydrogen bonds with the GlcNAc residue of the sugar nucleotide and dvB_1505. The ligand is shown in stick representation with the UDP moiety shown in magenta, the linker in green and the peptide in blue.

Closer examination of the dvB_1505:OGT structure revealed that the acetamido group occupies nearly the same space as the N-acetyl group of the GlcNAc moiety of the nucleotide sugar analogue UDP-5S-GlcNAc [231], with a maximum atomic shift of 0.9 Å between the amine of the acetamido and N-acetyl groups. The structure also revealed that the acetamido group does not have an optimal orientation to form a hydrogen bond with His^{498}, as seen for the N-acetyl group of the sugar nucleotide [241] but may still form hydrogen bonds. The linker adopts a different conformation than ThioGoblin 1 with a dihedral angle of 86.7° (S-C_1-C_2-C_3), adopting a synclinal conformation similar to that of Goblin 1. It is possible that the introduction of an additional favourable interaction with OGT through the 2-acetamido group is negated by this unfavourable linker geometry.

It was shown previously that the nucleotide forms an intricate network of interactions with hOGT [240,241]. The most significant reductions in binding affinity were seen when the UDP moiety is modified. Introduction of either a UMP-phosphonoacetate, a 2'amino-2'deoxy-UDP or a uridine methylene bisphosphonate group in place of UDP leads to reductions in potency (Table 3.2), independent of the acceptor residue. This steep drop in inhibitory efficacy most likely occurs due changes in the in the pyrophosphate linkage, changing its electrostatic properties and preventing the formation of a phosphorus oxyanion that can interact with the electric dipole of the active site α-helix [240,241].

The uracil ring forms stacking interactions with His^{901} and hydrogen bonds with the carbonyl oxygen of Ala^{896} via its ring amine and the side chain of Arg^{904} via the 4-carbonyl oxygen but carries no functional groups at position C_5 and C_6 to form additional interactions. Fortuitously, the C_5 carbon of the uracil ring is positioned only 3.8 Å away from a hydrophobic pocket formed by residues Phe^{837},

Leu[866], Phe[868] and Val[895], creating the possibility of fluorinating C_5 to introduce additional non-polar interactions. However, fluorination of the C_5 position of the uracil ring resulted in a 10-fold drop of binding affinity compared to ThioGoblin 1 (Table 3.2). A possible explanation is that the π-electron system of the uracil ring is perturbed by the presence of fluorine, the most electronegative element, which could alter the stacking interactions with His[901].

Small modifications of physiological interactions were shown to have adverse effects on the potency of bisubstrate hOGT inhibitors. However, it also reveals that the combination of a modular synthetic approach for the synthesis of bisubstrate conjugates, fluorescence polarimetry and high-resolution macromolecular x-ray crystallography creates a platform for further expansion into this area of research.

3.5. The mechanism inspired bisubstrate conjugate ThioGoblin 1 inhibits OGT in cell lysates

Although ThioGoblin 1 is a low µM inhibitor of hOGT *in vitro*, it does not necessarily follow that it would also show inhibition in the context of a complex cell lysate. However, due to the polar nature of these bisubstrate conjugates it is unlikely that they would penetrate the cell membrane, without the help of shuttling agents such as cell-penetrating peptides and cholesterol [324]. As an alternative, the inhibition of hOGT by ThioGoblin 1 in a complex cell lysate was investigated. Here, HeLa cell lysates were pre-treated with the promiscuous O-GlcNAc hydrolase *Cp*OGA [185], to strip it of existing O-GlcNAc. Subsequently recombinantly expressed full-length hOGT to a final concentration of 1 µM and UDP-GlcNAc to a final concentration of 1 mM were added to the lysates and incubated for 1.5 h at 37 °C, in the presence and absence of ThioGoblin 1 and *L*-ThioGoblin 1. *Cp*OGA activity was inhibited by the addition of GlcNAcstatin G [218], a pM inhibitor of *Cp*OGA. Dynamic O-GlcNAcylation was visualised via Western blot analysis and quantified by integration of the fluorescent signal of a fluorescently-labelled secondary antibody (Fig. 3.7) Experiments were performed by Andrii Gorelik.

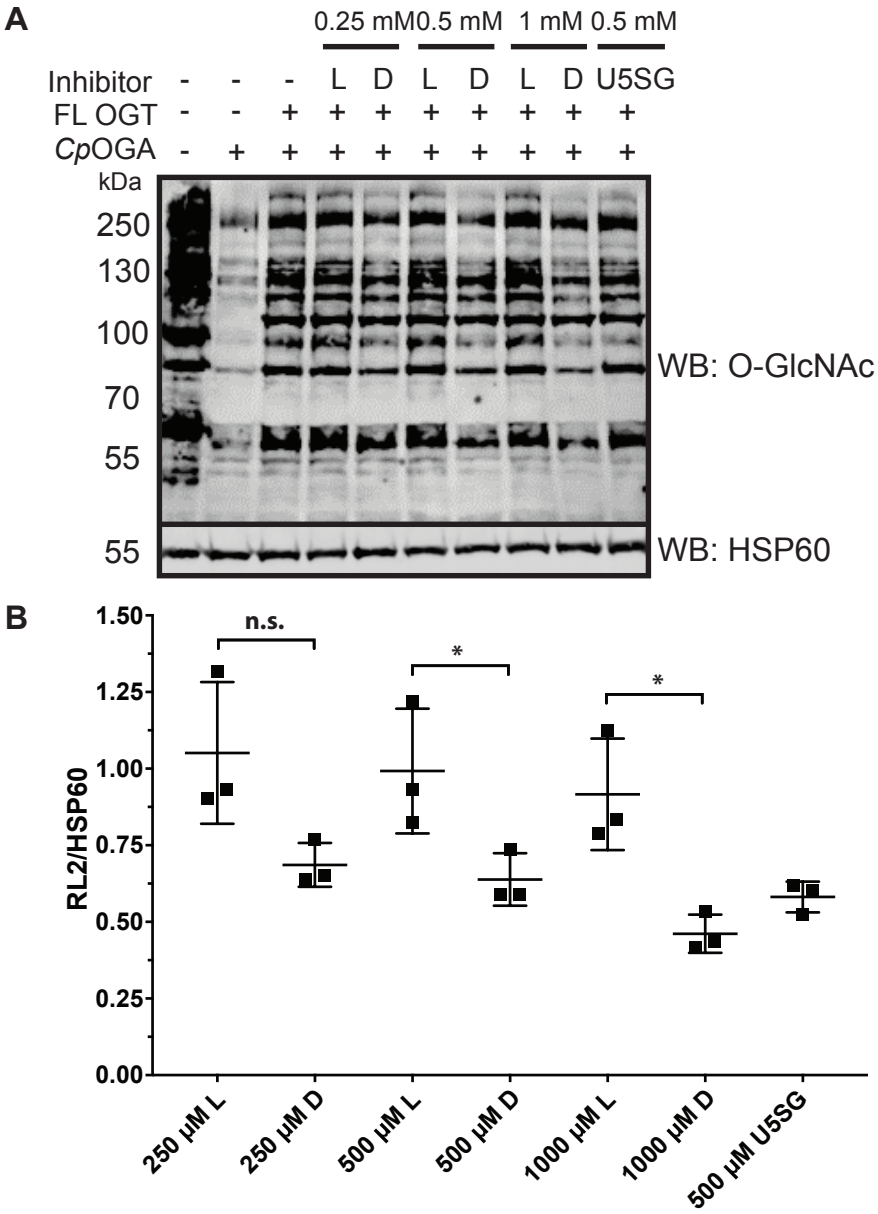

Figure 3.7: ThioGoblin 1 inhibits hOGT in a complex cell lysate setting. (A) Western blot analysis showing total O-GlcNAc signal of CpOGA treated cell lysates after exposure to recombinant full length hOGT and UDP-GlcNAc in the presence and absence of varying concentrations L = *L*-ThioGoblin 1 and D = ThioGoblin 1. UDP-5S-GlcNAc was used as a positive control. Equal loading was verified by immunoblotting for HSP60. **(B)** Quantification of total O-GlcNAc signal in the presence of L, D and UDP-5S-GlcNAc. Signal was normalised to the cell lysate treated with full length OGT in the absence of inhibitors. * = $P < 0.05$. these Experiments were performed by Andrii Gorelik.

Under the chosen experimental condition O-GlcNAcylation of multiple cellular proteins in the cell lysate could be observed, which was inhibited in a dose dependent manner by ThioGoblin 1, however not by its stereoisomer *L*-ThioGoblin 1, even at concentrations up to 1 mM. Even though ThioGoblin 1 was added up to a concentration 1000-fold higher than its K_d (Table 3.2), only a ~ 50% reduction of global O-GlcNAc could be seen, compared to the control lane. In a complex cell lysate ThioGoblin 1 has to compete with the endogenous pool of UDP-GlcNAc for binding to hOGT. Under physiological conditions part of the cellular UDP-GlcNAc pool is funnelled into the endoplasmatic reticulum (ER) and the Golgi apparatus for the N-glycosylation of proteins and the biosynthesis of complex glycans [3], reducing the amount of UDP-GlcNAc that can bind to OGT. Under the experimental conditions used here, the majority of endogenous GTs are likely inactive and UDP-GlcNAc formerly localised to the ER and Golgi is liberated, adding to the exogenous UDP-GlcNAc present. Furthermore, eukaryotic cell lysis is performed in the presence of detergents, which could affect the diffusion of ThioGoblin 1. Taken together, the higher level of sugar nucleotide and artificial environment might explain the low inhibitory potency seen here. However, ThioGoblin 1 does perform as well as the known hOGT inhibitor UDP-5S-GlcNAc, at comparable concentration both reducing global O-GlcNAcylation by ~ 40%. Thus, the mechanism inspired bisubstrate inhibitor ThioGoblin 1 inhibits hOGT in a complex cell lysate.

3.6. Concluding remarks

The attachment of β-N-acetylglucosamine to nucleocytoplasmic proteins is an abundant and essential post-translational modification, yet its precise biological role is unknown [31,338]. The development and design of potent, selective OGT inhibitors is a prerequisite for dissecting the biological role of O-GlcNAc and its role in disease formation and progression. Targeting multisubstrate enzymes such as kinases and glycosyl transferases can rarely be achieved with compounds that compete with a single substrate. Bisubstrate inhibitors are assembled using two conjugated fragments, each targeting a different binding site of a bisubstrate enzyme. The concept of bisubstrate inhibitors dates back to the early 1970s and was developed in parallel to the transition-transition state analogue concept [339–341]. Early bisubstrate inhibitors of kinases, comprised of a peptide and nucleotide-pyrophosphate, were reported over three decades ago and have since allowed the development of low nM cell-penetrant bisubstrate inhibitors [342–346]. Early pioneering work on the development of bisubstrate inhibitors targeting glycosyl transferases was introduced by Palcic *et al.*, reporting the design of a bisubstrate fucosyl transferase inhibitor [347] and have since then been improved on [348,349].

Recently two OGT bisubstrate inhibitors were reported [333]. In the present study, advances in understanding the mechanism of action of this class of inhibitors and resulting improvements in their design are reported. The development of a modular synthetic approach using TEC induced formation of thiol-carbon linkages, has greatly accelerated the synthesis of bisubstrate OGT inhibitors. Using high-resolution macromolecular crystallography and biophysical characterisation, the inhibitory potency of a small panel of 36 linear bisubstrate inhibitors were investigated (Table 3.2). Combination of the modular synthesis

and structural and biophysical characterisation, yielded a handful of bisubstrate conjugates capable of inhibiting OGT in the low μM range *in vitro* (Table 3.2), just short of inhibitory potencies seen for recently reported kinase bisubstrate inhibitors [350].

Substitution of the linker serine with a cysteine resulted in the ThioGoblin 1 bisubstrate inhibitor that was found to be a 10-fold more potent inhibitor than its serine linked progenitor Goblin 1 [333], making it the most potent inhibitor of OGT known to date [285] (Fig. 3.2, Table 3.2). Intrigued by this sudden increase in potency, structural studies were performed to investigate changes in the binding mode of the bisubstrate inhibitor. High-resolution diffraction data of hOGT in complex with either ThioGoblin 1 or Goblin 1 allowed the dissection of small changes in the binding conformation between both ligands (Fig. 3.3). Goblin 1 and ThioGoblin 1 are capable of mimicking a pseudo-Michaelis complex between hOGT and its substrates [231] retaining the alignment of the UDP and peptide moieties, with a maximum atomic shift of the UDP moieties between 0.7-0.8 Å and no measurable atomic shift for the peptide, recapitulating the binding seen for a previously published low-resolution Goblin 1:hOGT complex [333]. However, introduction of a thiol-linkage into the bisubstrate scaffold results in a change of dihedral angles from a synclinal to an antiperiplanar configuration, which is energetically more favourable and leads to a binding of ThioGoblin 1 to hOGT.

Modification of the peptide sequence/length, linker type or nucleotide led to measurable changes in potency of the bisubstrate OGT inhibitors. Results showed that a proline in the -2 position of the peptide, in respect to the linker residue, benefits binding of the bisubstrate conjugate and results in more potent inhibitors, which is in good agreement with previously reported sequence preference of hOGT [43,231,247,250,252]. Recent reports and results presented in

chapter 2 of this work, describe the recognition of peptide substrates inside the TPR superhelix [10,237], suggesting a C-terminal extension of our bisubstrate inhibitors would be beneficial, assuming that this is a common recognition mode for all substrates of hOGT. However, a significant reduction in potency was seen, when residues were added to the C-terminus of the bisubstrate peptide (Table 3.2) facing the TPR domain, likely introducing steric clashes unfavourable for binding and indicating that only a subset of OGT peptides can be recognised in the superhelical fold of the TPR domain. Introduction of additional residues on the N-terminus of the bisubstrate inhibitors had no measurable effect on potency (Table 3.2).

The most significant effect on potency was seen when the pyrophosphate bond of the nucleotide was altered (Table 3.2), leading to a sharp decrease in inhibition potency. Structural characterisation of OGT in complex with UDP or the substrate analogue UDP-5S-GlcNAc revealed involvement of the pyrophosphate group in the formation of multiple interactions with hOGT [240,241]. One important interaction is the formation of a phosphorus oxyanion, capable of interacting with the induced dipole moment of a proximal active site α-helix. Unsurprisingly, changes in the electronic configuration of the pyrophosphate that lead to a loss of the oxyanion may impede binding to OGT. Previously published substrate and product analogues capable of inhibiting OGT also suffered from a drop in potency when the pyrophosphate bond was altered [285].

OGT binds its substrates in an ordered bi-bi mechanism, with the sugar nucleotide binding first, followed by the acceptor substrate [240]. The uracil ring binds in a deep pocket in the active site, forming polar and stacking interactions with hOGT. The orientation of the uracil ring in the active site positions its C_5 carbon so it faces a hydrophobic pocket formed by Phe^{837}, Leu^{866}, Phe^{868} and

Val[895], making it an interesting target for modifications with hydrophobic groups. However, a bisubstrate conjugate carrying a fluorine on the C_5 carbon of the uracil ring resulted in a 10-fold decrease of binding affinity towards hOGT. Fluorine is the most electronegative element known and has been shown to alter the π-electron system of aromatic structures [351,352]. This perturbation may negatively affect the stacking interaction of the uracil ring and His[901] of OGT potentially leading to a decrease in potency of the bisubstrate inhibitor that could counteract the potential interaction of the fluorine with the targeted hydrophobic pocket. Introduction of a less electronegative methyl group could be considered in the design of future bisubstrate inhibitors, to target the hydrophobic, without altering the π-electron system of the uracil ring. A stereoisomer of ThioGoblin 1 decorated with an *L*-UDP instead of *D*-UDP significantly reduced the binding affinity of the bisubstrate inhibitor and is not able to inhibit OGT activity *in vitro* (Fig 3.2, Table 3.2), making it a suitable negative control in cell biology experiments to study any OGT independent phenotypes caused by treatment with ThioGoblin 1.

Determination of the inhibition potency of bisubstrate conjugates with different linker length, revealed a preference for a three-carbon linker over a four-carbon linker (Table 3.2). Measurement of the distance between the acceptor hydroxyl-group and the oxygen linking the anomeric carbon and the β-phosphate, revealed a distance of 4.5 Å [231], which is comparable to the length of a three chain carbon (~ 4.62 Å) and may eplain the preference of a three-carbon linker over a four-carbon linker. Structural characterisation of the ThioGoblin 1:hOGT complex revealed a potential modification point of the C_2 carbon in the linker moiety of the bisubstrate scaffold. The attachment of a acetamido group to the linker C_2 carbon did not result in an increase in inhibition potency of the bisubstrate inhibitor (Table 3.2, Fig. 3.6). Structural characterisation of the dvB_1505:hOGT complex

revealed that the acetamido group occupies a similar space to the N-acetyl group of the GlcNAc residue of the nucleotide sugar analogue UDP-5S-GlcNAc [231], with a maximum atomic shift of 0.9 Å between the acetamido/GlcNAc amine groups. However, the orientation of the acetamido group is less favourable for the formation of a hydrogen bond with His[498] of hOGT. Previous affinity measurements showed that the GlcNAc residue does not contribute significantly to the binding of the nucleotide sugar to OGT [241], which taken together with the orientation of the acetamido group seen in the dvB_1505:hOGT structure could explain comparable binding affinity to ThioGoblin 1.

The use of fluorescence polarimetry has greatly accelerated the discovery of novel protein binders and has been used extensively recently [283,296,331,353–355]. A similar system for the measurement of OGT binders has been described before [296]. However, Gross *et al.* introduced a fluorescein molecule to the N-acetyl group of UDP-GlcNAc via a three-carbon linker and cannot exclude possible steric clashes between the probe and hOGT. Moreover, is likely to be unstable for long periods of time as it can undergo OGT independent auto-hydrolysis or OGT catalysed transfer of the GlcNAc onto water, which could affect read-out of the FP signal. The development of a FP assay system using a fluorescently tagged ThioGoblin 1 derivative has allowed the rapid determination of binding affinities for OGT binding ligands (Table 3.2). Furthermore, it addresses the potential instability issues that could arise from using the fluorescent probe reported earlier [296] by replacing it with the non-hydrolysable F-ThioGoblin 1.

Finally, using HeLa cell lysates the inhibition of hOGT in a complex cell lysate by ThioGoblin 1 was investigated, showing a measurable reduction in global O-GlcNAcylation comparable to the lysates treated with UDP-5S-GlcNAc,

suggesting that ThioGoblin 1 is capable of inhibiting hOGT in a complex cell lysate.

The use of bisubstrate conjugates to target OGT addresses pitfalls of currently available OGT inhibitors such as their a specific nature and low potency [243,356]. The introduction of a thioether-linkage into the bisubstrate scaffold has significantly increased their potency towards hOGT and has moved them on par with the to date strongest known OGT binder UDP. In combination with the use of OGT specific substrates the bisubstrate conjugates represent the most potent and specific OGT inhibitors to date. However, these bisubstrate scaffolds are unlikely to cross cell membranes and are therefore unsuitable for cell biology studies. This lack of cell-penetrance could be addressed by attaching cell-penetrating peptide sequences to the inhibitors to promote cellular uptake. The design of cell-penetrant bisubstrate OGT inhibitors is covered in chapter 4.

4. Development of bisubstrate conjugates that target the human O-GlcNAc transferase in cellulo

4.1 Introduction

In recent years, the use of high molecular weight biomolecules as novel pharmacophores has gained traction [357,358]. However, due to their polar / hydrophilic nature, large biomolecules have a low membrane permeability and are thus poor at penetrating the physical barrier of a eukaryotic cell membrane. Nearly two decades ago a new class of peptides was reported capable of promoting cellular uptake of large hydrophilic macromolecules [359]. Derossi *et al.* reported the identification of a 16-mer peptide sequence (RQIKIWFQNRRMKWKK) derived from the third-helix of the *Drosophila melanogaster* Antennapedia transcription factor homeodomain (pAntp) [359]. The homeodomain protein pAntp belongs to a class of trans-activating transcription factors involved in various morphological processes [360–363] and was shown to translocate across neuronal membranes and is conveyed to the nucleus where it binds to the DNA to control the expression of *homeotic* genes [364,365]. Antennapedia, together with a peptide derived from the HIV-1 Tat protein (YGRKKRRQRRR) [366], were the first examples of such cell-penetrating peptides (CPP). Since their discovery various CPPs, derived from natural, hybrid and synthetic sources have been reported (Table 4.1), reviewed in [367].

Table 4.1: List of commonly used cell-penetrating peptides.

CPP	Sequence	Source	Reference
Antennapedia / Penetratin	**RQIKIWFQNRRMKWKK**	*Drosophila melanogaster*	359
Tat	**YGRKKRRQRRR**	HIV-1	366,368,369
transportan	**GWTLNSAGYLLGPHIDNHRSFHDKYGLA**	Neuropeptide galanin + wasp venom mastoparan	370
polyArginine	**R_6, R_7, R_8, R_9, R_{12}**	Synthetic	371–374

The discovery and characterisation of CPPs enabled the transport of non-cell permeable biologically active cargos across cell membranes, from peptides [372,375,376] to large proteins [377–379] and *oligo*-nucleotides [380–382] (Fig. 4.1).

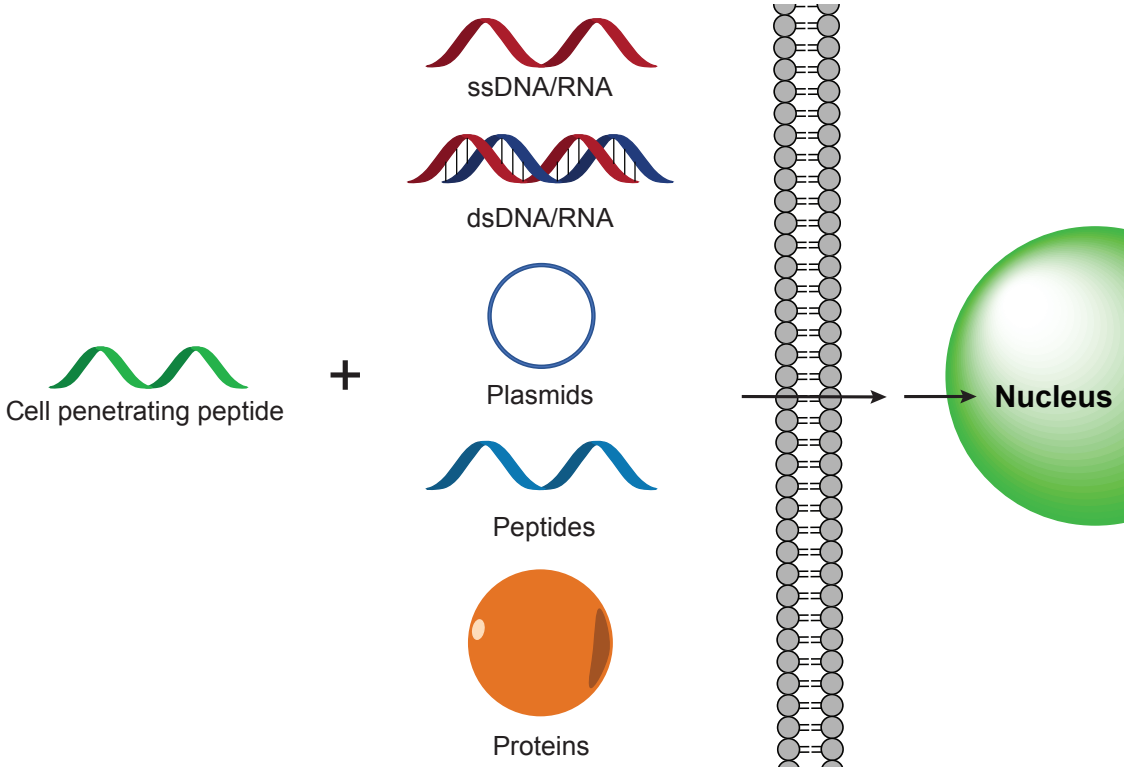

Figure 4.1: Schematic representation of various cargos that have been successfully shuttled into the via cell-penetrating peptides across the cell membrane. Figure adapted from [367].

The molecular mechanism by which CPPs translocate across cellular membranes is still elusive. The common understanding is that CPPs induce energy-independent endocytosis of cargo by binding to anionic heparin sulfates or anionic surface lipids [383], followed by a release of the cargo into the cytosol by acidification of the endosome [384]. However, the broad use of these cargo shuttling peptides is limited by two major factors: (i) the delivery is tissue unspecific and (ii) limited endosomal escape after formation of the endosome [385]. More recently, a new generation of cyclised CPPs (cCPP) received attention due to their increased stability and ability to escape the early endosome [386–388]. Using cCPPs

Lian *et al.* have successfully targeted the protein tyrosine phosphatase 1B and the peptidyl-prolyl cis-trans isomerase Pin1 using bicyclic peptide inhibitors [388].

In the following chapter, attempts to achieve cell penetrance of the bisubstrate hOGT inhibitors discussed in chapter 3, are discussed. Translocation of bisubstrate conjugates across the cellular membrane is attempted by attaching CPP sequences to the N-terminal tail of the peptide part of the bisubstrate scaffold. Using variations of the penetratin/antennapedia, Tat and polyArg sequences (Table 3.2), as well as cyclising the bisubstrate inhibitor scaffolds with a sequence of residues that promote cellular uptake (Table 3.2), limited degrees of endosomal uptake and release into the cytoplasm were achieved.

4.2 Materials and Methods

Chemical synthesis was performed by Dr. Vladimir S. Borodkin. Fluorescence microscopy was performed by Dr. Riccardo Trapannone and Andrii Gorelik. Western blot experiments performed by Dr. Riccardo Trapannone are indicated as such.

4.2.1 Crystallography

Human OGT (312-1031) was recombinantly expressed as a cleavable GST-fusion protein and purified as described previously [241]. A solution of 8 mg/mL hOGT and 1 mM bisubstrate conjugate in 25 mM Tris-HCl pH 8.0, 20 mM NaCl, 0.5 mM TCEP was incubated on ice for 15. Hanging-drop crystallisation experiments were performed by mixing drops containing 1 µL of protein-ligand solution and 0.5 µL reservoir solution [1.3 M DL-Malic acid, 0.1 M Bis-Tris propane pH 6.4] supplemented with hOGT crystal seeds, grown in the same condition as for the dvB_1399:OGT / dvB_1523:OGT complex. Crystals grew O/N and were cryo-protected by short immersion in 2.5 M sodium malonate pH 7.0, supplemented with 1 mM of respective bisubstrate conjugate and flash-frozen in liquid nitrogen. Data were collected at the European Synchrotron Radiation Facility (ESRF) on beamline ID30A-3 (dvB_1399) and ID23-2 (dvB_1523), processed with XDS [325] and scaled to 2.60 Å (dvB_1399) and 2.10 Å (dvB_1523) using aimless [326]. 5 % of total reflections were set aside as an R_{free} test set. Crystals belonged to space group F222 with one molecule per asymmetric unit, a solvent content of 61.9/62.5% and a Matthews coefficient of 3.23/3.28 for dvB_1339 and dvB_1523, respectively. The structures were solved with MOLREP [327], using chain A of PDB 3PE4 [240] as a search model. The

structure was fully refined using iterative cycles of Refmac5 [328] and COOT [329]. Ligand topologies were generated with PRODRG [330].

4.2.2 Fluorescence polarimetry

The binding affinity of dvB_1566 was determined by titrating hOGT at a fixed concentration of F-ThioGoblin 1, described in chapter 3, (125 nM) in 0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5 mM TCEP and 5 % DMSO in a total volume of 25 µL and incubated in the dark for 30 min before read-out.

Binding affinities for ligands were determined by F-ThioGoblin 1 displacement. Reactions contained 0.8 µM hOGT and 0.75 µM F-ThioGoblin 1 in 0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5 mM TCEP and 5 % DMSO in a total volume of 25 µL. Reaction mixtures were incubated with varying concentrations of inhibitor for 0.5 h in the dark. Fluorescence polarimetry was measured on a PHERAStar plate reader (BMG LABTECH) at excitation and emission wavelengths of 485 nm and 520 nm, respectively. Data analysis was performed in GraphPad Prism 7. Binding constants for ligands in displacement experiments were determined using a 4-parameter non-linear regression curve fit and the equation reported by Nikolovska-Coleska *et al.*[331].

4.2.3 hOGT activity measurements

Human OGT (312-1031) was recombinantly expressed as a His_6-fusion protein and was purified as described previously [231]. OGT activity was determined using reactions containing 50 nM His_6-hOGT in 50 mM Tris-HCl pH 7.5, 0.1 mg/mL BSA, 10 µM sodium dithionate and 10 µM of peptide (KKENSPAVTPVSTA) in a total volume of 100 µL. Reaction mixtures were pre-incubated with varying concentrations of ligands for 15 min and initiated by addition of UDP-GlcNAc to a final concentration of 3.2 µM. The reaction was stopped after 30 minutes 50 nM hOGT at 21 °C by addition of 200 µL of 25 mM HEPES pH 7.4, 10 mM NaCl, 75 µM pyrocatechol violet, 50 % (v/v) MeOH and 15 µM fluorophore [231,319,334]. UDP formation was detected fluorimetrically on a SpectraMax i3x multimode-plate reader (Molecular Devices) at excitation and emission wavelengths of 485 nm and 530 nm, respectively. Data analysis was performed with GraphPad Prism 7. Turnover did not exceed 10% of substrate.

4.2.4 HEK293 cell treatment and Western blot analysis

HEK293 cells were grown in Dulbecco's Modified Eagle's Medium supplemented with 10 % FBS, 2 mM L-Glutamine and 1 % Penicillin (100 U/mL) / Streptavidin (100 µg/mL) at 37 °C and 5 % CO_2. Cells were grown in a 6-well plate to 70 % confluency before treating with varying concentrations of dvB_1565, _1566 or Ac_4-5S-GlcNAc (positive control) for 24 h at 37 °C / 5 % CO_2. All ligands were dissolved in 100 % DMSO with a final DMSO concentration in the assay of max. 0.2 %. After 24 h cells were washed 3x times with ice-cold PBS buffer before addition of 50 µL lysis buffer (50 mM Tris-HCl pH 7.5, 0.1 mM EGTA, 1 mM EDTA, 1% Triton-X 100, 1 mM Na_3VO_4, 50 mM NaF, 5 mM $Na_4P_2O_7$, 0.27 M sucrose, 1 mM β-Mercaptoethanol, 1 mM benzamidine, 0.2 mM PMSF, 5 µM leupeptin) to

each well before transferring into a 1.5 mL vial and incubated on ice for 10 minutes. Cell lysates were spun down at 17000x g for 10 minutes and the supernatant transferred into fresh tubes. Protein concentration was quantified using the Bradford assay. 4x LDS loading dye was added to a final concentration of 1x LDS and samples boiled for 5 minutes at 95 °C. A total of 25 µg of lysate was loaded onto an SDS-PAGE gel (6 % / 8 % / 4-12 % Bis-Tris NuPAGE Invitrogen) and proteins resolved by electrophoretic mobility at 120 V for 120 minutes. Proteins were then transferred onto nitrocellulose membranes (GE Healthcare) by wet-transfer at 35 V for 90 minutes. Loading was controlled by Ponceau-S protein dye staining before blocking in 5 % BSA in 0.1 M Tris-HCl pH 7.4, 0.15 M NaCl buffer supplemented with 0.1 % Tween (TBST). Membranes were then incubated with anti-O-GlcNAc (RL2, 1: 1000) antibody and anti-HSP90β (1:1000) in 5 % BSA in TBST over night at 4 °C. Membranes were subsequently washed 3x in TBST before incubating with secondary fluorescently-labelled antibodies (anti-O-GlcNAc – anti-mouse IR680 1:10000, anti-HSP90β – anti-rabbit IR800 1:10000) in 5 % BSA in TBST for 45 minutes at room temperature. Blots were imaged using the Li-Cor Odyssey infrared imaging system (Li-Cor, Lincoln, NE) and fluorescent signal quantified using the ImageStudioLite (Licor) software.

4.3 Results and Discussion

4.3.1 An N-terminal extension of a bisubstrate conjugate with a CPP sequence does not impair binding to hOGT and promotes endosomal uptake

Potent and selective inhibition of multisubstrate enzymes, such as OGT, is rarely achieved with ligands that bind solely to a single site. The conjugation of two substrates/substrate analogues, targeting different binding sites of the same bisubstrate enzyme, allows the generation of bisubstrate inhibitors that, in theory, exhibit high selectivity and potency towards the target enzyme [350]. However, bisubstrate inhibitors containing peptide sequences or polar groups are unlikely to cross the cellular membrane. Achieving cell penetrance is one of the major hurdles that need to be overcome for bisubstrate inhibitors to become suitable tools to study cell biological processes. The characterisation of multiple bisubstrate hOGT inhibitors discussed in the previous chapter revealed the potential modification of the scaffolds on the N-terminus of the peptide, without a reduction in inhibitory potency. Bisubstrate conjugates carrying Antennapedia/Penetratin- and Tat- sequences were successfully synthesised and showed no reduction of hOGT inhibition *in vitro* (Table 4.2, Fig. 4.2), when compared to their parent scaffold ThioGoblin 1 (IC_{50} = 2 μM, Fig 4.2).

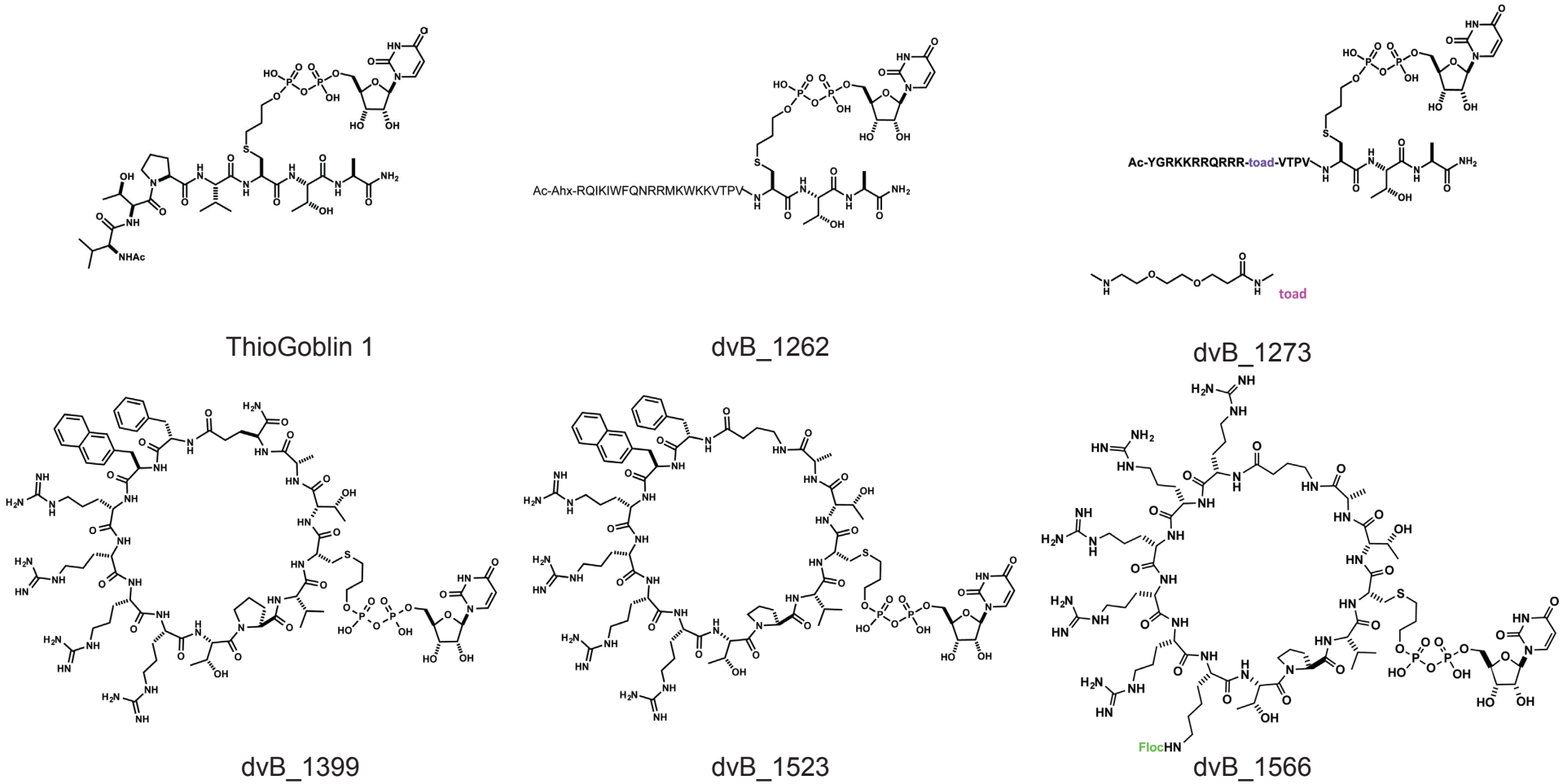

Figure 4.2: Chemical structures of bisubstrate inhibitors discussed in the present chapter. ThioGoblin 1 (Chapter 3), , dvB_1262, dvB_1273, dvB_1399, dvB_1523, dvB_1566.

Table 4.2: Overview of bisubstrate conjugates designed and tested in the course of the study. DvB codes are inhouse ID numbers, the acceptor residue in each peptide is highlighted in red. Residues used for cyclisation are marked in orange and a lower case c. Resolutions in the final column represent the highest resolution limit. Unless stated otherwise, all IC_{50} values were obtained using the fluorometric steady-state assay system and all K_i values via the fluorescence polarimetry assay. N.a. = not available, n.d. = not determinable, * = value amibgous, Floc = 5(6) fluorescein carboxamide, Ahx = 6-aminohexanoic acid, tod = 4-(2-(2-(methylamino)ethoxy)ethoxy)butan-2-one, ndx = N-(6-(2,5-dioxopyrrolidin-1-yl)hexyl, Φ = D-Naphtilalanine.

dvB_Code	Peptide sequence	Linker	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure [y/n]
1059	KKKKKKVTPVSTA	C3	Ser	UDP	19	n.a.	N
1262	Antennapedia-*toad*-VTPVCTA	C3	Cys	UDP	n.a.	n.a.	N
1273	TAT-*toad*-VTPVCTA	C3	Cys	UDP	5	n.a.	N
1399	cCTA(iQ)FΦRRRRTPVc	C3	Cys	UDP	n.a.	8	Y / 2.6 Å

dvB_Code	Peptide sequence	Linker	Acceptor residue	Nucleotide	IC_{50} / µM	K_i / µM	Crystal structure [y/n]
1507	cCTAF(Φ)RRRRTPVc	C3	Cys	UDP	n.a.	28	N
1523	cCTA(*Abu*)FΦRRRRTPVc	C3	Cys	UDP	n.a.	< 1*	Y / 2.1 Å
1526	cCTAGKVTPVc	C3	Cys	UDP	n.a.	50	N
1548	cC(*ndx*Floc)RRRRΦFc	C3	Cys	---	n.a.	n.a.	N
1549	cC(*ndx*Floc)TA(*Abu*)FΦRRRRTPVc	C3	Cys	---	n.a.	n.a.	N
1550	cRRRRΦFCc(*nsAbuAhx*)VTPVCTA	C3	Cys	UDP	n.a.	1	N
1566	cCTA(*Abu*)RRRRRK(Floc)TPVc	C3	Cys	UDP	n.a.	n.a.	N

Cellular uptake of bisubstrate scaffolds decorated with linear CPP sequences (Table 4.2) was measured by determining global O-GlcNAc levels of HEK293 cells treated with dvB_1262 and 1273 (Table 4.2, Fig. 4.2, work done by Dr. Riccardo Trapannone).

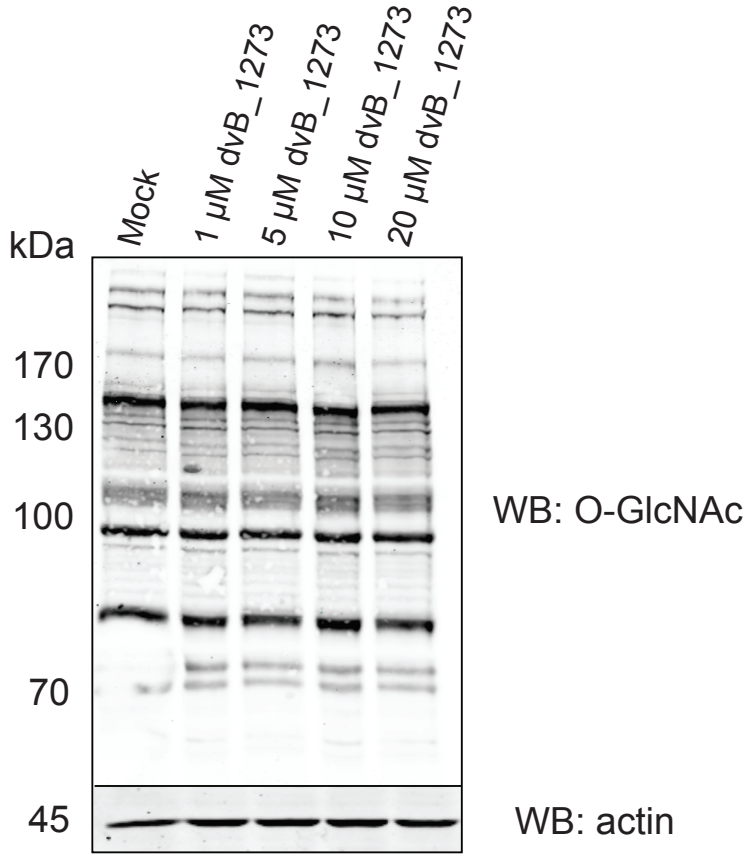

Figure 4.3: Western blot analysis of HEK293 cells treated with dvB_1273. Global O-GlcNAc levels were visualised using an O-GlcNAc specific antibody (RL2, Abcam, Cat#: ab2739). Work done by Dr. Riccardo Trapannone.

Western blot analysis of dvB_1273 treated HEK293 did not show a reduction in global O-GlcNAc levels (Figure 4.3) nor was a reduction seen for HEK293 cells treated with dvB_1262 (data not shown here, work done by Dr. Riccardo Trapannone). Intrigued by these results, fluorescence microscopy studies of HeLa cells treated with an N-terminally fluorescein modified dvB_1262 were performed to investigate whether the addition of a CPP promoted cellular uptake (work done by Dr. Riccardo Trapannone) (Fig. 4.3).

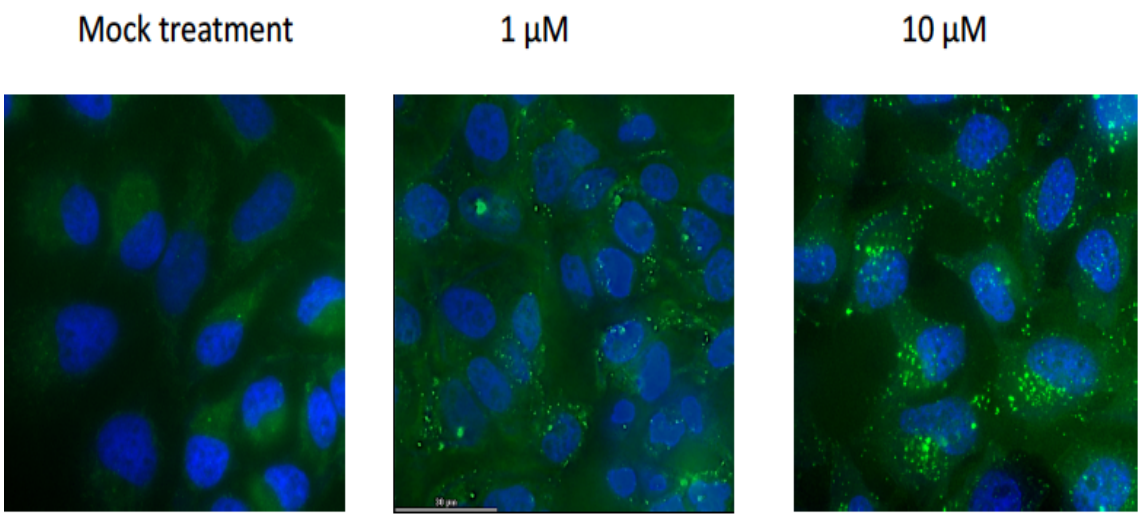

Figure 4.4: DvB_1262 is readily taken up by cells, however does not escape the endosome. Deconvolution microscopy images showing HeLa cells treated with 1 and 10 µM of an N-terminal fluorescein modified dvB_1262 (Table 4.2). Blue = DAPI, Green = fluorescein.

Deconvolution microscopy images showed concentration of fluorescein signal in multiple punctae throughout the cell, indicating that whilst dvB_1262 is readily taken up via endocytosis it appears to be unable to escape the endosome (Fig. 4.4). Lack of or limited endosomal escape of linear CPP-carrying bisubstrate conjugates into the cytoplasm may explain the lack of hOGT inhibition in dvB_1262 and dvB_1273 treated cells (Fig. 4.3).

4.3.2 Cyclised bisubstrate scaffolds bind to hOGT with low µM affinity and bind to hOGT by mimicking a pseudo-Michaelis complex

Limited endosomal escape of biomolecules with linear CPP sequences is a limitation that has been described previously [385]. Recently, Qian *et al.* reported the synthesis and application of cyclic CPP peptides that translocate large peptides and proteins across cellular membranes and promote early endosomal escape, using a unique CPP peptide motif (RRRRΦF, Φ = naphthylalanine) [388]. The authors also showed increased stability of the cyclic scaffolds in serum

containing cell media compared to linear CPP sequences [388]. Incorporation of the CPP RRRRΦF and cyclisation of the bisubstrate inhibitor scaffolds could have additional advantages such as (i) protection against protease degradation and (ii) an increase in binding affinity by reduction of the entropic penalty. To investigate whether the cellular uptake and early endosomal escape of the reported cyclic CPP scaffolds reported by Qian *et al.* could be reproduced, a small panel of fluorescently labelled cyclic CPP peptides was synthesised and their cellular uptake followed by fluorescence microscopy experiments (work done by Dr. Riccardo Trapannone) (Fig. 4.5).

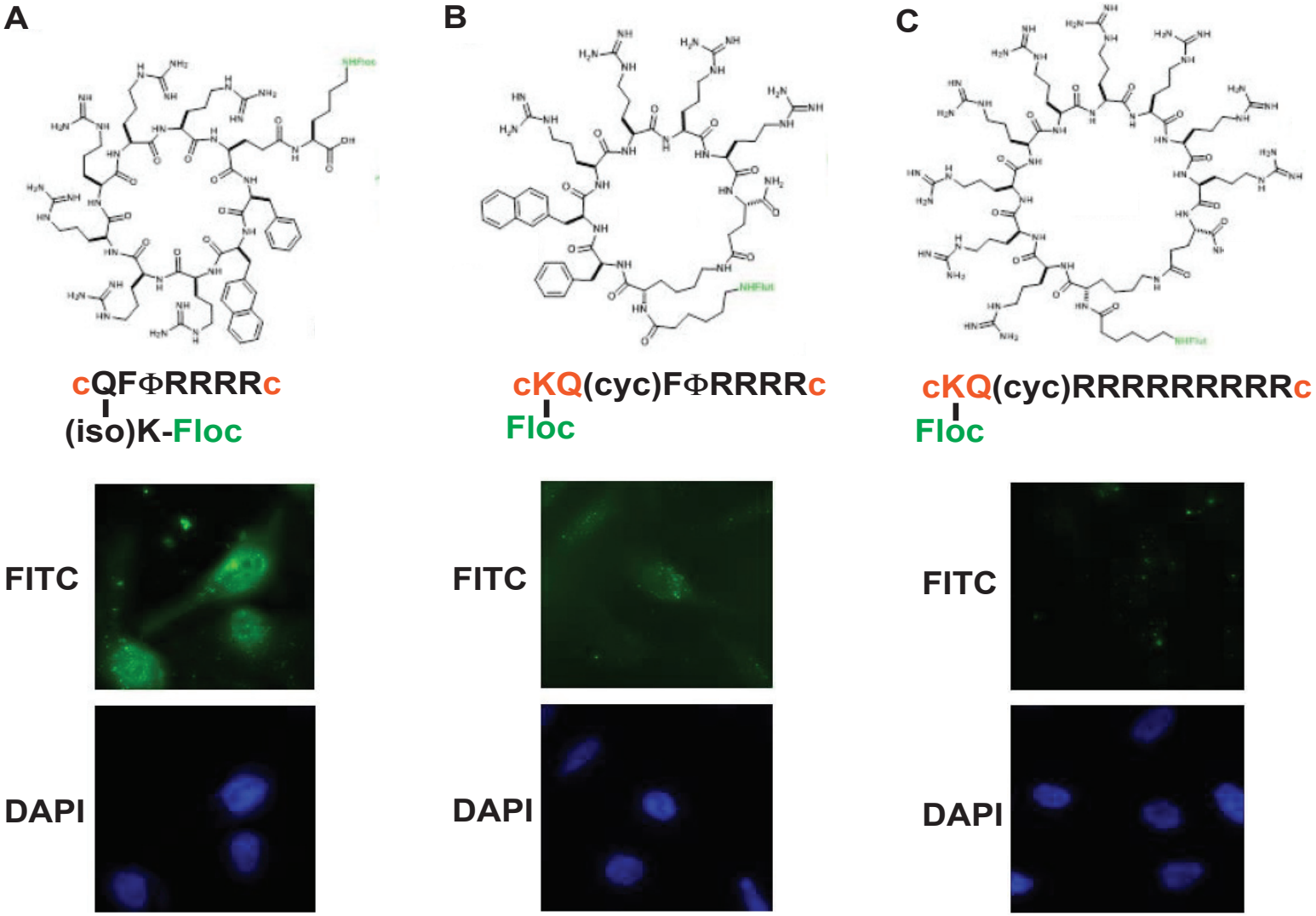
A
B
C
cQFΦRRRRc
(iso)K-Floc
cKQ(cyc)FΦRRRRc
Floc
cKQ(cyc)RRRRRRRRRRc
Floc
FITC
DAPI
FITC
DAPI
FITC
DAPI

Figure 4.5: Different cyclic cell penetrating peptides are taking up by cells with varying success. (A) Cyclised peptide with the sequence QFΦRRRR. **(B)** Cyclised peptide with the sequence QFΦRRRRK. The lysine and glutamine side chains were introduced as part of the cycle. **(C)** Cyclised peptide with the sequence QRRRRRRRRK. The lysine and glutamine side chains are incorporated into the cycle **(A-C)** Microscopy images show HeLa cells treated with 10 µM of respective cyclic CPP for 2 h. This work was done by Dr. Riccardo Trapannone. FITC = Fluorescein isothiocyanate, DAPI = 2-(4-amidinophenyl)-1H -indole-6-carboxamidine, Floc = 5(6)-carboxyfluorescein.

Deconvolution microscopy images of HeLa cells treated with 10 µM of fluorescein labelled cCPPs showed diffused fluorescent signal in the nucleocytoplasm to a varying degree (Fig. 4.5). A cyclic-CPP with the sequence RRRRΦF showed strong, diffused signal in the nucleocytoplasm, suggesting successful translocation across the cellular membrane and endosomal escape (Fig. 4.5 A), recapitulating earlier studies [386]. However, cCPPs with the sequences QFΦRRRRK and QRRRRRRRRRK that incorporated the glutamine and lysine into the cycle, showed little to no translocation across the cell membrane, respectively (Fig. 4.5 B, C).

To investigate whether a cyclic bisubstrate inhibitor based on the ThioGoblin 1 scaffold (Fig 4.2) is able to bind to hOGT, a cyclic peptide, containing the CPP sequence QRRRRΦF and the peptide sequence TPVCTAQ, was successfully synthesised. Cyclisation was induced by *iso*peptide bond formation between the C-terminal glutamate side chain and the N-terminal phenylalanine amine, followed by attachment of a UDP-allyl via TEC induced S-C bond formation with the cysteine (chapter 3), culminating in the cyclic bisubstrate scaffold dvB_1399 (Fig. 4.2). Binding to hOGT of dvB_1399 was determined using the fluorescence polarimetry assay system described in chapter 3 (Fig. 4.6).

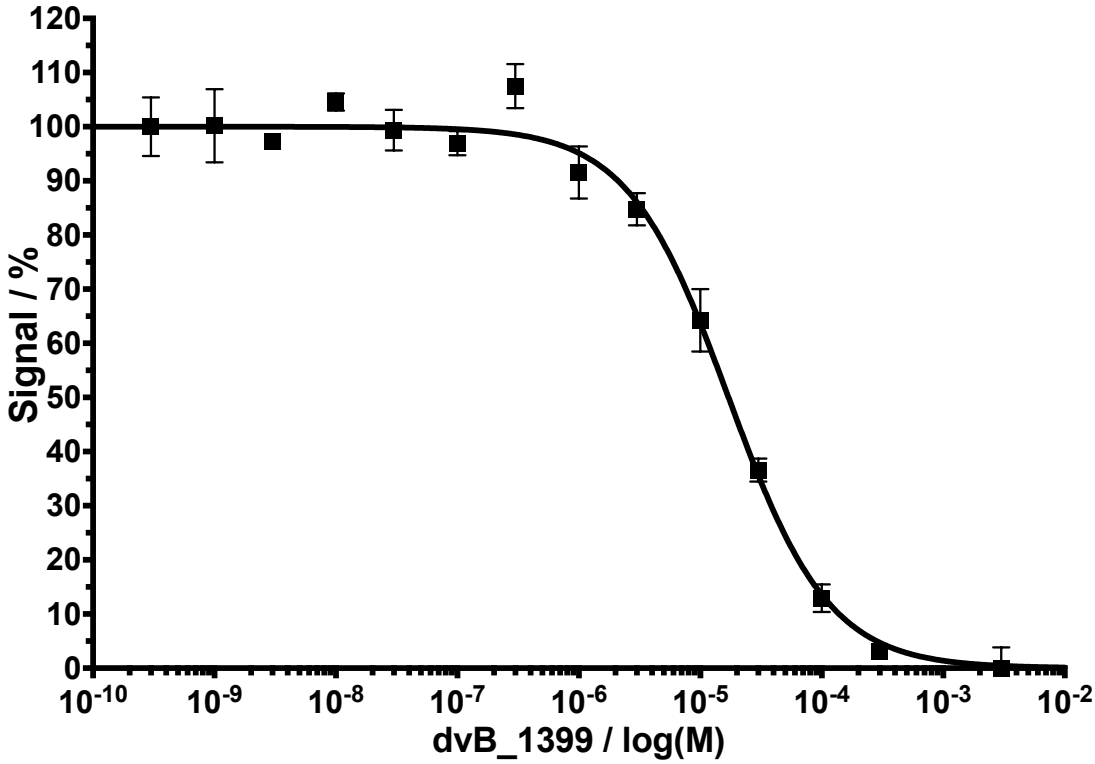

Figure 4.6: A cyclic bisubstrate conjugate (dvB_1399) binds to hOGT with low µM affinity. *K_i determination of dvB_1399.* Errors represent the *s.e.m of three replicates.*

The results of the binding assay show that dvB_1399 is capable of binding to hOGT with low µM affinity (Fig. 4.6). To characterise the binding mode of dvB_1399 to hOGT and to dissect any differences in the binding mode of ThioGoblin 1 the structure of hOGT in complex with dvB_1399 was solved. High-resolution synchrotron diffraction data were collected to 2.6 Å (R_{work}/R_{free} = 0.18/0.24, Table 4.3) and subsequent structure solution by molecular replacement, refinement and model building revealed continuous electron density that allowed the unambiguous placement of the ligand (Fig. 4.7).

Table 4.3: X-ray diffraction data collection and structure refinement statistics for hOGT:dvB_1399 and hOGT:dvB_1523 complexes. Values for the highest resolution shell are shown in parenthesis.

	hOGT + dvB_1399	hOGT + dvB_1523
Data collection		
Beamline, wavelength	ID30-A3, 0.967 Å	ID23-2, 0.873 Å
Space group	*F222*	*F*222
Cell dimensions (Å)	*a*=138.29, *b*=150.94, *c*=199.73, $\alpha=\beta=\gamma=90°$	*a*=138.72, *b*=152.1, *c*=200.75, $\alpha=\beta=\gamma=90°$
Resolution (Å)	46.01-2.60 (2.72-2.60)	46.33-2.10 (2.15-2.10)
R_{merge}	0.147 (0.961)	0.131 (1.815)
$I/\sigma I$	10.4 (2.1)	11.1 (123)
$CC_{1/2}$	0.99 (0.76)	0.998 (0.49)
R_{meas}	0.172 (1.128)	0.155 (2.176)
R_{pim}	0.066 (0.429)	0.059 (0.821)
Completeness (%)	100 (100)	100 (100)
Redundancy	6.7 (6.8)	6.8 (6.9)
Refinement		
Resolution (Å)	46.01 – 2.60	46.33 - 2.10Å
No. total reflections	215889	419482
No. unique reflections	32203	61612
R_{work}, R_{free}	0.182/0.239	0.206 / 0.243
No. atoms		
Protein	5498	5471
Ligand	96	83
Water	254	30
B-factor average		
Protein	49.87	44.46
Ligand	58.17	39.28
R.m.s. deviations		
Bond lengths (Å)	0.015	0.012
Bond angles (°)	1.9	1.55

Structural characterisation of the hOGT:dvB_1399 complex revealed that the cyclic bisubstrate conjugate binds in the active site of hOGT mimicking a pseudo-Michaelis complex reported earlier [231] and binds similarly to the linear ThioGoblin 1, described in chapter 3. The electron density only allowed placement of the bisubstrate inhibitor part corresponding to the ThioGoblin 1 scaffold, the first two N-terminal arginine residues and the peptide backbone of the glutamate, suggesting that the remaining residues are disordered (Fig. 4.7). The overall deviation of the UDP moiety and the bisubstrate linker in between dvB_1399 and ThioGoblin 1 is negligible with a maximum atomic shift of 0.4 Å between the sulphur atoms of the linker cysteine residue. Notably, the C_{α}-carbon of the glutamate residue of dvB_1399 and Gln^{399} of hOGT are only ~ 9.3 Å apart, suggesting the presence of steric clashes between the disordered region of dvB_1399 and the hOGT TPR domain (Fig. 4.7 B). Differences between the spatial occupation of terminal residues of ThioGoblin 1 and the corresponding residues in dvB_1399 were observed, showing a maximum atomic shift of 2.8 Å between carbonyl carbon of the acetylated N-terminus of ThioGoblin 1 and the corresponding carbonyl carbon of the dvB_1399 amide bond (Fig. 4.7 C, D). The amide nitrogen of the dvB_1399 glutamate showed an atomic shift of 3.3 Å, when compared to the aminated C-terminus of ThioGoblin 1 (Fig. 4.7 C, D). Changes

in the positioning of the “termini residues” of dvB_1399 compared to those of ThioGoblin 1 are likely caused by cyclisation.

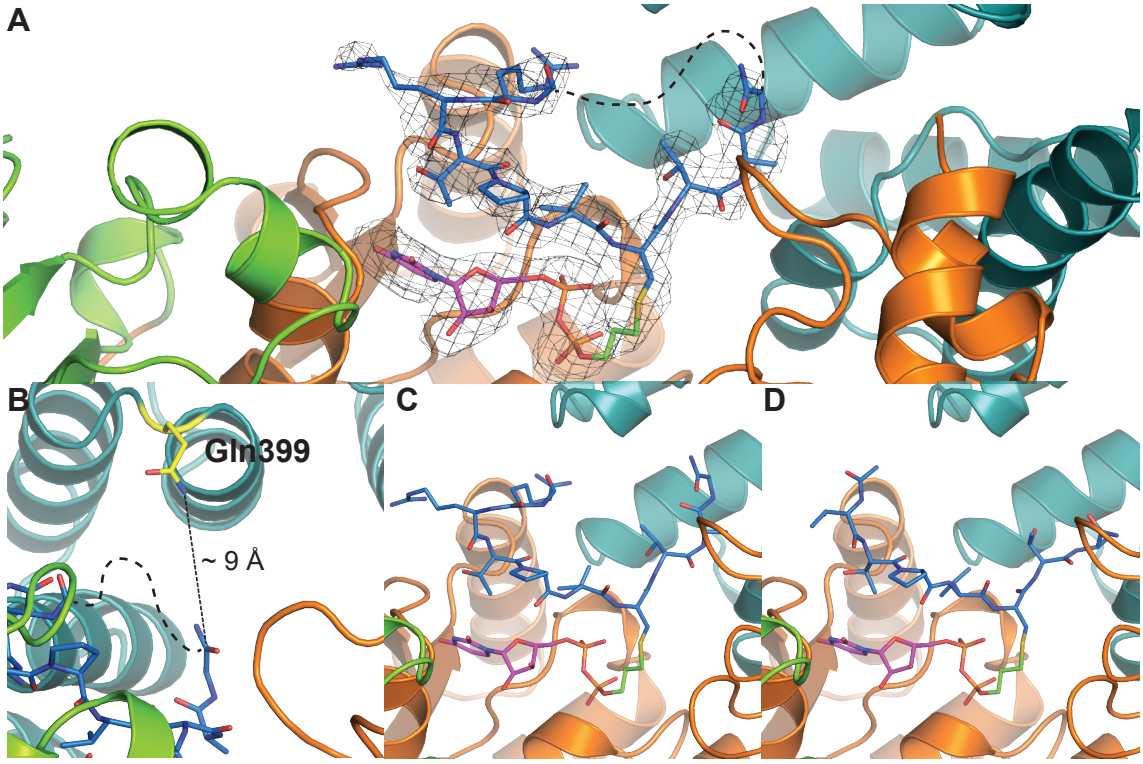

Figure 4.7: A cyclic bisubstrate inhibitor binds in the active site of OGT. (A) Structure of hOGT in complex with dvB_1399. **(B)** Close-up view of dvB_1399 and Gln^{399}, represented as yellow sticks, the straight dashed black line shows the ~ 9 Å between them. **(C) + (D)** Side-by-side view of hOGT in complex with dvB_1399 **(C)** and ThioGoblin 1 **(D)**. For **(A) - (D)** hOGT is shown in cartoon representation with the TPR domain shown in teal, the catalytic lobes in orange and the intervening domain in green. DvB_1399 is shown as sticks, with the UDP moiety shown in magenta, the linker green and the peptide in blue. The unbiased $|F_o\text{-}F_c|$ map is shown as black mesh, contoured to 2.25 σ. Missing residues in the disordered part of dvB_1399 are represented as a dashed black line.

In the initial synthetic strategy of the peptide chains used in this study the presence of a glutamate was required, as a point of attachment on the resin for solid phase peptide synthesis and as point of cyclisation. However, the synthesis strategy turned out to be labour intensive and yielded poor amounts of the cyclic scaffolds. In an attempt to optimise the synthetic strategy and reduce potential steric clashes of the glutamate linker with TPR domain of hOGT a γ-butyric acid (GABA) linker was introduced yielding the cyclic bisubstrate conjugate dvB_1523 (Fig. 4.2).

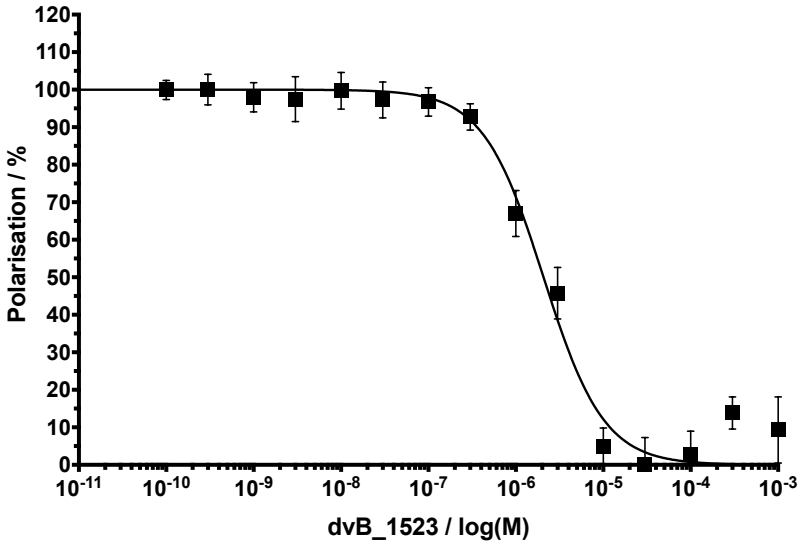

Figure 4.8: The cyclic bisubstrate conjugate dvB_1523 binds hOGT with sub µM affinity. K_i determination of dvB_1523. Error bars represent the s.e.m. of three replicates.

Determination of the K_i of dvB_1523 revealed potential sub µM binding affinity to hOGT, however could not be accurately determined as the binding affinity of the fluorescent probe (F-ThioGoblin 1, Chapter 3) used in the FP assay is lower than of dvB_1523 (Fig. 4.8). The resolution of a fluorescence polarimetry assay is determined by the binding affinity of the fluorescent probe and is not suitable to measure the binding affinity of tighter binding ligands [331].

Structural studies of hOGT in complex with dvB_1523 were performed in order to investigate changes in the binding mode of the ligand that promote a tighter binding to hOGT, compared to dvB_1399. High-resolution synchrotron diffraction data were collected to 2.1 Å (R_{work}/R_{free} = 0.206 / 0.243, Table 4.3) and subsequent structure solution by molecular replacement, refinement and model building revealed continuous electron density and allowed the unambiguous placement of the ligand (Fig. 4.9).

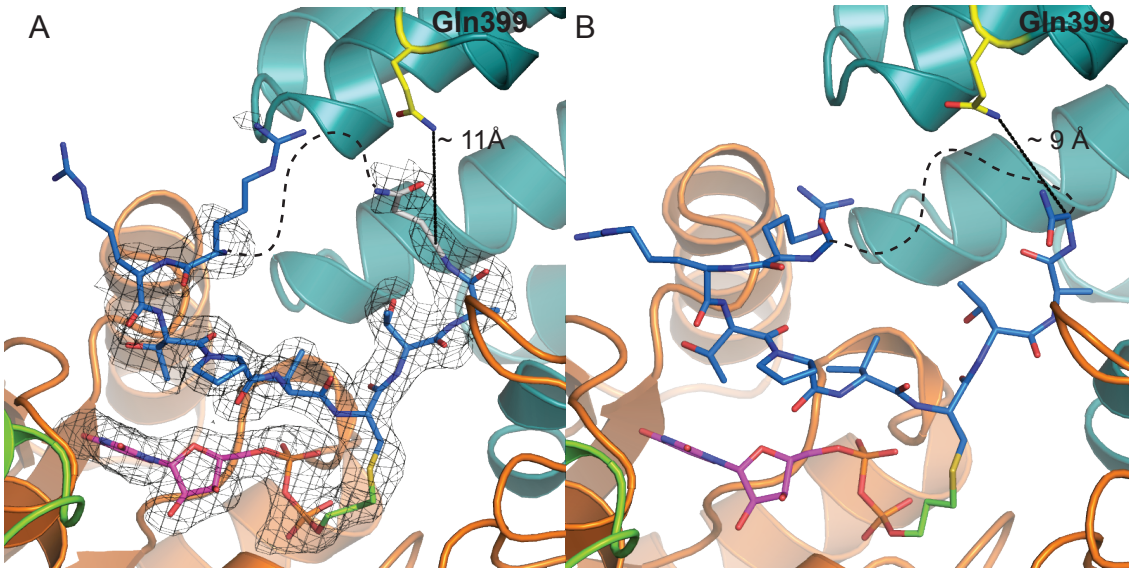

Figure 4.9: The cyclic bisubstrate conjugate dvB_1523 binds to hOGT similar to a pseudo-Michaelis complex. (A) Structure of hOGT in complex with dvB_1523. (B) Structure of hOGT in complex with dvB_1399. **(A) + (B)** Missing residues that could not be built are represented by a dashed black line. hOGT is shown in cartoon representation with the TPR domain shown in teal, the catalytic lobes in orange and the intervening domain in green. Gln^{399} is shown as yellow sticks. The cyclic ligands are shown as sticks, with the UDP moiety shown in magenta, the linker in green, the peptide in blue and the GABA linker **(A)** in grey. The distance between the C_α of the dvB_1399 glutamate **(B)** and the corresponding C in the GABA linker **(A)** are shown as dashed black line. **(A)** The unbiased $|F_o-F_c|$ map is shown as a black mesh contoured to 2.25 σ.

Inspection of the dvB_1523:hOGT structure revealed that dvB_1523 binds similarly to the binding mode of ThioGoblin 1 and dvB_1399 (Fig. 4.9).

Differences in the positioning of the dvB_1523 UDP and peptide moiety and the corresponding dvB_1399 moieties are negligible, with a maximum atomic shift 0.8 Å seen between amine of the threonine-alanine peptide bond. However, the linker moiety of dvB_1523, connecting the peptide and UDP, adopted an energetically unfavourable synclinal conformation, with a torsion angle of 37.2° (S-C-C-C). Furthermore, the C_{α} of the glutamate residue of dvB_1399 and the corresponding carbon in the GABA (C_{GABA}) linker of dvB_1523 occupy a different space, showing an atomic shift of 2.1 Å, resulting in a 11 Å distance between C_{GABA} and Gln^{399} of hOGT (Fig. 4.9). The shift in positioning of the GABA linker in respect to hOGT could alleviate potential steric clashes with the hOGT TPR domain and may explain the difference in potency to dvB_1399 (*K*i = 8 μM, Table 4.2). Furthermore, the unbiased $|F_o\text{-}F_c|$ map of the dvB_1523:hOGT complex allowed model building of the GABA linker, whereas lack of electron density did not allow model building of the glutamate side chain of dvB_1399. The lack of defined electron density, suggests that dvB_1523 forms a more stable complex with hOGT.

4.3.3 Cyclic CPP containing bisubstrate inhibitors readily cross the cellular membrane

To investigate translocation across the cellular membrane of the cyclic bisubstrate conjugates the cyclic peptide scaffold of dvB_1523 was labelled with a fluorescein at the cysteine residue usually harbouring the UDP moiety, yielding the cyclic peptide scaffold dvB_1549 (Table 4.3).

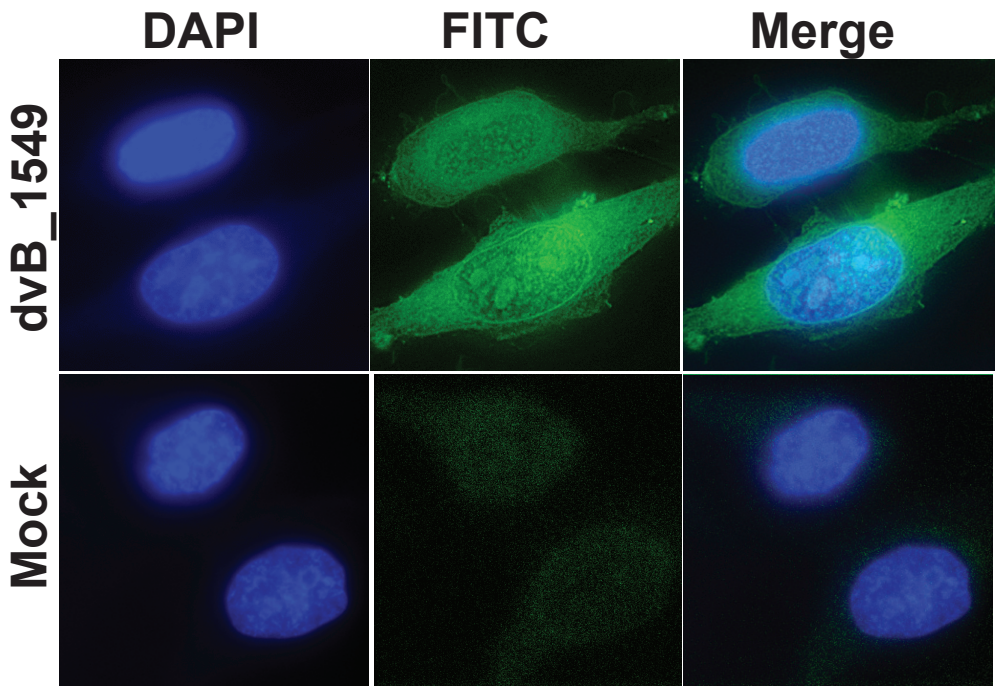

Figure 4.10: Deconvolution microscopy studies of dvB_1549 treated HeLa cells. Cells were treated for 2 h with 100 µM dvB_1549. FITC = fluorescein isothiocyanate, DAPI = 2-(4-amidinophenyl)-1H -indole-6-carboxamidine, Floc = 5(6)-carboxyfluorescein. Experiments were performed by Andrii Gorelik.

Deconvolution microscopy experiments revealed fluorescent signal in the nucleocytoplasm of treated HeLa cells, suggesting successful translocation of dvB_1549 across the cell membrane. Furthermore, cellular organelles could be distinguished over the background and indicate translocation of dvB_1549 across multiple membranes.

However, *in cellulo* experiments using dvB_1523 were inconclusive due to low solubility of the compound in aqueous buffers. Solubility issues were not encountered in *in vitro* binding experiments due to the amount of DMSO present (5%) and were not seen when using dvB_1549 in fluorescence microscopy studies. Attempts to facilitate solubilisation of dvB_1523 using co-solvents (PEG400, glycerol) and/or preheating did not yield satisfactory results (data not shown, experiments performed by Andrii Gorelik). To address solubility issues the aromatic residues of the CPP sequence were replaced by hydrophilic arginine residues, as poly-arginine sequences were reported to facility cellular uptake of biological molecules [389–392]. Substitution of the aromatic naphtilalanine and phenylalanine by arginine residues and introduction of a lysine into the cyclic scaffold yielded the bisubstrate scaffold dvB_1566 and was readily labelled by fluorescein at the primary amine of the lysine residue (Fig. 4.2, Table 4.3).

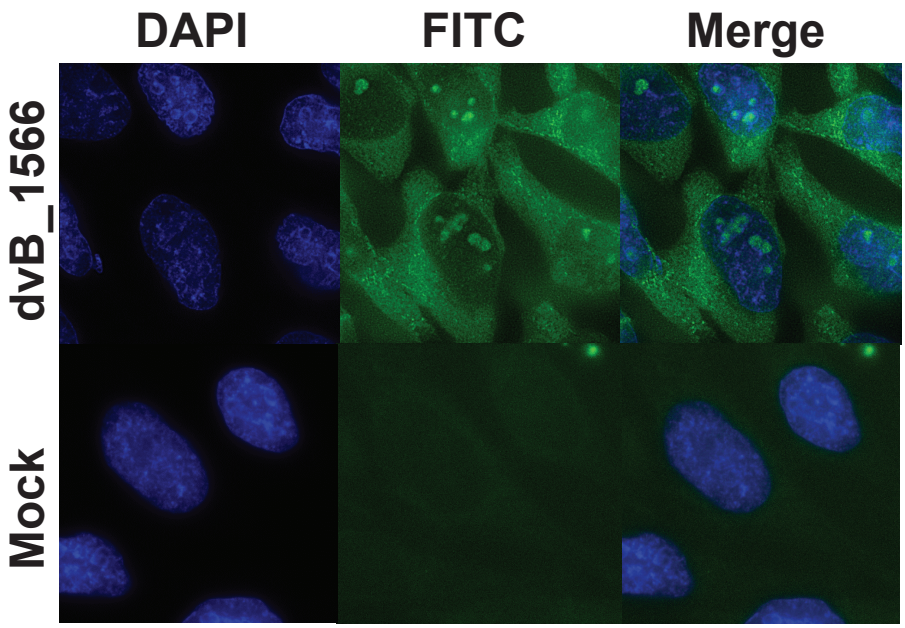

Figure 4.11: Deconvolution microscopy experiment with dvB_1399 treated HeLa cells. Cells were treated for 2 h with 1 µM of dvB_1566. FITC = Fluorescein isothiocyanate, DAPI = 2-(4-amidinophenyl)-1H-indole-6-carboxamidine, Floc = 5(6)-carboxyfluorescein. Work done by Andrii Gorelik.

Fluorescence microscopy images of dvB_1566 treated HeLa cells revealed fluorescein signal in the nucleocytoplasm, suggesting translocation across the membrane. Similar to dvB_1549 treated cells, cellular organelles could be distinguished and indicate translocation of dvB_1566 across multiple membranes (Fig. 4.11).

4.3.4 A cyclic pentaArg bisubstrate scaffold inhibits OGT in cellulo

To approximate the *in vitro* potency of dvB_1566, the unlabelled precursor dvB_1565 (Table 4.2) was subjected to *in vitro* characterisation of its potency to inhibit hOGT (Fig. 4.13).

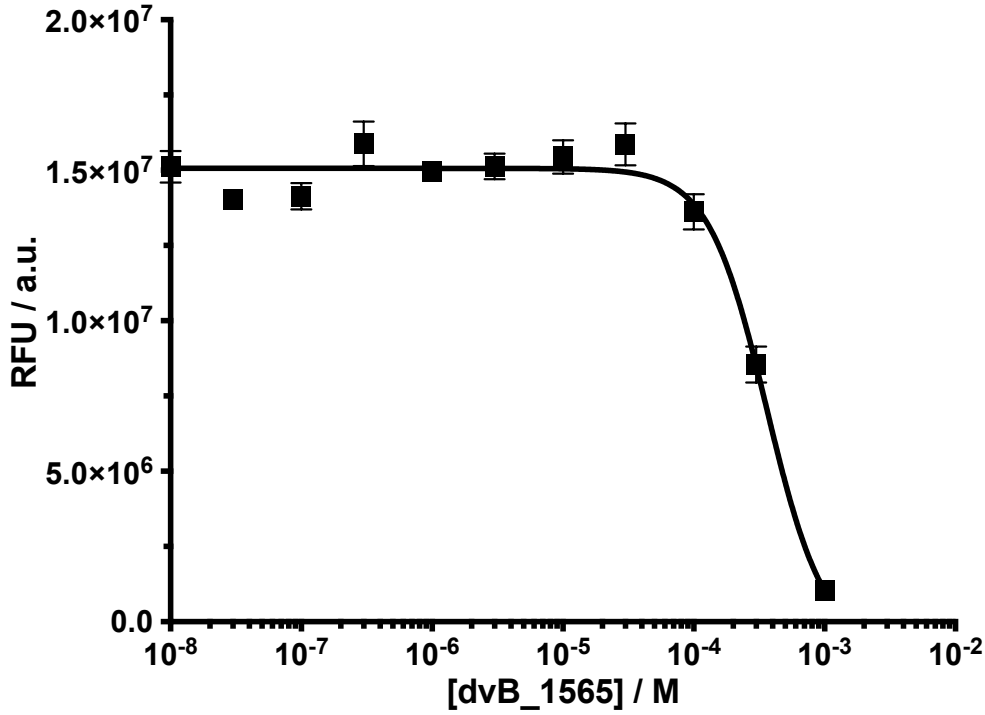

Figure 4.12: hOGT is inhibited by dvB_1565 in vitro. The dose-dependent activity of hOGT was measured in the presence of increasing concentrations of dvB_1565. Error represents the s.e.m. of three replicates.

Experimental determination of hOGT activity in the presence of increasing concentrations of inhibitor revealed weak inhibition by dvB_1565 (Fig. 4.13), supporting the results of the Western blot analysis (Fig. 4.12). The reduction in inhibition potency of dvB_1566 compared to dvB_1523 may be explained by substitution of the phenylalanine residue with an arginine residue. Structural analysis of hOGT in complex with the cyclic bisubstrate inhibitors dvB_1399 and dvB_1523 suggest the presence steric clashes between the aromatic residue of the inhibitors and the TPR domain of hOGT (Fig. 4.7 & 4.9). Replacement of phenylalanine by an arginine increases the effective space the bisubstrate

conjugate occupies, as the side chain length of phenylalanine (C_{α}-C_{ϑ} = 5.1 Å) is smaller than the arginine side chain (C_{α}-$N_{\eta 1/2}$ = 6.8 Å). This 1.7 Å difference in size could negate the reduction in steric clashes gained by introduction of a GABA linker into the cyclic bisubstrate structure.

To determine the efficiency of hOGT activity inhibition by dvB_1566 *in cellulo* cell culture experiments were performed. Human embryonic kidney 293 cells were treated with 100 µM dvB_1566 or Ac_4-5S-GlcNAc for 24 h before cell lysis and global O-GlcNAc levels visualised by Western blot analysis (Fig. 4.12).

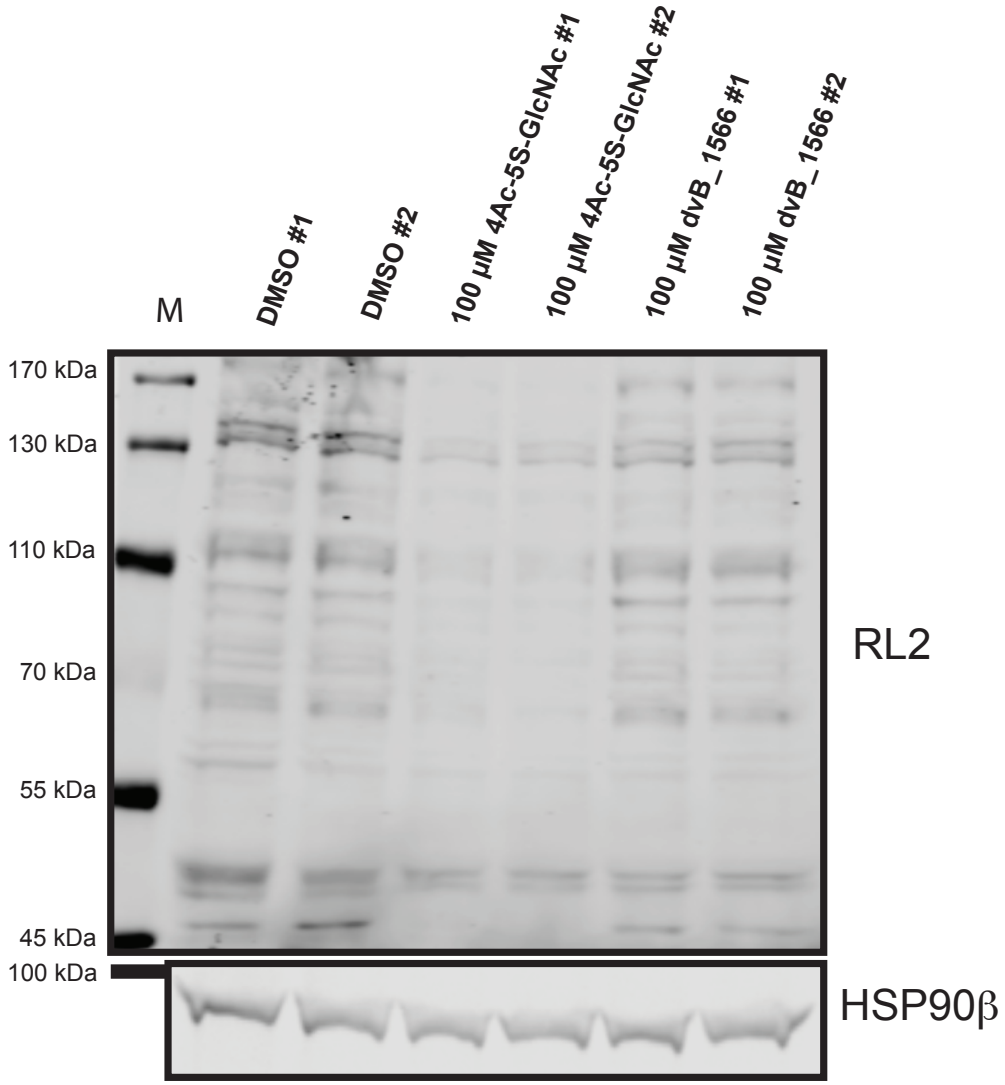

Figure 4.13: The cyclic bisubstrate scaffold dvB_1566 inhibits hOGT in cellulo. Western blot analysis showing global O-GlcNAc levels (RL2) of treated and untreated HEK293 cells. Cells were treated for 24 h with 100 µM Ac_4-5S-GlcNAc (positive control) and dvB_1566 before lysis and resolved on a 6% SDS-PAGE before transfer onto a nitrocellulose membrane. HSP90β was used as a loading control.

Western blot analysis of dvB_1566 treated HEK293 cells showed some reduction in global O-GlcNAc levels compared to DMSO treated controls, suggesting inhibition of cellular hOGT (Fig. 4.12). However, reduction of global O-GlcNAc levels by dvB_1566 were not as pronounced as in cells treated with the UDP-5S-

GlcNAc precursor Ac_4-5S-GlcNAc, a known *in cellulo* inhibitor of hOGT, and is in agreement with the results of the *in vitro* experiments.

4.4 Concluding remarks

O-GlcNAc is an abundant and essential modification of nucleocytoplasmic proteins in metazoa [338]. However, the precise role of this post-translational modification is still elusive and the lack of potent selective inhibitors of OGT has greatly slowed the advances made to elucidate its biological function. Recently a novel class of bisubstrate inhibitors was reported capable of inhibiting hOGT *in vitro* but are unsuitable for *in cellulo* experiments as they are unlikely to cross cellular membranes [333]. The recent discovery of cell penetrating peptide sequences, such as antennapedia [359], Tat [366] and polyArgs [392] have allowed the translocation of large biomolecules across cellular membranes [367].

The aim of the work presented in this chapter was to capitalise on the advances made in the understanding of bisubstrate hOGT inhibitors (discussed in chapter 3) and achieve cell penetrance by attachment of cell penetrating peptide (CPP) sequences. ThioGoblin 1 (Fig. 4.2) a low μM inhibitor of hOGT (IC_{50} = 2 μM, chapter 3, Table 3.2) was decorated with known CPPs, such as antennapedia and Tat, and its cellular uptake and *in cellulo* activity investigated. Western blot analysis of HEK293 cells treated with the antennapedia/Tat-decorated bisubstrate inhibitors did not show reduction in global O-GlcNAc levels and were subsequently shown to fail endosomal escape with consequent loss of bioavailability, a phenomenon reported previously for linear cell penetrating peptides [377,385]. The enhanced early endosomal escape of biomolecular cargo delivered by cyclic CPP was recently reported, showing a 3.7-12-fold more efficient delivery [386,388]. The cyclisation of bisubstrate scaffolds in conjunction with a cell penetrating RRRRΦF [386,388] sequence culminated in the synthesis of the three cyclic hOGT bisubstrate inhibitors dvB_1399, dvB_1523 and dvB_1565/1566, capable of translocating across the cellular membrane.

Structural characterisation of hOGT in complex with dvB_1399 or dvB_1523 revealed binding of the ligands in the active site, mimicking pseudo-Michaelis complexes reported before [231,240,241]. Most notably, cyclisation affected the positioning of the TPR proximal ligand part, suggesting the presence of steric clashes between hOGT and the aromatic residues of the cyclic bisubstrate scaffolds, which were alleviated by the introduction of a GABA linker in the ring scaffold instead of the less flexible glutamate, allowing the repositioning of the ligand and increasing the distance to the TPR by ~ 2 Å. Determination of the binding affinity of dvB_1523 to hOGT suggested a sub μM affinity, likely making it more potent than UDP (IC_{50} = 1.8 μM) [285]. Further biophysical and thermodynamic characterisation of the compound was impeded due to low availability of the compound and solubility issues encountered in cell culture studies (Fig. 4.2, chapter 3). Cyclisation of the bisubstrate scaffold could reduce the entropic penalty of ligand binding to hOGT, an effect resulting from the adoption of a defined structure, restricting the motion of the ligand, which results in a reduction in free energy of the system [393,394]. Finally, preliminary experiments showed a weak but detectable reduction of global O-GlcNAc levels in HEK293 cells treated with the pentaArg cyclic bisubstrate inhibitors dvB_1566, comparable to the positive control Ac_4-5S-GlcNAc [298].

The present chapter describes the research that has culminated in a new class of cyclic bisubstrate hOGT inhibitors, combining the selectivity of a hOGT substrate peptide sequence and the inhibition potency of UDP with cell penetrating peptide sequences capable of inhibiting hOGT *in vitro* and potentially *in cellulo*. Further, the work presented here addresses the common issues of aspecificity and low potency of commonly used inhibitors such as Ac_4-5S-GlcNAc, BZX and OSMI-1 [296,298,303]. Future work will be directed at a more in-

depth biophysical and thermodynamic characterisation of the cyclic bisubstrate inhibitors and optimisation of their design to address solubility and potency issues encountered in the present research. The design of a soluble, selective and highly potent cyclic bisubstrate hOGT inhibitor (CBSI) would allow elucidation of the biological role of OGT in cellular processes as well as dissection of the OGT dependent formation and propagation of disease pathologies. Initial *in cellulo* experiments would investigate the effect of cyclic bisubstrate inhibitor treatment on global O-GlcNAc levels, changes in surface glycosylation and cellular toxicity across various cell lines. A stereoisomer of the CBSI would be used as a negative control to account for any OGT-independent phenotypes. Future cell culture experiments would investigate the OGT-dependent role of O-GlcNAc signalling in wound healing [395,396], *Drosophila melanogaster* development [132] and prevention of Tau-hyperphosphorylation [96,156,397].

Concluding remarks and future perspective

The O-GlcNAc modification of nucleocytoplasmic proteins is an abundant and essential [47,48,50,74] modification of nucleocytoplasmic proteins and has been reported in all eukaryotic organisms apart from yeast, yet the precise cellular function and biological role of this post-translational modification is elusive [67,73,338,398]. The addition of GlcNAc residues of intracellular proteins is regulated by two enzymes, the O-GlcNAc transferase (OGT) that transfers the sugar onto proteins and the O-GlcNAc hydrolase (OGA), which removes it. However, the precise mechanism by which OGA and OGT recognise and bind their plethora of substrates is still elusive. Early studies on the domain structure of OGT revealed it to be a multi-domain protein consisting of an N-terminal tetratricopeptide repeat (TPR) domain, proposed to be important for substrate recognition and binding, and a C-terminal catalytic domain, responsible for the transfer of the GlcNAc residue onto acceptor serine and threonine residues from the donor substrate UDP-GlcNAc [72,227,228,232]. Early structural characterisation of the TPR domain revealed a superhelical fold, reminiscent of the armadillo-repeats of importin α [235] and β-catenin [233], and describe the presence of regularly spaced asparagine residues thought to be involved in substrate recognition [230]. A recent structural study investigating the complex formation between the human O-GlcNAc transferase (hOGT) and a short peptide derived from the proteolytic substrate HCF-1, revealed the involvement of multiple asparagine residues on the concave surface of the TPR domain in the recognition of the proteolytic substrate. In chapter 2 of the present work, a linear fusion approach to trap OGT in complex with its glycosylation substrate TAB1, a protein involved innate immune signalling, is described to investigate the role of TPR mediated recognition of

glycosylation substrates. Analysis of the reported linear fusion HCF-1:hOGT and TAB1:hOGT structures revealed the recognition of the substrate peptide in the superhelical fold of the TPR domain, by formation of asparagine mediated, sequence independent interactions with the peptide backbone, abolished by mutation of the asparagine residues. Observation of an extended conformation of the substrate peptide in the active site recapitulated the results published in a recent study [231] suggesting that OGT might bind its substrates through a combination of mechanisms. Results of work presented in chapter 3 of this thesis however, suggest that the recognition of substrates inside the superhelical fold of the TPR domain may only be true for a subset of substrates. As the addition of residues to the C-terminus of peptide part of the bisubstrate scaffold was shown to result in a decrease in binding affinity, suggesting the presence of steric clashes between the TPR domain of OGT and a glycosylation substrates. Taken together the results described in chapter 2, combined with recently published results, may suggest that the linear fusion of OGT to its substrates is a viable approach to study substrate recognition and binding. Moreover, a linear fusion approach may be adopted for whole protein glycosylation substrates to investigate a potential additional recognition of substrates via the catalytic domain of OGT.

Recent studies link dysregulation of O-GlcNAc to the development and propagation of metabolic and neurodegenerative diseases [92,399,400], yet the precise role of O-GlcNAc in these disease phenotypes is poorly understood. The development of potent and selective inhibitors of OGT is a prerequisite to elucidate the biological role of O-GlcNAc and its involvement in diseases. However, commonly used OGT inhibitors suffer from aspecificity and low potency *in cellulo*, raising concerns about the interpretation of data acquired involving

inhibitor treatments [243,356]. BZX is a covalent inhibitor of hOGT, which crosslinks the active site residues Lys^{842} and Cys^{917} rendering the enzyme inactive, however this particular pairing of residues is a common motif for various kinases making them a likely target for BZX [296,308]. The inhibitor Ac_4-5S-GlcNAc inhibitor, which is taken up via the salvage pathway of the hexosamine biosynthetic pathway and metabolised into UDP-5S-GlcNAc, a donor substrate analogue of UDP-GlcNAc, was shown to inhibit OGT in the low μM range (EC_{50} = 5 μM [298]). However, UDP-5S-GlcNAc is likely to target any UDP-GlcNAc utilising enzyme and may have adverse effects on N-glycosylation and glycan biosynthesis.

Chapters 2 and 3 of this thesis describe the advances that were made to optimise a recently reported class of selective, μM bisubstrate hOGT inhibitors [333]. The effect modifications have on either of the three components of the bisubstrate inhibitors the peptide, the linker and the UDP moiety was investigated by the biophysical and structural characterisation of a diverse panel of bisubstrate scaffolds described in chapter 2. Substitution of the linker serine by cysteine culminated in the identification of ThioGoblin 1, a 10-fold more potent derivative (IC_{50} = 2 μM) than its progenitor Goblin 1 (IC_{50} = 18 μM [333]), exhibiting similar potencies to the strongest OGT binder known UDP (IC_{50} = 1.8 μM) [285]. The results further supported previous studies describing a preferred sequence motif of OGT, which describes the beneficial effects of having a proline present in the -2 position, in respect to the acceptor residue [231]. However, the results also indicate that a C-terminal extension of the peptide sequence results in a decreased binding affinity. Changes of the UDP moiety showed drastic reductions in potency of the bisubstrate scaffolds, supporting earlier structural studies that describe the significant interactions formed by the nucleotide and OGT and the importance of the pyrophosphate bond for binding [240,241,285]. Structural analysis of an acetamido

derivative of the bisubstrate inhibitor revealed that this group binds in a similar position to the GlcNAc residue of UDP-5S-GlcNAc found in a recently reported structure [231], forming the same hydrogen bond with His[920]. Moreover, the structure revealed a close proximity of the acetamido group of the bisubstrate conjugate dvB_1505 to Cys[917], a residue targeted by the covalent OGT inhibitor BZX. Decorating the acetamido of dvB_1505 with an electrophilic group may allow the covalent modification of the nucleophilic cysteine, forming a covalent link between the bisubstrate scaffold and OGT. Furthermore, a UDP-GlcNAc analogue harbouring an electrophilic group may be used to elegantly trap an OGT-substrate complex, which would be beneficial for crystallisation and subsequent structural characterisation and supersede the need of a covalent attachment of the OGT and substrate as described for the linear fusion approach. In the final results chapter of the present thesis, the low cell-permeability of the bisubstrate scaffolds was addressed by incorporation of cell-penetrating peptide (CPP) sequences. Bisubstrate conjugates decorated with linear CPPs were not suitable as tools to study OGT inhibition *in cellulo*, due to their limited endosomal escape. However, in preliminary experiment, cyclisation of CPP containing bisubstrate inhibitors culminated in a small set of highly cell permeable scaffolds capable of translocating into the nucleocytoplasm across the cell membrane and inhibiting OGT to varying degrees. Preliminary binding affinity experiments suggest sub μM potency for the cyclic OGT inhibitor dvB_1523, comparable to affinities reported for other cyclic peptide inhibitors [386–390,401], however further characterisation is needed and in future experiments. The results presented in chapter 4 of this thesis describe the initial characterisation of a novel class of potent and selective cyclic hOGT bisubstrate inhibitors that can be developed into suitable tools to study the biological role of OGT and O-GlcNAc as well as their

role in disease progression. Furthermore, a recent developmental study of *Drosophila melanogaster* suggests the possibility of a catalytic and non-catalytic, scaffolding function of OGT [132], which may be dissected using these inhibitors.

References

1. Cohen, P. The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem. Sci.* **25,** 596–601 (2000).
2. Komander, D. & Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **81,** 203–229 (2012).
3. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13,** 448–462 (2012).
4. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. 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 (2012).
5. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21,** 255–261 (2003).
6. Wang, Y.-C., Peterson, S. E. & Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **24,** 143–160 (2014).
7. Nadolski, M. J. & Linder, M. E. Protein lipidation. *FEBS J.* **274,** 5202–5210 (2007).
8. Keembiyehetty, C. N., Krzeslak, A., Love, D. C. & Hanover, J. A. A lipid-droplet-targeted O-GlcNAcase isoform is a key regulator of the proteasome. *J. Cell Sci.* **124,** 2851–60 (2011).
9. Ahearn, I. M., Haigis, K., Bar-Sagi, D. & Philips, M. R. Regulating the regulator: post-translational modification of RAS. *Nat. Rev. Mol. Cell Biol.* **13,** 39–51 (2011).

10. Zhu, Y. *et al.* O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. *Nat. Chem. Biol.* **11,** 319–325 (2015).

11. Scott, A., Song, J., Ewing, R. & Wang, Z. Regulation of protein stability of DNA methyltransferase 1 by post-translational modifications Regulation of DNMT1 Protein Stability by Acetylation and Ubiquitination Regulation of DNMT1 Protein Stability by Phosphorylation and Methylation. *Acta Biochim Biophys Sin* **46,** 199–203 (2014).

12. Bienvenut, W. V. *et al.* Dynamics of post-translational modifications and protein stability in the stroma of Chlamydomonas reinhardtii chloroplasts. *Proteomics* **11,** 1734–1750 (2011).

13. Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.* **21,** 140–146 (2009).

14. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Comput. Biol.* **11,** 1–23 (2015).

15. Spiro, R. G. Glycoproteins. *Adv. Protein Chem.* **27,** 349–467 (1973).

16. Lechner, J. & Wieland, F. Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.* **58,** 173–94 (1989).

17. Messner, P. Bacterial glycoproteins. *Glycoconj. J.* **14,** 3–11 (1997).

18. Lodish, H. F. *Molecular cell biology*. (W.H. Freeman, 2000).

19. JOHANSEN, P. G., MARSHALL, R. D. & NEUBERGER, A. Carbohydrates in protein. 3 The preparation and some of the properties of a glycopeptide from hen's-egg albumin. *Biochem. J.* **78,** 518–27 (1961).

20. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3,** 97–130 (1993).

21. Haltiwanger, R. & Lowe, J. Role Of Glycosylation In Development. *Annu.*

Rev. Biochem. **73,** 491–537 (2004).

22. Xu, C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nat. Rev. Mol. Cell Biol.* **16,** 742–752 (2015).
23. Rothman, J. E. & Fine, R. E. Coated vesicles transport newly synthesized membrane glycoproteins from endoplasmic reticulum to plasma membrane in two successive stages. *Proc. Natl. Acad. Sci. U. S. A.* **77,** 780–4 (1980).
24. Roth, J. Protein N -Glycosylation along the Secretory Pathway: Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions. *Chem. Rev.* **102,** 285–304 (2002).
25. Cummings, R. D. *et al.* The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5,** 1087 (2009).
26. Cummings, R. D. & Pierce, J. M. *Handbook of glycomics*. (Academic Press/Elsevier, 2009).
27. Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: Transcript profiling of glycan-related genes. *J. Biol. Chem.* **283,** 17298–17313 (2008).
28. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37,** D233-8 (2009).
29. Ikezawa, H. Glycosylphosphatidylinositol (GPI)-anchored proteins. *Biol. Pharm. Bull.* **25,** 409–417 (2002).
30. Springer, S. A. & Gagneux, P. Glycomics: Revealing the dynamic ecology and evolution of sugar molecules. *J. Proteomics* **135,** 90–100 (2016).
31. Torres, C. & Hart, G. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* **259,** 3308–17 (1984).

32. Holts, G. D. & Hart, G. W. The Subcellular Distribution of Terminal N-Acetylglucosamine Moieties. *J. Biol. Chem.* **261,** 8049–4305 (1986).

33. Wells, L. Mapping Sites of O-GlcNAc Modification Using Affinity Tags for Serine and Threonine Post-translational Modifications. *Mol. Cell. Proteomics* **1,** 791–804 (2002).

34. Hahne, H. *et al.* Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J. Proteome Res.* **12,** 927–936 (2013).

35. Maury, J. J. P., Ng, D., Bi, X., Bardor, M. & Choo, A. B.-H. Multiple reaction monitoring mass spectrometry for the discovery and quantification of O-GlcNAc-modified proteins. *Anal. Chem.* **86,** 395–402 (2014).

36. Vosseller, K. *et al.* O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* **5,** 923–34 (2006).

37. Cole, R. N. & Hart, G. W. Cytosolic O-glycosylation is abundant in nerve terminals. *J. Neurochem.* **79,** 1080–1089 (2001).

38. Khidekel, N. *et al.* Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **3,** 339–48 (2007).

39. Chalkley, R. J., Thalhammer, A., Schoepfer, R. & Burlingame, A. L. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8894–9 (2009).

40. Qin, W. *et al.* Quantitative time-resolved chemoproteomics reveals that stable *O* -GlcNAc regulates box C/D snoRNP biogenesis. *Proc. Natl. Acad. Sci.* 201702688 (2017). doi:10.1073/pnas.1702688114

41. Zhao, P. *et al.* The human embryonic stem cell proteome revealed by multidimensional fractionation followed by tandem mass spectrometry. *Proteomics* **15,** 554–566 (2015).

42. Ruan, H.-B., Nie, Y. & Yang, X. Regulation of protein degradation by O-GlcNAcylation: crosstalk with ubiquitination. *Mol. Cell. Proteomics* **12,** 1–24 (2013).

43. Trinidad, J. C. *et al.* Global Identification and Characterization of Both *O* -GlcNAcylation and Phosphorylation at the Murine Synapse. *Mol. Cell. Proteomics* **11,** 215–229 (2012).

44. Selvan, N. *et al.* The early metazoan Trichoplax adhaerens possesses a functional O-GlcNAc system. *J. Biol. Chem.* **290,** jbc.M114.628750 (2015).

45. Bond, M. R., Ghosh, S. K., Wang, P. & Hanover, J. A. Conserved Nutrient Sensor O-GlcNAc Transferase Is Integral to C. elegans Pathogen-Specific Immunity. *PLoS One* **9,** e113231 (2014).

46. Ghosh, S. K. *et al.* Disruption of O-GlcNAc cycling in C. elegans perturbs nucleotide sugar pools and complex glycans. *Front. Endocrinol. (Lausanne).* **5,** 1–7 (2014).

47. Gambetta, M. C., Oktaba, K. & Müller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* **325,** 93–6 (2009).

48. Ingham, P. W. A gene that regulates the bithorax complex differentially in larval and adult cells of Drosophila. *Cell* **37,** 815–823 (1984).

49. Yang, Y. R. *et al.* O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. *Aging Cell* **11,** 439–448 (2012).

50. Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse

ontogeny. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5735–9 (2000).

51. Yu, Y. *et al.* Differential effects of an o-glcnacase inhibitor on tau phosphorylation. *PLoS One* **7,** 1–8 (2012).

52. Slawson, C. & Hart, G. O-GlcNAc signalling: implications for cancer cell biology. *Nat. Rev. Cancer* **11,** 678–84 (2011).

53. Schirm, M. *et al.* Flagellin from Listeria monocytogenes is glycosylated with beta-O-linked N-acetylglucosamine. *J. Bacteriol.* **186,** 6721–6727 (2004).

54. Shen, A., Kamp, H. D., Gründling, A. & Higgins, D. E. A bifunctional O-GlcNAc transferase governs flagellar motility through anti-repression. *Genes Dev.* **20,** 3283–3295 (2006).

55. Hanover, J. a, Krause, M. W. & Love, D. C. Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nat. Rev. Mol. Cell Biol.* **13,** 312–21 (2012).

56. Sinclair, D. a R. *et al.* Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13427–32 (2009).

57. Daou, S. *et al.* 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).

58. Geng, F., Zhu, W., Anderson, R. a, Leber, B. & Andrews, D. W. Multiple post-translational modifications regulate E-cadherin transport during apoptosis. *J. Cell Sci.* **125,** 2615–25 (2012).

59. Lefebvre, T. *et al.* Evidence of a balance between phosphorylation and O-GlcNAc glycosylation of Tau proteins - A role in nuclear localization. *Biochim. Biophys. Acta - Gen. Subj.* **1619,** 167–176 (2003).

60. O'Donnell, N., Zachara, N. E., Hart, G. W. & Marth, J. D. Protein

Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability Ogt -Dependent X-Chromosome-Linked Protein Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability. *Mol Cell Biol* **24,** 1680–1690 (2004).

61. Fang, B. & Miller, M. W. Use of galactosyltransferase to assess the biological function of O-linked N-acetyl-d-glucosamine: a potential role for O-GlcNAc during cell division. *Exp. Cell Res.* **263,** 243–253 (2001).

62. Slawson, C., Shafii, S., Amburgey, J. & Potter, R. Characterization of the O-GlcNAc protein modification in Xenopus laevis oocyte during oogenesis and progesterone-stimulated maturation. *Biochim. Biophys. Acta - Gen. Subj.* **1573,** 121–129 (2002).

63. Slawson, C. *et al.* Perturbations in O-linked ??-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. *J. Biol. Chem.* **280,** 32944–32956 (2005).

64. Housley, M. P. *et al.* O-GlcNAc regulates FoxO activation in response to glucose. *J. Biol. Chem.* **283,** 16283–16292 (2008).

65. Kang, E. S. *et al.* O-GlcNAc modulation at Akt1 Ser473 correlates with apoptosis of murine pancreatic ?? cells. *Exp. Cell Res.* **314,** 2238–2248 (2008).

66. Walgren, J. L. E. *et al.* High glucose and insulin promote O- GlcNAc modification of proteins , including α -tubulin High glucose and insulin promote O- GlcNAc modification of proteins , including ␣ -tubulin. **29425,** 424–434 (2011).

67. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the ‘O-GlcNAc code’. *Sci STKE* **2005,** re13 (2005).

68. Walgren, J. L. E., Vincent, T. S., Schey, K. L. & Buse, M. G. High glucose

and insulin promote *O*- GlcNAc modification of proteins, including α-tubulin. *Am. J. Physiol. - Endocrinol. Metab.* **284,** E424–E434 (2003).

69. Ding, F. *et al.* O-GlcNAcylation involvement in high glucose-induced cardiac hypertrophy via ERK1/2 and cyclin D2. *Amino Acids* (2013). doi:10.1007/s00726-013-1504-2

70. Liu, J., Marchase, R. B. & Chatham, J. C. Glutamine-induced protection of isolated rat heart from ischemia/reperfusion injury is mediated via the hexosamine biosynthesis pathway and increased protein O-GlcNAc levels. *J. Mol. Cell. Cardiol.* **42,** 177–185 (2007).

71. Swamy, M. *et al.* Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy. *Nat. Immunol.* 1–11 (2016). doi:10.1038/ni.3439

72. Kreppel, L. K. & Hart, G. W. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *J. Biol. Chem.* **274,** 32015–22 (1999).

73. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* **80,** 825–858 (2011).

74. Watson, L. J. *et al.* Cardiomyocyte ogt is Essential for Postnatal Viability. *Am. J. Physiol. Heart Circ. Physiol.* ajpheart.00438.2013- (2013). doi:10.1152/ajpheart.00438.2013

75. Webster, D. M. *et al.* O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. *BMC Dev. Biol.* **9,** 28 (2009).

76. Comer, F. I. & Hart, G. W. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II.

Biochemistry **40,** 7845–7852 (2001).

77. Comer, F. I. & Hart, G. W. O-glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate. *J. Biol. Chem.* **275,** 29179–29182 (2000).

78. Yi, W. *et al.* Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science* **337,** 975–80 (2012).

79. Hardie, D. G., Ross, F. A. & Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat. Rev. Mol. Cell Biol.* **13,** 251–262 (2012).

80. Bullen, J. W. *et al.* Cross-talk between two essential nutrient-sensitive enzymes O-GlcNAc Transferase (OGT) and Amp-Activated Protein Kinase (AMPK). *J. Biol. Chem.* **289,** 10592–10606 (2014).

81. Hawley, S. A. *et al.* Calmodulin-dependent protein kinase kinase-?? is an alternative upstream kinase for AMP-activated protein kinase. *Cell Metab.* **2,** 9–19 (2005).

82. Wang, Z., Pandey, A. & Hart, G. W. Dynamic Interplay between *O* -Linked *N* -Acetylglucosaminylation and Glycogen Synthase Kinase-3-dependent Phosphorylation. *Mol. Cell. Proteomics* **6,** 1365–1379 (2007).

83. Wang, Z. *et al.* Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal.* **3,** ra2 (2010).

84. Chou, T. Y., Hart, G. W. & Dang, C. V. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *Journal of Biological Chemistry* **270,** 18961–18965 (1995).

85. Cheng, X. & Hart, G. W. Alternative O-glycosylation/O-phosphorylation of Serine-16 in murine estrogen receptor ??: Post-translational regulation of turnover and transactivation activity. *J. Biol. Chem.* **276,** 10570–10575

(2001).

86. Slawson, C., Lakshmanan, T., Knapp, S. & Hart, G. W. A mitotic GlcNAcylation/phosphorylation signaling complex alters the posttranslational state of the cytoskeletal protein vimentin. *Mol. Biol. Cell* **19,** 4130–40 (2008).

87. Bias, W. B., Cheung, W. D., Wang, Z. & Hart, G. W. Regulation of calcium/calmodulin-dependent kinase IV by O-GlcNAc modification. *J. Biol. Chem.* **284,** 21327–21337 (2009).

88. Yang, W. H. *et al.* Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat. Cell Biol.* **8,** 1074–83 (2006).

89. Housley, M. P. *et al.* A PGC-1α- *O* -GlcNAc Transferase Complex Regulates FoxO Transcription Factor Activity in Response to Glucose. *J. Biol. Chem.* **284,** 5148–5157 (2009).

90. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12,** 43R–56R (2002).

91. Hanover, J. A. *et al.* Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch. Biochem. Biophys.* **409,** 287–297 (2003).

92. Slawson, C., Copeland, R. J. & Hart, G. W. O-GlcNAc signaling: a metabolic link between diabetes and cancer? *Trends Biochem. Sci.* **35,** 547–55 (2010).

93. Banerjee, P. S., Ma, J. & Hart, G. W. Diabetes-associated dysregulation of O -GlcNAcylation in rat cardiac mitochondria. **2015,** (2015).

94. Vosseller, K., Wells, L., Lane, M. D. & Hart, G. W. 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,** 5313–8 (2002).

95. McLarty, J. L., Marsh, S. a & Chatham, J. C. Post-translational protein modification by O-linked N-acetyl-glucosamine: its role in mediating the adverse effects of diabetes on the heart. *Life Sci.* **92,** 621–7 (2013).

96. Kang, M. J. *et al.* Synapsin-1 and tau reciprocal O-GlcNAcylation and phosphorylation sites in mouse brain synaptosomes. *Exp. Mol. Med.* **45,** e29 (2013).

97. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C.-X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10804–9 (2004).

98. Förster, S. *et al.* Increased O-GlcNAc levels correlate with decreased O-GlcNAcase levels in Alzheimer disease brain. *Biochim. Biophys. Acta* (2014). doi:10.1016/j.bbadis.2014.05.014

99. Ma, Z., Vocadlo, D. J. D. & Vosseller, K. Hyper-O-GlcNAcylation is anti-apoptotic and maintains constitutive NF-κB activity in pancreatic cancer cells. *J. Biol. Chem.* **288,** 1–19 (2013).

100. Itkonen, H. M. *et al.* O-GlcNAc transferase integrates metabolic pathways to regulate the stability of c-MYC in human prostate cancer cells. *Cancer Res.* **73,** 5277–87 (2013).

101. Li, Z. & Yi, W. Regulation of cancer metabolism by O-GlcNAcylation. *Glycoconj. J.* (2013). doi:10.1007/s10719-013-9515-5

102. Ferrer, C. M. *et al.* O-GlcNAcylation regulates breast cancer metastasis via SIRT1 modulation of FOXM1 pathway. *Oncogene* 1–11 (2016). doi:10.1038/onc.2016.228

103. D'Onofrio, M. *et al.* Partial cDNA sequence encoding a nuclear pore protein

modified by O-linked N-acetylglucosamine. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 9595–9 (1988).

104. Park, M. K., D'Onofrio, M., Willingham, M. C. & Hanover, J. a. A monoclonal antibody against a family of nuclear pore proteins (nucleoporins): O-linked N-acetylglucosamine is part of the immunodeterminant. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 6462–6466 (1987).

105. Holt, G. D. *et al.* Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine. *J. Cell Biol.* **104,** 1157–1164 (1987).

106. Starr, C. M., D'Onofrio, M., Park, M. K. & Hanover, J. A. Primary sequence and heterologous expression of nuclear pore glycoprotein p62. *J. Cell Biol.* **110,** 1861–1871 (1990).

107. Starr, M. & Hanover, J. A. of Nuclear Pore Protein ~ 62. **265,** (1990).

108. Zhu, Y. *et al.* Post-translational O-GlcNAcylation is essential for nuclear pore integrity and maintenance of the pore selectivity filter. *J. Mol. Cell Biol.* **8,** 2–16 (2016).

109. Labokha, A. A. *et al.* Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. *EMBO J.* **32,** 204–218 (2012).

110. Lewis, B. A. & Hanover, J. A. O-GlcNAc and the epigenetic regulation of gene expression. *J. Biol. Chem.* **289,** 34440–34448 (2014).

111. Fujiki, R. *et al.* GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* **480,** 557–560 (2011).

112. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493,** 561–4 (2013).

113. Vella, P. *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* **49,** 645–56

(2013).

114. Lee, J. E., Park, J. H., Moon, P. G. & Baek, M. C. Identification of differentially expressed proteins by treatment with PUGNAc in 3T3-L1 adipocytes through analysis of ATP-binding proteome. *Proteomics* **13,** 2998–3012 (2013).

115. Nandi, A. *et al.* Global identification of O-GlcNAc-modified proteins. *Anal. Chem.* **78,** 452–458 (2006).

116. Ninomiya-Tsuji, J. *et al.* The kinase TAK1 can activate the NIK-I kappaB as well as the MAP kinase cascade in the IL-1 signalling pathway. *Nature* **398,** 252–256 (1999).

117. Pathak, S. *et al.* O-GlcNAcylation of TAB1 modulates TAK1-mediated cytokine release. *EMBO J.* **31,** 1394–404 (2012).

118. Yang, X. *et al.* Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* **451,** 964–9 (2008).

119. Chen, G., Liu, P., Thurmond, D. C. & Elmendorf, J. S. Glucosamine-induced insulin resistance is coupled to O-linked glycosylation of Munc18c. *FEBS Lett.* **534,** 54–60 (2003).

120. Jackson, S. P. & Tjian, R. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55,** 125–133 (1988).

121. Han, I. & Kudlow, J. E. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. *Mol. Cell. Biol.* **17,** 2550–2558 (1997).

122. Zachara, N. E. & Hart, G. W. O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim. Biophys. Acta* **1673,** 13–28 (2004).

123. Zachara, N. E. *et al.* Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J. Biol. Chem.* **279,** 30133–42 (2004).

124. Laczy, B., Marsh, S. A., Brocks, C. a, Wittmann, I. & Chatham, J. C. Inhibition of O-GlcNAcase in perfused rat hearts by NAG-thiazolines at the time of reperfusion is cardioprotective in an O-GlcNAc-dependent manner. *Am. J. Physiol. Heart Circ. Physiol.* **299,** H1715-27 (2010).

125. Ngoh, G. A., Facundo, H. T., Zafir, A. & Jones, S. P. O-GlcNAc signaling in the cardiovascular system. *Circ. Res.* **107,** 171–185 (2010).

126. Champattanachai, V., Marchase, R. B. & Chatham, J. C. Glucosamine protects neonatal cardiomyocytes from ischemia-reperfusion injury via increased protein O-GlcNAc and increased mitochondrial Bcl-2. *Am. J. Physiol. Cell Physiol.* **294,** C1509–C1520 (2008).

127. Fülöp, N., Zhang, Z., Marchase, R. B. & Chatham, J. C. Glucosamine cardioprotection in perfused rat hearts associated with increased O-linked N-acetylglucosamine protein modification and altered p38 activation. *Am. J. Physiol. Hear. Circ. Physiol.* **292,** H2227-36 (2007).

128. Guo, B. *et al.* O-GlcNAc-modification of SNAP-29 regulates autophagosome maturation. *Nat. Cell Biol.* **16,** 1215–1226 (2014).

129. Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151,** 1256–1269 (2012).

130. Park, S. *et al.* O-GlcNAc modification is essential for the regulation of autophagy in Drosophila melanogaster. *Cell. Mol. Life Sci.* (2015). doi:10.1007/s00018-015-1889-z

131. Forsythe, M. E. *et al.* Caenorhabditis elegans ortholog of a diabetes

susceptibility locus: oga-1 (O-GlcNAcase) knockout impacts O-GlcNAc cycling, metabolism, and dauer. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11952–7 (2006).

132. Mariappa, D. *et al.* Dual functionality of O-GlcNAc transferase is required for Drosophila development. *Open Biol.* **5,** (2015).

133. Gambetta, M. C. & Müller, J. O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. *Dev. Cell* **31,** 629–39 (2014).

134. Love, D. C. *et al.* Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 7413–8 (2010).

135. Hanover, J. A. *et al.* A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 11266–71 (2005).

136. Kenwrick, S., Amaya, E. & Papalopulu, N. Pilot Morpholino Screen in Xenopus tropicalis Identifies a Novel Gene Involved in Head Development. *Dev. Dyn.* **229,** 289–299 (2004).

137. Keembiyehetty, C. *et al.* Conditional knock-out reveals a requirement for O-Linked N-Acetylglucosaminase (O-GlcNAcase) in metabolic homeostasis. *J. Biol. Chem.* **290,** 7097–7113 (2015).

138. Olivier-Van Stichelen, S., Wang, P., Comly, M., Love, D. C. & Hanover, J. A. Nutrient-driven O-linked N-acetylglucosamine (O-GlcNAc) cycling impacts neurodevelopmental timing and metabolism. *J. Biol. Chem.* **292,** 6076–6085 (2017).

139. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. repeats (Wilson et al. 1993a, 1995; Kristie et al. 1995). These cleavages result in stable N- (HCF-1. *Genes Dev.* 896–911 (2003).

doi:10.1101/gad.252103.conserved

140. Ruan, H.-B. *et al.* O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1α stability. *Cell Metab.* **16,** 226–37 (2012).

141. Iyer, S. P. N. & Hart, G. W. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J. Biol. Chem.* **278,** 24608–16 (2003).

142. Ito, R. *et al.* TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* **19,** 52–65 (2014).

143. Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32,** 645–55 (2013).

144. Mariappa, D., Pathak, S. & van Aalten, D. M. F. A sweet TET-à-tête-synergy of TET proteins and O-GlcNAc transferase in transcription. *EMBO J.* **32,** 612–613 (2013).

145. Speakman, C. M. *et al.* Elevated O-GlcNAc levels activate epigenetically repressed genes and delay mouse ESC differentiation without affecting na??ve to primed cell transition. *Stem Cells* **32,** 2605–2615 (2014).

146. Hayakawa, K. *et al.* Epigenetic switching by the metabolism-sensing factors in the generation of orexin neurons from mouse embryonic stem cells. *J. Biol. Chem.* **288,** 17099–17110 (2013).

147. März, P. *et al.* Ataxin-10 interacts with O-linked beta-N-acetylglucosamine transferase in the brain. *J. Biol. Chem.* **281,** 20263–70 (2006).

148. Akimoto, Y. *et al.* Localization of the O-GlcNAc transferase and O-GlcNAc-modified proteins in rat cerebellar cortex. *Brain Res.* **966,** 194–205 (2003).

149. Alfaro, J. F. *et al.* Tandem mass spectrometry identi fi es many mouse brain

O -GlcNAcylated proteins including EGF domain-specific O -GlcNAc transferase targets. *Pnas* **109,** 7280–7285 (2012).

150. Gong, C.-X., Liu, F., Grundke-Iqbal, I. & Iqbal, K. Impaired brain glucose metabolism leads to Alzheimer neurofibrillary degeneration through a decrease in tau O-GlcNAcylation. *J. Alzheimers. Dis.* **9,** 1–12 (2006).

151. Dias, W. B. & Hart, G. W. O-GlcNAc modification in diabetes and Alzheimer's disease. *Mol. Biosyst.* **3,** 766 (2007).

152. Wang, P. *et al.* O-GlcNAc cycling mutants modulate proteotoxicity in Caenorhabditis elegans models of human neurodegenerative diseases. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17669–74 (2012).

153. Yoshida, H. & Goedert, M. Phosphorylation of microtubule-associated protein tau by AMPK-related kinases. *J. Neurochem.* **120,** 165–176 (2012).

154. Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J. & Crowther, R. A. Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J.* **8,** 393–9 (1989).

155. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. & Crowther, R. A. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3,** 519–526 (1989).

156. Morris, M. *et al.* Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nat. Neurosci.* (2015). doi:10.1038/nn.4067

157. Graham, D. L. *et al.* Increased O-GlcNAcylation reduces pathological tau without affecting its normal phosphorylation in a mouse model of tauopathy.

Neuropharmacology **79,** 307–313 (2014).

158. Ropers, H. H. Genetics of Early Onset Cognitive Impairment. *Annu. Rev. Genomics Hum. Genet.* **11,** 161–187 (2010).

159. Bouazzi, H., Lesca, G., Trujillo, C., Alwasiyah, M. K. & Munnich, A. Nonsyndromic X-linked intellectual deficiency in three brothers with a novel *MED12* missense mutation [c.5922G>T (p.Glu1974His)]. *Clin. Case Reports* **3,** 604–609 (2015).

160. Vaidyanathan, K. *et al.* Identification and characterization of a missense mutation in the O-linked β-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J. Biol. Chem.* **292,** 8948–8963 (2017).

161. Willems, A. P. *et al.* Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. *J. Biol. Chem.* **292,** 12621–12631 (2017).

162. Niranjan, T. S. *et al.* Affected Kindred Analysis of Human X Chromosome Exomes to Identify Novel X-Linked Intellectual Disability Genes. *PLoS One* **10,** e0116454 (2015).

163. Zhang, Z., Tan, E. P., VandenHull, N. J., Peterson, K. R. & Slawson, C. O-GlcNAcase expression is sensitive to changes in O-GlcNAc homeostasis. *Front. Endocrinol. (Lausanne).* **5,** 1–8 (2014).

164. de Queiroz, R. M., Carvalho, E. & Dias, W. B. O-GlcNAcylation: The Sweet Side of the Cancer. *Front. Oncol.* **4,** 132 (2014).

165. Sekine, O., Love, D. C., Rubenstein, D. S. & Hanover, J. A. Blocking O-linked GlcNAc cycling in Drosophila insulin-producing cells perturbs glucose-insulin homeostasis. *J. Biol. Chem.* **285,** 38684–38691 (2010).

166. Maurer, P. H. English (3), Warburg.

167. Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276,** 9838–45 (2001).

168. Dong, D. L. & Hart, G. W. Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **269,** 19321–30 (1994).

169. Heckel, D. *et al.* Novel immunogenic antigen homologous to hyaluronidase in meningioma. *Hum. Mol. Genet.* **7,** 1859–72 (1998).

170. Farook, V. S., Bogardus, C. & Prochazka, M. Analysis of MGEA5 on 10q24.1-q24.3 encoding the beta-O-linked N-acetylglucosaminidase as a candidate gene for type 2 diabetes mellitus in Pima Indians. *Mol. Genet. Metab.* **77,** 189–93 (2002).

171. Wells, L. *et al.* Dynamic O-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic beta-N-acetylglucosaminidase, O-GlcNAcase. *J. Biol. Chem.* **277,** 1755–61 (2002).

172. Yoshida, K. ichiro. Demonstration and some properties of N-acetyl-β,d-hexosaminidase (HEX) C isoenzyme in human renal tissues: relative increase of HEX C activity in renal cell carcinoma. *Clin. Chim. Acta* **226,** 55–65 (1994).

173. Overdijk, B., Van Der Kroef, W. M. J., Van Steijn, G. & Lisman, J. J. W. Isolation and further characterization of bovine brain hexosaminidase C. *Biochim. Biophys. Acta - Enzymol.* **659,** 255–266 (1981).

174. BEUTLEE, E. & KUHL, W. The tissue distribution of hexosaminidase S and hexosaminidase C. *Ann. Hum. Genet.* **41,** 163–167 (1977).

175. Comtesse, N., Maldener, E. & Meese, E. Identification of a nuclear variant of {MGEA5,} a cytoplasmic hyaluronidase and a {beta-N-acetylglucosaminidase.}. *Biochem. Biophys. Res. Commun.* **283,** 634–640 (2001).

176. Macauley, M. S. & Vocadlo, D. J. Enzymatic characterization and inhibition of the nuclear variant of human O-GlcNAcase. *Carbohydr. Res.* **344,** 1079–1084 (2009).

177. Toleman, C., Paterson, A. J., Whisenhunt, T. R. & Kudlow, J. E. Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *J. Biol. Chem.* **279,** 53665–73 (2004).

178. Toleman, C. A., Paterson, A. J. & Kudlow, J. E. The histone acetyltransferase NCOAT contains a zinc finger-like motif involved in substrate recognition. *J. Biol. Chem.* **281,** 3918–3925 (2006).

179. Butkinaree, C. *et al.* Characterization of ??-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis. *J. Biol. Chem.* **283,** 23557–23566 (2008).

180. He, Y., Roth, C., Turkenburg, J. P. & Davies, G. J. Three-dimensional structure of a Streptomyces sviceus GNAT acetyltransferase with similarity to the C-terminal domain of the human GH84 O-GlcNAcase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70,** 186–195 (2014).

181. Rao, F. V *et al.* Structure of a bacterial putative acetyltransferase defines the fold of the human O-GlcNAcase C-terminal domain. *Open Biol.* **3,** 130021 (2013).

182. Cheung, W. D., Sakabe, K., Housley, M. P., Dias, W. B. & Hart, G. W. O-linked ??-N-acetylglucosaminyltransferase substrate specificity is

regulated by myosin phosphatase targeting and other interacting proteins. *J. Biol. Chem.* **283,** 33935–33941 (2008).

183. Whisenhunt, T. R. *et al.* Disrupting the enzyme complex regulating O-GlcNAcylation blocks signaling and development. *Glycobiology* **16,** 551–563 (2006).

184. Beausoleil, S. A. *et al.* Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12130–5 (2004).

185. Rao, F. V *et al.* Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J.* **25,** 1569–78 (2006).

186. Schimpl, M., Borodkin, V. S., Gray, L. J. & van Aalten, D. M. F. Synergy of peptide and sugar in O-GlcNAcase substrate recognition. *Chem. Biol.* **19,** 173–8 (2012).

187. Schimpl, M., Schüttelkopf, A. W., Borodkin, V. S. & van Aalten, D. M. F. Human OGA binds substrates in a conserved peptide recognition groove. *Biochem. J.* **432,** 1–7 (2010).

188. Li, B., Li, H., Lu, L. & Jiang, J. Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. *Nat. Struct. Mol. Biol.* (2017). doi:10.1038/nsmb.3390

189. Roth, C. *et al.* Structural and functional insight into human O-GlcNAcase. *Nat. Chem. Biol.* 1–5 (2017). doi:10.1038/nchembio.2358

190. Elsen, N. L. *et al.* Insights into activity and inhibition from the crystal structure of human O-GlcNAcase. *Nat. Chem. Biol.* 1–5 (2017). doi:10.1038/nchembio.2357

191. Ostrowski, A., Gundogdu, M., Ferenbach, A. T., Lebedev, A. A. & van Aalten, D. M. F. Evidence for a Functional O-Linked N-Acetylglucosamine (O-GlcNAc) System in the Thermophilic Bacterium Thermobaculum

terrenum. *J. Biol. Chem.* **290,** 30291–305 (2015).

192. Dennis, R. J. *et al.* Structure and mechanism of a bacterial β-glucosaminidase having O-GlcNAcase activity. *Nat. Struct. Mol. Biol.* **13,** 365–371 (2006).

193. Laskowski, R. A. & Swindells, M. B. LigPlot + : Multiple Ligand À Protein Interaction Diagrams for Drug Discovery. 2778–2786 (2011). doi:10.1021/ci200227u

194. Gloster, T. M. & Vocadlo, D. J. Mechanism, Structure, and Inhibition of O-GlcNAc Processing Enzymes. *Curr Signal Transduct Ther* **5,** 74–91 (2010).

195. Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D. & Vocadlo, D. J. O-GlcNAcase uses substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired inhibitors. *J. Biol. Chem.* **280,** 25313–22 (2005).

196. Knapp, S. *et al.* NAG-thiazoline , An N -Acetyl- -hexosaminidase Inhibitor That Implicates Acetamido Participation Departments of Chemistry Rutgers s The State Uni V ersity of New Jersey New Brunswick , New Jersey 08903 Uni V ersity of British Columbia , Vancou V er Recei. *J. Am. Chem. Soc.* **118,** 6804–6805 (1996).

197. Horsch, M., Hoesch, L., Vasella, a & Rast, D. M. N-acetylglucosaminono-1,5-lactone oxime and the corresponding (phenylcarbamoyl)oxime. Novel and potent inhibitors of beta-N-acetylglucosaminidase. *Eur. J. Biochem.* **197,** 815–818 (1991).

198. VAVRA, J. J., DEBOER, C., DIETZ, A., HANKA, L. J. & SOKOLSKI, W. T. Streptozotocin, a new antibacterial antibiotic. *Antibiot. Annu.* **7,** 230–5 (1960).

199. Yanagisawa, M. & Yu, R. K. O-linked beta-N-acetylglucosaminylation in

mouse embryonic neural precursor cells. *J. Neurosci. Res.* **87,** 3535–45 (2009).

200. Kaasik, K. *et al.* Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. *Cell Metab.* **17,** 291–302 (2013).

201. Pantaleon, M., Tan, H. Y., Kafer, G. R. & Kaye, P. L. Toxic effects of hyperglycemia are mediated by the hexosamine signaling pathway and o-linked glycosylation in early mouse embryos. *Biol. Reprod.* **82,** 751–758 (2010).

202. Filhoulaud, G., Guillemain, G. & Scharfmann, R. The hexosamine biosynthesis pathway is essential for pancreatic beta cell development. *J. Biol. Chem.* **284,** 24583–24594 (2009).

203. Haltiwanger, R. S., Grove, K. & Philipsberg, G. A. Modulation of O -Linked N -Acetylglucosamine Levels on Nuclear and Cytoplasmic Proteins in Vivo Using the Peptide O -GlcNAc- - N - acetylglucosaminidase Inhibitor O - (2-Acetamido-2-deoxy- D - glucopyranosylidene) amino- N - phenylcarbamate *. **273,** 3611–3617 (1998).

204. Gao, Y., Parker, G. J. & Hart, G. W. Streptozotocin-Induced β-Cell Death Is Independent of Its Inhibition of O-GlcNAcase in Pancreatic Min6 Cells. *Arch. Biochem. Biophys.* **383,** 296–302 (2000).

205. Macauley, M. S., Bubb, A. K., Martinez-Fleites, C., Davies, G. J. & Vocadlo, D. J. Elevation of global O-GlcNAc levels in 3T3-L1 adipocytes by selective inhibition of O-GlcNAcase does not induce insulin resistance. *J. Biol. Chem.* **283,** 34687–34695 (2008).

206. Mehdy, A. *et al.* PUGNAc treatment leads to an unusual accumulation of free oligosaccharides in CHO cells. *J. Biochem.* **151,** 439–446 (2012).

207. Dorfmueller, H. C. *et al.* GlcNAcstatin: a picomolar, selective O-GlcNAcase inhibitor that modulates intracellular O-glcNAcylation levels. *J. Am. Chem. Soc.* **128,** 16484–5 (2006).

208. Beer, D., Maloisel, J.-L., Rast, D. M. & Vasella, A. Synthesis of 2-Acetamido-2-deoxy-D-gluconhydroximolactone- and Chitobionhydroximolactone-DerivedN-Phenylcarbamates, Potential Inhibitors of ?-N-Acetylglucosaminidase. *Helv. Chim. Acta* **73,** 1918–1922 (1990).

209. RAKIETEN, N., RAKIETEN, M. L. & NADKARNI, M. V. Studies on the diabetogenic action of streptozotocin (NSC-37917). *Cancer Chemother. reports* **29,** 91–8 (1963).

210. Schnedl, W. J., Ferber, S., Johnson, J. H. & Newgard, C. B. STZ transport and cytotoxicity: Specific enhancement in GLUT2-expressing cells. *Diabetes* **43,** 1326–1333 (1994).

211. Liu, K., Paterson, a J., Chin, E. & Kudlow, J. E. Glucose stimulates protein modification by O-linked GlcNAc in pancreatic beta cells: linkage of O-linked GlcNAc to beta cell death. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2820–2825 (2000).

212. Konrad, R. J., Mikolaenko, I., Tolar, J. F., Liu, K. & Kudlow, J. E. The potential mechanism of the diabetogenic action of streptozotocin: inhibition of pancreatic beta-cell O-GlcNAc-selective N-acetyl-beta-D-glucosaminidase. *Biochem. J.* **356,** 31–41 (2001).

213. Toleman, C., Paterson, A. J., Shin, R. & Kudlow, J. E. Streptozotocin inhibits O-GlcNAcase via the production of a transition state analog. *Biochem. Biophys. Res. Commun.* **340,** 526–534 (2006).

214. Pathak, S., Dorfmueller, H. C., Borodkin, V. S. & van Aalten, D. M. F.

Chemical dissection of the link between streptozotocin, O-GlcNAc, and pancreatic cell death. *Chem. Biol.* **15,** 799–807 (2008).

215. Yamamoto, H., Uchigata, Y. & Okamoto, H. Streptozotocin and alloxan induce DNA strand breaks and poly(ADP–ribose) synthetase in pancreatic islets. *Nature* **294,** 284–286 (1981).

216. Turk, J., Corbett, J. A., Ramanadham, S., Bohrer, A. & Mcdaniel, M. L. Biochemical Evidence for Nitric Oxide Formation from Streptozotocin in Isolated Pancreatic Islets. *Biochem. Biophys. Res. Commun.* **197,** 1458–1464 (1993).

217. Yuzwa, S. a *et al.* A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. *Nat. Chem. Biol.* **4,** 483–90 (2008).

218. Dorfmueller, H. C. *et al.* Cell-penetrant, nanomolar O-GlcNAcase inhibitors selective against lysosomal hexosaminidases. *Chem. Biol.* **17,** 1250–1255 (2010).

219. Dorfmueller, H. C., Borodkin, V. S., Schimpl, M. & van Aalten, D. M. F. GlcNAcstatins are nanomolar inhibitors of human O-GlcNAcase inducing cellular hyper-O-GlcNAcylation. *Biochem. J.* **420,** 221–7 (2009).

220. Goldberg, H., Whiteside, C. & Fantus, I. G. O -linked N - N - acetylglucosamine supports p38 MAPK activation by high glucose in glomerular mesangial cells. 2–4 (2011). doi:10.1152/ajpendo.00108.2011.

221. Mi, W. *et al.* O-GlcNAcylation is a novel regulator of lung and colon cancer malignancy. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1812,** 514–519 (2011).

222. Kiss, H. *et al.* Complete genome sequence of ‘Thermobaculum terrenum’ type strain (YNP1). *Stand. Genomic Sci.* **3,** 153–62 (2010).

223. Borodkin, V. S. & Van Aalten, D. M. F. An efficient and versatile synthesis

of GlcNAcstatins - Potent and selective O-GlcNAcase inhibitors built on the tetrahydroimidazo[1,2-a]pyridine scaffold. *Tetrahedron* **66,** 7838–7849 (2010).

224. Hart, G. W., Haltiwanger, R. S., Holt, G. D. & Kelly, W. G. Nucleoplasmic and cytoplasmic glycoproteins. *Ciba Found. Symp.* **145,** 102–12, NaN-8 (1989).

225. Haltiwanger, R. . S. ., Holt, G. . D. . & Hart, G. . W. . Enzymatic addition of O-GlcNAc to Nuclear and Cytoplasmic Proteins. *J. Biol. Chem.* **265,** 2563–2568 (1990).

226. Haltiwanger, R. S., Blomberg, M. A. & Hart, G. W. Glycosylation of nuclear and cytoplasmic proteins: Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide ??-N-acetylglucosaminyltransferase. *J. Biol. Chem.* **267,** 9005–9013 (1992).

227. Kreppel, L. K., Blomberg, M. a & Hart, G. W. Dynamic Glycosylation of Nuclear and Cytosolic Proteins. *J. Biol. Chem.* **272,** 9308–9315 (1997).

228. Lubas, W. A., Frank, D. W., Hanover, J. A. & Krause, M. Cell Biology and Metabolism : O -Linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats O -Linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats *. **272,** 9316–9324 (1997).

229. Roos, M. D. & Hanover, J. a. Structure of O-linked GlcNAc transferase: mediator of glycan-dependent signaling. *Biochem. Biophys. Res. Commun.* **271,** 275–80 (2000).

230. Jínek, M. *et al.* The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat. Struct. Mol. Biol.* **11,** 1001–1007 (2004).

231. Pathak, S. *et al.* The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nat. Struct. Mol. Biol.* (2015). doi:10.1038/nsmb.3063

232. Lubas, W. & Hanover, J. Functional expression of O-linked GlcNAc transferase Domain structure and substrate specificity. *J. Biol. Chem.* **275,** 10983–8 (2000).

233. Xing, Y. *et al.* Crystal Structure of a Full-Length β-Catenin. *Structure* **16,** 478–487 (2008).

234. Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T. & Adam, S. A. Importin ??: A multipurpose nuclear-transport receptor. *Trends Cell Biol.* **14,** 505–514 (2004).

235. Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. *Cell* **94,** 193–204 (1998).

236. Huber, A. H. & Weis, W. I. The Structure of the B-Catenin/E-Cadherin Complex and the Molecular Basis of Diverse Ligand Recognition by B-Catenin. *Cell* **105,** 391–402 (2001).

237. Lazarus, M. B. *et al.* HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* **342,** 1235–9 (2013).

238. Clarke, A. J. *et al.* Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J.* **27,** 2780–8 (2008).

239. Martinez-Fleites, C. *et al.* Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. *Nat. Struct. Mol. Biol.* **15,** 764–5 (2008).

240. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature*

469, 564–7 (2011).

241. Schimpl, M. *et al.* O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nat. Chem. Biol.* **8,** 969–74 (2012).

242. Lazarus, M. B. *et al.* Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nat. Chem. Biol.* **8,** 966–968 (2012).

243. Trapannone, R., Rafie, K. & van Aalten, D. M. F. O-GlcNAc transferase inhibitors: current tools and future challenges. *Biochem. Soc. Trans.* **44,** 88–93 (2016).

244. Selvan, N. *et al.* A mutant O-GlcNAcase enriches Drosophila developmental regulators. *Nat. Chem. Biol.* 4–11 (2017). doi:10.1038/nchembio.2404

245. Jochmann, R., Holz, P., Sticht, H. & Stürzl, M. Validation of the reliability of computational O-GlcNAc prediction. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 416–421 (2014).

246. Wu, H.-Y. *et al.* Characterization and identification of protein O-GlcNAcylation sites with substrate specificity. *BMC Bioinformatics* **15,** S1 (2014).

247. Leavy, T. M. & Bertozzi, C. R. A high-throughput assay for O-GlcNAc transferase detects primary sequence preferences in peptide substrates. *Bioorganic Med. Chem. Lett.* **17,** 3851–3854 (2007).

248. Myers, S. A., Panning, B. & Burlingame, A. L. Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **108,** 9490–9495 (2011).

249. Wang, J., Torii, M., Liu, H., Hart, G. W. & Hu, Z.-Z. dbOGAP - an integrated

bioinformatics resource for protein O-GlcNAcylation. *BMC Bioinformatics* **12,** 91 (2011).

250. Hahne, H., Gholami, A. M. & Kuster, B. Discovery of *O* -GlcNAc-modified Proteins in Published Large-scale Proteome Data. *Mol. Cell. Proteomics* **11,** 843–850 (2012).

251. Britto-borges, T. & Barton, G. J. A study of the structural properties of sites modified by the O -linked 6-N- acetylglucosamine transferase. 1–25 (2017).

252. Shi, J., Sharif, S., Ruijtenbeek, R. & Pieters, R. J. Activity based high-throughput screening for novel O-GlcNAc transferase substrates using a dynamic peptide microarray. *PLoS One* **11,** 1–14 (2016).

253. Liu, X. *et al.* A peptide panel investigation reveals the acceptor specificity of O-GlcNAc transferase. *FASEB J.* 1–11 (2014). doi:10.1096/fj.13-246850

254. Zhao, P. *et al.* Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome Res.* **10,** 4088–4104 (2011).

255. Teo, C. F. *et al.* Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nat. Chem. Biol.* **6,** 338–343 (2010).

256. Vosseller, K., Wells, L. & Hart, G. W. Nucleocytoplasmic O-glycosylation: O-GlcNAc and functional proteomics. *Biochimie* **83,** 575–581 (2001).

257. Zaro, B. W., Yang, Y.-Y., Hang, H. C. & Pratt, M. R. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc. Natl. Acad. Sci.* **108,** 8146–8151 (2011).

258. Wang, Z. *et al.* Enrichment and Site Mapping of *O* -Linked *N* -Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass

Spectrometry. *Mol. Cell. Proteomics* **9,** 153–160 (2010).

259. Kelly, W., Dahmus, M. & Hart, G. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J. Biol. Chem.* **268,** 10416–24 (1993).

260. Lu, L. *et al.* Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. *Biochemistry* **55,** 1149–1158 (2016).

261. Hang, H. C., Yu, C., Pratt, M. R. & Bertozzi, C. R. Probing Glycosyltransferase Activities with the Staudinger Ligation. *J. Am. Chem. Soc.* **126,** 6–7 (2004).

262. Pratt, M. R. *et al.* Deconvoluting the functions of polypeptide N-alpha-acetylgalactosaminyltransferase family members by glycopeptide substrate profiling. *Chem. Biol.* **11,** 1009–16 (2004).

263. Jochmann, R. *et al.* O-GlcNAc transferase inhibits KSHV propagation and modifies replication relevant viral proteins as detected by systematic O-GlcNAcylation analysis. *Glycobiology* **23,** 1114–30 (2013).

264. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics* **11,** M111.014050 (2012).

265. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7,** 548–548 (2014).

266. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

267. Zhao, X., Ning, Q., Chai, H., Ai, M. & Ma, Z. PGlcS: Prediction of protein O-GlcNAcylation sites with multiple features and analysis. *J. Theor. Biol.*

380, 524–529 (2015).

268. Gupta, R. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symp. Biocomput.* **322,** 310–322 (2002).

269. Kao, H.-J. *et al.* A two-layered machine learning method to identify protein O-GlcNAcylation sites with O-GlcNAc transferase substrate motifs. *BMC Bioinformatics* **16,** S10 (2015).

270. Forman, J. R. & Clarke, J. Mechanical unfolding of proteins: insights into biology, structure and folding. *Curr. Opin. Struct. Biol.* **17,** 58–66 (2007).

271. Wilson, A. C., LaMarco, K., Peterson, M. G. & Herr, W. The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74,** 115–125 (1993).

272. Zargar, Z. & Tyagi, S. Role of Host Cell Factor-1 in cell cycle regulation. *Transcription* **3,** 187–192 (2012).

273. Wilson, A. C., Freiman, R. N., Goto, H., Nishimoto, T. & Herr, W. VP16 targets an amino-terminal domain of HCF involved in cell cycle progression. *Mol. Cell. Biol.* **17,** 6139–6146 (1997).

274. Mangone, M., Myers, M. P. & Herr, W. Role of the HCF-1 basic region in sustaining cell proliferation. *PLoS One* **5,** 1–9 (2010).

275. Kristie, T. M. & Sharp, P. A. Purification of the cellular C1 factor required for the stable recognition of the Oct-1 homeodomain by the herpes simplex virus ??-trans-induction factor (VP16). *J. Biol. Chem.* **268,** 6525–6534 (1993).

276. Wilson, A. C., Peterson, M. G. & Herr, W. The HCF repeat is an unusual proteolytic cleavage signal. *Genes Dev.* **9,** 2445–58 (1995).

277. Wilson, A. C., Boutros, M., Johnson, K. M. & Herr, W. HCF-1 Amino-and

Carboxy-Terminal Subunit Association through Two Separate Sets of Interaction Modules: Involvement of Fibronectin Type 3 Repeats. *Mol. Cell. Biol.* **20,** 6721–6730 (2000).

278. Julien, E. & Herr, W. Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *EMBO J.* **22,** 2360–2369 (2003).

279. Capotosti, F. *et al.* O-GlcNAc Transferase Catalyzes Site-Specific Proteolysis of HCF-1. *Cell* **144,** 376–388 (2011).

280. Janetzko, J., Trauger, S. ., Lazarus, M. . & Walker, S. How the glycosyltransferase OGT catalyzes amide bond cleavage. *Nat.Chem.Biol.* (2016). doi:10.2210/PDB5HGV/PDB

281. Howerton, C. L. & Bale, T. L. Targeted placental deletion of OGT recapitulates the prenatal stress phenotype including hypothalamic mitochondrial dysfunction. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 9639–44 (2014).

282. Lagerlöf, O., Hart, G. W. & Huganir, R. L. O-GlcNAc transferase regulates excitatory synapse maturity. *Proc. Natl. Acad. Sci.* 201621367 (2017). doi:10.1073/pnas.1621367114

283. Mariappa, D. *et al.* A mutant O-GlcNAcase as a probe to reveal global dynamics of protein O-GlcNAcylation during Drosophila embryonic development. *Biochem. J.* **470,** 255–62 (2015).

284. Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D. & Vocadlo, D. J. O-GlcNAcase uses substrate-assisted catalysis: Kinetic analysis and development of highly selective mechanism-inspired inhibitors. *J. Biol. Chem.* **280,** 25313–25322 (2005).

285. Dorfmueller, H. C. *et al.* Substrate and product analogues as human O-

GlcNAc transferase inhibitors. *Amino Acids* **40,** 781–92 (2011).

286. Konrad, R. J. *et al.* Alloxan is an inhibitor of the enzyme O-linked N-acetylglucosamine transferase. *Biochem. Biophys. Res. Commun.* **293,** 207–12 (2002).

287. Lenzen, S. The mechanisms of alloxan- and streptozotocin-induced diabetes. *Diabetologia* **51,** 216–226 (2008).

288. Wu, T. *et al.* Cardioprotection of salidroside from ischemia/reperfusion injury by increasing N-acetylglucosamine linkage to cellular proteins. *Eur. J. Pharmacol.* **613,** 93–99 (2009).

289. Dehennaut, V. *et al.* O-linked N-acetylglucosaminyltransferase inhibition prevents G2/M transition in Xenopus laevis oocytes. *J. Biol. Chem.* **282,** 12527–36 (2007).

290. Tallent, M. K. *et al.* In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. *J. Biol. Chem.* **284,** 174–181 (2009).

291. Liu, J. *et al.* Cardiovascular effects of endomorphins in alloxan-induced diabetic rats. *Peptides* **26,** 607–614 (2005).

292. Xu, C., Liu, G., Liu, X. & Wang, F. O-GlcNAcylation under hypoxic conditions and its effects on the blood-retinal barrier in diabetic retinopathy. *Int. J. Mol. Med.* 624–632 (2013). doi:10.3892/ijmm.2013.1597

293. Lee, T. N., Alborn, W. E., Knierman, M. D. & Konrad, R. J. Alloxan is an inhibitor of O-GlcNAc-selective N-acetyl-??-d-glucosaminidase. *Biochem. Biophys. Res. Commun.* **350,** 1038–1043 (2006).

294. Lenzen, S. & Munday, R. Thiol-group reactivity, hydrophilicity and stability of alloxan, its reduction products and its N-methyl derivatives and a comparison with ninhydrin. *Biochem. Pharmacol.* **42,** 1385–91 (1991).

295. Zhang, H., Gao, G. & Brunk, U. T. Extracellular reduction of alloxan results in oxygen radical-mediated attack on plasma and lysosomal membranes. *APMIS* **100,** 317–25 (1992).

296. Gross, B. J., Kraybill, B. C. & Walker, S. Discovery of O-GlcNAc transferase inhibitors. *J. Am. Chem. Soc.* **127,** 14588–9 (2005).

297. Hennebicq-Reig, S. *et al.* Permanent exposure of mucin-secreting HT-29 cells to benzyl-N-acetyl-alpha-D-galactosaminide induces abnormal O-glycosylation of mucins and inhibits constitutive and stimulated MUC5AC secretion. *Biochem. J.* **334 (Pt 1,** 283–295 (1998).

298. Gloster, T. M. *et al.* Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. *Nat. Chem. Biol.* **7,** 174–81 (2011).

299. Donovan, K., Alekseev, O., Qi, X., Cho, W. & Azizkhan-Clifford, J. O-GlcNAc Modification of Transcription Factor Sp1 Mediates Hyperglycemia-Induced VEGF-A Upregulation in Retinal Cells. *Invest. Ophthalmol. Vis. Sci.* **55,** 7862–7873 (2014).

300. Perez-Cervera, Y. *et al.* Insulin signaling controls the expression of O-GlcNAc transferase and its interaction with lipid microdomains. *FASEB J.* **27,** 3478–3486 (2013).

301. Ma, Z., Vocadlo, D. & Vosseller, K. Hyper-O-GlcNAcylation is anti-apoptotic and maintains constitutive NF-κB activity in pancreatic cancer cells. *J. Biol. Chem.* 1–19 (2013). doi:10.1074/jbc.M113.470047

302. Olivier-Van Stichelen, S. *et al.* Serum-stimulated cell cycle entry promotes ncOGT synthesis required for cyclin D expression. *Oncogenesis* **1,** e36 (2012).

303. Ortiz-meoz, R. F. *et al.* A small molecule that inhibits OGT activity in cells

Department of Microbiology and Immunobiology , Harvard Medical School , Boston , National Center for Advancing Translational Sciences , National Institutes of Health , General Methods : Reagents : Gener. 1–22

304. Jeon, J. H., Suh, H. N., Kim, M. O., Ryu, J. M. & Han, H. J. Glucosamine-induced OGT activation mediates glucose production through cleaved Notch1 and FoxO1, which coordinately contributed to the regulation of maintenance of self-renewal in mouse embryonic stem cells. *Stem Cells Dev.* **23,** 1–42 (2014).

305. Suh, H. N., Lee, Y. J., Kim, M. O., Ryu, J. M. & Han, H. J. Glucosamine-Induced Sp1 O-GlcNAcylation Ameliorates Hypoxia-Induced SGLT Dysfunction in Primary Cultured Renal Proximal Tubule Cells. *J. Cell. Physiol.* **229,** 1557–1568 (2014).

306. Caldwell, S. a *et al.* Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1. *Oncogene* **29,** 2831–2842 (2010).

307. Lima, V. V. *et al.* O-GlcNAcylation contributes to the vascular effects of ET-1 via activation of the RhoA/Rho-kinase pathway. *Cardiovasc. Res.* **89,** 614–622 (2011).

308. Jiang, J., Lazarus, M. B., Pasquina, L., Sliz, P. & Walker, S. A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. *Nat. Chem. Biol.* **8,** 72–7 (2012).

309. Ortiz-Meoz, R. F. *et al.* A small molecule that inhibits OGT activity in cells. *ACS Chem. Biol.* **10,** 150309142052000 (2015).

310. Angelova, M., Ortiz-Meoz, R. F., Walker, S. & Knipe, D. M. Inhibition of O-Linked *N* -Acetylglucosamine Transferase Reduces Replication of Herpes Simplex Virus and Human Cytomegalovirus. *J. Virol.* **89,** 8474–8483

(2015).

311. Borodkin, V. S. *et al.* Bisubstrate UDP-peptide conjugates as human O-GlcNAc transferase inhibitors. *Biochem. J.* (2013). doi:10.1042/BJ20131272

312. Kristie, T. M., Pomerantz, J. L., Twomey, T. C., Parent, S. A. & Sharp, P. A. The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. *J. Biol. Chem.* **270,** 4387–94 (1995).

313. Shibuya, H. *et al.* TAB1: an activator of the TAK1 MAPKKK in TGF-beta signal transduction. *Science* **272,** 1179–82 (1996).

314. Yamaguchi, K. *et al.* Family as a Potential Mediator of TGF- | Signal Transduction. *Science (80-.).* **270,** 2008–2011 (1995).

315. Conner, S. H. *et al.* TAK1-binding protein 1 is a pseudophosphatase. *Biochem. J.* **399,** 427–34 (2006).

316. Cheung, P. C. F., Campbell, D. G., Nebreda, A. R. & Cohen, P. Feedback control of the protein kinase TAK1 by SAPK2a/p38alpha. *EMBO J.* **22,** 5793–805 (2003).

317. Mendoza, H. *et al.* Roles for TAB1 in regulating the IL-1-dependent phosphorylation of the TAB3 regulatory subunit and activity of the TAK1 complex. *Biochem. J.* **409,** 711–22 (2008).

318. Wolf, A. *et al.* Identification and functional characterization of novel phosphorylation sites in TAK1-binding protein (TAB) 1. *PLoS One* **6,** (2011).

319. Ojida, A., Takashima, I., Kohira, T., Nonaka, H. & Hamachi, I. Turn-on fluorescence sensing of nucleoside polyphosphates using a xanthene-based Zn(II) complex chemosensor. *J. Am. Chem. Soc.* **130,** 12095–101 (2008).

320. Tarrant, M. K. *et al.* Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. *Nat. Chem. Biol.* **8,** 262–269 (2012).

321. Wang, P. *et al.* OGT Mediated Histone H2B S112 GlcNAcylation Regulates DNA Damage Response. *J. Genet. Genomics* **42,** 467–475 (2015).

322. Rao, X. *et al.* O-GlcNAcylation of G6PD promotes the pentose phosphate pathway and tumor growth. *Nat. Commun.* **6,** 8468 (2015).

323. Stewart, K. M., Horton, K. L. & Kelley, S. O. Cell-penetrating peptides as delivery vehicles for biology and medicine. *Org. Biomol. Chem.* **6,** 2242 (2008).

324. Koren, E. & Torchilin, V. P. Cell-penetrating peptides: Breaking through to the other side. *Trends Mol. Med.* **18,** 385–393 (2012).

325. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

326. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67,** 235–42 (2011).

327. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 22–5 (2010).

328. Murshudov, G. N., Vagin, a a & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53,** 240–55 (1997).

329. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 486–501 (2010).

330. Schüttelkopf, A. W. & van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.*

Sect. D Biol. Crystallogr. **60,** 1355–63 (2004).

331. Nikolovska-Coleska, Z. *et al.* Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal. Biochem.* **332,** 261–73 (2004).

332. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

333. Borodkin, V. S. *et al.* Bisubstrate UDP–peptide conjugates as human O-GlcNAc transferase inhibitors. *Biochem. J.* **457,** 497–502 (2014).

334. Rafie, K. *et al.* Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR repeats. *Open Biol.* **7,** (2017).

335. Gross, B. J., Swoboda, J. G. & Walker, S. A strategy to discover inhibitors of O-linked glycosylation. *J. Am. Chem. Soc.* **130,** 440–1 (2008).

336. Kim, E. J. In Vitro Biochemical Assays for O-GlcNAc Processing Enzymes. *Chembiochem* (2017). doi:10.1002/cbic.201700138

337. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **22,** 3099–108 (1973).

338. Hart, G., Housley, M. & Slawson, C. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446,** 1017–22 (2007).

339. Wolfenden, R. Analog Approaches to the Structure of the Transition State in Enzyme Reactions. *Acc. Chem. Res.* **5,** 10–18 (1972).

340. Lienhard, G. E. & Secemski, I. I. P 1 ,P 5 -Di(adenosine-5')pentaphosphate, a potent multisubstrate inhibitor of adenylate kinase. *J. Biol. Chem.* **248,** 1121–3 (1973).

341. Collins, K. D. & Stark, G. R. Collins and george. *J. Biol. Chem.* **246,** 6599–6605 (1971).

342. Bogoyevitch, M. A., Barr, R. K. & Ketterman, A. J. Peptide inhibitors of protein kinases-discovery, characterisation and use. *Biochim. Biophys. Acta* **1754,** 79–99 (2005).

343. Starke, K. *et al. Handbook of Experimental Pharmacology Volume 167*. *Handbook of Experimental Pharmacology* **11,** (1990).

344. Kayser-Bricker, K. J. *et al.* Non-peptidic substrate-mimetic inhibitors of Akt as potential anti-cancer agents. *Bioorg. Med. Chem.* **17,** 1764–71 (2009).

345. Kayser, K. J., Glenn, M. P., Sebti, S. M., Cheng, J. Q. & Hamilton, A. D. Modifications of the GSK3beta substrate sequence to produce substrate-mimetic inhibitors of Akt as potential anti-cancer therapeutics. *Bioorg. Med. Chem. Lett.* **17,** 2068–73 (2007).

346. Parang, K. & Cole, P. A. Designing bisubstrate analog inhibitors for protein kinases. *Pharmacol. Ther.* **93,** 145–157 (2002).

347. Palcic, M. M., Heerze, L. D., Srivastava, O. P. & Hindsgaul, O. A bisubstrate analog inhibitor for alpha(1----2)-fucosyltransferase. *J. Biol. Chem.* **264,** 17174–81 (1989).

348. Izumi, M., Kaneko, S., Yuasa, H. & Hashimoto, H. Synthesis of bisubstrate analogues targeting α-1,3-fucosyltransferase and their activities. *Org. Biomol. Chem.* **4,** 681–690 (2006).

349. Qiao, L., Murray, B. W., Shimazaki, M., Schultz, J. & Wong, C. H. Synergistic inhibition of human α-1,3-fucosyltransferase V. *J. Am. Chem. Soc.* **118,** 7653–7662 (1996).

350. Lavogina, D., Enkvist, E. & Uri, A. Bisubstrate inhibitors of protein kinases: from principle to practical applications. *ChemMedChem* **5,** 23–34 (2010).

351. Cozzi, F. & Siegel, J. S. Interaction between stacked aryl groups in 1,8-diarylnaphthalenes: Dominance of polar/π over charge-transfer effects. *Pure Appl. Chem.* **67,** 683–689 (1995).

352. Cozzi, F., Ponzini, F., Annunziata, R., Cinquini, M. & Siegel, J. S. Polar Interactions between Stackedπ Systems in Fluorinated 1,8-Diarylnaphthalenes: Importance of Quadrupole Moments in Molecular Recognition. *Angew. Chemie Int. Ed. English* **34,** 1019–1020 (1995).

353. Huang, X. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J. Biomol. Screen.* **8,** 34–8 (2003).

354. Drake, E. J., Duckworth, B. P., Neres, J., Aldrich, C. C. & Gulick, A. M. Biochemical and structural characterization of bisubstrate inhibitors of BasE, the self-standing nonribosomal peptide synthetase adenylate-forming enzyme of acinetobactin synthesis. *Biochemistry* **49,** 9292–305 (2010).

355. Hu, Y., Helm, J., Chen, L. & Ginsberg, C. Identification of selective inhibitors for the glycosyltransferase MurG via high-throughput screening. *Chem. Biol.* **11,** 703–11 (2004).

356. Ostrowski, A. & van Aalten, D. M. F. Chemical tools to probe cellular O-GlcNAc signalling. *Biochem. J.* **456,** 1–12 (2013).

357. Duncan, R. The dawning era of polymer therapeutics. *Nat. Rev. Drug Discov.* **2,** 347–60 (2003).

358. Maeda, H. Macromolecular therapeutics in cancer treatment: The EPR effect and beyond. *J. Control. Release* **164,** 138–144 (2012).

359. Derossi, D., Joliot, A. H., Chassaing, G. & Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological

membranes. *J. Biol. Chem.* **269,** 10444–10450 (1994).

360. Gehring, W. J. the Study Development. **75,** (1987).

361. Plaza, S. *et al.* Molecular basis for the inhibition of Drosophila eye development by Antennapedia. *EMBO J.* **20,** 802–11 (2001).

362. Schneuwly, S., Klemenz, R. & Gehring, W. J. Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. *Nature* **325,** 816–818 (1987).

363. Schneuwly, S., Kuroiwa, A. & Gehring, W. J. Molecular analysis of the dominant homeotic Antennapedia phenotype. *EMBO J.* **6,** 201–206 (1987).

364. Joliot, A., Pernelle, C., Deagostini-Bazin, H. & Prochiantz, A. Antennapedia homeobox peptide regulates neural morphogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1864–8 (1991).

365. Joliot, A. H., Triller, A., Volovitch, M., Pernelle, C. & Prochiantz, A. alpha-2,8-Polysialic acid is the neuronal surface receptor of antennapedia homeobox peptide. *New Biol.* **3,** 1121–34 (1991).

366. Vivès, E., Brodin, P. & Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biol. Chem.* **272,** 16010–7 (1997).

367. Mäe, M. & Langel, Ü. Cell-penetrating peptides as vectors for peptide, protein and oligonucleotide delivery. *Curr. Opin. Pharmacol.* **6,** 509–514 (2006).

368. Erazo-Oliveras, A. *et al.* Protein delivery into live cells by incubation with an endosomolytic agent. *Nat. Methods* **11,** 861–867 (2014).

369. Nischan, N. *et al.* Covalent attachment of cyclic TAT peptides to GFP results in protein delivery into live cells with immediate bioavailability. *Angew. Chemie - Int. Ed.* **54,** 1950–1953 (2015).

370. Pooga, M., Hällbrink, M., Zorko, M. & Langel, U. Cell penetration by transportan. *FASEB J.* **12,** 67–77 (1998).

371. Rothbard, J. B. *et al.* Conjugation of arginine oligomers to cyclosporin A facilitates topical delivery and inhibition of inflammation. *Nat. Med.* **6,** 1253–1257 (2000).

372. Jiang, T. *et al.* Tumor imaging by means of proteolytic activation of cell-penetrating peptides. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 17867–72 (2004).

373. Turner, J. J. *et al.* Cell-penetrating peptide conjugates of peptide nucleic acids (PNA) as inhibitors of HIV-1 Tat-dependent trans-activation in cells. *Nucleic Acids Res.* **33,** 6837–6849 (2005).

374. Khafagy, E. S. & Morishita, M. Oral biodrug delivery using cell-penetrating peptide. *Adv. Drug Deliv. Rev.* **64,** 531–539 (2012).

375. Snyder, E. L. *et al.* Enhanced targeting and killing of tumor cells expressing the CXC chemokine receptor 4 by transducible anticancer peptides. *Cancer Res.* **65,** 10646–10650 (2005).

376. Fotin-Mleczek, M. *et al.* Cationic cell-penetrating peptides interfere with TNF signalling by induction of TNF receptor internalization. *J. Cell Sci.* **118,** 3339–3351 (2005).

377. Wadia, J. S., Stan, R. V & Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **10,** 310–315 (2004).

378. Jain, M. *et al.* Penetratin improves tumor retention of single-chain antibodies: A novel step toward optimization of radioimmunotherapy of solid tumors. *Cancer Res.* **65,** 7840–7846 (2005).

379. Gros, E. *et al.* A non-covalent peptide-based strategy for protein and peptide nucleic acid transduction. *Biochim. Biophys. Acta - Biomembr.*

1758, 384–393 (2006).

380. Turner, J. J., Arzumanov, A. A. & Gait, M. J. Synthesis, cellular uptake and HIV-1 Tat-dependent trans-activation inhibition activity of oligonucleotide analogues disulphide-conjugated to cell-penetrating peptides. *Nucleic Acids Res.* **33,** 27–42 (2005).

381. Shiraishi, T., Pankratova, S. & Nielsen, P. E. Calcium ions effectively enhance the effect of antisense peptide nucleic acids conjugated to cationic Tat and oligoarginine peptides. *Chem. Biol.* **12,** 923–929 (2005).

382. Zielinski, J. *et al.* In vivo identification of ribonucleoprotein-RNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1557–62 (2006).

383. Magzoub, M. & Gräslund, A. Cell-penetrating peptides : small from inception to application. *Q. Rev. Biophys.* **37,** 147–195 (2004).

384. Magzoub, M., Pramanik, A. & Gräslund, A. Modeling the endosomal escape of cell-penetrating peptides: Transmembrane pH gradient driven translocation across phospholipid bilayers. *Biochemistry* **44,** 14890–14897 (2005).

385. Feng, J. & Tang, L. The cell-type specificity and endosomal escape of cell-penetrating peptides. *Curr. Pharm. Des.* **21,** 1351–6 (2015).

386. Qian, Z. *et al.* Early endosomal escape of a cyclic cell-penetrating peptide allows effective cytosolic cargo delivery. *Biochemistry* **53,** 4034–4046 (2014).

387. Cascales, L. *et al.* Identification and Characterization of a New Family of Cell-penetrating Peptides. *J. Biol. Chem.* **286,** 36932–36943 (2011).

388. Lian, W., Jiang, B., Qian, Z. & Pei, D. Cell-permeable bicyclic peptide inhibitors against intracellular proteins. *J. Am. Chem. Soc.* **136,** 9830–9833 (2014).

389. Ramos-Molina, B. *et al.* Cationic cell-penetrating peptides are potent furin inhibitors. *PLoS One* **10,** 1–11 (2015).

390. Qian, Z. *et al.* Efficient delivery of cyclic peptides into mammalian cells with short sequence motifs. *ACS Chem. Biol.* **8,** 423–431 (2013).

391. Oh, D. *et al.* Enhanced cellular uptake of short polyarginine peptides through fatty acylation and cyclization. *Mol. Pharm.* **11,** 2845–2854 (2014).

392. El-Sayed, A., Futaki, S. & Harashima, H. Delivery of macromolecules using arginine-rich cell-penetrating peptides: ways to overcome endosomal entrapment. *AAPS J.* **11,** 13–22 (2009).

393. McGrath, S. *et al.* Structure-Enabled Discovery of a Stapled Peptide Inhibitor to Target the Oncogenic Transcriptional Repressor TLE1. *Chem. - A Eur. J.* **23,** 9577–9584 (2017).

394. Chu, Q. *et al.* Towards understanding cell penetration by stapled peptides. *Med. Chem. Commun.* **6,** 111–119 (2015).

395. Runage, K., Bektas, M., Berkowitz, P. & Rubenstein, D. S. Targeting O-Glycosyltransferase (OGT) to promote healing of diabetic skin wounds. *J. Biol. Chem.* **289,** 5462–5466 (2014).

396. Huang, S. M. *et al.* High-glucose environment induced intracellular O-GlcNAc glycosylation and reduced galectin-7 expression in keratinocytes: Implications on impaired diabetic wound healing. *J. Dermatol. Sci.* **87,** 168–175 (2017).

397. Yuzwa, S. A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **8,** 393–399 (2012).

398. Hart, G. W. Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu. Rev. Biochem.* **66,** 315–35 (1997).

399. Yuzwa, S. a & Vocadlo, D. J. O-GlcNAc and neurodegeneration:

biochemical mechanisms and potential roles in Alzheimer's disease and beyond. *Chem. Soc. Rev.* (2014). doi:10.1039/c4cs00038b

400. Singh, J. P., Zhang, K., Wu, J. & Yang, X. O-GlcNAc Signaling in Cancer Metabolism and Epigenetics. *Cancer Lett.* 1–7 (2014). doi:10.1016/j.canlet.2014.04.014

401. Georghiou, G., Kleiner, R. E., Pulkoski-Gross, M., Liu, D. R. & Seeliger, M. a. Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles. *Nat. Chem. Biol.* **8,** 366–74 (2012).