Elucidating the role of *O*-GlcNAc glycosylation in neurobiology and neurodegeneration

Thesis by

Elizabeth Hwang Jensen

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

Biochemistry and Molecular Biophysics

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2017

Defended November 13, 2017

Elizabeth Hwang Jensen
ORCID: 0000-0002-6177-4304

for PSSETS, HwangLink, and the Jensensations

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Linda Hsieh-Wilson, for providing me with the support, freedom, and patience to pursue this project. I would also like to thank my committee, Peter Dervan, Ellen Rothenberg, and David Prober, for their wisdom, support, and advice. Furthermore, I would like to thank the collaborators who made this work possible: Igor Antoshechkin, Vijaya Kumar, Rachael Neve, Syed Ahsan, Kwan Lee, Natalie Verduzco, Shirley Pease, Mona Shahgholi, Dan Mason, Eric Peters, and Sonja Hess. In particular, I would like to especially thank Igor who has consistently been encouraging and supportive of our work.

In addition, I would like to thank the members of the Hsieh-Wilson lab that contributed to this work including Peter Clark, Jess Rexach, Andrew Wang, Matt Griffin, Yelena Koldobskaya, Yao Xiao, and Priya Choudhry. Specifically, I would like to thank Matt for being a generous statistically significant other and Greg for being a pillar of support through our last year together. Finally, I would next like to thank Yao for being a supportive friend and colleague. I would also like to thank everyone else in the Hsieh-Wilson lab; it has been a privilege and honor to work alongside these hardworking and passionate scientists.

Outside of my lab, I would like to thank my friends throughout grad school including Emzo de los Santos, Naeem Husain, Naomi Kreamer, Adam Shai, Betty Wong, Liz Wang, Gloria Sheng, Tri Vu, Joe Levine, Chandra Bhattacharya, Yaqian Liu, and Hung Lu. You all have enriched my life in more ways than I can expound, and I will forever cherish the memories we have made together. In addition, I would like to thank all the dedicated members of the GSC, Active Minds, and the ultimate groups. I would also like to thank the Caltech gym, health center staff, Adrienne, and Mariel for providing me with much-needed Chicken soup for the soul.

I would like to thank those who nurtured my scientific and personal growth before and during my time at Caltech. I would like to thank Dr. Walker for showing me through example how to be a empowering and motivating mentor. I would like to thank my extended and immediate family for their love and support. To my sisters, Sine and Tasha, thank you for being equal parts the wind behind my sails and mooring in turbulent times. I must also thank Mama Jensen who raised and taught us through her grit and gumption.

Last, but not least, I would like to thank my partner, Sandy. I have loved going through this journey with you, and I am so excited for our next great adventure and life together.

ABSTRACT

O-GlcNAc glycosylation is a dynamic, inducible post-translational modification (PTM) essential for neuronal homeostasis and found on proteins associated with neurodegenerative diseases such as α-synuclein, amyloid precursor protein, and tau. Intracellularly, *O*-GlcNAc modification is cycled by two enzymes in mammalian cells: *O*-GlcNAc transferase (OGT) appends *O*-GlcNAc to serine or threonine residues and O-GlcNAcase (OGA) removes *O*-GlcNAc. OGT modifies over 1000 different proteins, but the lack of a well-defined consensus sequence or substrate structural constraints has hampered efforts to predict sites *a priori*. Furthermore, the identification of O-GlcNAc modification sites has been obstructed by the difficulty of enriching and detecting *O*-GlcNAc using traditional biochemical methods. Here, we established and employed biological and chemical tools to illuminate the role of *O*-GlcNAc in neuronal function.

In Chapter 2, we sought to determine the role of *O*-GlcNAc in learning, memory, and neurodegeneration. Deletion of the OGT gene causes early postnatal lethality in mice, complicating efforts to study *O*-GlcNAc glycosylation in mature neuronal function and dysfunction. We demonstrated that the loss of OGT in the forebrain of adult mice (OGT cKO) leads to progressive neurodegeneration, including neuronal death, neuroinflammation, hyperphosphorylated tau, amyloidogenic Aβ-peptides, and memory deficits. In the hippocampus, we showed that OGT ablation lead to the upregulation of neuroinflammatory genes and the downregulation of cholesterol biosynthetic genes. Additionally, a gene network analysis (WGCNA), qPCR, and immunohistochemistry (IHC) revealed that loss of *O*-GlcNAc perturbed cell cycle progression in the hippocampal neurons. In the hippocampus, we identified increased neuroinflammatory gene transcription

in OGT cKO mice and both tau neurofibrillary tangle (NFT)-forming and amyloid-forming Alzheimer's disease (AD) mouse models. However, only OGT cKO and NFT-forming mice displayed decreased synaptic gene expression, suggesting that NFT formation and OGT cKO compromise hippocampal synaptic transcription. These studies indicate that *O*-GlcNAcylation regulates pathways vital for the maintenance of neuronal health and suggest that dysfunctional *O*-GlcNAc signaling may be an important contributor to neurodegenerative diseases.

In order to understand the critical *O*-GlcNAc-mediated neuronal functions that underlie OGT cKO dysfunction, we next developed and utilized novel biological and chemical tools in order to identify key OGT interactors and substrates in the brain in Chapter 3. Due to the lack of a well-defined OGT substrate sequence and structural constraints, OGT is believed to obtain its substrate specificity through its interactome where specific interactors target OGT to specific substrates. In order to identify these interactors, we used CRISPR/Cas9 to generate a novel mouse with a minimally tagged OGT in order to identify the endogenous OGT brain interactome using tandem affinity purification and MS methods. The preliminary OGT brain interactome consisted of previously identified OGT interactors and substrates as well as novel interactors. The identified OGT interactors were enriched for ribosomal and cytoskeletal proteins in addition to axonal, dendritic, and neuronal cell body proteins, implicating OGT as a pivotal mediator of neuronal structure and function.

In addition to the OGT interactome, we sought to uncover OGT's substrates or the *O*-GlcNAcome. We developed an improved approach to quantitatively label and enrich *O*-GlcNAcylated proteins for site identification. Chemoenzymatic labeling followed by Cu(I)-

catalyzed azide-alkyne cycloaddition (CuAAC) installed a new MS-compatible linker designed for facile purification and release of *O*-GlcNAcylated proteins for downstream MS analysis. We validated the approach by identifying several established *O*-GlcNAc sites on the proteins α-crystallin and OGT as well as discovering new, previously unreported sites on both proteins. Notably, these novel sites on OGT lie in key functional domains of OGT, underscoring how this site identification method can reveal important biological insights into protein activity and regulation.

Finally, in Chapters 4 and 5, we focus on the post-translational modification (PTM) code on a specific transcription factor (TF), CREB (cAMP response element binding protein). CREB regulates memory formation through its transcriptional control of neuronal metabolism, activity, differentiation, development, and survival. CREB phosphorylation at serine 133 has been previously shown to enhance CREB-mediated transcription while CREB glycosylation at serine 40 has been shown to decrease CREB-mediated transcription. However, the exact gene networks modulated by and potential interplay between CREB glycosylation and phosphorylation have not been explored. Through differential expression analysis with glycosylation-deficient (S40A) and phosphorylation-deficient (S133A) CREB mutants, we showed that CREB *O*-GlcNAcylation is important for neuronal activity and excitability while phosphorylation at serine 133 regulated the expression of genes involved in neuronal differentiation. Using WGCNA, we demonstrated that CREB *O*-GlcNAcylation at serine 40 and phosphorylation at serine 133 mediate mutually exclusive gene networks. The glycosylation-deficient mutant enhanced neuronal activity- and excitotoxicity-related gene networks while the phosphorylation-deficient mutant perturbed neuronal differentiation and amino and fatty acid metabolism-related

gene networks. Our work sheds light on the regulation of CREB through PTMs to modulate neuronal function and delineate the roles of *O*-GlcNAcylation and phosphorylation in modulating neuronal excitability and neuronal development and metabolism respectively. Altogether, these studies demonstrate that *O*-GlcNAc modification is a critical mediator of neuronal homeostasis and neurodegeneration.

PUBLISHED CONTENT AND CONTRIBUTIONS

Wang, AC, **Jensen EH**, Rexach JE, Vinters HV, and Hsieh-Wilson LC. "Loss of *O*-GlcNAc glycosylation in forebrain excitatory neurons induces neurodegeneration." *PNAS*. **2016**, *113*(52):15120-15125. doi: 10.1073/pnas.1606899113.

E.H.J. participated in the execution of the experiments including the microarray, qPCR, IHC, and western blotting experiments, data analysis, and in the writing of the manuscript. The article, including figures, is reproduced in part within Chapter 2 with permission under the *PNAS* rights and permissions.

Griffin ME, **Jensen EH**, Mason DE, Jenkins CL, Stone SE, Peters EC, Hsieh-Wilson LC. "Comprehensive mapping of O-GlcNAc modification sites using a chemically cleavable tag." Mol. Biosys. 2016, 12 (6): 1756-1759. doi: 10.1039/c6mb00138f. Research article.

E.H.J. participated in the execution of the experiments including the peptide and protein chemoenzymatic labeling and preparation for mass spectrometry and in writing of the manuscript. The article, including figures, is reproduced in part within Chapter 6 with permission under a Creative Commons Attribution 3.0 Unported License.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Number *Page*

NOMENCLATURE

A or Ala	alanine or adenosine
a.a.	amino acids
ACN	acetonitrile
AD	Alzheimer's disease
ALS	amyotrophic lateral sclerosis (a.k.a.Lou-Gehrig's disease or motor neuron disease)
AMPA	
Anti-anti	anti-mycotic, anti-microbial
B	biotin/biotinylated
BCA	bicinchoninic acid
BDNF	brain-derived neurotrophic factor
BEMAD	β-elimination Michael addition
BG	E11.5 basal ganglia
BM	bone marrow
Bp	base pair(s)
BSA	bovine serum albumin
C or Cys	cysteine or cytosine
CBP	CREB-binding protein
CID	collision-induced dissociation
CNs	cortical neurons (either E15.5 or E16.5)
Cre	Cre recombinase
CRE	consensus sequence for CREB binding (Full site: TGACGTCA, Half site: TGACG/CGTCA)
CREB	cAMP response element binding protein
CRISPR	clustered regularly interspaced short palindromic repeats
CuAAC	copper-catalyzed azide-alkyne cycloaddition
Cu(I)	copper (I), Cu^{+}
$CuSO_4$	copper sulfate
DAPI	4',6-diamidino-2-phenylindole
Dde	1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl
Ddv/ivDde	1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)isovaleryl
DE	differentially expressed
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(β-aminoethyl ether)tetraacetic acid
E#	embryonic day #

EOGT	EGF domain-specific *O*-linked *N*-acetylglucosamine transferase (extracellular)
ESC(s)	embryonic stem cell(s)
ETD	electron transfer dissociation
EtOH	ethanol
FBS	fetal bovine serum
FDR	false discovery rate
FLAG	protein tag with the sequence DYKDDDDK
FTD(P)	frontotemporal dementia (with Parkinsonism)
G or Gly	glycine or guanosine
Gal	galactose
GalN	galactosamine
GalNAc	*N*-acetylgalactosamine
GalNAz	*N*-azidoacetylgalactose
(Y289L) GalT	(Y289L) β-1,4-galactosyltransferase
gDNA	genomic DNA
GFP	green fluorescent protein
GlcNAc	*N*-acetylglucosamine
GlcNAz	*N*-azidoacetylglucosamine
HA	hemagglutinin tag (YPYDVPDYA)
HBSS	Hank's buffered saline solution
HCD	higher-energy collisional dissociation
HD	Huntington's disease
HDR	homology-directed repair/recombination
HEK293	human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSV	replication-deficient herpes simplex virus
ICC	immunocytochemistry
IF	immunofluorescence
IHC	immunohistochemistry
Ile (or I)	isoleucine
IRDye800	infrared dye 800
Leu (or L)	leucine
IgG	immunoglobulin domain G
IgM	immunoglobulin domain M
©KO	(conditional) knockout
lncRNA	long-noncoding ribonucleic acid
LTM	long term memory
LTQ	linear trap quadrupole
LSM	laser scanning microscope
LV	lentivirus
LWAC	lectin weak affinity chromatography
MAPK	mitogen-activated protein kinase
MEF(s)	mouse embryonic fibroblast(s)
MeOH	methanol

Met (or M)	methionine
MS	mass spectrometry
NaAsc	sodium ascorbate
NFTs	neurofibrillary tangles
NHEJ	non-homologous end joining
NMR	nuclear magnetic resonance
NPCs	neural progenitor cells
nt	nucleotide
O-GlcNAc	*O*-linked *N*-acetylglucosamine
OGA	*O*-GlcNAcase
(s)OGT	(short isoform of) *O*-GlcNAc transferase
PAM	protospacer motif
PARP1	poly-ADP ribose polymerase 1
PBS(T)	phosphate buffered saline (Tween 20)
PC	photocleavable
PC12	pheochromocytoma 12 (cell line)
PCR	polymerase chain reaction
PD	Parkinson's disease
PEG	polyethylene glycol
Phe (or F)	phenylalanine
PI3K	phosphatidylinositol 3-kinase
P/S	penicillin/streptomycin
PSP	progressive supranuclear palsy
PTM	post-translational modification
PUGNAc	*O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate
qRT-PCR (qPCR)	quantitative reverse transcription- polymerase chain reaction
RapiGest	sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid
rRNA	ribosomal ribonucleic acid
RNAi	RNA interference
RT	room temperature
siRNA	short interfering RNA
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate protein acrylamide gel electrophoresis
S.E.M	standard error of the mean
Ser (or S)	serine
ssODN	single-stranded oligonucletide
STM	short term memory
TAE	40 mM Tris base, 20 mM acetic acid, 1 mM EDTA, pH 8.3
TBS(T)	Tris buffered saline (Tween 20)
TFA	trifluoroacetic acid

THPTA	tris(3-hydroxypropyltriazolylmethyl)amine
Thr (or T)	threonine
TNFα	tumor necrosis factor alpha
Trp (or W)	tryptophan
Tyr (or Y)	tyrosine
UDP	uridine diphosphate
(3’ or 5’)UTR	(3’ or 5’) untranslated region of RNA
UV	ultraviolet
(s)WGA	(succinylated) wheat germ agglutinin
Wnt	wingless-type MMTV integration site family member