

New Tools for Studying *O*-GlcNAc  
Glycosylation and Chondroitin Sulfate Proteoglycans  
and  
Studies on the Roles of *O*-GlcNAc Glycosylation  
on the Transcription Factor CREB

Thesis by

Peter Michael Clark

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2010

(Defended December 21, 2010)

© 2010

Peter Michael Clark

All Rights Reserved

## Acknowledgements

First I would like to thank Prof. Linda Hsieh-Wilson for her support, advice, guidance, and scientific mentorship. I would also like to thank Jessica Rexach, my lifelong collaborator, without whose constant encouragement, support, ideas, (and cooking) much of this work would not have been accomplished. In addition, I would like to thank members of the Hsieh-Wilson lab, past and present, who have made my time at Caltech both enjoyable and productive, and especially those who taught me many of the techniques I still use today: Nelly Khidekel, Nathan Lamarre-Vincent, and Sarah Tully. Finally, I would like to thank my committee members — Prof. Dennis Dougherty, Prof. Peter Dervan, and Prof. Mary Kennedy — for their insightful scientific ideas and advice.

## Abstract

The addition and removal of the monosaccharide *N*-acetyl-D-glucosamine (GlcNAc) to serine and threonine residues of proteins has emerged as a critical regulator of cellular processes. However, studies of *O*-GlcNAc in such complex systems as the brain have been limited, in part due to the lack of tools. Here we report the development of new tools for studying *O*-GlcNAc, and the application of these and other tools for studying the roles of *O*-GlcNAc in the brain.

Working from a previously established chemoenzymatic method, we designed an isotopic labeling strategy for probing the dynamics of *O*-GlcNAc glycosylation using quantitative proteomics. With this tool, we show that *O*-GlcNAc is dynamically modulated on specific proteins by excitatory stimulation of the brain *in vivo*. Separately, we improved this chemoenzymatic strategy by integrating [3+2] azide-alkyne cycloaddition chemistry to attach biotin and fluorescent tags to *O*-GlcNAc residues. These tags allow for the direct fluorescence detection, proteomic analysis, and cellular imaging of *O*-GlcNAc modified proteins. With this strategy, we identified over 146 novel glycoproteins from the mammalian brain.

The transcription factor cAMP-response element binding protein (CREB) is critical for numerous functions in the brain, including neuronal survival, neuronal development, synaptic plasticity, and long-term memory. We show that CREB is highly glycosylated in the brain and discover new glycosylation sites on CREB in neurons. One

of these sites is dynamically modulated by neuronal activity and is important for regulating CREB. Removal of this glycosylation site accelerates axon and dendrite development *in vitro* and long-term memory consolidation *in vivo*. These studies are the first demonstration that *O*-glycosylation at a specific site on a specific protein is critical for neuronal function and behavior.

Chondroitin sulfates (CS) are sulfated linear polysaccharides important in neuronal development and viral invasion. Depending on their sulfation patterns, CS molecules differ dramatically in their functions. We developed a computational method to model the structure and function of CS. Using this approach, we show that different CS tetrasaccharides have distinct solution structures. We also modeled the CS binding site on a variety of proteins and discovered that CS may be important in modulating the interaction between specific growth factors and their receptors.

## Table of Contents

Chapter 1: Chemical Approaches to Understanding <i>O</i> -GlcNAc Glycosylation in the Brain	... 1
Chapter 2: The Roles of <i>O</i> -GlcNAc Glycosylation in the Brain	... 21
Chapter 3: Probing the Dynamics of <i>O</i> -GlcNAc Glycosylation in the Brain Using Quantitative Proteomics	... 39
Chapter 4: Direct In-Gel Fluorescence Detection and Cellular Imaging of <i>O</i> -GlcNAc-Modified Proteins	... 70
Chapter 5: CREB — A Key Transcription Factor in the Nervous System	... 111
Chapter 6: Dynamic <i>O</i> -Glycosylation Regulates CREB-Mediated Neuronal Gene Expression and Memory Formation	... 128
Chapter 7: Computational Modeling of Glycosaminoglycans	... 165
Appendix I: <i>O</i> -GlcNAc Glycosylation Regulates CREB Activity in Pancreatic Beta Cells	... 210

## List of Illustrations and Tables

### Chapter 1: Chemical Approaches to Understanding *O*-GlcNAc Glycosylation in the Brain

<b>Figure 1:</b> <i>O</i> -GlcNAc glycosylation is the addition of $\beta$ - <i>N</i> -acetylglucosamine to serine or threonine residues of proteins	... 2
<b>Figure 2:</b> Strategy for chemically tagging <i>O</i> -GlcNAc proteins	... 4
<b>Figure 3:</b> The BEMAD strategy for mapping glycosylation sites	... 6
<b>Figure 4:</b> Small molecule inhibitors of OGT and OGA	... 11
<b>Figure 5:</b> Chemical tools for monitoring <i>O</i> -GlcNAc dynamics	... 14

### Chapter 2: The Roles of *O*-GlcNAc Glycosylation in the Brain

<b>Figure 1:</b> <i>O</i> -GlcNAc glycosylation is the addition of $\beta$ - <i>N</i> -acetylglucosamine to serine or threonine residues of proteins	... 22
<b>Figure 2:</b> <i>O</i> -GlcNAc proteome from rodent brain	... 27

### Chapter 3: Probing the Dynamics of *O*-GlcNAc Glycosylation in the Brain Using Quantitative Proteomics

<b>Scheme 1:</b> QUIC-Tag strategy for quantitative <i>O</i> -GlcNAc proteomics	... 41
<b>Figure 1:</b> Accurate quantification of known <i>O</i> -GlcNAc peptides from complex mixtures using the QUIC-Tag approach	... 43
<b>Table 1a:</b> Mean ratios of individual peptides from $\alpha$ -crystallin and OGT	... 44
<b>Table 1b:</b> Mean ratios of all peptides	... 44
<b>Figure 2:</b> <i>O</i> -GlcNAc glycosylation is reversible in cultured cortical neurons	... 45
<b>Figure 3:</b> Sequencing of tagged <i>O</i> -GlcNAc peptides regulated by	... 46

PUGNAc treatment using CAD	
<b>Figure 4:</b> Sequencing of tagged <i>O</i> -GlcNAc peptides regulated by PUGNAc treatment using ETD	... 49
<b>Table 2:</b> Identification and quantification of changes in <i>O</i> -GlcNAc glycosylation induced by PUGNAc	... 50
<b>Figure 5:</b> Quantification of <i>O</i> -GlcNAc glycosylation on intact proteins by immunoblotting and infrared imaging detection	... 51
<b>Figure 6:</b> <i>O</i> -GlcNAc glycosylation is dynamically modulated by robust excitatory stimulation of the brain <i>in vivo</i> using kainic acid	... 52
<b>Table 3:</b> Identification and quantification of changes in <i>O</i> -GlcNAc glycosylation induced by kainic acid	... 54
<b>Figure 7:</b> Expression levels of EGR-1, GRASP55, and eIF4G following kainic acid treatment of rats	... 54
<b>Table 4:</b> <i>O</i> -GlcNAc glycosylated proteins identified from the cerebral cortex of kainic acid-stimulated rats	... 55
<b>Chapter 4: Direct In-Gel Fluorescence Detection and Cellular Imaging of <i>O</i>-GlcNAc-Modified Proteins</b>	
<b>Figure 1:</b> Chemoenzymatic labeling of <i>O</i> -GlcNAc proteins using [3+2] cycloaddition chemistry	... 72
<b>Figure 2:</b> Selective labeling of $\alpha$ -crystallin	... 73
<b>Figure 3:</b> Detection sensitivity of the chemoenzymatic approach	... 73
<b>Figure 4:</b> Enrichment and in-gel fluorescence detection of <i>O</i> -GlcNAc-modified proteins in 1D (top and middle) and 2D (bottom) gels	... 74
<b>Figure 5:</b> Functional classification of <i>O</i> -GlcNAc proteins from rat brain identified by MS	... 76
<b>Figure 6:</b> Direct detection of changes in <i>O</i> -GlcNAc glycosylation levels upon cellular stimulation	... 76
<b>Figure 7:</b> Detection of <i>O</i> -GlcNAc levels on chromatin	... 78

<b>Figure 8:</b> <i>O</i> -GlcNAc levels are distributed evenly across the mouse chromosomes	... 78
<b>Figure 9:</b> Fluorescence imaging of <i>O</i> -GlcNAc proteins in HeLa cells or cortical neurons	... 79
<b>Table 1.</b> <i>O</i> -GlcNAc glycosylated proteins identified by mass spectrometry	... 81
<b>Table 2:</b> Promoters enriched in <i>O</i> -GlcNAc levels	... 90

## **Chapter 6: Dynamic *O*-Glycosylation Regulates CREB-Mediated Neuronal Gene Expression and Memory Formation**

<b>Figure 1:</b> CREB is <i>O</i> -GlcNAc glycosylated in neurons.	... 131
<b>Figure 2:</b> CREB is <i>O</i> -GlcNAc glycosylated at Ser40 and Thr227 or Thr228.	... 131
<b>Figure 3:</b> CREB glycosylation sites	... 132
<b>Figure 4:</b> CREB is <i>O</i> -GlcNAc glycosylated at Ser40 in response to neuronal activity	... 132
<b>Figure 5:</b> Kinetics of endogenous CREB glycosylation and Ser133 phosphorylation upon depolarization of cortical neurons	... 133
<b>Figure 6:</b> The protein synthesis inhibitor cycloheximide does not block depolarization-induced glycosylation of CREB	... 133
<b>Figure 7:</b> CREB glycosylation is modulated by specific kinase pathways	... 134
<b>Figure 8:</b> Glycosylation is not induced on S133A CREB	... 134
<b>Figure 9:</b> Forskolin induces CREB phosphorylation but not CREB glycosylation in cortical neurons	... 136
<b>Figure 10:</b> Chemoenzymatic labeling of endogenous CREB for visualizing phosphorylation and glycosylation within the same protein molecule and for quantifying the levels of each modification within distinct posttranslationally modified subpopulations	... 136

<b>Figure 11:</b> CREB is simultaneously phosphorylated at Ser133 and glycosylated at Ser40	... 137
<b>Figure 12:</b> Kinetics of Ser133 phosphorylation for specific post-translationally modified subpopulations of endogenous CREB	... 137
<b>Figure 13:</b> Kinetics of glycosylation for specific post-translationally modified subpopulations of endogenous CREB	... 137
<b>Figure 14:</b> Knockdown of endogenous CREB and overexpression of shRNA-resistant, Flag-tagged WT or S40A CREB in neuro2a cells	... 138
<b>Figure 15:</b> Glycosylation at Ser40 represses CREB activity	... 139
<b>Figure 16:</b> Glycosylation at Ser40 represses CREB activity	... 139
<b>Figure 17:</b> WT and S40A CREB show similar binding to the CRE promoter <i>in vitro</i>	... 139
<b>Figure 18:</b> Glycosylation at Ser40 blocks the interaction between CREB and CRTC2	... 139
<b>Figure 19:</b> Glycosylation at Ser40 represses CREB activity via a CRTC-dependent mechanism in neurons	... 140
<b>Figure 20:</b> S40A CREB produces approximately a 2.5–3.6-fold increase in mRNA expression relative to WT CREB	... 141
<b>Figure 21:</b> Both CREB and OGT occupy the <i>BDNF</i> exon IV, <i>Arc</i> , <i>Cdk5</i> , <i>c-fos</i> , and <i>Wnt2</i> promoters	... 142
<b>Figure 22:</b> WT and S40A CREB show comparable levels of promoter occupancy on <i>BDNF</i> exon IV, <i>Arc</i> , <i>Cdk5</i> , <i>c-fos</i> , and <i>Wnt2</i> promoters	... 142
<b>Figure 23:</b> S40A CREB regulates a subset of CREB-mediated genes	... 142
<b>Figure 24:</b> Blocking glycosylation enhances <i>BDNF IV</i> and <i>c-fos</i> expression even more after depolarization	... 143
<b>Figure 25:</b> WT and S40A Flag-tagged CREB are expressed at similar levels in cortical neurons	... 143

<b>Figure 26:</b> CREB glycosylation at Ser40 represses dendritic growth	... 144
<b>Figure 27:</b> Expression of S40A CREB enhances axon outgrowth	... 144
<b>Figure 28:</b> OGT expression represses while OGT knockout enhances axon length	... 145
<b>Figure 29:</b> OGT knockout decreases <i>O</i> -GlcNAc levels, and OGT overexpression increases <i>O</i> -GlcNAc levels	... 145
<b>Figure 30:</b> CREB glycosylation at Ser40 represses dendritic growth through overexpression on Wnt2	... 146
<b>Figure 31:</b> Depolarization-induced dendrite growth is blocked by Wnt2 siRNA, Dkk-1, and Ncad(intra)	... 147
<b>Figure 32:</b> The enhanced axon outgrowth caused by expression of S40A CREB is blocked by TrkB-Fc or CRTCI knockdown	... 148
<b>Figure 33:</b> Herpes simplex virus (HSV) infection of the lateral amygdala of mice	... 150
<b>Figure 34:</b> CREB glycosylation at Ser40 modulates long-term conditioned fear memory	... 150
<b>Figure 35:</b> Inhibition of protein synthesis immediately after training blocks the memory enhancement of S40A CREB at 2 h	... 150
<b>Figure 36:</b> Inhibition of protein synthesis 2 h after training blocks the memory enhancement of WT CREB at 24 h, but not that of S40A CREB	... 151
<b>Figure 37:</b> CREB glycosylation is induced following activation of neurons <i>in vivo</i>	... 151
<b>Table 1:</b> Primers used for RT-PCR	... 159
<b>Table 2:</b> Primers used for ChIP PCR	... 159
 <b>Chapter 7: Computational Modeling of Glycosaminoglycans</b>	
<b>Figure 1:</b> CS-A, CS-C, CS-E, and CS-R tetrasaccharides	... 169

<b>Figure 2:</b> <i>Top:</i> The lowest energy structures of CS-A, CS-C, CS-E, and CS-R tetrasaccharides. <i>Bottom:</i> Electrostatic representations of these structures	... 172
<b>Figure 3:</b> <b>a,</b> Predicted heparin binding site of FGF-2 from 1BFB using default ScanBindSite parameters. <b>b,</b> Predicted heparin binding site of FGF-2 from 1BFB using modified ScanBindSite parameters	... 173
<b>Figure 4:</b> <b>a,</b> Heparin binding site on FGF-2 (from 1BFB). <b>b,</b> Predicted heparin binding site (from 1BFB crystal structure). <b>c,</b> Predicted heparin binding site (from 1BLA crystal structure)	... 174
<b>Table 1:</b> Residues that interact with heparin from the 1BFB crystal structure and in the predicted heparin binding site from FGF-2 in the 1BFB and 1BLA crystal structure	... 175
<b>Figure 5:</b> <b>a,</b> Heparin binding site on FGF-1 (from 2AXM). <b>b,</b> Predicted heparin binding site	... 177
<b>Table 2:</b> Residues that interact with heparin from the 2AXM crystal structure and in the predicted heparin binding site from FGF-1 in the 2AXM crystal structure	... 178
<b>Figure 6:</b> <b>a,</b> Residues important for CS-A binding to DLB3X (left) and DLB6 (right), as previously determined by mutagenesis experiments. <b>b,</b> Predicted CS-A binding site on DLB3X (left) and DLB6 (right)	... 179
<b>Table 3:</b> Predicted CS-A binding site on DBL3X and residues experimentally determined as important for CS-A binding	... 180
<b>Table 4:</b> Predicted CS-A binding site on DBL6 and residues experimentally determined as important for CS-A binding	... 180
<b>Figure 7:</b> <b>a,</b> CS-A binding site on Cathepsin K (from 3C9E). <b>b,</b> Predicted CSA binding site	... 181
<b>Table 5:</b> Residues that interact with CS-A from the 3C9E crystal structure and in the predicted CS-A binding site from Cathepsin A in the 3C9E crystal structure	... 182
<b>Figure 8:</b> <b>a,</b> Crystal structure of TNF trimer (from 1TNF). <b>b,</b> Predicted CS-E binding site on TNF trimer. <b>c,</b> Overlay between predicted CS-E binding site (slate) and predicted TNF-R1 (green, from 1TNR) complex.	... 183

<b>d</b> , Overlay between predicted CS-E binding site (slate) and residues important for TNF-R2 (green) binding	
<b>Figure 9: a</b> , BDNF monomer crystal structure (from 1BND). <b>b</b> , CS-E binding site (slate). <b>c</b> , Homology model of BDNF dimer crystal structure (wheat and cyan). <b>d</b> , CS-E binding site (slate)	... 184
<b>Figure 10: a</b> , NT-3 monomer crystal structure (from 1BND). <b>b</b> , CS-E binding site (slate). <b>c</b> , NT-3 dimer crystal structure (from 1NT3). <b>d</b> , CS-E binding site (slate)	... 185
<b>Table 6: Predicted CS-E binding sites on BDNF, NGF, NT-4/5, and NT-3</b>	... 186
<b>Figure 11: Cationic amino acids (yellow) in the respective CS-E binding sites CS-E binding site comparison</b>	... 187
<b>Figure 12: a</b> , NGF monomer crystal structure (from 2IFG). <b>b</b> , CS-E binding site (slate). <b>c</b> , NGF dimer crystal (from 2IFG). <b>d</b> , CS-E binding site (slate)	... 187
<b>Figure 13: a</b> , NT-4/5 monomer crystal structure (from 1HCF). <b>b</b> , CS-E binding site (slate). <b>c</b> , NT-4/5 dimer crystal (from 1HCF). <b>d</b> , CS-E binding site (slate)	... 188
<b>Figure 14: CS-E binding sites on the Trk family of receptors. a</b> , TrkA. <b>b</b> , TrkB. <b>c</b> , TrkC	... 190
<b>Table 7: Predicted CS-E binding sites on TrkA, TrkB, and TrkC</b>	... 191
<b>Figure 15: CS-E binding sites for the neurotrophins (slate) and Trk receptors (green), projected onto the neurotrophin (wheat) - receptor (cyan) complexes. a</b> , NGF–TrkA. <b>b</b> , BDNF–TrkB. <b>c</b> , NT-4/5–TrkB. <b>d</b> , NT-3–TrkC	... 192
<b>Figure 16: a</b> , Midkine (from 1MKN). <b>b</b> , Predicted CS-E binding site (slate)	... 194
<b>Table 8: Residues that interact with heparin as previously determined by NMR and the predicted CS-E binding site</b>	... 194
<b>Figure 17: a</b> , GDNF crystal structure (from 2V5E). <b>b</b> , Predicted CS-E binding site on GDNF from <b>a</b> (slate and yellow). <b>c</b> , GDNF crystal structure (from Chain D 3FUB). <b>d</b> , Predicted CS-E binding site on GDNF	... 196

(slate and yellow) from **c**. Residues predicted to be in the CS-E binding site for both GDNF crystal structures are yellow; those predicted to be in the CS-E binding site for only one GDNF crystal structure are colored slate

**Figure 18:** **a**, GDNF protein–receptor complex (wheat and cyan, ... 197 respectively, from 3FUB). **b**, Predicted CS-E binding site (slate). **c**, Binding site of heparin mimic sucrose octasulfate (slate) as determined from the co-crystal

**Table 9:** CS-E binding site on GDNF monomer (3FUB, 2V5E ... 198 crystal structure), GDNF receptor, and Nogo

**Figure 19:** CS-E binding sites (slate) on GDNF (wheat) and CS-E binding ... 199 sites (yellow) on the GDNF receptor (cyan) as mapped onto the GDNF–GDNF receptor complex (3FUB)

**Figure 20:** **a**, Nogo receptor crystal structure (from 1P8T). **b**, CS-E ... 200 binding site (slate)

## **Appendix I: O-GlcNAc Glycosylation Regulates CREB Activity in Pancreatic Beta Cells**

**Figure 1:** Hyperglycemia induces CREB glycosylation ... 214

**Figure 2:** CREB is dynamically glycosylated at Ser40 in HIT-T15 cells ... 216

**Figure 3:** O-GlcNAc glycosylation represses CREB transcriptional ... 217 activity

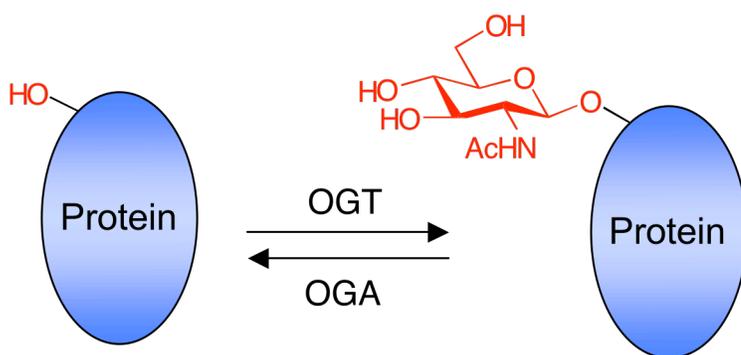
**Figure 4:** The *IRS-2* gene is a direct target for regulation by CREB and ... 219 OGT

## **Chapter 1: Chemical Approaches to Understanding *O*-GlcNAc Glycosylation in the Brain**

Portions of this chapter are from Rexach, J.E., Clark, P.M. & Hsieh-Wilson, L.C. Chemical approaches to understanding *O*-GlcNAc glycosylation in the brain. *Nat. Chem. Biol.* **4**, 97-106 (2008).

Over the past five years, new chemical approaches have been developed to meet the specific challenges associated with studying *O*-GlcNAc glycosylation in the brain. Methods for chemically tagging and identifying *O*-GlcNAc proteins have vastly expanded the number of neuronal proteins known to be *O*-GlcNAc modified. New pharmacological inhibitors of the *O*-GlcNAc enzymes have been discovered that may overcome some of the limitations of genetic approaches. Quantitative proteomics strategies designed to accommodate post-mitotic neurons and brain tissue have been used to monitor changes in glycosylation levels in response to neuronal stimuli. Combined with recent advances in mass spectrometry, these powerful tools have provided an unprecedented opportunity to explore the *O*-GlcNAc proteome, manipulate glycosylation levels, and study the dynamics of this modification *in vivo*.

*O*-GlcNAc glycosylation is the covalent attachment of the monosaccharide  $\beta$ -*N*-acetyl-D-glucosamine to the hydroxyl group of serine and threonine residues (Fig. 1).



**Figure 1:** *O*-GlcNAc glycosylation is the addition of  $\beta$ -*N*-acetylglucosamine to serine or threonine residues of proteins.

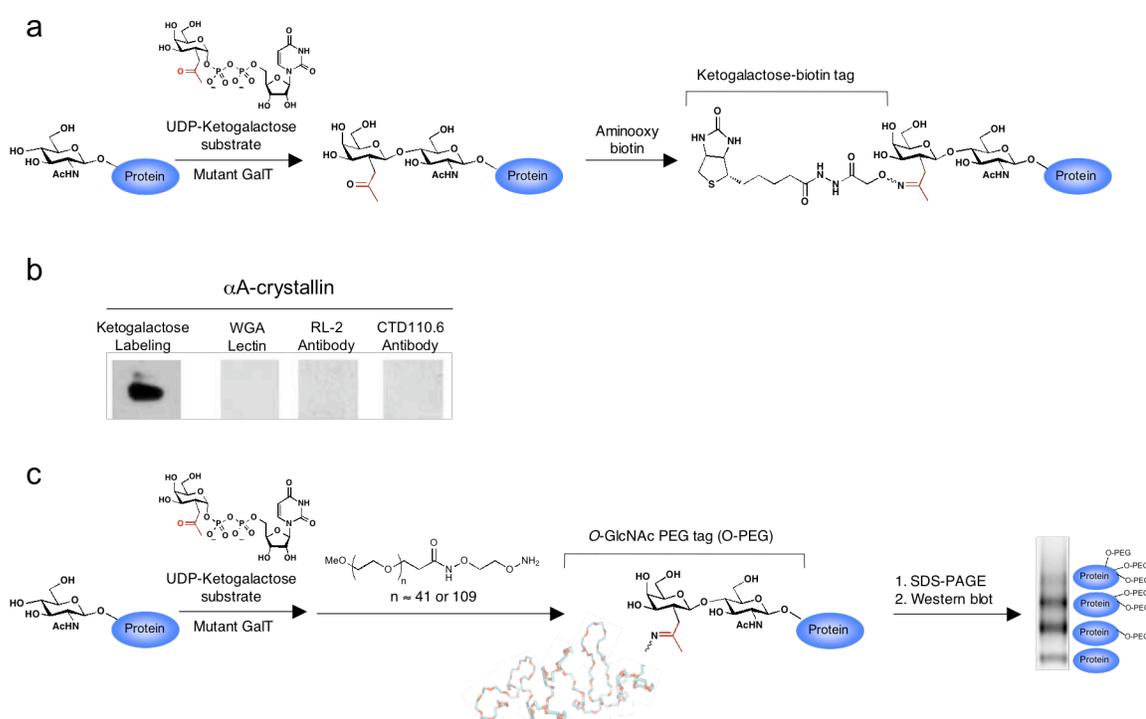
Recent advances in our understanding of the cellular functions of *O*-GlcNAc have been accelerated by the development of chemical tools for studying the modification. Although *O*-GlcNAc glycosylation was

discovered over twenty years ago<sup>1</sup>, its extent and functional significance are only beginning to be fully understood. A major challenge has been the difficulty of detecting and studying the modification *in vivo*. Like other PTMs, *O*-GlcNAc is often dynamic, substoichiometric, targeted to specific subcellular compartments, and prevalent on low-abundance regulatory proteins<sup>2,3</sup>. The sugar is also enzymatically and chemically labile, being subject to reversal by cellular glycosidases and cleavage on a mass spectrometer<sup>2</sup>. As with many protein kinases, the lack of a well-defined consensus sequence for OGT has precluded the determination of *in vivo* modification sites based on primary sequence. Here, we highlight some chemical approaches designed to overcome these challenges and advance a fundamental understanding of *O*-GlcNAc.

### **Rapid, sensitive detection**

Traditional methods for detecting *O*-GlcNAc glycosylation include the use of wheat germ agglutinin (WGA) lectin<sup>4</sup>, pan-specific *O*-GlcNAc antibodies<sup>5,6</sup>, or radiolabeling using  $\beta$ -1,4-galactosyltransferase (GalT)<sup>7</sup>, which transfers [<sup>3</sup>H]-Gal from UDP-[<sup>3</sup>H]-galactose to terminal GlcNAc groups. Although these approaches have greatly facilitated studies of the modification, many *O*-GlcNAc proteins still elude their detection. For instance, tritium labeling suffers from low sensitivity, often necessitating exposure times of several days to months. *O*-GlcNAc antibodies and lectins have limited binding affinity and specificity, and thus are best suited for highly glycosylated proteins with multiple modification sites. Recently, chemical approaches have been developed to tag *O*-GlcNAc proteins with reporter groups such as biotin to enable more rapid, sensitive detection. Khidekel and coworkers designed an unnatural substrate for GalT containing a

ketone moiety at the C2 position of UDP-galactose (**Fig. 2a**)<sup>8</sup>. An engineered enzyme with a single Y289L mutation that enlarges the binding pocket<sup>9</sup> was used to transfer the ketogalactose sugar onto *O*-GlcNAc proteins of interest. Once transferred, the ketone functionality was reacted with an aminoxy biotin derivative and detected by chemiluminescence using streptavidin conjugated to horseradish peroxidase. This approach provides a significant improvement in sensitivity, enabling detection of proteins



**Figure 2:** Strategy for chemically tagging *O*-GlcNAc proteins. (a) The UDP-ketogalactose substrate is used with an engineered mutant GalT enzyme to transfer the ketogalactose sugar onto *O*-GlcNAc proteins. Once transferred, the ketone functionality is reacted with an aminoxy biotin derivative. Tagged *O*-GlcNAc glycoproteins are then detected by chemiluminescence or isolated by streptavidin affinity chromatography. (b) Detection of  $\alpha$ A-crystallin using the ketogalactose-biotin tagging approach and comparison with other methods. The chemical tagging approach provides significantly improved sensitivity relative to antibodies and lectins. (c) The aminoxy biotin tag can be replaced with an aminoxy PEG tag to afford rapid detection of *O*-GlcNAc stoichiometries and dynamics.

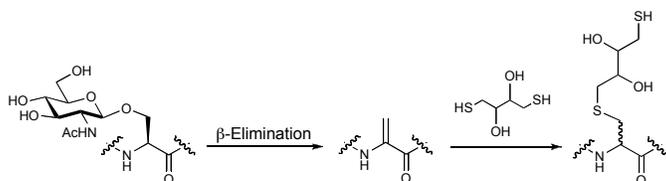
beyond the reach of traditional methods (**Fig. 2b**). The biotin handle also facilitates isolation of the glycoproteins from cell or tissue extracts, thus circumventing the need to develop purification procedures for each protein of interest. As such, virtually any protein can be readily interrogated for the modification, and comparisons can be made across specific functional classes. For instance, Tai and coworkers have applied the approach to demonstrate that numerous transcriptional regulatory proteins are *O*-GlcNAc glycosylated *in vivo*, including transcription factors, transcriptional coactivators and corepressors<sup>10</sup>. A variant of this method has recently been developed in which the biotin handle is replaced by a heavy poly-ethylene glycol tag. This tag increase the overall weight of the *O*-GlcNAc-modified fractions on a protein, allowing for rapid detection of *O*-GlcNAc stoichiometry and dynamics (**Fig. 2c**)<sup>11</sup>.

A complementary strategy involves tagging *O*-GlcNAc modified proteins through metabolic labeling of living cells with *N*-azidoacetylglucosamine<sup>12</sup>. Vocadlo and coworkers showed that the azido sugar is processed by enzymes in the hexosamine salvage pathway and incorporated into *O*-GlcNAc modified proteins. The bio-orthogonal azide moiety can then be derivatized with a FLAG peptide or biotin tag using the Staudinger ligation. Treatment of Jurkat cells with peracetylated *N*-azidoacetylglucosamine led to labeling of the highly glycosylated nuclear pore protein p62, as detected by immunoprecipitation of p62 and Western blotting. Although the labeling efficiency and sensitivity of this approach could not be evaluated and may depend on cell type, the ability to incorporate bio-orthogonal groups selectively into *O*-GlcNAc proteins represents a powerful strategy for the detection of *O*-GlcNAc proteins in living cells.

### Mapping *O*-GlcNAc glycosylation sites

Mapping the sites of *O*-GlcNAc glycosylation within proteins is essential for elucidating the functional roles of *O*-GlcNAc in specific biological contexts. Unfortunately, direct observation of the *O*-GlcNAc moiety by mass spectrometry during collision-induced dissociation (CID) is difficult as the glycosidic linkage is labile and readily cleaved, providing little peptide fragmentation<sup>13</sup>. Moreover, although OGT seems to favor sequences rich in proline, serine, and threonine residues<sup>13</sup>, there is no apparent consensus sequence that directs the action of OGT.

The development of chemical tools coupled to mass spectrometry has greatly facilitated the localization of *O*-GlcNAc to short peptide sequences within proteins and



**Figure 3:** The BEMAD strategy for mapping glycosylation sites. In this approach, the GlcNAc sugar undergoes a  $\beta$ -elimination reaction. Michael addition with dithiothreitol produces a sulfide adduct that is stable to MS/MS analysis.

can sometimes be used to determine exact glycosylation sites. One approach, which relies on  $\beta$ -elimination followed by Michael addition with dithiothreitol (BEMAD), results in replacement of the labile GlcNAc moiety with a

more stable sulfide adduct (**Fig. 3**)<sup>14</sup>. As this adduct is not cleaved upon CID, sites of glycosylation can be more readily determined. However, selectivity controls must be performed to distinguish *O*-GlcNAc from *O*-phosphate and other *O*-linked carbohydrates. Extension of BEMAD to proteomic studies for the high-throughput mapping of *O*-GlcNAc sites is an important future goal.

A second approach capitalizes on the ability to selectively biotinylate *O*-GlcNAc proteins using ketone- or azido-containing UDP-galactose sugars as described above (**Fig. 2**). In addition to isolating intact proteins, the biotin handle can be used to enrich *O*-GlcNAc peptides following proteolytic digestion<sup>13</sup>. This enrichment step is essential for mass spectrometric detection as the *O*-GlcNAc-modified species represents only a small fraction of the total peptides. Unlike BEMAD, the approach enables direct detection of the *O*-GlcNAc moiety by mass spectrometry, with the GlcNAc-ketogalactose-biotin tag providing a unique fragmentation pattern for unambiguous identification of *O*-GlcNAc peptides<sup>13</sup>. Notably, the strategy has been applied to both individual proteins and complex mixtures to localize the modification to short sequences within over 50 different proteins. However, mapping exact glycosylation sites remains challenging in most cases due to the lability of the *O*-glycosidic linkage. As a potential solution to this problem, Khidekel and coworkers demonstrated that this ketogalactose-biotin tagging approach could be combined with BEMAD to identify specific glycosylation sites on the HIV-1 Rev binding protein<sup>13</sup>. Similarly, Wang and coworkers recently used azidogalactose-biotin tagging in conjunction with BEMAD to map glycosylation sites on vimentin<sup>15</sup>.

Newer mass spectrometry approaches such as electron transfer dissociation (ETD) and electron capture dissociation (ECD) promise to greatly accelerate our ability to identify *O*-GlcNAc modification sites. ETD and ECD produce sequence-specific fragmentation of peptides without the loss of labile PTMs such as phosphorylation and glycosylation<sup>16</sup>. As such, these approaches are ideal for mapping *O*-GlcNAc glycosylation sites. Vosseller and coworkers demonstrated that ECD could be employed

to identify modification sites on several neuronal proteins following enrichment of the glycopeptides by lectin weak affinity chromatography (LWAC)<sup>17</sup>. Khidekel and coworkers applied the ketogalactose-biotin tagging approach in conjunction with ETD to map glycosylation sites on proteins from rat brain lysates (discussed in detail next chapter). More recently, Wang and coworkers have combined ETD with a photocleavable biotin tag to map glycosylation sites from a tau-enriched sample from rat brain<sup>18</sup> and HeLa cells<sup>19</sup>, and Chalkley and coworkers have combined ETD with LWAC to map glycosylation sites from the post-synaptic density of mouse brain<sup>20</sup>. With further methodological refinements and advances in database search algorithms for fragment ions, we anticipate that ETD and ECD will become an increasingly powerful tool for the study of *O*-GlcNAc glycosylation.

### **Proteome-wide analyses**

As described earlier, the analysis of *O*-GlcNAc on a proteome level has begun to reveal exciting functional roles for *O*-GlcNAc in the brain. Chemical strategies to tag, enrich, and detect *O*-GlcNAc peptides and proteins have been instrumental in this regard, enabling the first proteome-wide studies of *O*-GlcNAc. Specifically, the ketogalactose-biotin tagging approach has been applied in combination with high-throughput LC-MS/MS analysis to identify over 25 *O*-GlcNAc glycosylated proteins from the mammalian brain<sup>13</sup>. Direct detection of the modified species was observed in each case, which permitted mapping of the glycosylation sites to short peptide sequences. Notably, many of the proteins identified participate in the regulation of gene expression (e.g., CCR4-NOT, Sox2, HCF, TLE-4), neuronal signaling (WNK-1, bassoon, PDZ-GEF), and

synaptic plasticity (synaptopodin), suggesting that *O*-GlcNAc may contribute to neuronal communication processes. A similar approach employing a photo-cleavable biotin tag was used to identify 8 and 141 *O*-GlcNAc glycosylation sites from a tau-enriched sample from rat brain<sup>18</sup> and HeLa cells<sup>19</sup>, respectively. In another approach, Nandi and co-workers metabolically labeled HeLa cells with *N*-(2-azidoacetyl)glucosamine<sup>21</sup>. Tryptic digestion of the captured proteins led to the identification of 199 putative *O*-GlcNAc-modified proteins. As the presence of the GlcNAc moiety was inferred rather than detected directly, independent confirmation of the modification by immunoprecipitation was required and demonstrated for 23 of the proteins.

Recently, biochemical tools such as antibodies and lectins have been exploited for proteomic analyses of *O*-GlcNAc. Vosseller and co-workers used a uniquely packed, 39-foot lectin affinity column to enrich *O*-GlcNAc peptides and identified 18 proteins from the postsynaptic density fraction of murine brain<sup>17</sup>. Antibody affinity chromatography using traditional *O*-GlcNAc antibodies successfully identified 45 putative *O*-GlcNAc proteins from COS7 cells<sup>15</sup>. Additionally, Wells and co-workers developed new *O*-GlcNAc antibodies and identified over 200 *O*-GlcNAc proteins from HEK293T cells and rat liver<sup>22</sup>. The challenge with both approaches is that the weak binding affinity of antibodies and lectins necessitates gentle washing conditions and can lead to false positives, such as interacting proteins or non-specific binding proteins. In many cases, further confirmation of the presence of the modification can be provided by evaluating individually immunoprecipitated proteins or by directly detecting *O*-GlcNAc peptides by mass spectrometry analysis.

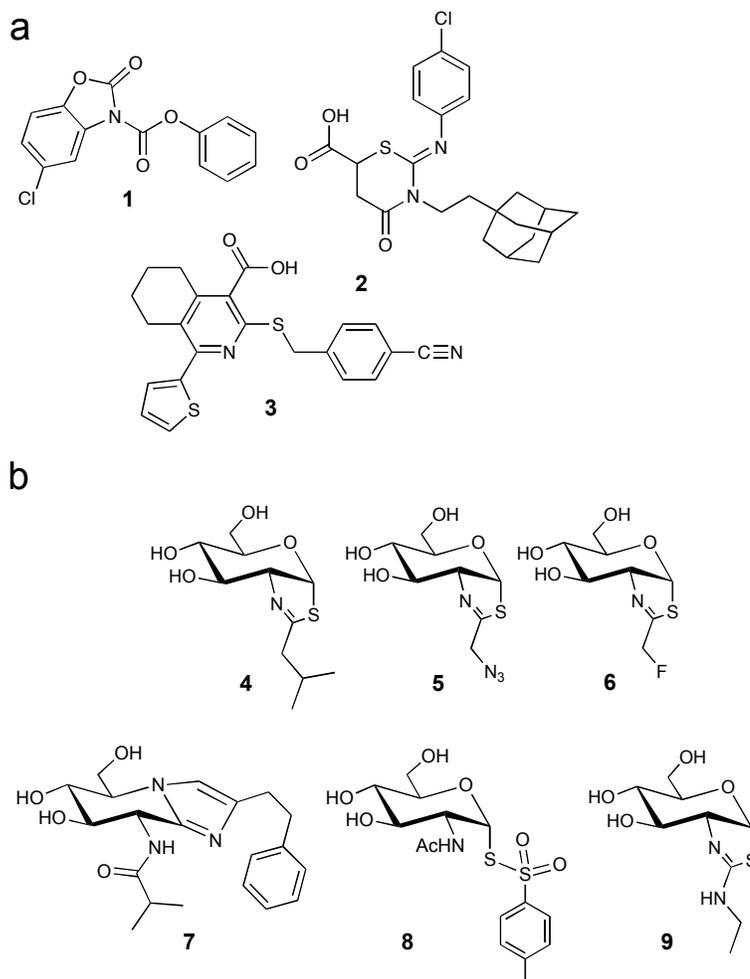
Together, proteome-wide studies have revealed that *O*-GlcNAc is a ubiquitous and abundant modification, with considerable functional significance and breadth. In less than 5 years, the number of known *O*-GlcNAc proteins in the brain has expanded from tens to hundreds, with many more proteins yet to be discovered.

### **Small molecule inhibitors of OGT and OGA**

Historically, genetic manipulation of OGT and OGA activity *in vivo* has proven difficult. Whole animal and conditional deletions of the OGT gene have revealed that OGT is essential for cell survival and mouse embryogenesis<sup>23,24</sup>. Modulation of OGT using siRNA is complicated by the long half-life of the protein and often produces only partial knockdown of OGT<sup>25</sup>. Although deletion of OGA in *C. elegans* leads to metabolic changes and increased dauer formation<sup>26</sup>, no mammalian knockout of OGA has yet been reported. These features have challenged efforts to study *O*-GlcNAc glycosylation and have limited the use of conventional genetic tools to elucidate its role in cellular processes.

A complementary approach involves the generation of pharmacological agents to inhibit OGT and OGA. A well-established inhibitor of OGT, alloxan, exhibits multiple non-specific effects, such as inhibition of OGA<sup>27</sup>, glucokinase inhibition<sup>28</sup>, and formation of superoxide radicals<sup>29</sup>. By screening a 64,416-member library of compounds using a novel fluorescence-based assay, Gross and coworkers identified several promising new compounds that inhibit OGT activity *in vitro* (**Fig. 4a**, Compounds **1–3**)<sup>30</sup>. These compounds were selective in inhibiting OGT but not another related enzyme MurG, which also uses UDP-GlcNAc as a substrate. Recently, a benzoxazolinone compound

(Fig. 4a, Compound 1) was shown to inhibit OGT activity in oocytes where it prevents meiotic progression<sup>31</sup>. In the future, these numerous, chemically-distinct OGT inhibitors might be used in parallel to distinguish the potential non-specific effects of individual inhibitors *in vivo*.



**Figure 4:** Small molecule inhibitors of OGT and OGA. (a) OGT inhibitors (1–3) identified by screening a 64,416-member library of compounds. The benzoxazinone compound 1 inhibits OGT activity in oocytes and prevents meiotic progression. (b) Representative OGA inhibitors (4–9) with enhanced selectivity for OGA over  $\beta$ -hexosaminidase

As with OGT, many of the early OGA inhibitors such as PUGNAc exhibit non-specific activity toward  $\beta$ -hexosaminidase<sup>32</sup> and only partially inhibit the short isoform of OGA<sup>33</sup>. Recently, several labs have designed more selective OGA inhibitors that are either structural variants of early hexosaminidase inhibitors and/or have been rationally designed using information about the OGA active site. For instance, Macauley et al.

and Knapp et al. functionalized the non-specific hexosaminidase inhibitor GlcNAc-thiazoline with longer alkyl chains, fluoro, or azido groups to generate new OGA inhibitors that show greater than 3000-fold selective inhibition of OGA over  $\beta$ -hexosaminidase (**Fig. 4b**, Compounds **4–6**)<sup>34,35</sup>. In other studies, extending the *N*-acyl group of PUGNAc led to the creation of new inhibitors with greater than 10-fold selectivity for OGA over  $\beta$ -hexosaminidase<sup>32,36</sup>. A drawback of these compounds is that they inhibit OGA more weakly than PUGNAc. In response to this problem, Dorfmeuller and coworkers designed a nagstatin derivative based on the crystal structure of NagJ, a bacterial homologue of human OGA (**Fig. 4b**, Compound **7**)<sup>37</sup>. The isobutanamido group at the N8 position provides improved selectivity by fitting into a pocket that is larger than the corresponding pockets in other hexosaminidases, while the phenethyl group at the C2 position offered stronger inhibition ( $K_i = 4.6 \pm 0.1$  pM) for the bacterial homologue of OGA by interacting with a solvent-exposed tryptophan residue. Nonetheless, a related nagstatin derivative still showed weaker inhibition toward human OGA than PUGNAc<sup>38</sup>. More recently, Kim and coworkers designed a compound that strongly inhibits the short isoform of OGA (**Fig. 4**, Compound **8**)<sup>33</sup>, an isoform that is only partially inhibited by 1 mM PUGNAc. However, selectivity may be an issue with this compound as thiosulphonate moieties have been shown to react with exposed cysteine residues of proteins<sup>39</sup>. Notably, compounds **4** and **7** have been tested in cell culture and were shown to increase overall cellular *O*-GlcNAc levels although none of these compounds were shown to have efficacy *in vivo*. Using a mechanism-inspired approach, Yuzwa and coworkers developed thiamet-G a nanomolar inhibitor ( $K_i = 21 \pm 3$  nM) of OGA with low or no activity towards human lysosomal  $\beta$ -hexosaminidase and other glycoside

hydrolases and, important, showed that it increases *O*-GlcNAc levels in the brain *in vivo* (**Fig. 4b**, Compound **9**)<sup>40</sup>.

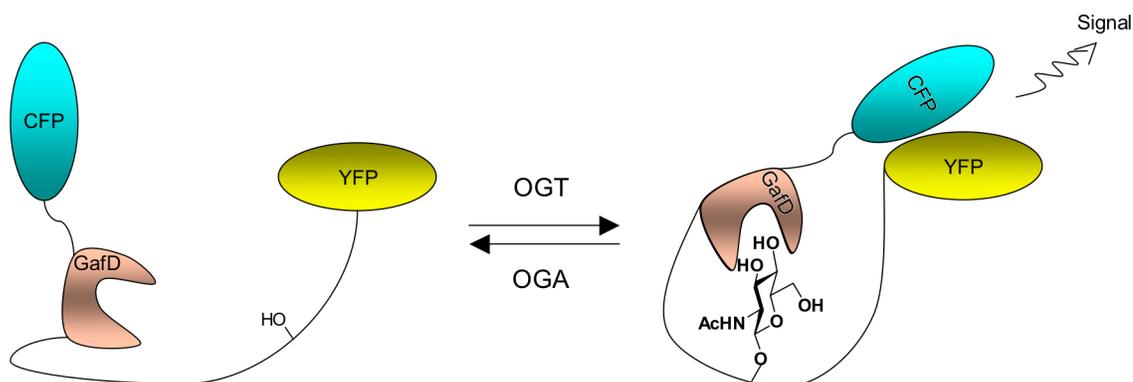
The discovery of potent, selective inhibitors of OGT and OGA provides powerful tools for perturbing *O*-GlcNAc glycosylation in cells and *in vivo*. Application of these inhibitors in specific contexts should reveal new insights into the functional roles of *O*-GlcNAc and cellular mechanisms for the regulation of OGT and OGA.

### **Monitoring *O*-GlcNAc dynamics**

The dynamic nature of *O*-GlcNAc is a unique characteristic that distinguishes it from other forms of glycosylation. As described earlier, this feature has important implications for the regulation of protein structure and function and the interplay with other PTMs. An exciting challenge in the future will be to understand the cellular dynamics of the modification, as well as the signaling pathways and mechanisms by which *O*-GlcNAc is regulated on specific proteins. Toward this end, several groups have developed chemical approaches to monitor changes in *O*-GlcNAc glycosylation levels in response to cellular stimuli.

Carrillo and coworkers designed a FRET-based sensor for the detection of *O*-GlcNAc dynamics in living cells<sup>41</sup>. Their sensor consisted of two fluorophores (enhanced cyan and yellow fluorescent protein) separated by a known peptide substrate for OGT and an *O*-GlcNAc lectin (GafD). Upon *O*-GlcNAc glycosylation, the lectin was expected to bind the glycosylated peptide substrate and bring the two fluorophores in close proximity, producing a stronger FRET signal (**Fig. 5**). As anticipated, an increase in FRET (7–30%) was observed upon stimulation of transfected HeLa cells with

PUGNAc and glucosamine. In the future, it will be interesting to examine changes in OGT activity in response to a variety of cellular stimuli.



**Figure 5:** Chemical tools for monitoring O-GlcNAc dynamics. A FRET-based sensor designed to detect the dynamics of *O*-GlcNAc glycosylation in living cells. Upon glycosylation, binding of the GafD lectin to the *O*-GlcNAc moiety induces a conformational change and produces a stronger FRET signal. CFP, enhanced cyan fluorescent protein; YFP, yellow fluorescent protein

In addition to monitoring OGT activity, identifying the intracellular signaling pathways and dynamics of *O*-GlcNAc glycosylation on specific protein substrates represents an important, challenging goal. General *O*-GlcNAc antibodies (CTD110.6 and RL-2) have been extremely valuable for measuring global changes in glycosylation in response to cellular stimuli<sup>5,6</sup>. However, a limitation of these antibodies is that they detect only a small subset of the *O*-GlcNAc-modified proteins<sup>8,42</sup>. Moreover, it remains difficult to identify the specific proteins that undergo changes in glycosylation. In response to these challenges, Khidekel and coworkers have developed a method to probe the dynamics of *O*-GlcNAc glycosylation *in vivo* using quantitative proteomics (discussed in more detail next chapter). More recently, Wang and coworkers used immunoaffinity chromatography or chemoenzymatic labeling in conjunction with SILAC (stable isotope labeling with amino acids in cell culture), a well-established method for

quantitative proteomics<sup>43</sup>, to study the dynamic interplay between *O*-GlcNAc and phosphorylation by glycogen synthase kinase-3 (GSK-3)<sup>15</sup> or cytokinesis<sup>19</sup>. In the GSK-3 study, heavy and light isotope-labeled cells were treated with LiCl to inhibit GSK-3, and *O*-GlcNAc proteins of interest were isolated by affinity chromatography using a general *O*-GlcNAc antibody. Forty-five putative *O*-GlcNAc glycosylated proteins were identified by mass spectrometry, ten of which showed enrichment after LiCl treatment, suggesting that these proteins underwent increases in *O*-GlcNAc glycosylation. In four of the cases, the glycosylation increases that were observed indirectly by mass spectrometry were confirmed by immunoprecipitation. Interestingly, other proteins exhibited no change or decreases in glycosylation, suggesting that a complex interplay likely exists between phosphorylation and *O*-GlcNAc glycosylation within signaling networks.

With these new tools, in-depth studies of the dynamics of *O*-GlcNAc within cells are now within reach. Understanding the molecular mechanisms by which this dynamic signaling comes about and regulates specific proteins is a future challenge that promises to propel the field in exciting new directions.

### **Conclusion and Future Challenges**

Many of the new discoveries in the field of *O*-GlcNAc have been accelerated by the development of new chemical tools, such as those for detecting the modification in cells and for inhibiting OGT and OGA. Combined with advances in mass spectrometry, these new technologies have provided an unprecedented opportunity to define the *O*-

GlcNAc proteome, manipulate *O*-GlcNAc enzymes, and explore the dynamics of this modification *in vivo*.

The development of new chemical tools to produce homogeneously glycosylated proteins will represent an important step toward this goal. Unlike phosphorylation, *O*-GlcNAc glycosylation cannot be readily mimicked by any naturally occurring amino acid. The current state-of-the-art involves alanine mutagenesis of specific glycosylation sites within proteins to effect changes in function. Chemical methods such as native chemical ligation<sup>44</sup> may allow for the construction of *O*-GlcNAc glycosylated proteins *in vitro* and in cells. In the future, these methods may provide new insights into whether and how *O*-GlcNAc glycosylation affects protein structure, modulates protein-protein interactions, and influences other post-translational modifications.

One of the central challenges of neuroscience is to understand the unique molecular and cellular heterogeneity of the brain as it relates to systems level phenomena, such as learning and memory. Sensitive methods to detect the modification on small subpopulations of cells or proteins will be required to dissect the role of *O*-GlcNAc in fear, addiction, and other complex learning and memory models. Despite significant progress, faster, higher-throughput methods are still needed to identify *O*-GlcNAc proteins and study *O*-GlcNAc dynamics *in vivo*. For instance, the ability to directly monitor the glycosylation status of specific proteins using chemical tagging approaches or site-specific *O*-GlcNAc antibodies will be essential. To facilitate the production of *O*-GlcNAc antibodies, facile synthetic routes to access *O*-GlcNAc-modified peptides are needed. Moreover, the continued development of methods to precisely map

glycosylation sites, particularly on small quantities of material and on selected proteins of interest, will be critical for any functional studies.

New sensitive and selective OGT and OGA inhibitors will be important tools for finely dissecting the role of each enzyme in neuronal function and dysfunction. Given the diversity of OGT and OGA substrates and the lethality of deleting the OGT gene in mice, creative new genetic or chemical approaches are still needed to more selectively target functional subsets of OGT and OGA by interfering, for instance, with the enzymes in certain subcellular compartments.

From the time of its discovery, the appeal of *O*-GlcNAc has been both the intrigue of understanding its unique biology and the great technical challenges associated with its study. Over the last five years, we have seen a surge of new chemistry designed to meet these obstacles. Strengthened by an arsenal of chemical tools, the future of *O*-GlcNAc is primed for new and exciting discoveries.

## References

1. Torres, C.R. & Hart, G.W. 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).
2. Hart, G.W., Housley, M.P. & Slawson, C. Cycling of *O*-linked beta-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-22 (2007).
3. Love, D.C. & Hanover, J.A. The hexosamine signaling pathway: deciphering the "*O*-GlcNAc code". *Sci STKE* **2005**, re13 (2005).
4. Cieniewski-Bernard, C. et al. Identification of *O*-linked *N*-acetylglucosamine proteins in rat skeletal muscle using two-dimensional gel electrophoresis and mass spectrometry. *Mol. Cell. Proteomics* **3**, 577-85 (2004).
5. 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. USA* **97**, 2820-5 (2000).
6. 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).

7. Roquemore, E.P., Chou, T.Y. & Hart, G.W. Detection of *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) on cytoplasmic and nuclear proteins. *Methods Enzymol.* **230**, 443-60 (1994).
8. Khidekel, N. et al. A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J. Am. Chem. Soc.* **125**, 16162-3 (2003).
9. Ramakrishnan, B. & Qasba, P.K. Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient *N*-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity. *J. Biol. Chem.* **277**, 20833-9 (2002).
10. Tai, H.C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel identification of *O*-GlcNAc-modified proteins from cell lysates. *J. Am. Chem. Soc.* **126**, 10500-1 (2004).
11. Rexach, J.E. et al. Quantitation of *O*-Glycosylation Stoichiometry and Dynamics using Resolvable Mass Tags. *Nat. Chem. Biol.* **In press**(2010).
12. Vocadlo, D.J., Hang, H.C., Kim, E.J., Hanover, J.A. & Bertozzi, C.R. A chemical approach for identifying *O*-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci. USA* **100**, 9116-21 (2003).
13. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. USA* **101**, 13132-7 (2004).
14. Wells, L. et al. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **1**, 791-804 (2002).
15. 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**, 1365-79 (2007).
16. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* **101**, 9528-33 (2004).
17. 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).
18. 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**, 153-60 (2010).
19. Wang, Z. et al. Extensive crosstalk between *O*-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal* **3**, ra2 (2010).
20. 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. USA* **106**, 8894-9 (2009).
21. Nandi, A. et al. Global identification of *O*-GlcNAc-modified proteins. *Anal. Chem.* **78**, 452-8 (2006).
22. Teo, C.F. et al. Glycopeptide-specific monoclonal antibodies suggest new roles for *O*-GlcNAc. *Nat. Chem. Biol.* **6**, 338-43 (2010).

23. O'Donnell, N., Zachara, N.E., Hart, G.W. & Marth, J.D. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol. Cell. Biol.* **24**, 1680-90 (2004).
24. 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. USA* **97**, 5735-9 (2000).
25. Dauphinee, S.M., Ma, M. & Too, C.K. Role of *O*-linked beta-*N*-acetylglucosamine modification in the subcellular distribution of alpha4 phosphoprotein and Sp1 in rat lymphoma cells. *J. Cell. Biochem.* **96**, 579-88 (2005).
26. 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. USA* **103**, 11952-7 (2006).
27. Lee, T.N., Alborn, W.E., Knierman, M.D. & Konrad, R.J. Alloxan is an inhibitor of *O*-GlcNAc-selective *N*-acetyl-beta-D-glucosaminidase. *Biochem. Biophys. Res. Commun.* **350**, 1038-43 (2006).
28. Meglasson, M.D., Burch, P.T., Berner, D.K., Najafi, H. & Matschinsky, F.M. Identification of glucokinase as an alloxan-sensitive glucose sensor of the pancreatic beta-cell. *Diabetes* **35**, 1163-73 (1986).
29. Szkudelski, T. The mechanism of alloxan and streptozotocin action in B cells of the rat pancreas. *Physiol. Res.* **50**, 537-46 (2001).
30. Gross, B.J., Kraybill, B.C. & Walker, S. Discovery of *O*-GlcNAc transferase inhibitors. *J. Am. Chem. Soc.* **127**, 14588-9 (2005).
31. Dehennaut, V. et al. *O*-linked *N*-acetylglucosaminyltransferase inhibition prevents G2/M transition in *Xenopus laevis* oocytes. *J. Biol. Chem.* **282**, 12527-36 (2007).
32. Kim, E.J., Perreira, M., Thomas, C.J. & Hanover, J.A. An *O*-GlcNAcase-specific inhibitor and substrate engineered by the extension of the *N*-acetyl moiety. *J. Am. Chem. Soc.* **128**, 4234-5 (2006).
33. Kim, E.J. et al. Distinctive inhibition of *O*-GlcNAcase isoforms by an alpha-GlcNAc thiolsulfonate. *J. Am. Chem. Soc.* **129**, 14854-5 (2007).
34. 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).
35. Knapp, S. et al. Tautomeric modification of GlcNAc-thiazoline. *Org. Lett.* **9**, 2321-4 (2007).
36. Stubbs, K.A., Zhang, N. & Vocadlo, D.J. A divergent synthesis of 2-acyl derivatives of PUGNAc yields selective inhibitors of *O*-GlcNAcase. *Org. Biomol. Chem.* **4**, 839-45 (2006).
37. Dorfmüller, 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).
38. Shanmugasundaram, B. et al. Inhibition of *O*-GlcNAcase by a gluco-configured nagstatin and a PUGNAc-imidazole hybrid inhibitor. *Chem. Commun. (Camb.)* **13**, 4372-4 (2006).

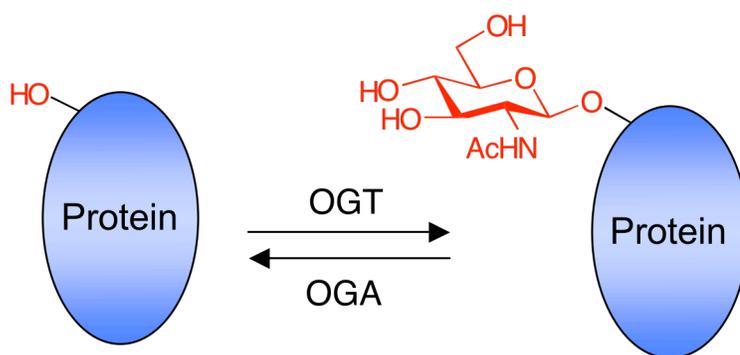
39. Davis, B.G., Lloyd, R.C. & Jones, J.B. Controlled site-selective glycosylation of proteins by a combined site-directed mutagenesis and chemical modification approach. *J. Org. Chem.* **63**, 9614-9615 (1998).
40. Yuzwa, S.A. et al. A potent mechanism-inspired *O*-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. *Nat. Chem. Biol.* **4**, 483-90 (2008).
41. Carrillo, L.D., Krishnamoorthy, L. & Mahal, L.K. A cellular FRET-based sensor for beta-*O*-GlcNAc, a dynamic carbohydrate modification involved in signaling. *J. Am. Chem. Soc.* **128**, 14768-9 (2006).
42. Khidekel, N. et al. Probing the dynamics of *O*-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **3**, 339-48 (2007).
43. Ong, S.E., Mittler, G. & Mann, M. Identifying and quantifying *in vivo* methylation sites by heavy methyl SILAC. *Nat. Methods* **1**, 119-26 (2004).
44. Shin, Y. et al. Fmoc-based synthesis of peptide-thioesters: application to the total chemical synthesis of a glycoprotein by native chemical ligation. *J. Am. Chem. Soc.* **121**, 11684-11689 (1999).

## **Chapter 2: The Roles of *O*-GlcNAc Glycosylation in the Brain**

Portions of this chapter are from Rexach, J.E., Clark, P.M. & Hsieh-Wilson, L.C.  
Chemical approaches to understanding *O*-GlcNAc glycosylation in the brain. *Nat. Chem. Biol.* **4**, 97-106 (2008).

***O*-GlcNAc glycosylation is a unique, dynamic form of glycosylation found on intracellular proteins of all multicellular organisms. Studies suggest that *O*-GlcNAc represents a key regulatory modification in the brain, contributing to transcriptional regulation, neuronal communication, and neurodegenerative disease. Here, we highlight some of the emerging roles for *O*-GlcNAc in the nervous system and describe the challenges in understanding and studying the biology behind *O*-GlcNAc.**

*O*-GlcNAc glycosylation, the covalent attachment of  $\beta$ -*N*-acetyl-D-glucosamine to serine or threonine residues of proteins, is an unusual form of protein glycosylation (Fig. 1)<sup>1</sup>. Unlike other types of glycosylation, this single sugar modification occurs on intracellular proteins and is not elaborated further into complex glycans. The *O*-GlcNAc transferase (OGT) enzyme is a soluble protein that is found in the cytosol, nucleus, and mitochondria<sup>2</sup> rather than in the endoplasmic reticulum or Golgi. The dynamics of *O*-GlcNAc are also unique among sugar modifications, being cycled on a time scale shorter



**Figure 1:** *O*-GlcNAc glycosylation is the addition of  $\beta$ -*N*-acetylglucosamine to serine or threonine residues of proteins

than protein turnover<sup>3</sup>.

Thus, in many respects *O*-GlcNAc is more akin to phosphorylation than to conventional forms of glycosylation.

Several reviews have described the roles of *O*-

GlcNAc in cellular processes, such as transcription<sup>2,4</sup>, the stress response<sup>5,6</sup>, apoptosis<sup>7,8</sup>, signal transduction<sup>2,9</sup>, glucose-sensing<sup>5,10</sup>, and proteasomal degradation<sup>5</sup>. Only a few reviews have highlighted the importance of *O*-GlcNAc glycosylation in the nervous system, and those reports have focused on its potential impact on neurodegenerative diseases<sup>11,12</sup>. However, multiple lines of evidence suggest that *O*-GlcNAc plays critical roles in both neuronal function and dysfunction. The enzymes responsible for the modification are most highly expressed in the brain<sup>13,14</sup> and are enriched at neuronal synapses<sup>15,16</sup>. Neuron-specific deletion of the OGT gene in mice leads to locomotor defects and neuronal dysfunction, resulting in neonatal death<sup>17</sup>. The *O*-GlcNAc modification is abundant in the brain and present on many proteins important for transcription, neuronal signaling, and synaptic plasticity, such as cAMP-responsive element binding protein (CREB)<sup>18</sup>, synucleins<sup>19</sup>, and  $\beta$ -amyloid precursor protein (APP)<sup>20</sup>. An intriguing interplay between *O*-GlcNAc and phosphorylation has been observed in cerebellar neurons, wherein activation of certain kinase pathways reduced *O*-GlcNAc levels on cytoskeletal-associated proteins<sup>21</sup>. Finally, recent studies suggest that *O*-GlcNAc can modulate calcium signaling and affect long-term potentiation<sup>22,23</sup>.

Here we will describe emerging functions for *O*-GlcNAc glycosylation in the nervous system.

### **The enzymes OGT and OGA**

OGT and  $\beta$ -*N*-acetylglucosaminidase (OGA or *O*-GlcNAcase) catalyze the reversible addition and removal of *O*-GlcNAc, respectively. Both enzymes are most highly expressed in the brain and exist as multiple different isoforms<sup>2,24</sup>. Three distinct

isoforms of OGT have been identified, including a 110-kDa and 78-kDa isoform that can assemble into a multimer<sup>25,26</sup>, and a smaller mitochondrial isoform. Each isoform contains the C-terminal catalytic domain, but differs in the number of tetratricopeptide repeats (TPRs) within its N-terminal domain. The TPRs serve as protein-protein interaction modules that appear to target OGT to accessory proteins and potential substrates, such as the GABA<sub>A</sub> receptor interacting factor-1 (GRIF-1)<sup>27</sup> and the related *O*-GlcNAc transferase interacting protein (OIP106)<sup>27</sup>, which have been implicated in mitochondrial trafficking to synapses<sup>28,29</sup>, and the transcriptional repressor complex mSin3A-histone deacetylase 1 (HDAC1)<sup>30</sup>. In addition, OGT forms a complex with protein phosphatase-1 (PP1) in the brain<sup>31</sup>. The association between OGT and PP1 is particularly intriguing as it may provide a direct mechanism to couple *O*-GlcNAc glycosylation to dephosphorylation of specific substrates. Although OGT is found in the nucleus, cytosol, and mitochondria, it is particularly enriched in the nucleus<sup>15</sup> and the soluble synaptic compartment<sup>16</sup>.

Like OGT, OGA appears to be highly active at neuronal synapses<sup>16</sup>, and it is also found in the nucleus and cytosol<sup>32</sup>. OGA contains an N-terminal glycosidase domain and a putative C-terminal histone acetyltransferase (HAT) domain<sup>33</sup>. Two distinct isoforms of OGA exist, a 130-kDa and 75-kDa variant, which share the same catalytic domain but differ in their C-terminus<sup>34</sup>. The potential HAT activity of OGA may provide an intriguing mechanism for coupling deglycosylation of nuclear proteins to transcriptional activation. As with OGT, OGA has been shown to interact with specific proteins, including calcineurin/protein phosphatase-2B, amphiphysin, and dihydropyrimidinase-related protein 2 (DRP-2)<sup>32</sup>.

## Transcriptional regulation

Early studies revealed that the *O*-GlcNAc modification is enriched on chromatin<sup>35</sup> and is found on RNA polymerase II and a large number of its transcription factors<sup>36</sup>. As described in several reviews<sup>2,4,37</sup>, *O*-GlcNAc glycosylation has been shown both to enhance and suppress the activity of transcription factors. *O*-GlcNAc can function to disrupt protein-protein interactions, as in the case of Sp1, whose glycosylation represses transcription at Sp1-driven promoters<sup>38 39</sup>. In other cases, it can promote protein-protein interactions, as in the case of STAT5A, whose glycosylation enhances its activity by recruiting the transcriptional coactivator CREB-binding protein (CBP)<sup>40</sup>. *O*-GlcNAc may also play a more general role in transcriptional repression through a mechanism involving the targeting of OGT to an HDAC1 complex by the corepressor mSin3A<sup>30</sup>. In addition to altering protein-protein interactions, *O*-GlcNAc can affect posttranslational modifications. For instance, glycosylation stabilized the tumor suppressor protein p53 by decreasing its phosphorylation and subsequent degradation by the proteasome<sup>41</sup>.

Much less information is known about the roles of *O*-GlcNAc in regulating transcription in the brain. However, CREB, a transcription factor important for neuronal survival, long-term memory storage, and drug addiction<sup>42,43</sup>, was shown to be *O*-GlcNAc glycosylated in the rodent brain<sup>18</sup>. Glycosylation occurred at two major sites within the Q2 transactivation domain of CREB and disrupted binding of CREB to TAF<sub>II</sub>130, a component of the basal transcriptional machinery. As a result, glycosylation repressed the transcription of CRE-mediated genes both *in vitro* and in cells<sup>18</sup>. It will be interesting to investigate whether glycosylation of CREB is dynamically regulated in neurons and

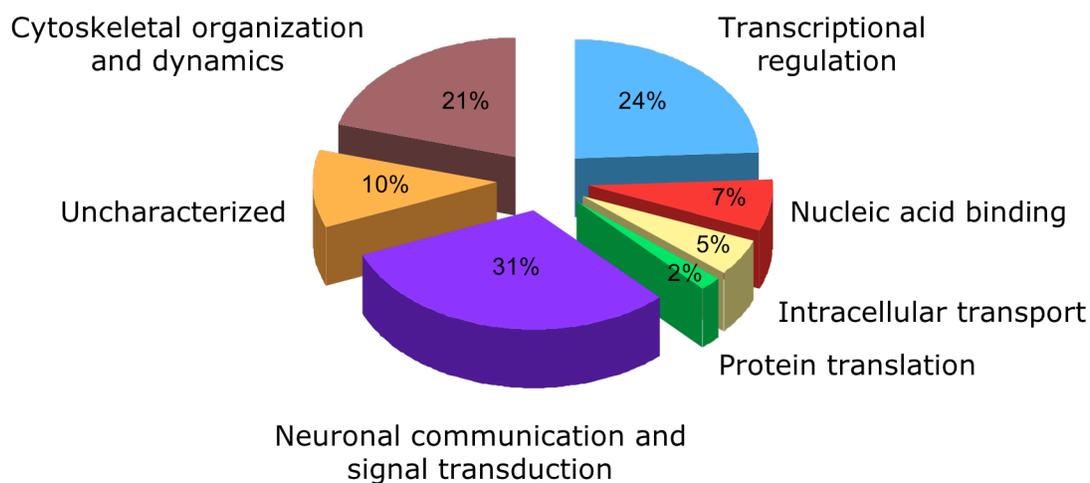
whether it down-regulates specific genes associated with memory storage and cell survival.

Proteomic studies of *O*-GlcNAc modified-proteins from the brain have also underscored the importance of *O*-GlcNAc in regulating transcription. Approximately one-quarter of the neuronal *O*-GlcNAc proteins known to date are transcriptional regulatory proteins (**Fig. 2**)<sup>44</sup>. This includes numerous transcription factors (e.g., Sox2, ATF-2), as well as transcriptional coactivators (SRC-1), repressors (MeCP2<sup>19</sup>, p66 $\beta$ , BHC80) and corepressors (TLE-4, CCR4-NOT). For instance, Sox2 is a member of the high mobility group box (HMG) superfamily of minor groove DNA-binding proteins, and it functions to regulate transcription on different promoters depending on its interactions with different protein partners<sup>45</sup>. Sox2 interacts with proteins through its highly conserved HMG DNA-binding domain, which also contains its *O*-GlcNAc modification site<sup>44</sup>. One of the well-established functions of Sox2 is its critical role in the maintenance of embryonic stem cell pluripotency in partnership with OCT3/4<sup>46</sup>. In the adult rat brain, Sox2 expression has been reported to occur in actively dividing adult neuronal precursor cells and in neurogenic astrogliia<sup>47</sup>. Another example of the expanding role of *O*-GlcNAc in transcription is the modification of two proteins (including a ubiquitin ligase) in carbon catabolite repression 4-negative on TATA-less (CCR4-NOT)<sup>44</sup>, a large protein complex involved in mRNA metabolism and the global control of gene expression<sup>48</sup>. Together with earlier studies demonstrating glycosylation of RNA polymerase II, these findings suggest that *O*-GlcNAc may participate in regulating multiple aspects of transcription.

## Synaptic proteins and neuronal communication

Consistent with the observation that OGT and OGA are highly active at synapses, proteomic studies have uncovered a significant number of synaptic proteins in the *O*-GlcNAc proteome (**Fig. 2**)<sup>44,49-51</sup>. Many of these proteins are enriched in the postsynaptic density where they participate in the regulation of dendritic spine morphology and associate with the cytoskeleton. For instance, synaptopodin<sup>44</sup>, SH3 and multiple ankyrin repeat domains protein 2 (shank2)<sup>49</sup> are critical for the normal formation of dendritic spine apparatuses<sup>52-54</sup>. Synaptopodin and  $\delta$ -catenin have been shown to play important roles in learning and memory<sup>52,55</sup>.

*O*-GlcNAc modifications are also highly abundant in presynaptic terminals. Several proteins involved in neurotransmitter release or synaptic vesicle endocytosis, such as bassoon<sup>44</sup>, piccolo<sup>49</sup>, synapsin<sup>49</sup>, and clathrin assembly protein (AP180)<sup>56</sup>, are *O*-



**Figure 2:** *O*-GlcNAc proteome from rodent brain. Approximately 24% of the known *O*-GlcNAc proteins participate in transcriptional regulation, 31% are involved in neuronal communication and signaling, and 21% are associated with forming cytoskeletal structures. Proteins were classified according to categories described by Schoof et al.<sup>119</sup>

GlcNAc glycosylated. The *O*-GlcNAc-modified protein, collapsin response mediator protein 2 (CRMP-2)<sup>16</sup>, plays key roles in axon formation, elongation, and branching<sup>57</sup>. Moreover, many cytoskeletal proteins themselves are known to be glycosylated, including tau<sup>58</sup>, the neurofilament proteins NF-H<sup>59</sup>, NF-L<sup>49</sup> and NF-M<sup>49</sup>, and the microtubule-associated proteins MAP1B<sup>44</sup> and 2B<sup>44</sup>.

Recently *O*-GlcNAc has been shown to regulate and modify processes important for neuronal communication. Inducing glycosylation by inhibiting OGA decreases the number of axonal filopodia whereas decreasing glycosylation by overexpressing OGA increases the number of filopodia as well as the percentage of neurons exhibiting axon branching in cultured primary chicken forebrain neurons<sup>60</sup>. Furthermore the *O*-GlcNAc modification on neuronal inositol 1,4,5-trisphosphate receptor type 1 decrease channel activity<sup>23</sup>. Finally elevation or reduction of *O*-GlcNAc levels enhances or blocks long term potentiation in acute hippocampal slices<sup>22</sup>.

Additional functional studies are needed to define the mechanisms by which *O*-GlcNAc regulates these proteins. Nonetheless, the prevalence of *O*-GlcNAc on proteins intricately involved in neurotransmitter release and cytoskeletal rearrangements underlying synaptic plasticity suggests roles for the modification in regulating key neuronal functions. As described below, emerging evidence indicates that *O*-GlcNAc levels can be dynamically modulated in response to neuronal stimuli. Moreover, the potential interplay between *O*-GlcNAc and kinase pathways in neurons may provide a powerful means to control protein function and modulate neuronal communication processes.

## Neurodegenerative disease

*O*-GlcNAc glycosylation has been implicated in several neurodegenerative diseases, such as Alzheimer's<sup>58,61,62</sup> and amyotrophic lateral sclerosis (ALS)<sup>63</sup>. The genes encoding OGA and OGT map to chromosomal regions associated with late-onset Alzheimer's disease<sup>64</sup> and dystonia-Parkinsonism syndrome<sup>65</sup>, respectively. Moreover, *O*-GlcNAc levels are abnormally altered in the brains of Alzheimer's disease patients, although the magnitude and direction of the change appears to depend on the subcellular protein fraction<sup>61,62</sup>.

In the pathology of Alzheimer's disease, the microtubule protein tau becomes hyperphosphorylated, which in turn, causes it to aggregate and form neurofibrillary tangles that are hallmarks of the disease<sup>66</sup>. Tau is extensively *O*-GlcNAc glycosylated in the adult rat brain, although the estimated 12 or more modification sites have yet to be mapped<sup>58</sup>. Importantly, several studies suggest that *O*-GlcNAc glycosylation of tau negatively regulates its ability to be phosphorylated. For instance, inducing tau glycosylation with OGA inhibitors or by overexpression of OGT decreases tau phosphorylation at specific sites<sup>62,67,68</sup>. Conversely, stimulation of hyperphosphorylated tau using the phosphatase inhibitor okadaic acid leads to hypoglycosylated tau in human neuroblastoma cells<sup>69</sup>. Neuron-specific deletion of the OGT gene in mice<sup>17</sup> or inhibition of *O*-GlcNAc biosynthesis in rats<sup>70</sup> induces hyperphosphorylated tau similar to that found in Alzheimer's disease. As impaired glucose uptake/metabolism has been linked to Alzheimer's disease and appears to worsen as the disease progresses<sup>71</sup>, one theory is that tau glycosylation becomes reduced in Alzheimer's patients and leads to hyperphosphorylated tau. Consistent with this view, mouse models of starvation that

mimic this impaired glucose metabolism display reduced tau glycosylation and a corresponding increase in tau phosphorylation at specific sites<sup>62,72</sup>.

Abnormal *O*-GlcNAc glycosylation may also contribute to neurodegenerative diseases in more diverse ways. The amyloid precursor protein (APP), which forms the  $\beta$ -amyloid plaques characteristic of the disease, is both *O*-GlcNAc glycosylated and phosphorylated<sup>20</sup>. In an animal model of ALS, the *O*-GlcNAc levels of neurofilament protein M are decreased at the same time as its phosphorylation levels are increased<sup>63</sup>. Finally, *O*-GlcNAc glycosylation has been demonstrated to inhibit the proteasome<sup>73</sup>, thus providing a mechanism to couple ubiquitin-mediated protein degradation to the general metabolic state of the cell. Blocking the removal of *O*-GlcNAc from the proteasome leads to increased protein ubiquitination<sup>73</sup> and possibly neuronal apoptosis<sup>74</sup>. Proteasomal dysfunction and ubiquitinated inclusion bodies are found in the diseased tissue of ALS, Parkinson's, Huntington's, and Alzheimer's disease patients<sup>75</sup>. Thus, aberrations in glucose metabolism and the *O*-GlcNAc glycosylation of specific proteins have been associated with several neurodegenerative disorders. It will be important in the future to determine the extent to which these changes are critical to the development and progression of such diseases.

### ***O*-GlcNAc dynamics and cycling**

A unique feature of *O*-GlcNAc glycosylation is its ability to undergo dynamic cycling in contrast to other, more static forms of protein glycosylation. Studies have shown that *O*-GlcNAc levels are altered by extracellular stimuli on a time scale similar to phosphorylation. For instance, a transient increase in glycosylation of the transcription

factor nuclear factor activated T-cells (NFAT) was observed within 5 minutes after T or B cell activation<sup>76</sup>.

*O*-GlcNAc levels are highly responsive to glucose concentrations and influx through the hexosamine biosynthesis pathway (HBP) in neurons and other cell types<sup>77,78</sup>. Approximately 2-5% of all cellular glucose is metabolized through the HBP pathway to generate UDP-GlcNAc<sup>79</sup>. As OGT activity is exquisitely sensitive to UDP-GlcNAc concentrations<sup>80</sup>, *O*-GlcNAc glycosylation may act as a sensor for the general metabolic state of the cell. Consistent with this notion, *O*-GlcNAc appears to be intricately linked to cell survival<sup>17</sup> and is induced by many forms of cell stress<sup>81</sup>.

In the brain, phosphorylation serves as a central mechanism for neuronal communication by regulating ion channels, neurotransmitter receptors, gene transcription, and synaptic vesicle release<sup>82,83</sup>. Protein kinases and phosphatases work together to coordinate different forms of synaptic plasticity, and they are necessary for the induction and maintenance of postsynaptic long-term potentiation and long-term depression<sup>84</sup>. Thus, the potential interplay between *O*-GlcNAc glycosylation and phosphorylation has exciting implications for many neuronal functions. Early studies showed that activation of protein kinase C (PKC) or cAMP-dependent protein kinase (PKA) significantly decreased overall *O*-GlcNAc glycosylation levels in the cytoskeletal protein fraction of cultured cerebellar neurons<sup>21</sup>. Conversely, inhibition of PKC, PKA, cyclin-dependent protein kinases or S6 kinase increased overall *O*-GlcNAc levels in these fractions. A more complex relationship was observed with tyrosine kinases and phosphatases. Inhibition of tyrosine phosphatases led to a decrease in overall *O*-GlcNAc levels, while inhibition of tyrosine kinases induced both increases and decreases in *O*-

GlcNAc, depending on the protein fraction. More recent studies showed that elevation of *O*-GlcNAc levels in the brain increased in activating-phosphorylation sites on ERK 1/2 and CaMKII<sup>22</sup> and elevation of *O*-GlcNAc in culture affected the phosphorylation of PKA substrates in response to forskolin<sup>60</sup>.

Together, emerging evidence suggests that *O*-GlcNAc represents a key regulatory modification in the brain. Not only is it present on a large number of functionally important neuronal proteins, it appears to be reversible, differentially regulated, and responsive to neuronal activity. Further studies are needed to elucidate the molecular mechanisms involved and how activation of specific signaling pathways leads to the regulation of OGT and OGA. Moreover, changes in *O*-GlcNAc glycosylation have been monitored only on a global level, and the specific proteins undergoing dynamic changes in glycosylation as well as how those changes affect the protein function remain largely unknown.

### **Conclusion and Future Challenges**

Over the past decade, a surge of discoveries in *O*-GlcNAc glycosylation has revealed new roles for this modification in the nervous system. *O*-GlcNAc is abundant in the brain and present on many diverse proteins involved in transcription, neuronal signaling, and synaptic plasticity. Indeed, recent studies have begun to uncover the functional roles of *O*-GlcNAc, its complex dynamics in the brain, and the interplay between *O*-GlcNAc and phosphorylation.

Although the pace and scope of understanding *O*-GlcNAc has expanded considerably, much still remains to be discovered. Due to the challenge of studying the

modification, evidence linking *O*-GlcNAc to specific biological functions has often been indirect or correlative. This is particularly true in the brain, where the complexity of the nervous system and its unique technical challenges (e.g., post-mitotic cells, multiple cell types, blood-brain barrier, complex organization) render *O*-GlcNAc more difficult to investigate. Nonetheless, in-depth functional studies on proteins will be essential in the future to determine the roles of *O*-GlcNAc in neuronal-specific contexts.

## References

1. Torres, C.R. & Hart, G.W. 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).
2. Love, D.C. & Hanover, J.A. The hexosamine signaling pathway: deciphering the "*O*-GlcNAc code". *Sci. STKE* **312**, re13 (2005).
3. Roquemore, E.P., Chevrier, M.R., Cotter, R.J. & Hart, G.W. Dynamic *O*-GlcNAcylation of the small heat shock protein alpha B-crystallin. *Biochemistry* **35**, 3578-86 (1996).
4. Comer, F.I. & Hart, G.W. *O*-GlcNAc and the control of gene expression. *Biochim. Biophys. Acta.* **1473**, 161-71 (1999).
5. 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).
6. Fulop, N., Marchase, R.B. & Chatham, J.C. Role of protein *O*-linked *N*-acetylglucosamine in mediating cell function and survival in the cardiovascular system. *Cardiovasc. Res.* **73**, 288-97 (2007).
7. Wells, L. & Hart, G.W. *O*-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar. *FEBS Lett.* **546**, 154-8 (2003).
8. Wells, L., Whelan, S.A. & Hart, G.W. *O*-GlcNAc: a regulatory post-translational modification. *Biochem. Biophys. Res. Commun.* **302**, 435-41 (2003).
9. Slawson, C. & Hart, G.W. Dynamic interplay between *O*-GlcNAc and *O*-phosphate: the sweet side of protein regulation. *Curr. Opin. Struct. Biol.* **13**, 631-6 (2003).
10. Wells, L., Vosseller, K. & Hart, G.W. A role for *N*-acetylglucosamine as a nutrient sensor and mediator of insulin resistance. *Cell. Mol. Life Sci.* **60**, 222-8 (2003).
11. Lefebvre, T. et al. Does *O*-GlcNAc play a role in neurodegenerative diseases? *Expert Rev. Proteomics* **2**, 265-75 (2005).

12. Lefebvre, T., Caillet-Boudin, M.L., Buee, L., Delacourte, A. & Michalski, J.C. O-GlcNAc glycosylation and neurological disorders. *Adv. Exp. Med. Biol.* **535**, 189-202 (2003).
13. Kreppel, L.K., Blomberg, M.A. & Hart, G.W. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9308-15 (1997).
14. 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).
15. 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).
16. Cole, R.N. & Hart, G.W. Cytosolic O-glycosylation is abundant in nerve terminals. *J. Neurochem.* **79**, 1080-9 (2001).
17. O'Donnell, N., Zachara, N.E., Hart, G.W. & Marth, J.D. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol. Cell. Biol.* **24**, 1680-90 (2004).
18. Lamarre-Vincent, N. & Hsieh-Wilson, L.C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J. Am. Chem. Soc.* **125**, 6612-3 (2003).
19. 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**, 153-60 (2010).
20. Griffith, L.S., Mathes, M. & Schmitz, B. Beta-amyloid precursor protein is modified with O-linked N-acetylglucosamine. *J. Neurosci. Res.* **41**, 270-8 (1995).
21. Griffith, L.S. & Schmitz, B. O-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to perturbations of phosphorylation. *Eur. J. Biochem.* **262**, 824-31 (1999).
22. Tallent, M.K. et al. In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. *J. Biol. Chem.* **284**, 174-81 (2009).
23. Rengifo, J., Gibson, C.J., Winkler, E., Collin, T. & Ehrlich, B.E. Regulation of the inositol 1,4,5-trisphosphate receptor type I by O-GlcNAc glycosylation. *J. Neurosci.* **27**, 13813-21 (2007).
24. Lazarus, B.D., Love, D.C. & Hanover, J.A. Recombinant O-GlcNAc transferase isoforms: identification of O-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology* **16**, 415-21 (2006).
25. 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).
26. Jinek, M. et al. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat. Struct. Mol. Biol.* **11**, 1001-7 (2004).

27. Iyer, S.P., Akimoto, Y. & Hart, G.W. Identification and cloning of a novel family of coiled-coil domain proteins that interact with *O*-GlcNAc transferase. *J. Biol. Chem.* **278**, 5399-409 (2003).
28. Brickley, K., Smith, M.J., Beck, M. & Stephenson, F.A. GRIF-1 and OIP106, members of a novel gene family of coiled-coil domain proteins: association in vivo and in vitro with kinesin. *J. Biol. Chem.* **280**, 14723-32 (2005).
29. Stowers, R.S., Megeath, L.J., Gorska-Andrzejak, J., Meinertzhagen, I.A. & Schwarz, T.L. Axonal transport of mitochondria to synapses depends on milton, a novel *Drosophila* protein. *Neuron* **36**, 1063-77 (2002).
30. Yang, X., Zhang, F. & Kudlow, J.E. Recruitment of *O*-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein *O*-GlcNAcylation to transcriptional repression. *Cell* **110**, 69-80 (2002).
31. Wells, L., Kreppel, L.K., Comer, F.I., Wadzinski, B.E. & Hart, G.W. *O*-GlcNAc transferase is in a functional complex with protein phosphatase 1 catalytic subunits. *J. Biol. Chem.* **279**, 38466-70 (2004).
32. 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).
33. 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).
34. 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**, 634-40 (2001).
35. Kelly, W.G. & Hart, G.W. Glycosylation of chromosomal proteins: localization of *O*-linked *N*-acetylglucosamine in *Drosophila* chromatin. *Cell* **57**, 243-51 (1989).
36. Jackson, S.P. & Tjian, R. *O*-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**, 125-33 (1988).
37. Zachara, N.E. & Hart, G.W. Cell signaling, the essential role of *O*-GlcNAc! *Biochim. Biophys. Acta* **1761**, 599-617 (2006).
38. Yang, X. et al. *O*-linkage of *N*-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability. *Proc. Natl. Acad. Sci. USA* **98**, 6611-6 (2001).
39. Roos, M.D., Su, K., Baker, J.R. & Kudlow, J.E. *O* glycosylation of an Sp1-derived peptide blocks known Sp1 protein interactions. *Mol. Cell. Biol.* **17**, 6472-80 (1997).
40. Gewinner, C. et al. The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *J. Biol. Chem.* **279**, 3563-72 (2004).
41. 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).
42. Carlezon, W.A., Jr., Duman, R.S. & Nestler, E.J. The many faces of CREB. *Trends Neurosci.* **28**, 436-45 (2005).
43. Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell Biol.* **2**, 599-609 (2001).

44. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. USA* **101**, 13132-7 (2004).
45. Wilson, M. & Koopman, P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr. Opin. Genet. Dev.* **12**, 441-6 (2002).
46. Rodda, D.J. et al. Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731-7 (2005).
47. Komitova, M. & Eriksson, P.S. Sox-2 is expressed by neural progenitors and astroglia in the adult rat brain. *Neurosci. Lett.* **369**, 24-7 (2004).
48. Collart, M.A. Global control of gene expression in yeast by the Ccr4-Not complex. *Gene* **313**, 1-16 (2003).
49. 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).
50. Wells, L. et al. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **1**, 791-804 (2002).
51. 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. USA* **106**, 8894-9 (2009).
52. Deller, T. et al. Synaptopodin-deficient mice lack a spine apparatus and show deficits in synaptic plasticity. *Proc. Natl. Acad. Sci. USA* **100**, 10494-9 (2003).
53. Vazquez, L.E., Chen, H.J., Sokolova, I., Knuesel, I. & Kennedy, M.B. SynGAP regulates spine formation. *J. Neurosci.* **24**, 8862-72 (2004).
54. Sala, C. et al. Regulation of dendritic spine morphology and synaptic function by Shank and Homer. *Neuron* **31**, 115-30 (2001).
55. Israely, I. et al. Deletion of the neuron-specific protein delta-catenin leads to severe cognitive and synaptic dysfunction. *Curr. Biol.* **14**, 1657-63 (2004).
56. Murphy, J.E., Hanover, J.A., Froehlich, M., DuBois, G. & Keen, J.H. Clathrin assembly protein AP-3 is phosphorylated and glycosylated on the 50-kDa structural domain. *J. Biol. Chem.* **269**, 21346-52 (1994).
57. Yoshimura, T. et al. GSK-3beta regulates phosphorylation of CRMP-2 and neuronal polarity. *Cell* **120**, 137-49 (2005).
58. Arnold, C.S. et al. The microtubule-associated protein tau is extensively modified with *O*-linked N-acetylglucosamine. *J. Biol. Chem.* **271**, 28741-4 (1996).
59. Dong, D.L., Xu, Z.S., Hart, G.W. & Cleveland, D.W. Cytoplasmic *O*-GlcNAc modification of the head domain and the KSP repeat motif of the neurofilament protein neurofilament-H. *J. Biol. Chem.* **271**, 20845-52 (1996).
60. Francisco, H. et al. *O*-GlcNAc post-translational modifications regulate the entry of neurons into an axon branching program. *Dev. Neurobiol.* **69**, 162-73 (2009).
61. Griffith, L.S. & Schmitz, B. *O*-linked N-acetylglucosamine is upregulated in Alzheimer brains. *Biochem. Biophys. Res. Commun.* **213**, 424-31 (1995).
62. 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. USA* **101**, 10804-9 (2004).

63. Ludemann, N. et al. *O*-glycosylation of the tail domain of neurofilament protein M in human neurons and in spinal cord tissue of a rat model of amyotrophic lateral sclerosis (ALS). *J. Biol. Chem.* **280**, 31648-58 (2005).
64. Van Tine, B.A., Patterson, A.J. & Kudlow, J.E. Assignment of *N*-acetyl-D-glucosaminidase (Mgea5) to rat chromosome 1q5 by tyramide fluorescence in situ hybridization (T-FISH): synteny between rat, mouse and human with Insulin Degradation Enzyme (IDE). *Cytogenet. Genome Res.* **103**, 202B (2003).
65. Nolte, D. & Muller, U. Human *O*-GlcNAc transferase (OGT): genomic structure, analysis of splice variants, fine mapping in Xq13.1. *Mamm. Genome* **13**, 62-4 (2002).
66. Ballatore, C., Lee, V.M. & Trojanowski, J.Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* **8**, 663-72 (2007).
67. Robertson, L.A., Moya, K.L. & Breen, K.C. The potential role of tau protein *O*-glycosylation in Alzheimer's disease. *J. Alzheimers Dis.* **6**, 489-95 (2004).
68. 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).
69. 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* **1619**, 167-76 (2003).
70. Liu, F. et al. Reduced *O*-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease. *Brain* **132**, 1820-32 (2009).
71. Alexander, G.E., Chen, K., Pietrini, P., Rapoport, S.I. & Reiman, E.M. Longitudinal PET Evaluation of Cerebral Metabolic Decline in Dementia: A Potential Outcome Measure in Alzheimer's Disease Treatment Studies. *Am. J. Psychiatry* **159**, 738-45 (2002).
72. Li, X., Lu, F., Wang, J.Z. & Gong, C.X. Concurrent alterations of *O*-GlcNAcylation and phosphorylation of tau in mouse brains during fasting. *Eur. J. Neurosci.* **23**, 2078-86 (2006).
73. Zhang, F. et al. *O*-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **115**, 715-25 (2003).
74. Liu, K. et al. Accumulation of protein *O*-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high *O*-GlcNAc metabolism. *J. Neurochem.* **89**, 1044-55 (2004).
75. Ciechanover, A. & Brundin, P. The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron* **40**, 427-46 (2003).
76. Golks, A., Tran, T.T., Goetschy, J.F. & Guerini, D. Requirement for *O*-linked *N*-acetylglucosaminyltransferase in lymphocytes activation. *EMBO J.* **20**, 20 (2007).
77. 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. USA* **97**, 2820-5 (2000).
78. Rex-Mathes, M. et al. *O*-GlcNAc expression in developing and ageing mouse brain. *Biochimie* **83**, 583-90 (2001).
79. Marshall, S., Bacote, V. & Traxinger, R.R. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role

- of hexosamine biosynthesis in the induction of insulin resistance. *J. Biol. Chem.* **266**, 4706-12 (1991).
80. 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 beta-*N*-acetylglucosaminyltransferase. *J. Biol. Chem.* **267**, 9005-13 (1992).
  81. 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).
  82. Greengard, P., Valtorta, F., Czernik, A.J. & Benfenati, F. Synaptic vesicle phosphoproteins and regulation of synaptic function. *Science* **259**, 780-5 (1993).
  83. Soderling, T.R. CaM-kinases: modulators of synaptic plasticity. *Curr. Opin. Neurobiol.* **10**, 375-80 (2000).
  84. Colbran, R.J. Protein phosphatases and calcium/calmodulin-dependent protein kinase II-dependent synaptic plasticity. *J. Neurosci.* **24**, 8404-9 (2004).
  85. Schoof, H. et al. MIPS Arabidopsis thaliana Database (MAtdB): an integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res.* **30**, 91-3 (2002).

## **Chapter 3: Probing the Dynamics of *O*-GlcNAc Glycosylation in the Brain Using Quantitative Proteomics**

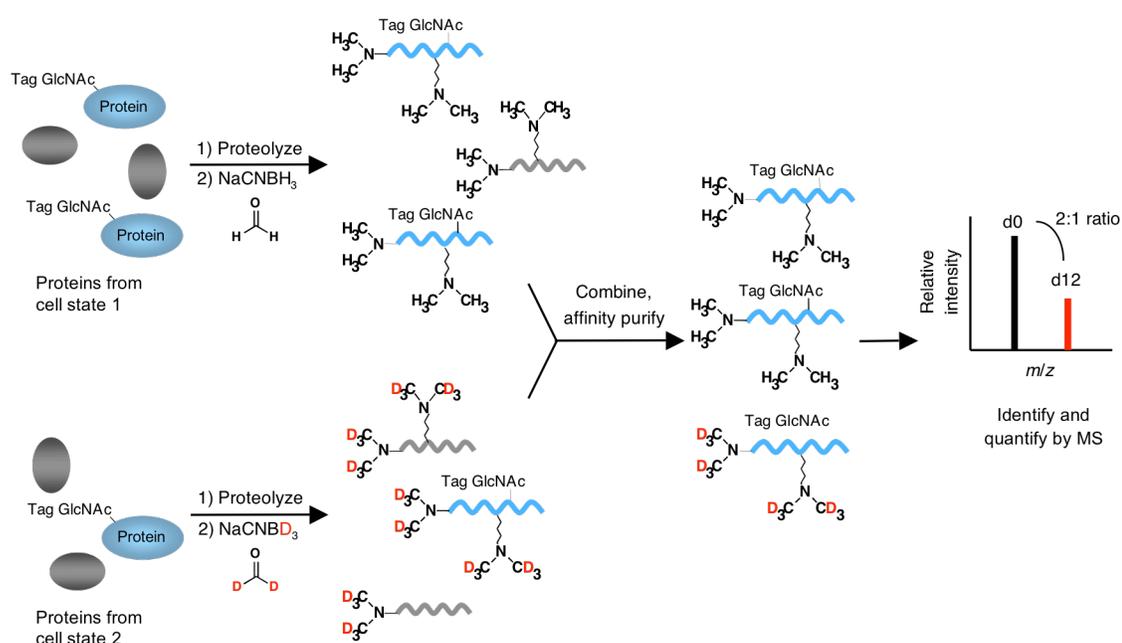
Portions of this chapter are from Khidekel, N., Ficarro, S.B., Clark, P.M., Bryan, M.C., Swaney, D.L., Rexach, J.E., Sun, Y.E., Coon, J.J., Peters, E.C. & Hsieh-Wilson, L.C. Probing the dynamics of *O*-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **3**, 339-48 (2007).

The addition of the monosaccharide  $\beta$ -*N*-acetyl-D-glucosamine to proteins (*O*-GlcNAc glycosylation) is an intracellular, post-translational modification that shares features with phosphorylation. Here, we demonstrate a new strategy for monitoring the dynamics of *O*-GlcNAc glycosylation using quantitative mass spectrometry-based proteomics. Our method, termed QUIC-tag, combines selective, chemoenzymatic tagging of *O*-GlcNAc proteins with an efficient isotopic labeling strategy. A key advantage of the approach is that it can be applied to post-mitotic cells such as neurons after *in vivo* stimulation. Using the method, we detect changes in *O*-GlcNAc glycosylation on several proteins involved in the regulation of transcription and mRNA translocation. We also provide the first evidence that *O*-GlcNAc glycosylation is dynamically modulated by excitatory stimulation of the brain *in vivo*. Finally, we employ electron transfer dissociation (ETD) mass spectrometry to identify exact sites of *O*-GlcNAc modification. Together, our studies suggest that *O*-GlcNAc glycosylation occurs reversibly in neurons and, akin to phosphorylation, may play important roles in mediating the communication between neurons.

### **The QUIC-Tag Strategy for *O*-GlcNAc Peptide Identification and Quantification**

As the majority of peptides from a biological sample are not post-translationally modified, detection of a specific modification by MS requires an enrichment strategy to isolate peptides containing the modification of interest from other species. We reasoned that our chemoenzymatic strategy (**Chapter 1, Fig. 2a**)<sup>1</sup> could be combined with differential isotopic labeling to allow for the first direct, high-throughput quantification of

*O*-GlcNAc dynamics on specific proteins. In this approach, which we have termed Quantitative Isotopic and Chemoenzymatic Tagging (QUIC-Tag), lysates from two cellular states (e.g., stimulated vs. unstimulated, diseased vs. normal) were chemoenzymatically labeled and proteolytically digested (**Scheme 1**). A modified dimethyl labeling strategy<sup>2</sup> incorporated stable isotopes into peptide N-terminal amines and  $\epsilon$ -amino groups of lysine residues by reductive amination for subsequent MS quantification. Treatment with either formaldehyde/NaCNBH<sub>3</sub> or deuterated formaldehyde/NaCNBD<sub>3</sub> created mass differences of 6 x n between the peptides from the two cell populations, where n is the number of primary amine functionalities in the peptide. This allowed for complete resolution of isotopic envelopes even at higher charge states (i.e., +4) during MS analysis. Following isotopic labeling, we combined



**Scheme 1:** QUIC-Tag strategy for quantitative *O*-GlcNAc proteomics. *O*-GlcNAc proteins from two different cell states are selectively tagged, proteolyzed and differentially labeled with ‘light’ or ‘heavy’ isotopes. The mixtures are combined, and *O*-GlcNAc peptides of interest are specifically enriched by avidin chromatography for selective quantification by LC-MS.

and enriched the peptides from both populations by affinity chromatography for the presence of *O*-GlcNAc. Relative quantification of *O*-GlcNAc glycosylation in the two cellular states was accomplished by calculation of the chromatographic peak area as determined by the MS response to each eluting glycosylated pair of peptide ions.

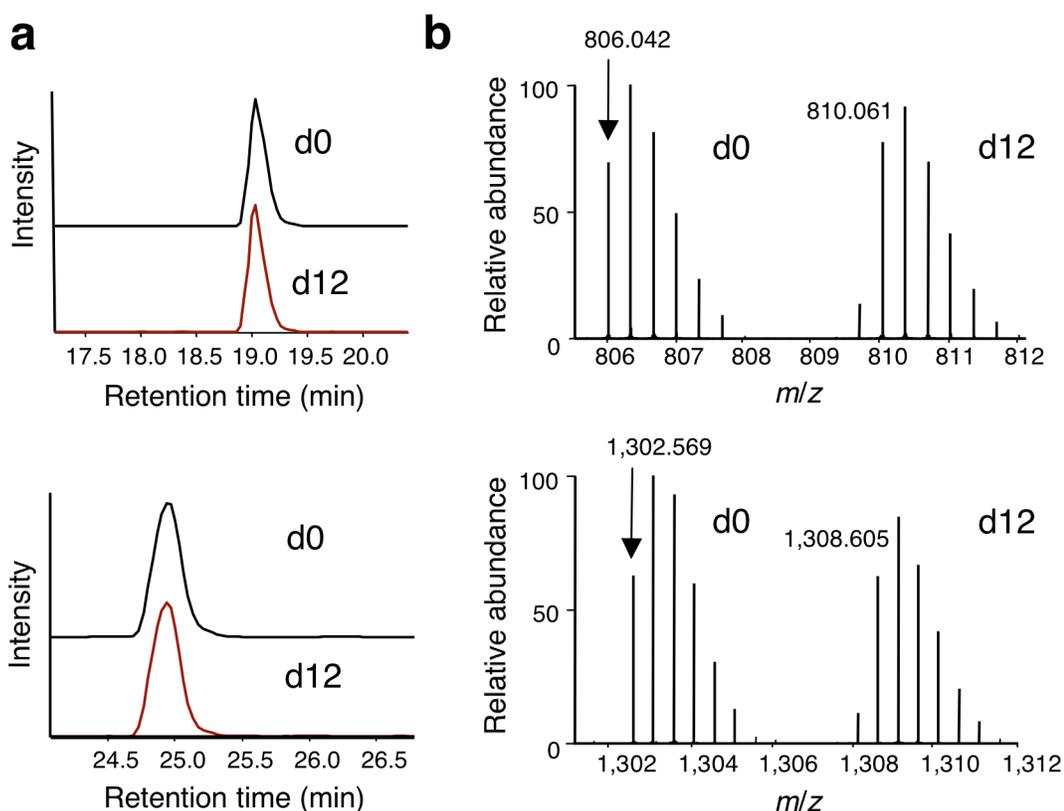
### **Quantification of Known *O*-GlcNAc Peptides from Complex Mixtures**

Nelly Khidekel first evaluated the effectiveness of the dimethyl labeling strategy using the model protein  $\alpha$ -casein.  $\alpha$ -casein was digested with trypsin, and the resulting peptides were reacted with formaldehyde and NaCNBH<sub>3</sub> at pH values ranging from 5-8. Liquid chromatography-mass spectrometry (LC-MS) analysis of the labeled peptides indicated that reductive amination proceeded quantitatively for both lysine and N-terminal primary amines in less than 10 min at pH 7 (data not shown). In contrast to previous studies<sup>2</sup>, we observed that higher pH values were necessary to achieve complete labeling of basic lysine residues.

Having established the optimal conditions for dimethyl labeling, Nelly investigated our ability to capture and quantify known *O*-GlcNAc peptides<sup>3, 4</sup> from complex mixtures. Known amounts of the proteins  $\alpha$ -crystallin (ca. 300 pmol) and OGT (ca. 10 pmol) were added to two samples of rat brain lysate. We chose to examine  $\alpha$ -crystallin because of its low stoichiometry of glycosylation (<10%) and because it has represented a formidable challenge for detection by several methods<sup>1, 5</sup>. The samples were chemoenzymatically labeled, proteolytically digested, isotopically labeled and combined as described in **Scheme 1**. Following avidin capture of the *O*-GlcNAc peptides, Scott Ficarro performed relative quantification of glycosylated peptide pairs

using an orbitrap mass spectrometer<sup>6</sup>, which provided accurate mass (<20 ppm) and high resolution (100,000 at  $m/z$  400) ion measurements. Precursor peptide cations that exhibited the signature loss of the labile ketogalactose-biotin and GlcNAc-ketogalactose-biotin groups during MS/MS were subjected to further fragmentation via MS<sup>4</sup>.

In these experiments, Nelly and Scott reproducibly captured and quantified 3  $\alpha$ -crystallin peptides that encompass all of the known glycosylation sites on both the A and B forms of  $\alpha$ -crystallin<sup>3,7</sup>. Additionally, Nelly captured 8 OGT peptides representing all



**Figure 1:** Accurate quantification of known *O*-GlcNAc peptides from complex mixtures using the QUIC-Tag approach. **(a)** Extracted ion chromatogram of the heavy and light forms of two representative *O*-GlcNAc glycosylated peptides,  $\alpha$ -crystallin peptide <sup>158</sup>AIPVSREEKPSSAPSS<sup>173</sup> (top) and OGT peptide <sup>390</sup>ISPTFADAYSNMGNTLK<sup>406</sup> (bottom). Co-elution by reversed-phase liquid chromatography was observed. **(b)** Quantification from the isotopic cluster of the heavy ( $m/z$  810.061) and light ( $m/z$  806.416) forms of the  $\alpha$ -crystallin peptide yields a heavy:light ratio of 0.97 – 0.09, 0.97 + 0.10 (g.s.d. of 1.10). Quantification of the heavy ( $m/z$  1308.605) and light ( $m/z$  1302.569) forms of the OGT peptide yields a heavy:light ratio of 0.93 – 0.12, 0.93 + 0.14 (g.s.d. of 1.15). Prior to labeling, both proteins were added to neuronal lysates at a ratio of 1:1.  $n = 7$ .

**Table 1a Mean ratios of individual peptides from  $\alpha$ -crystallin and OGT**

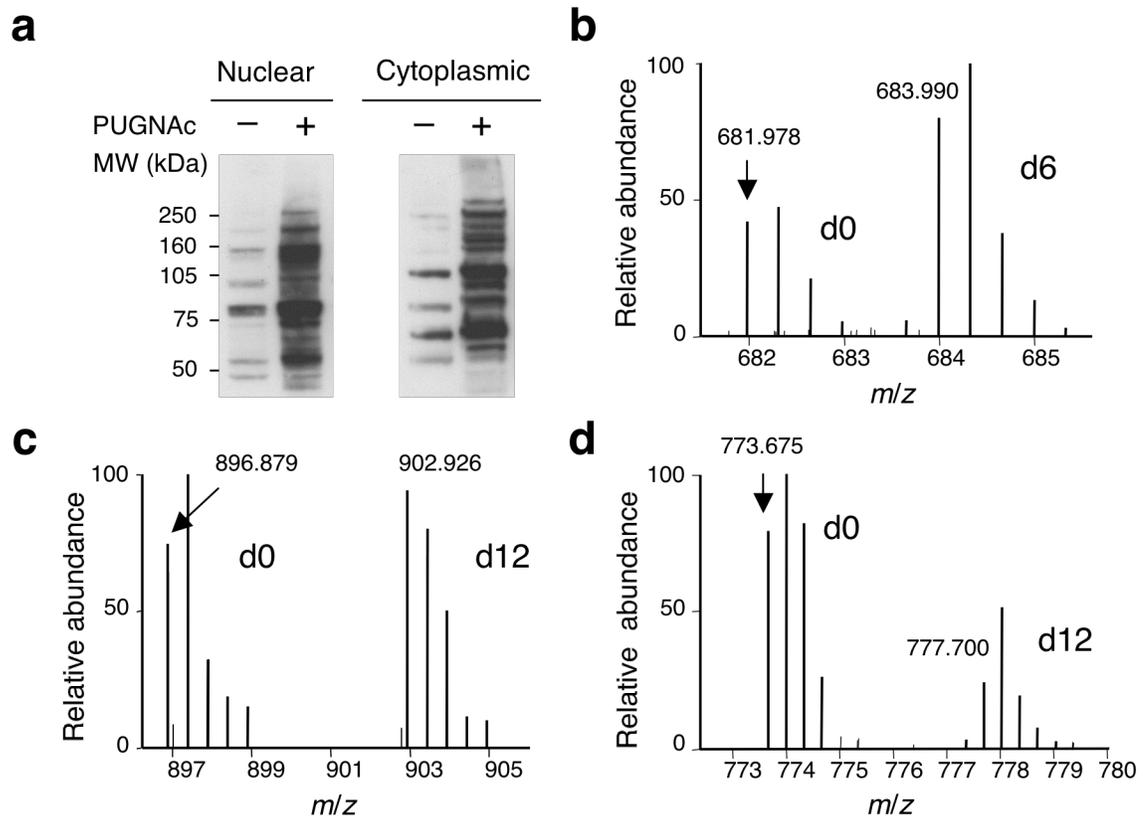
Protein	Peptide Sequence	n	Ratio <sup>a</sup>	s.d. <sup>b</sup>
crystallin 1	AIPVSREEKPSSAPSS	7	0.97 ±	0.10
crystallin 2	AIPVSREEKPSSAPS	7	0.90 ±	0.13
crystallin 3	EEKPVVTAAPK	4	0.81 ±	0.11
OGT 1	IKPVEVTESA	7	0.91 ±	0.33
OGT 2	AIQINPAFADAHSNLAHSLASIHK	7	0.77 ±	0.15
OGT 3	ISPTFADAYSNMGNLTK	7	0.93 ±	0.14
OGT 4	EMQDVQGALQCYTR	5	0.98 ±	0.11
OGT 5	AIQINPAFADAHSNLAHSLASIHKDSGNIPEAIASYSR	4	1.01 ±	0.29
OGT 6	AIQINPAFADAHSNLAHSLASIHKDSGNIPEAIAS	3	0.72 ±	0.15
OGT 7	AATGEEVPRTIIVTTR	7	0.96 ±	0.18
OGT 8	EAIRISPTFADAYSNMGNLTK	2	1.12 ±	0.18

<sup>a</sup> Geometric mean<sup>b</sup> Maximum absolute standard deviation (s.d.) calculated from g.s.d.**Table 1b Mean ratios of all peptides**

Experiment	Ratio <sup>a</sup>	s.d. <sup>b</sup>
1	0.83 ±	0.07
2	0.90 ±	0.28
3	0.80 ±	0.15
4	1.01 ±	0.27
5	0.94 ±	0.17
6	0.96 ±	0.20
7	0.89 ±	0.17

<sup>a</sup> Geometric mean<sup>b</sup> Maximum absolute standard deviation (s.d.) calculated from g.s.d.

of the known glycosylation sites on OGT<sup>4</sup>. The results for two such peptides, <sup>158</sup>AIPVSREEKPSSAPSS<sup>173</sup> from  $\alpha$ -crystallin and <sup>390</sup>ISPTFADAYSNMGNLTK<sup>406</sup> from OGT, are highlighted in **Figure 1**. The deuterated and non-deuterated peptides generally co-eluted during reversed-phase chromatography (**Fig. 1a**), minimizing the isotope resolution effects during LC previously reported to interfere with deuterium-labeled peptides<sup>2,8</sup>. To quantify the relative amounts of each peptide, Nelly compared the ratio of signal intensities from the heavy to the light forms, across the entire chromatographic profile of each peptide (**Fig. 1b**). She observed the  $\alpha$ -crystallin peptide at a mean heavy:light ratio of 0.97 – 0.09, 0.97 + 0.10 (geometric standard deviation (g.s.d) of 1.10) and the OGT peptide at a mean heavy:light ratio of 0.93 – 0.12, 0.93 + 0.14 (g.s.d. of 1.15). The geometric mean ratio and standard deviation obtained for each of the  $\alpha$ -crystallin and OGT peptides is found in **Table 1a**, and the mean ratio of all quantified peptides for each of seven independent experiments is shown in **Table 1b**. The mean ratio across all peptides over the seven experiments was 0.91 – 0.17, 0.91 + 0.21 (g.s.d. of 1.23), which compares favorably with the quantitative accuracy of other approaches such as iTRAQ and SILAC (mean observed ratios of 1.03 ± 0.16 and 1.03 ± 0.17 for an expected 1:1 ratio, respectively)<sup>9, 10</sup>.

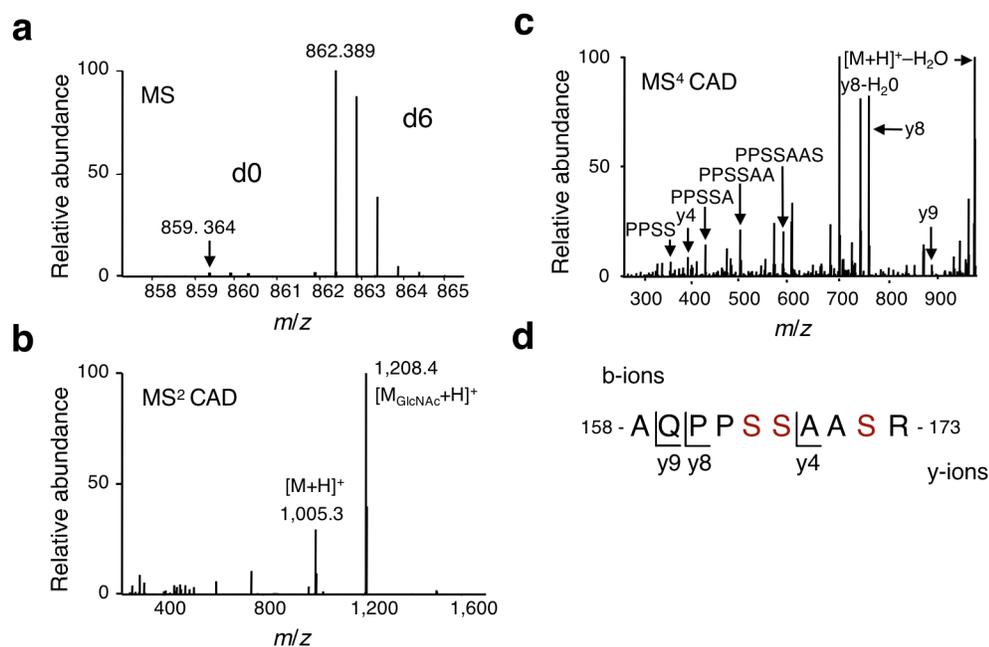


**Figure 2:** *O*-GlcNAc glycosylation is reversible in cultured cortical neurons. (a) Treatment of cortical neurons with the OGA inhibitor PUGNAc for 12 h enhances overall *O*-GlcNAc glycosylation levels in both nuclear and cytoplasmic fractions, as measured by immunoblotting with an anti-*O*-GlcNAc antibody. (b-d) Peptide mass spectra of three proteins displaying distinct activation profiles. *O*-GlcNAc glycosylation of the peptide in b was up-regulated in response to PUGNAc treatment, whereas the glycosylation level was unchanged for the peptide in c and was down-regulated for the peptide in d.

### Probing the Reversibility of *O*-GlcNAc Glycosylation in Neurons using QUIC-Tag

We next applied the approach to study the reversibility of the *O*-GlcNAc modification in neurons. Although studies have suggested that *O*-GlcNAc levels can be modulated in various cell types<sup>11, 12</sup>, the neuronal proteins that undergo reversible glycosylation are largely unknown. Nelly treated cultured cortical neurons from embryonic day-18 rats with the OGA inhibitor PUGNAc (*O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate)<sup>13</sup> for 12 h. PUGNAc has been shown to

up-regulate global *O*-GlcNAc levels in neutrophils<sup>11</sup>, kidney<sup>12</sup>, and other cells by preventing the de-glycosylation of *O*-GlcNAc proteins. Consistent with these studies, Nelly found that PUGNAc strongly enhanced the overall levels of *O*-GlcNAc glycosylation in both the nuclear and S100 cytoplasmic fractions of cortical neurons, as demonstrated by Western blotting with an anti-*O*-GlcNAc antibody (**Fig. 2a**). To identify the proteins undergoing changes, neurons stimulated with and without PUGNAc were lysed and treated as outlined in **Scheme 1**. Prior to chemoenzymatic labeling, Nelly added known quantities of the standards  $\alpha$ -crystallin and OGT into each lysate. Subsequent MS quantification focused on precursor ions that demonstrated characteristic



**Figure 3:** Sequencing of tagged *O*-GlcNAc peptides regulated by PUGNAc treatment using CAD. (a) MS spectrum of a representative peptide whose glycosylation level is significantly increased by PUGNAc treatment of cortical neurons. (b) MS/MS spectrum of the deuterated peak ( $m/z = 862.389$ ), showing loss of a ketogalactose-biotin moiety ( $m/z = 1208.4$ ) and GlcNAc-ketogalactose-biotin moiety ( $m/z = 1005.3$ ). (c-d) Fragmentation during MS<sup>4</sup> analysis yielded numerous internal cleavages and several prominent b and y ions that identified the peptide as <sup>158</sup>AQPPSSASSR<sup>173</sup> from eIF4G. The MS/MS spectrum of a derivatized synthetic peptide matched the MS<sup>4</sup>.

ketogalactose-biotin and GlcNAc-ketogalactose-biotin signature fragmentation patterns. To obtain the relative change in glycosylation on specific peptides, we corrected the heavy:light ratios using a normalization factor derived from the linear regression of the  $\alpha$ -crystallin and OGT standard ratios within each sample. Analysis of standard peptides suggests that we could detect 1.15-fold changes in the nuclear sample and 1.70-fold changes in the cytoplasmic sample with 95% confidence (see Methods for statistical analysis). The peptide standards formed a normal distribution around the mean standard ratio as measured by the D'Agostino-Pearson omnibus test, suggesting that ratios greater than 2 standard deviations ( $\sigma$ ) of the mean ratio are likely significant.

Using these criteria, 22 peptides from the nuclear sample and 11 peptides from the corresponding cytoplasmic sample showed an increase in *O*-GlcNAc glycosylation upon PUGNAc stimulation (**Fig. 2b**). Interestingly, we found that the presence of PUGNAc did not result in increased *O*-GlcNAc glycosylation on all proteins. For example, in the same nuclear sample, 4 *O*-GlcNAc peptides showed no measurable change in glycosylation, whereas in the cytoplasmic sample 16 peptides showed no measurable change (**Fig. 2c**). We also observed decreases in glycosylation on 5 nuclear and 4 cytoplasmic *O*-GlcNAc peptides (**Fig. 2d**). These site-dependent differences suggest differential regulation of the modification in cells, with some proteins being more susceptible to reversible cycling than others.

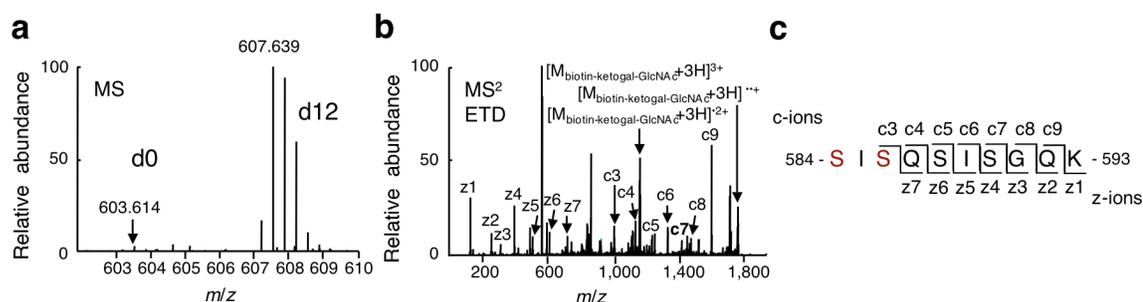
### **Identification of Proteins Subject to Reversible Glycosylation in Neurons**

To identify the neuronal proteins undergoing reversible glycosylation, Scott targeted a portion of the *O*-GlcNAc peptides for sequencing by MS<sup>4</sup> analysis. A

representative ESI-MS spectrum of an *O*-GlcNAc peptide whose glycosylation state was elevated upon PUGNAc treatment is shown (**Fig. 3a**). The CAD MS<sup>2</sup> spectrum of the deuterated, triply charged peptide ( $m/z = 862.389$ ) displays a characteristic loss of a ketogalactose-biotin moiety ( $m/z = 1208.4$ ) and GlcNAc-ketogalactose-biotin moiety ( $m/z = 1005.3$ ) (**Fig. 3b**). MS<sup>4</sup> analysis generated a series of b- and y-type product ions and internal cleavages that enabled definitive sequencing of the peptide (**Fig. 3c,d**). Database searching identified the peptide as belonging to the protein translation elongation initiation factor 4G (eIF4G).

To sequence *O*-GlcNAc-containing peptides and locate the exact sites of glycosylation, Nelly and Danielle Swaney also employed a recently reported fragmentation method, electron transfer dissociation (ETD)<sup>14, 15</sup>. ETD utilizes small molecule radical anions to deliver electrons to isolated peptide precursor cations. After receiving the electron, the odd-electron peptide cation undergoes backbone fragmentation with minimal cleavage of amino acid side chains. This results in the production of sequence-specific c- and z-type product ions without the loss of labile post-translational modification — dissociation pathways that can dominate CAD spectra. As ETD has been successfully used to elucidate exact sites of phosphorylation<sup>14</sup> and *N*-glycosylation<sup>16</sup>, we envisioned that it might be a powerful approach for mapping *O*-GlcNAc glycosylation sites. A representative ETD tandem mass spectrum of an *O*-GlcNAc-modified peptide whose glycosylation level was increased in the PUGNAc-treated sample is shown (**Fig. 4a**). ETD provided near complete sequence coverage for this peptide (**Fig. 4b**), belonging to the transcriptional repressor p66 $\beta$ . Importantly, the *O*-GlcNAc linkage was preserved during ETD fragmentation, and we observed the added mass corresponding to the tagged

*O*-GlcNAc moiety on the c-type product ion series. The tagged *O*-GlcNAc-modified c3 ion narrowed the *O*-GlcNAc glycosylation site to the N-terminal Ser-584 or Ser-586 of this peptide (**Fig. 4c**). ETD was highly effective for the fragmentation of lower *m/z* GlcNAc-ketogalactose-biotin peptide precursor cations (e.g., < ~800), but was less effective for precursors above this *m/z* value. Recent work suggests supplemental collisional activation of the electron transfer product species can help counter this problem<sup>17</sup>.



**Figure 4:** Sequencing of tagged *O*-GlcNAc peptides regulated by PUGNAc treatment using ETD. (a) MS spectrum of a second representative peptide whose glycosylation level is significantly enhanced in response to PUGNAc treatment of cortical neurons. (b, c) MS/MS analysis of the deuterated peak (*m/z* = 607.639) yielded c and z ions that identified the peptide as <sup>584</sup>SISQISIGQK<sup>593</sup> from the transcriptional repressor p66β. The presence of the tagged GlcNAc moiety on the c series of ions narrowed the site of glycosylation to Ser-584 or Ser-586.

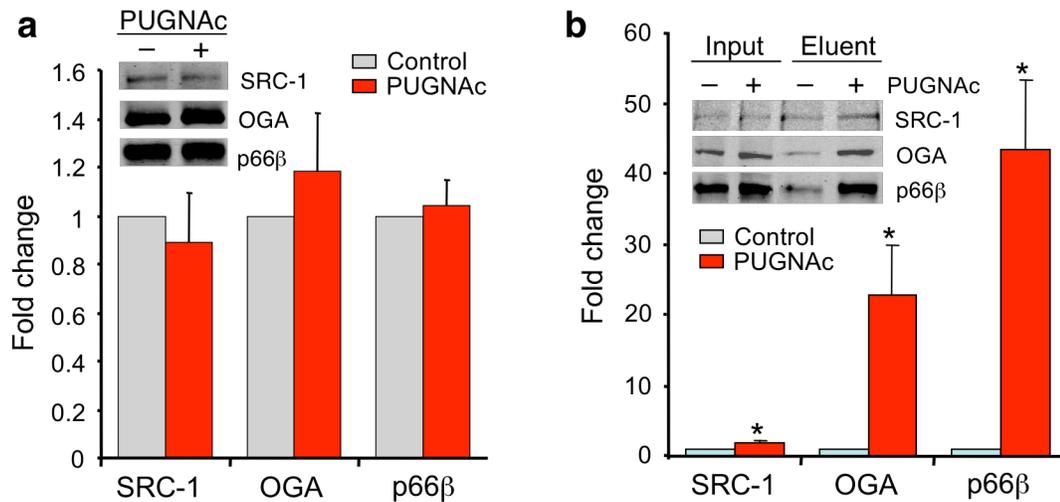
Using a combination of CAD and ETD, Scott and Danielle sequenced 7 of the *O*-GlcNAc peptides that undergo significant increases in glycosylation upon PUGNAc treatment (**Table 2**). In addition, Danielle identified another peptide by ETD that was not observed in the orbitrap MS analysis and thus could not be quantified. Among the *O*-GlcNAc proteins subject to reversible glycosylation are the transcriptional coactivator SRC-1 and the zinc finger RNA-binding protein, which we had previously identified as *O*-GlcNAc glycosylated<sup>3</sup>. Here, we extend those findings by identifying the exact site of glycosylation on both proteins using ETD and by showing that glycosylation at those

sites occurs reversibly in neurons. We also identified an *O*-GlcNAc peptide on the RNA-binding protein nucleoporin 153, which had been previously shown to be *O*-GlcNAc glycosylated<sup>18</sup>, but whose glycosylated peptides were unknown. In addition to these, we identified reversible sites of modification on several new proteins, including the transcriptional repressor p66 $\beta$ , translation factor eIF4G, and the neuron-specific transcriptional repressor BHC80. Finally, we found that the enzyme OGA is *O*-GlcNAc glycosylated in neurons, which is consistent with the ability of OGT to glycosylate OGA *in vitro*<sup>19</sup>. Inhibition of OGA using PUGNAc led to a robust increase in OGA glycosylation at Ser-405, raising the possibility that OGA activity may be regulated by OGT. Interestingly, OGT and OGA were recently shown to form a stable transcriptional regulatory complex, and Ser-405 is located within a region of OGA required for association with OGT<sup>20</sup>.

**Table 2 Identification and quantification of changes in *O*-GlcNAc glycosylation induced by PUGNAc**

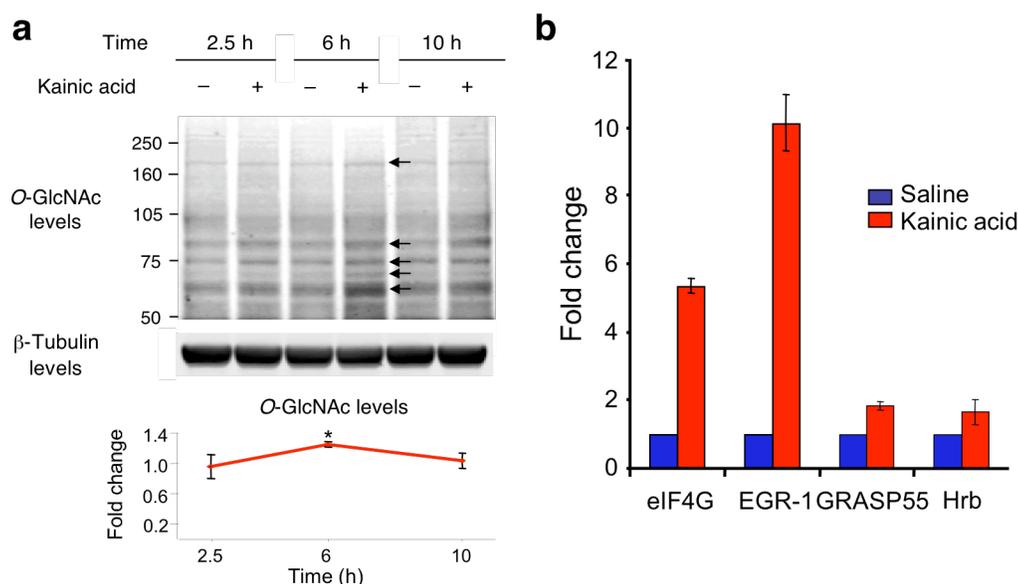
Protein	NCBI entry	Fold change <sup>a</sup>	Function	Peptide sequence <sup>b</sup>	Residues	MS method
BHC80	62645406	1.4	Neuronal gene repression, scaffolding	FTPTLPTSQNSIHPVR	284–300	ETD
eIF4G	62658155	33	Translation elongation scaffolding	AQPPSSAASR	63–72	MS4
Nucleoporin 153	1709215	4.7	RNA binding and transport	KEELPQSSSAG	1004–1114	MS4
OGA	18777747	28.7	<i>N</i> -acetyl-D-glucosaminidase	QVAHSGAK	401–408	MS4
p66 $\beta$	67846054	40.3	Transcriptional repression	SISQSIGGQK	584–593	ETD
SRC-1	34863079	1.5	Coactivation of nuclear receptor transcription	INPSVNPGISPAHGVTR	188–204	ETD
Zinc finger RNA-binding protein	34854400	24.6	RNA-binding protein	AGYSQGATQYTQAQQAR	58–74	ETD
RecQ protein-like 4	17313266	N/D	DNA helicase	KQAAFGGSGPR	378–388	ETD

<sup>a</sup> Fold change represents the observed heavy:light ratio averaged over all experiments. See Supplementary Methods for details on statistical analysis. <sup>b</sup> Potential glycosylation sites determined by ETD are shown in red. N/D, not detected



**Figure 5:** Quantification of *O*-GlcNAc glycosylation on intact proteins by immunoblotting and infrared imaging detection. **(a)** Minimal changes in the expression of SRC-1, OGA, and p66β were observed upon PUGNAc treatment of cortical neurons. Values represent quantification of 4-6 replicates, and a representative Western blot is shown for each protein. Data are mean ± standard deviation (s.d). **(b)** *O*-GlcNAc glycosylation of SRC-1, OGA and p66β was stimulated upon PUGNAc treatment by  $1.9 \pm 0.3$ -,  $22.8 \pm 7.0$ -, and  $43.3 \pm 9.8$ -fold, respectively. *O*-GlcNAc proteins from the lysates were chemoenzymatically labeled with the ketogalactose-biotin tag and selectively captured using streptavidin beads. Quantification was performed as described in the Methods, and values were corrected for any minor changes in protein expression levels shown in **Fig. 5a**. Data are mean ± standard deviation (s.d). Statistical analysis was performed using the Student's *t*-test,  $n = 3$ ,  $*P < 0.05$ . Input, lysates prior to streptavidin capture; Eluent, *O*-GlcNAc proteins captured by streptavidin

To rule out the possibility that the observed increases in *O*-GlcNAc glycosylation are due to altered protein expression, I immunoblotted cell lysates from neurons treated in the presence or absence of PUGNAc with all obtainable antibodies against the proteins of interest. Minimal changes in protein expression were detected upon PUGNAc treatment (**Fig. 5a**), suggesting that the observed changes are due to increased glycosylation. As further confirmation of our approach, I quantified the changes in *O*-GlcNAc levels using an alternative method. Specifically, I chemoenzymatically labeled *O*-GlcNAc proteins



**Figure 6:** *O*-GlcNAc glycosylation is dynamically modulated by robust excitatory stimulation of the brain *in vivo* using kainic acid. **(a)** Overall *O*-GlcNAc glycosylation levels on several proteins in the cerebral cortex (indicated by arrows) are elevated at 6 h post-injection and then return to basal levels after 10 h, as measured using an anti-*O*-GlcNAc antibody. Data are mean  $\pm$  standard deviation (s.d). Statistical analysis was performed using the Student's *t*-test,  $n = 3$ ,  $*P < 0.05$ . **(b)** Proteins identified using the QUIC-Tag method whose *O*-GlcNAc glycosylation levels increase by greater than 1.5-fold upon kainic acid stimulation. Cortical cell lysates were harvested at 6 h post-injection. Data are mean  $\pm$  s.d. Statistical analysis was performed using the Student's *t*-test,  $n = 2 - 4$ .

from cells treated with or without PUGNAc and captured the biotinylated proteins using streptavidin agarose. Following elution, I immunoblotted for specific proteins and quantified changes in *O*-GlcNAc based on the relative amounts of glycosylated protein captured by streptavidin. I found that PUGNAc treatment of neurons induced a  $1.9 \pm 0.3$ -fold increase in *O*-GlcNAc glycosylation of SRC-1, consistent with the results obtained using our quantitative proteomics approach (**Fig. 5b**). Similarly, *O*-GlcNAc glycosylation was stimulated approximately  $22.8 \pm 7.0$ -fold on OGA and  $43.3 \pm 9.8$ -fold on p66 $\beta$ . These results validate the quantitative proteomics methodology and highlight

the versatility of the chemoenzymatic platform for the detection of *O*-GlcNAc peptides or proteins by both MS and immunoblotting.

### ***O*-GlcNAc Glycosylation Is Regulated by Excitatory Stimulation *In Vivo***

Having demonstrated the reversibility of the *O*-GlcNAc modification in neurons, we next investigated whether *O*-GlcNAc glycosylation is induced *in vivo* by neuronal stimulation. Jessica Rexach and I intraperitoneally injected rats with kainic acid, a kainate-type glutamate receptor agonist that produces a robust excitatory stimulus of the brain. Kainic acid has been used to study excitatory pathways that induce gene expression and synaptic plasticity<sup>21</sup> and to invoke seizures as a well-characterized model for temporal lobe epilepsy<sup>22</sup>. We dissected the cerebral cortices of kainic acid-treated rats at distinct behavioral time points: 2.5 h post-injection at peak of seizure, 6 h post-injection when animals had resumed some normal resting behavior, and 10 h post-injection when animals showed nearly identical behavior to saline-injected controls. Global changes in *O*-GlcNAc levels were measured by immunoblotting the cortical cell lysate with an anti-*O*-GlcNAc antibody. I found that *O*-GlcNAc levels on several proteins were elevated at 6 h post-injection and returned to basal levels by 10 h post-injection (**Fig. 6a**).

To identify proteins undergoing changes in *O*-GlcNAc glycosylation in response to kainic acid, Nelly applied our quantitative proteomics strategy to cortical lysates obtained 6 h post-injection. Thirteen of 83 *O*-GlcNAc peptides detected by MS underwent a robust, reproducible increase in response to kainic acid stimulation of rats. Specifically, the changes for these peptides were greater than  $2\sigma$  over the mean of the 1:1

**Table 3 Identification and quantification of changes in *O*-GlcNAc glycosylation induced by kainic acid**

Protein	NCBI Entry	Fold Change <sup>a</sup>	s.d. <sup>b</sup>	n	Function	Peptide Sequence	Residues
EGR-1	6978799	10.1	± 0.9	2	gene transcription, stress response	ALVETSYPSQTR	87-99
eIF4G	62658155	5.3	± 0.2	2	translation elongation	AQPPSSAASR	63-72
GRASP55	51259254	1.8	± 0.1	2	membrane protein transport, golgi stacking	VPTTVEDR	423-430
Hrb	90101424	1.6	± 0.4	4	RNA trafficking	SSSADFGSFSTSQSHQTASTVSK	291-313
bassoon	9506427	1.3	± 0.3	4	synaptic vesicle cycling	SPSTSSTHISYGQPPTTANYGSQ- TEELPHAPSGPAGSGR <sup>c</sup>	1402-1440
bassoon	9506427	1.5	± 0.1	2	synaptic vesicle cycling	ASGAGGPPPELPAGGAR	2283-2300
inositol polyphosphate-4-phosphatase	13591898	1.2	± 0.6	4	lipid phosphatase	SDQQPPVTR	177-186

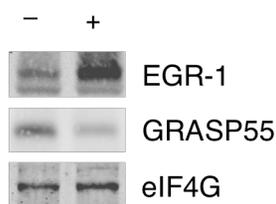
<sup>a</sup> Fold change represents the observed heavy:light ratio averaged over all experiments. See Supplementary Methods for details on statistical analysis.

<sup>b</sup> Maximum absolute standard deviation (s.d.) calculated from g.s.d.

<sup>c</sup> Peptide is also phosphorylated. See text for additional details.

standard peptides for multiple experiments. Using CAD tandem mass spectrometry, Scott successfully identified 4 of these proteins as eIF4G, the transcription factor early growth response-1 (EGR-1), the trafficking protein Golgi reassembly stacking protein 2 (GRASP55), and the HIV-1 Rev-binding protein (Hrb; **Fig. 6b** and **Table 3**). Interestingly, the same peptide of eIF4G that undergoes reversible glycosylation upon PUGNAc treatment also undergoes a change in glycosylation in response to kainic acid. Scott also sequenced 3 *O*-GlcNAc peptides that did not undergo reproducible changes in glycosylation (**Table 3**).

#### Kainic acid



**Figure 7:** Expression levels of EGR-1, GRASP55, and eIF4G following kainic acid treatment of rats. Cortical neuronal lysates were obtained 6 h post-injection of kainic acid or PBS. EGR-1 expression changed by  $1.8 \pm 0.2$ , GRASP55 expression by  $0.61 \pm 0.09$ , and eIF4G expression by  $1.5 \pm 0.1$ . Data represent the mean  $\pm$  s.d. for 3 experiments.

I confirmed that the observed increases in *O*-GlcNAc glycosylation were not due to enhanced protein expression by immunoblotting cortical lysates of kainic acid-treated or control PBS-treated rats with available antibodies against the proteins of interest. Consistent with previous reports that EGR-1 expression is upregulated approximately twofold in the cerebral cortex following kainic acid administration<sup>23</sup>, I found that EGR-1 expression was elevated  $1.8 \pm 0.2$ -fold at 6 h post-injection (**Fig. 7**). Given that *O*-GlcNAc glycosylation of EGR-1 is enhanced by 10.7-fold, protein expression changes

alone cannot account for the sizeable effect of kainic acid on EGR-1 glycosylation. Similarly, the change in eIF4G expression was modest ( $1.5 \pm 0.1$ ) relative to the change in its *O*-GlcNAc level ( $4.9 \pm 0.7$ ), and GRASP55 underwent a decrease in protein expression level with kainic acid treatment ( $0.61 \pm 0.09$ ). To our knowledge, these data represent the first demonstration that extracellular stimuli beyond glucose concentrations in the brain contribute to the dynamics of *O*-GlcNAc glycosylation.

### Expanding the *O*-GlcNAc Proteome of the Brain

In addition to obtaining quantitative information on the dynamics of *O*-GlcNAc glycosylation, we also identified 20 *O*-GlcNAc peptides corresponding to 6 new and 12 previously characterized *O*-GlcNAc proteins from the brain (**Table 4**). Although changes

**Table 4** *O*-GlcNAc glycosylated proteins identified from the cerebral cortex of kainic acid-stimulated rats

Protein	NCBI Entry	Function	Peptide Sequence	Residues
bassoon	9506427	synaptic vesicle cycling	VTQHFAK	1338-1444
CCR4-NOT4	34855140	global transcriptional regulation, mRNA metabolism	SNPVPISSNSHSAR	329-343
CRMP-2	599966	axonal guidance, neuronal polarity	TVTPASSAK <sup>a</sup>	512-520
erythrocyte protein band 4.1-like 1, isoform L	11067407	cytoskeletal protein	DVLTSTYGATAELSTSTTTHTVK	1460-1483
HCF	109511332	chromatin-associated factor	QPETYHTYTTNTPTTAR	1232-1248
LMP-1	62988302	contains PDZ and LIM domain	AQPAQSKPKQK	28-37
MAP2b	547890	dynamic assembly of microtubules at dendrites	VADVPVSEATTVLGDVHSPAEGFVGENISGEEK	380-413
<i>O</i> -GlcNAcase	18777747	<i>N</i> -acetyl-D-glucosaminidase	QVAHSGAK	401-408
PDZ-GEF	34857578	GTP/GDP exchange factor for RAP1/2	SSIVSNSSFDSVPSVLHDER	1215-1233
phosphatidylinositol-binding clathrin assembly protein	16758324	regulation of clathrin assembly	SSGDVHLPISSDVSTFTTR	436-454
Rab3 GDP/GTP exchange protein	1947050	regulation of GTP/GDP exchange for Rab3 subfamily G proteins	SSSSTTASSSPSTIVHGAHSEPADSTEVGDK	699-729
Rad23b	60422770	translocation ubiquitinated proteins	AAAATTTATTTTSGGHPLEFLR	176-198
SH3p8	2293466	SH3 domain binding protein, synaptic vesicle cycling	ITASSSFR	283-290
SRC-1	34863079	coactivation of nuclear receptor transcription	INPSVNPGISPAHGVTTR	188-204
SynGAP	34098355	inhibitory regulation of Ras pathway, synaptic strength regulation	QHSQTPSTLNTPMASER	1121-1138
Ythdf3	109466336	contains YTH domain	IGGDLTAAVTK	145-155

<sup>a</sup> Peptide is also phosphorylated. See text for additional details.

in their glycosylation levels could not be accurately quantified due to low signal-to-noise ratios, these proteins further expand the *O*-GlcNAc proteome of the brain and highlight the abundance of the *O*-GlcNAc modification in neurons. For instance, we identified a glycosylated peptide on the collapsin response mediator protein-2 (CRMP-2), a protein critical for proper axonal development in neurons. We also observed the *O*-GlcNAc modification on several peptides of the large presynaptic scaffolding protein bassoon as well as the phosphatidylinositol-binding clathrin assembly protein. Finally, we found

several new *O*-GlcNAc-modified proteins such as the Rab3 guanine nucleotide exchange protein.

We have developed the first quantitative proteomics method to study the dynamics of *O*-GlcNAc glycosylation *in vivo*. Our QUIC-Tag approach combines the ability to selectively biotinylate and capture *O*-GlcNAc-modified proteins with a simple and efficient isotopic labeling strategy. When combined with tandem mass spectrometry, the method enables unambiguous identification and simultaneous quantification of individual *O*-GlcNAc glycosylation sites. Notably, the chemoenzymatic tagging method does not perturb endogenous *O*-GlcNAc glycosylation levels, unlike previously reported metabolic labeling approaches<sup>24</sup>. The cells are rapidly lysed under denaturing conditions, and the physiological glycosylation state of proteins is preserved and captured by transfer of the ketogalactose-biotin tag. The isotopic labeling strategy has the advantage of being fast, high yielding and inexpensive relative to other methods. As it does not require metabolic labeling or multiple cell divisions for incorporation, the strategy can be readily applied to post-mitotic cells such as neurons or pancreatic islets, as well as to tissues harvested after *in vivo* stimulation. This enables *O*-GlcNAc glycosylation to be studied in more physiological settings and in key cell types where the modification is most highly abundant<sup>25, 26</sup>.

Our approach has distinct advantages over existing methods for monitoring *O*-GlcNAc glycosylation levels. Although a few examples of site-specific *O*-GlcNAc antibodies have been reported<sup>27, 28</sup>, such antibodies are limited in scope and are time-consuming and difficult to generate. As a result, many studies have utilized general *O*-

GlcNAc antibodies to detect global changes in *O*-GlcNAc glycosylation by immunoblotting<sup>11, 12</sup>. These general *O*-GlcNAc antibodies are powerful for many applications, but they have limited sensitivity and do not enable direct identification of specific proteins or sites of modification. Recently, BEMAD (beta-elimination followed by Michael addition with dithiothreitol), a chemical derivatization technique used to identify *O*-GlcNAc and phosphorylation sites, has been coupled to isotopic labeling to study phosphorylation sites in complex mixtures following phosphatase treatment<sup>29</sup>. However, the inherent promiscuity of  $\beta$ -elimination for any modified *O*-linked serine or threonine residues requires extensive internal controls to determine which *O*-linked species is being quantified. Overall, the scarcity of methods available for quantifying *O*-GlcNAc levels highlights the need for the development of new tools for identifying the proteins and pathways that regulate *O*-GlcNAc glycosylation.

In this study, we identified *O*-GlcNAc peptides of interest using two modes of peptide dissociation, CAD and ETD. By CAD, the chemoenzymatic tag produces a unique fragmentation pattern that permits definitive detection of *O*-GlcNAc-modified peptides. Peptides displaying the signature are then targeted for sequencing by MS<sup>4</sup>. In contrast to CAD, ETD generates product ions that retain the *O*-GlcNAc modification and thus can be used to identify exact sites of glycosylation within peptides. Moreover, because sequencing is conducted at the MS/MS stage, ETD forgoes the need for multiple additional stages of MS, which incur loss of signal at each stage. Unlike the related electron capture dissociation (ECD) strategy recently employed to map glycosylation sites that requires the use of FT instrumentation<sup>30</sup>, ETD may be performed directly in appropriately modified ion trap mass spectrometers whose speed, sensitivity, and

accessibility to most laboratories make ETD an ideal emerging technology. Here, we report the first use of ETD technology to study *O*-GlcNAc glycosylation and demonstrate both sequencing and site identification of *O*-GlcNAc peptides from complex mixtures.

Our studies indicate that PUGNAc treatment of cortical neurons induces dramatic changes in *O*-GlcNAc glycosylation on specific proteins. These results suggest that *O*-GlcNAc glycosylation is highly reversible and may be rapidly cycled within neurons. Notably, we found that only a fraction of the *O*-GlcNAc-modified proteins undergo reversible glycosylation. Thus, OGT and OGA may be subject to complex cellular regulation analogous to that of kinases and phosphatases, such as the influence of interacting partners, subcellular targeting and post-translational modifications. The cycling of *O*-GlcNAc on certain substrates, coupled with more inactive, perhaps constitutive, forms of *O*-GlcNAc glycosylation, may allow for the finely-tuned, selective regulation of protein function in response to neuronal stimuli.

One of the proteins whose glycosylation level is significantly increased by PUGNAc treatment is the transcriptional repressor p66 $\beta$ . p66 $\beta$  interacts with histone tails and mediates transcriptional repression by the methyl-CpG-binding domain protein MBD2<sup>31</sup>. Our observation that p66 $\beta$  is reversibly *O*-GlcNAc glycosylated reinforces growing evidence that *O*-GlcNAc plays an important role in the regulation of gene expression<sup>3, 4, 18, 25, 32, 33</sup>. As p66 $\beta$  appears to be sumoylated *in vivo* in a manner that affects its repression potential<sup>34</sup>, our results also highlight a growing network of post-translational modifications that may be fundamental for the regulation of transcription, and it provides a new target with which to study this process.

We also identified changes in glycosylation on several proteins involved in the transport and translocation of mRNA. Such processes are of particular interest in neurons, where regulated transport of mRNA from the cell body to dendrites and dendritic translation of mRNA are involved in changes in synaptic strength that give rise to synaptic plasticity<sup>35</sup>. In particular, we found reversible *O*-GlcNAc glycosylation on the zinc finger RNA-binding protein, which is associated with *staufen2* granules in neurons<sup>36</sup> and may be important in the early stages of RNA translocation from the nucleus to the dendrites. We also observed enhanced glycosylation of a peptide from the C-terminal domain of nucleoporin 153, a protein necessary for docking and trafficking of mRNA<sup>37</sup>.

In addition to studying the reversibility of *O*-GlcNAc in neurons, we demonstrated for the first time that *O*-GlcNAc glycosylation is regulated *in vivo* by robust excitatory stimulation. For example, we found that EGR-1, an immediate early gene and transcription factor important for long-term memory formation<sup>38</sup> and cell survival<sup>39</sup>, undergoes a tenfold increase in glycosylation upon kainic acid stimulation. As the site of glycosylation resides in the N-terminal transactivation domain of EGR-1, one possibility is that *O*-GlcNAc may influence the transactivation potential of EGR-1 and modulate the expression of genes such as the synapsins and proteasome components<sup>40</sup>, which play critical roles in synaptic plasticity.

We also observed an increase in *O*-GlcNAc glycosylation on the translation factor eIF4G upon kainic acid stimulation. As kainic acid treatment induces excitotoxicity in addition to synaptic potentiation<sup>41</sup> and suppressed translation is a known marker for neuronal excitotoxicity<sup>42</sup>, the potential regulation of eIF4G by *O*-GlcNAc glycosylation

may represent a stress-induced response. It will be important to examine whether other cellular stresses induce glycosylation of eIF4G and other proteins to modulate translation and neuronal survival. Consistent with this possibility, other components of the translational machinery have been shown to be *O*-GlcNAc modified, such as p67, which binds to the eukaryotic initiation factor 2 $\alpha$  (eIF2 $\alpha$ ) in its glycosylated form and promotes protein synthesis by preventing inhibitory phosphorylation of eIF2 $\alpha$ <sup>43</sup>.

The ability of *O*-GlcNAc to respond to specific extracellular stimuli suggests a potential role for the modification in mediating neuronal communication. This notion is supported by the identification of a growing number of *O*-GlcNAc glycosylated proteins involved in neuronal signaling and synaptic plasticity<sup>3,30</sup>. In the present study, we further expand the *O*-GlcNAc proteome of the brain to include proteins involved in synaptic vesicle trafficking, including Rab3 GEP, a protein involved in neurotransmitter release, and phosphatidylinositol clathrin protein, which mediates synaptic vesicle endocytosis. In keeping with recent work by Vosseller and colleagues<sup>30</sup>, we find that the presynaptic protein bassoon, which is necessary for the creation of stable synapses and proper neuronal communication, is *O*-GlcNAc modified. We also identify *O*-GlcNAc glycosylation on signal transduction proteins such as the kinase AAK1, which is involved in clathrin-mediated synaptic vesicle endocytosis, and the synaptic Ras GTPase activating protein SynGAP, which plays a critical role in AMPA (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptor trafficking and synapse formation.

Finally, our work highlights the emergent interplay between *O*-GlcNAc glycosylation and phosphorylation. For example, we identified a glycosylated peptide on bassoon that is likewise phosphorylated *in vivo*<sup>44</sup>. Moreover, the axonal guidance protein

CRMP-2 is phosphorylated at two residues within the glycopeptide identified in our studies<sup>45</sup>. Interestingly, when hyperphosphorylated within the residues of this peptide, CRMP-2 appears as a component of the neurofibrillary tangles associated with Alzheimer's disease (AD). This is reminiscent of the microtubule-associated protein tau, which is also *O*-GlcNAc glycosylated but exists in hyperphosphorylated form in the AD brain<sup>46</sup>. Deciphering the mechanisms that regulate the interplay of glycosylation and phosphorylation for these and other proteins may have important ramifications for the study of neuronal signaling and neurodegenerative disorders.

In summary, we demonstrate a new quantitative proteomics strategy for studying the dynamics of *O*-GlcNAc glycosylation. Our findings reveal that the *O*-GlcNAc modification is reversible and dynamically regulated in neurons, and is found on many proteins essential for synaptic function. These observations, along with the discovery that excitatory stimulation can induce *O*-GlcNAc glycosylation in the brain, suggest that *O*-GlcNAc may represent an important post-translational modification for the regulation of neuronal communication. We envision that further application of this methodology will significantly advance our understanding of the regulation of *O*-GlcNAc glycosylation in the nervous system.

## **Methods**

**PUGNAc treatment of cortical cultures.** Cortical neuronal cultures were prepared from embryonic day 18 or 19 Sprague Dawley rats as described<sup>47</sup>. Cells ( $8-12 \times 10^6$ ) were plated on 100-mm culture dishes coated with a  $0.1 \text{ mg ml}^{-1}$  sterile-filtered, aqueous

solution of poly-DL-lysine (Sigma). Cells were maintained for 4 days at 5% CO<sub>2</sub>/37 °C. The media was replaced on the second day and immediately prior to PUGNAc treatment. PUGNAc (Toronto Research Chemicals) was added to the cells at a final concentration of 100 μM (10 mM aqueous stock, sterile-filtered). After 12 h of incubation, the cells were scraped off the plates and pelleted. The media was removed by aspiration, and the cell pellet was washed with 1 ml of HEPES-buffered saline and lysed as described below. Basal neurons were treated identically, except that a water control was used instead of PUGNAc.

**Kainic acid administration.** Male Long Evans rats (7 weeks-old, 190–200 g) were injected intraperitoneally with either 10–11 mg kg<sup>-1</sup> of kainic acid (5 mg ml<sup>-1</sup> in phosphate buffered saline (PBS); Axxora) or PBS as a control. Animals were housed separately and closely monitored for behavioral changes characteristic of seizure activity. Animals were sacrificed at 3 time points, with paired animals demonstrating similar kainic-acid induced behavior: 2.5 h post-injection, when animals were displaying class 4 seizure behavior, 6 h post-injection, when seizure activity was subsiding and animals were displaying some similarity to controls, and 10 h post-injection, when animals were largely indistinguishable from controls. At each time point, the cortices were dissected, flash frozen in liquid N<sub>2</sub> and stored at -80 °C until further use. All animal protocols were approved by the Institutional Animal Care and Use Committee at Caltech, and the procedures were performed in accordance with the Public Health Service Policy on Humane Care and Use of Laboratory Animals.

**Dimethyl labeling.** Protein extracts from PUGNAc-treated cortical neurons and kainic-acid treated brain samples were prepared, chemoenzymatically labeled, and proteolytically digested as described below. Digested extracts were desalted using a Sep-Pak C18 cartridge (1 cc bed volume; Waters). Peptides were eluted in 500  $\mu$ l of 60% aqueous  $\text{CH}_3\text{CN}$ , concentrated by speedvac to a volume of 50  $\mu$ l, and diluted with 450  $\mu$ l of 1 M HEPES pH 7.5. To begin the reactions, the samples were mixed with 40  $\mu$ l of a 600 mM stock of  $\text{NaCNBH}_3$  or  $\text{NaCNBD}_3$  (Sigma) in water, followed by 40  $\mu$ l of 4% aqueous formaldehyde (Mallinckrodt Chemicals) or 40  $\mu$ l of 4% aqueous formaldehyde-d<sub>2</sub> (Sigma). The reactions were briefly vortexed, allowed to proceed for 10 min at room temperature, and then quenched by acidification with 100% AcOH to a pH <4.5. Dimethylated peptides were desalted using a Sep-Pak C18 cartridge (1 cc bed volume), and the eluents (500  $\mu$ l in 60% aqueous  $\text{CH}_3\text{CN}$ , 0.1% AcOH) were concentrated by speedvac to a volume of 100  $\mu$ l.

**Cation exchange and avidin chromatography.** Cation exchange chromatography (Applied Biosystems) was performed on dimethylated peptides as described by the manufacturer, except that peptides were eluted with a step gradient of 100 mM, 250 mM, and 350 mM KCl in 5 mM  $\text{KH}_2\text{PO}_4$  containing 25%  $\text{CH}_3\text{CN}$ . Fractionated peptides were enriched via monomeric avidin chromatography (Applied Biosystems) as follows: peptides were loaded onto the avidin column as described by the manufacturer and washed with 2 ml of 2X PBS (1X PBS final concentration: 10.1 mM  $\text{Na}_2\text{HPO}_4$ , 1.76 mM  $\text{KH}_2\text{PO}_4$ , 137 mM NaCl, 2.7 mM KCl, pH 6.7), 2 ml of 1X PBS, 1.5 ml of manufacturer

wash buffer 2 and 1 ml of ddH<sub>2</sub>O. Avidin-enriched peptides were eluted as described by the manufacturer.

**Orbitrap MS analysis and ETD analysis.** Automated nanoscale reversed-phase HPLC/ESI/MS was performed as described below and previously<sup>3</sup>. For data-dependent experiments, the mass spectrometer was programmed to record a full-scan ESI mass spectrum ( $m/z$  650–2000, ions detected in orbitrap mass spectrometer with a resolution set to 100000) followed by five data-dependent MS/MS scans (relative collision energy = 35%; 3.5 Da isolation window). Precursor ion masses for candidate glycosylated peptides were identified by a computer algorithm (Charge Loss Scanner; developed in-house with Visual Basic 6.0) that inspected product ion spectra for peaks corresponding to losses of the ketogalactose-biotin and GlcNAc-ketogalactose-biotin moieties. Up to eight candidate peptides at a time were analyzed in subsequent targeted MS<sup>4</sup> experiments to derive sequence information.

For all MS experiments, the electrospray voltage was set at 1.8 kV and the heated capillary was maintained at 250 °C. For database analysis to identify *O*-GlcNAc proteins, Bioworks Browser 3.2SR1 (ThermoElectron) software was used to create files from MS<sup>4</sup> data and ETD MS/MS data. These files were then directly queried, using the SEQUEST algorithm (ThermoElectron), against amino acid sequences in the NCBI rat/mouse protein database.

Quantification was conducted by generating single ion chromatograms from the orbitrap MS scans for candidate *O*-GlcNAc peptides. Peak areas of isotopic clusters were derived using Xcalibur 1.4 software (ThermoElectron) and relative ratios were

normalized against the mean relative ratio of standard peptides. Statistical analysis is described in detail below.

MS/MS experiments by ETD were conducted on a modified LTQ mass spectrometer. A chemical ionization source was added to the rear side of the LTQ to allow for the introduction of fluoranthene radical anions for ETD reactions. For data-dependent experiments, the mass spectrometer was programmed to record a full-scan ESI mass spectrum ( $m/z$  650–2000) followed by five data-dependent MS/MS scans (70–100 ms ETD activation; 3.5 Da isolation window). In some cases, targeted MS/MS was conducted on up to eight candidate peptides that had demonstrated the signature ketogalactose-biotin loss during CAD MS/MS. All sequenced peptides were manually verified.

#### **Chemoenzymatic labeling and streptavidin capture of *O*-GlcNAc proteins.**

Chemoenzymatic labeling was performed on neuronal lysates as described above. After reaction with the aminoxy biotin derivative, proteins were dialyzed (1 x 10 h, 2 x 3 h) into 7 M urea, 10 mM HEPES, pH 7.5 at room temperature followed by 10 mM HEPES pH 7.5, 100 mM NaCl, 0.2% Triton-X 100 (2 x 2 h, 1 x 10 h) at 4 °C. Fresh PMSF (1 mM) was added at each stage of dialysis. Proteins were captured on streptavidin beads as previously described<sup>4</sup> and probed by immunoblotting.

**Western blotting.** Lysates were resolved by SDS-PAGE, transferred to nitrocellulose membranes and immunoblotted as described previously<sup>4</sup>. Total *O*-GlcNAc levels were monitored using the anti-*O*-GlcNAc antibody CTD110.6 (Covance, 1:5000). The

following primary antibodies were also used: EGR-1 (Upstate Biotechnology, 1:1000), GRASP-55 (BD Transduction Laboratories, 1:1000), eIF4G (Santa Cruz, 1:100), OGA (a kind gift from Prof. Sidney Whiteheart, University of Kentucky, 1:1000), p66 $\beta$  (Upstate, 1:500), and SRC-1 (Santa Cruz, 1:100). After incubation with the secondary antibodies IRDye 800 goat anti-rabbit (Rockland Immunochemicals) or Alexa Fluor 680 goat anti-mouse (Molecular Probes), proteins were visualized and quantified using the Odyssey infrared imaging system (LI-COR Biosciences). To quantify differences in *O*-GlcNAc levels, we measured the relative intensities of the input bands (lysate prior to streptavidin capture) and eluent bands (lysate after streptavidin capture) using Odyssey imaging software (Version 2.1). For each sample, we normalized the eluent signals to the input signals, and the resulting values from control reactions lacking GalT were subtracted from those values obtained from reactions containing GalT to correct for any nonspecific background.

**Statistical analysis.** Quantification was conducted by generating single ion chromatograms from the orbitrap MS scans for candidate *O*-GlcNAc peptides. Peak areas of isotopic clusters were derived using Xcalibur 1.4 software. Mean values, standard deviations and confidence intervals were calculated using the program Excel on log-transformed ratios and reported in the original scale as previously described<sup>2,3</sup>. We used the geometric standard deviation (g.s.d.) to calculate maximum absolute standard deviations. Standard peptide ratios were tested for goodness of fit to the log-normal distribution via the D'Agostino-Pearson omnibus test and were used to determine the confidence with which changes in experimental peptides could be detected.

Experimental peptide ratios were normalized against the slope of the linear regression produced by the heavy vs. light forms of standard peptides within experiments.

## References

1. Khidekel, N. *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J. Am. Chem. Soc.* **125**, 16162-16163 (2003).
2. Hsu, J.L., Huang, S.Y., Chow, N.H. & Chen, S.H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843-6852 (2003).
3. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. USA* **101**, 13132-13137 (2004).
4. Tai, H.C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel identification of *O*-GlcNAc-modified proteins from cell lysates. *J. Am. Chem. Soc.* **126**, 10500-10501 (2004).
5. Chalkley, R.J. & Burlingame, A.L. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* **12**, 1106-1113 (2001).
6. Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **17**, 977-982 (2006).
7. Roquemore, E.P., Chevrier, M.R., Cotter, R.J. & Hart, G.W. Dynamic *O*-GlcNAcylation of the small heat shock protein alpha B-crystallin. *Biochemistry* **35**, 3578-3586 (1996).
8. Zhang, R., Sioma, C.S., Wang, S. & Regnier, F.E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **73**, 5142-5149 (2001).
9. Ong, S.E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1**, 119-126 (2004).
10. Ross, P.L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154-1169 (2004).
11. Kneass, Z.T. & Marchase, R.B. Neutrophils exhibit rapid agonist-induced increases in protein-associated *O*-GlcNAc. *J. Biol. Chem.* **279**, 45759-45765 (2004).
12. 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**, 30133-30142 (2004).
13. 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-beta-*N*-acetylglucosaminidase inhibitor *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate. *J. Biol. Chem.* **273**, 3611-3617 (1998).
14. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* **101**, 9528-9533 (2004).
15. Coon, J.J., Syka, J.E.P., Schwartz, J.C., Shabanowitz, J. & Hunt, D.F. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *Int. J. Mass Spectrom.* **236**, 33-42 (2004).

16. Hogan, J.M., Pitteri, S.J., Chrisman, P.A. & McLuckey, S.A. Complementary structural information from a tryptic *N*-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J. Proteome Res.* **4**, 628-632 (2005).
17. Swaney, D.L. *et al.* Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* **79**, 477-485 (2007).
18. Love, D.C. & Hanover, J.A. The hexosamine signaling pathway: deciphering the "*O*-GlcNAc code". *Sci. STKE* **2005**, re13 (2005).
19. Lazarus, B.D., Love, D.C. & Hanover, J.A. Recombinant *O*-GlcNAc transferase isoforms: identification of *O*-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology* **16**, 415-421 (2006).
20. Whisenhunt, T.R. *et al.* Disrupting the enzyme complex regulating *O*-GlcNAcylation blocks signaling and development. *Glycobiology* **16**, 551-563 (2006).
21. Nedivi, E., Hevroni, D., Naot, D., Israeli, D. & Citri, Y. Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature* **363**, 718-722 (1993).
22. Ben-Ari, Y. & Cossart, R. Kainate, a double agent that generates seizures: two decades of progress. *Trends Neurosci.* **23**, 580-587 (2000).
23. Beckmann, A.M., Davidson, M.S., Goodenough, S. & Wilce, P.A. Differential expression of Egr-1-like DNA-binding activities in the naive rat brain and after excitatory stimulation. *J. Neurochem.* **69**, 2227-2237 (1997).
24. Nandi, A. *et al.* Global identification of *O*-GlcNAc-modified proteins. *Anal. Chem.* **78**, 452-458 (2006).
25. Iyer, S.P. & Hart, G.W. Dynamic nuclear and cytoplasmic glycosylation: enzymes of *O*-GlcNAc cycling. *Biochemistry* **42**, 2493-2499 (2003).
26. Cole, R.N. & Hart, G.W. Cytosolic *O*-glycosylation is abundant in nerve terminals. *J. Neurochem.* **79**, 1080-1089 (2001).
27. Kamemura, K., Hayes, B.K., Comer, F.I. & Hart, G.W. Dynamic interplay between *O*-glycosylation and *O*-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J. Biol. Chem.* **277**, 19229-19235 (2002).
28. Ludemann, N. *et al.* *O*-glycosylation of the tail domain of neurofilament protein M in human neurons and in spinal cord tissue of a rat model of amyotrophic lateral sclerosis (ALS). *J. Biol. Chem.* **280**, 31648-31658 (2005).
29. Vosseller, K. *et al.* Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* **5**, 388-398 (2005).
30. 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**, 923-934 (2006).
31. Brackertz, M., Gong, Z., Leers, J. & Renkawitz, R. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res.* **34**, 397-406 (2006).
32. Lamarre-Vincent, N. & Hsieh-Wilson, L.C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J. Am. Chem. Soc.* **125**, 6612-6613 (2003).
33. Yang, X. *et al.* *O*-linkage of *N*-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability. *Proc. Natl. Acad. Sci. USA* **98**, 6611-6616 (2001).

34. Gong, Z., Brackertz, M. & Renkawitz, R. SUMO modification enhances p66-mediated transcriptional repression of the Mi-2/NuRD complex. *Mol. Cell. Biol.* **26**, 4519-4528 (2006).
35. Ule, J. & Darnell, R.B. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.* **16**, 102-110 (2006).
36. Elvira, G., Massie, B. & DesGroseillers, L. The zinc-finger protein ZFR is critical for Staufien 2 isoform specific nucleocytoplasmic shuttling in neurons. *J. Neurochem.* **96**, 105-117 (2006).
37. Bastos, R., Lin, A., Enarson, M. & Burke, B. Targeting and function in mRNA export of nuclear pore complex protein Nup153. *J. Cell Biol.* **134**, 1141-1156 (1996).
38. Jones, M.W. *et al.* A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories. *Nat. Neurosci.* **4**, 289-296 (2001).
39. Thiel, G. & Cibelli, G. Regulation of life and death by the zinc finger transcription factor Egr-1. *J. Cell. Physiol.* **193**, 287-292 (2002).
40. James, A.B., Conway, A.M. & Morris, B.J. Genomic profiling of the neuronal target genes of the plasticity-related transcription factor -- Zif268. *J. Neurochem.* **95**, 796-810 (2005).
41. Wang, Q., Yu, S., Simonyi, A., Sun, G.Y. & Sun, A.Y. Kainic acid-mediated excitotoxicity as a model for neurodegeneration. *Mol. Neurobiol.* **31**, 3-16 (2005).
42. Marin, P. *et al.* Glutamate-dependent phosphorylation of elongation factor-2 and inhibition of protein synthesis in neurons. *J. Neurosci.* **17**, 3445-3454 (1997).
43. Datta, R., Choudhury, P., Ghosh, A. & Datta, B. A glycosylation site, 60SGTS63, of p67 is required for its ability to regulate the phosphorylation and activity of eukaryotic initiation factor 2alpha. *Biochemistry* **42**, 5453-5460 (2003).
44. Collins, M.O. *et al.* Proteomic analysis of in vivo phosphorylated synaptic proteins. *J. Biol. Chem.* **280**, 5972-5982 (2005).
45. Gu, Y., Hamajima, N. & Ihara, Y. Neurofibrillary tangle-associated collapsin response mediator protein-2 (CRMP-2) is highly phosphorylated on Thr-509, Ser-518, and Ser-522. *Biochemistry* **39**, 4267-4275 (2000).
46. 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. USA* **101**, 10804-10809 (2004).
47. Gama, C.I. *et al.* Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat. Chem. Biol.* **2**, 467-473 (2006).

## **Chapter 4: Direct In-Gel Fluorescence Detection and Cellular Imaging of *O*-GlcNAc-Modified Proteins**

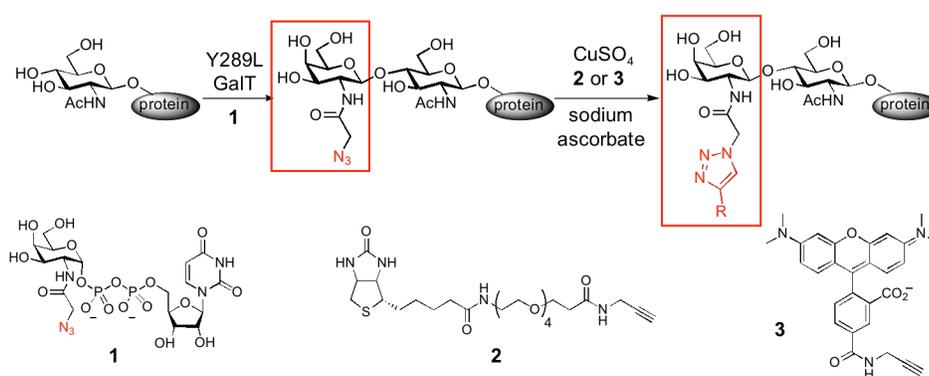
Portions of this chapter are from Clark, P.M., Dweck, J.F., Mason, D.E., Hart, C.R., Buck, S.B., Peters, E.C., Agnew, B.J. & Hsieh-Wilson, L.C. Direct in-gel fluorescence detection and cellular imaging of *O*-GlcNAc-modified proteins. *J. Am. Chem. Soc.* **130**, 11576-7 (2008).

**We report an advanced chemoenzymatic strategy for the direct fluorescence detection, proteomic analysis, and cellular imaging of *O*-GlcNAc-modified proteins. *O*-GlcNAc residues are selectively labeled with fluorescent or biotin tags using an engineered galactosyltransferase enzyme and [3+2] azide-alkyne cycloaddition chemistry. We demonstrate that this approach can be used for direct in-gel detection and mass spectrometric identification of *O*-GlcNAc proteins, identifying 146 novel glycoproteins from the mammalian brain. Furthermore, we show that the method can be exploited to quantify dynamic changes in cellular *O*-GlcNAc levels and to image *O*-GlcNAc glycosylated proteins within cells. As such, this strategy enables studies of *O*-GlcNAc glycosylation that were previously inaccessible and provides a new tool for uncovering the physiological functions of *O*-GlcNAc.**

Understanding posttranslational modifications to proteins is critical for elucidating the functional roles of proteins within the dynamic environment of cells. *O*-linked  $\beta$ -*N*-acetylglucosamine (*O*-GlcNAc) glycosylation has emerged as important for the regulation of diverse cellular processes, including transcription, cell division, and glucose homeostasis<sup>1-3</sup>. While new chemical tools have provided rapid, sensitive methods for detecting the modification and enabled better control over the activity of *O*-GlcNAc enzymes<sup>1, 4-10</sup>, significant challenges remain with regard to elucidating the functions of *O*-GlcNAc in cells. For instance, a robust method for the direct fluorescence detection of *O*-GlcNAc proteins in gels would permit monitoring of changes in glycosylation levels in response to cellular stimuli and greatly extend the reach of existing technologies. Furthermore, new tools for imaging *O*-GlcNAc glycosylated

proteins would enable the expression and dynamics of the modification to be monitored in cells and tissues. Here, we report an advanced chemoenzymatic labeling strategy that addresses these important needs.

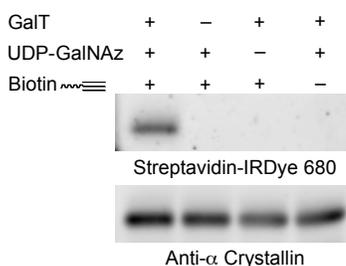
Previous studies have shown that an engineered  $\beta$ -1,4-galactosyltransferase enzyme (Y289L GalT) efficiently transfers a ketogalactose moiety from an unnatural UDP substrate selectively onto *O*-GlcNAc-modified proteins. Treatment with aminoxy-biotin followed by streptavidin capture and elution allowed for identification of *O*-



**Figure 1:** Chemoenzymatic labeling of *O*-GlcNAc proteins using [3+2] cycloaddition chemistry. R = biotin or TAMRA.

GlcNAc-modified proteins<sup>4, 11</sup>. However when I tried applying this strategy to more in-depth studies of *O*-GlcNAc proteins such as assaying *O*-GlcNAc dynamics across multiple conditions, I found that this was not an ideal system. First, the aminoxy-biotin reagent appeared to have strong nonspecific protein interactions. For example, robust streptavidin signal was detected in the control (- GalT) lane following *in vitro* labeling of  $\alpha$ -crystallin even after three days of dialysis to remove the excess aminoxy biotin. Similarly after considerable optimization, I found that quantitative streptavidin capture was achieved only when using a large excess of streptavidin beads, in our case equal to half the reaction volume, to capture the biotin-labeled proteins in the presence of the

excess free biotin, again even after three days of dialysis. Second, the aminoxy–ketone reaction required conditions of pH 4.5 for 24 hours in the presence of 5 M Urea. Yet in many cases, this prolonged reaction with low pH caused up to 50% of the proteins to precipitate out of solution.



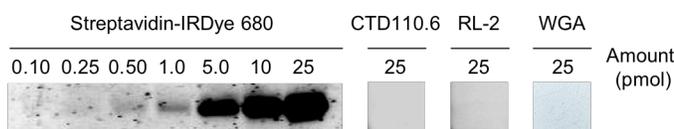
**Figure 2:** Selective labeling of  $\alpha$ -crystallin

at 25 °C. Analysis by gel electrophoresis and blotting with streptavidin conjugated to an IR680 dye showed robust, selective labeling of  $\alpha$ -crystallin, with no nonspecific labeling in the absence of GalT, **1** or **2** (Fig. 2). Notably, as little as 250 fmol of  $\alpha$ -crystallin (~25 fmol of glycosylated protein) was detectable, highlighting the sensitivity of the approach. In contrast, other

We therefore investigated whether Y289L GalT would accept the UDP-azidogalactose substrate **1** (UDP-GalNAz), which would allow for labeling of *O*-GlcNAc proteins using [3+2] azide-alkyne cycloaddition chemistry (Fig. 1)<sup>12-14</sup>. In addition to providing alternative dyes to potentially reduce nonspecific interactions, this Cu(I)-catalyzed cycloaddition reaction would have the advantage of being performed more rapidly and at physiological pH.

I tested the approach using  $\alpha$ -crystallin, a known *O*-GlcNAc-modified protein with a low extent (~10%) of glycosylation.  $\alpha$ -Crystallin was treated with **1** and Y289L

GalT, followed by reaction with  $\text{CuSO}_4$ , sodium ascorbate, and the biotin-alkyne derivative **2** for 1 h

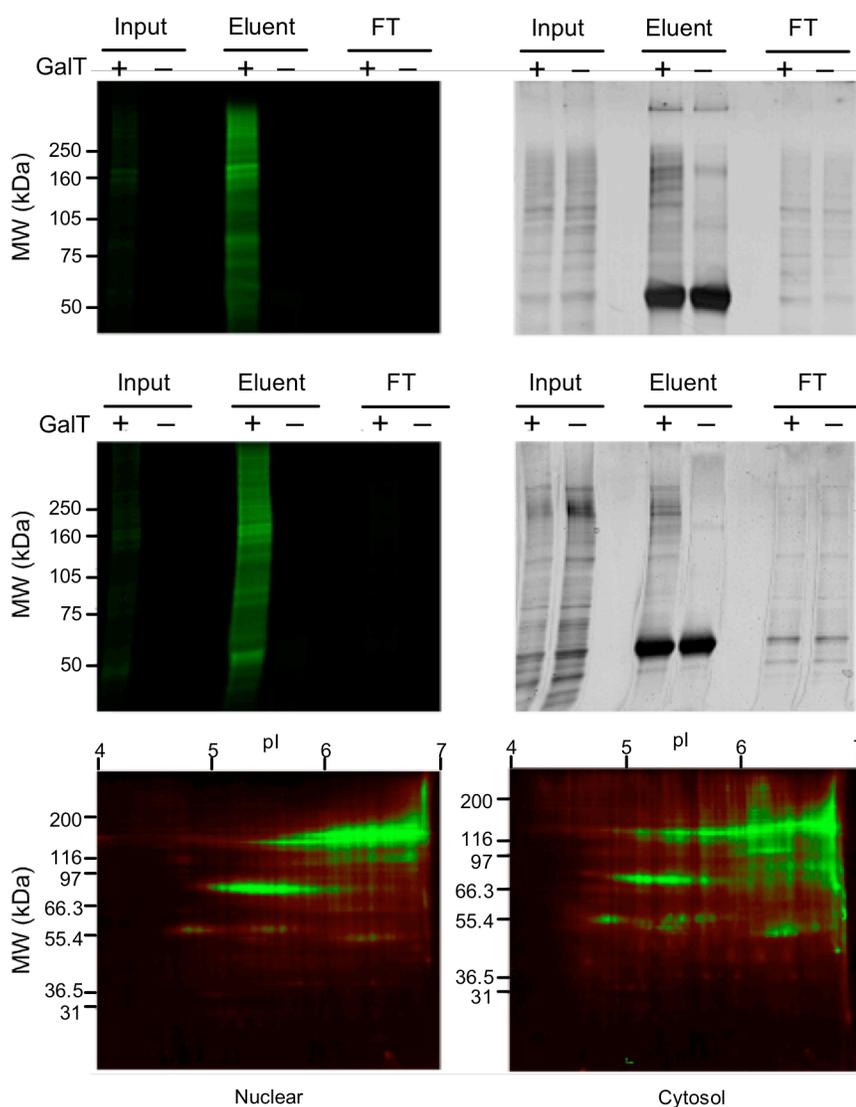


**Figure 3:** Detection sensitivity of the chemoenzymatic approach

at 25 °C. Analysis by gel electrophoresis and blotting with streptavidin conjugated to an IR680 dye showed robust, selective labeling of  $\alpha$ -crystallin, with no nonspecific labeling in the absence of GalT, **1** or **2** (Fig. 2). Notably, as little as 250 fmol of  $\alpha$ -crystallin (~25 fmol of glycosylated protein) was detectable, highlighting the sensitivity of the approach. In contrast, other

methods such as *O*-GlcNAc antibodies or lectins failed to detect the *O*-GlcNAc modification on  $\alpha$ -crystallin (**Fig. 3**)<sup>4</sup>.

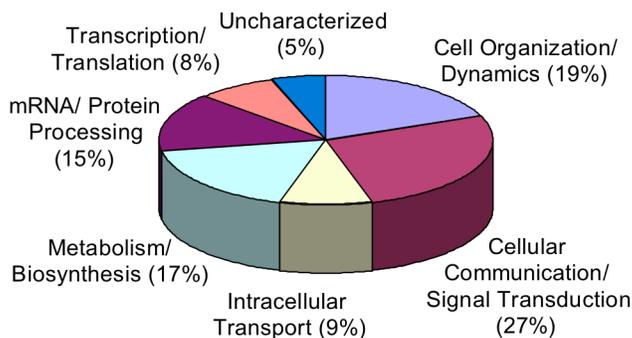
I next examined whether this approach could be used for direct in-gel fluorescence detection and proteome-wide analyses of *O*-GlcNAc glycosylated proteins. Nuclear and cytosolic protein fractions from rat forebrain were azide-labeled and then



**Figure 4:** Enrichment and in-gel fluorescence detection of *O*-GlcNAc-modified proteins in 1D (top and middle) and 2D (bottom) gels. For the 1D gels, 15  $\mu$ g of nuclear (top) or cytoplasmic (middle) protein was loaded in the input and FT lanes; material captured from 470  $\mu$ g of protein was loaded in the eluent lanes.

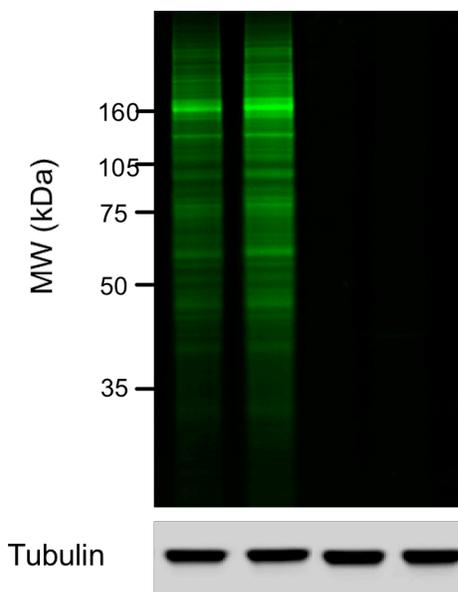
reacted with the tetramethyl-6-carboxyrhodamine (TAMRA)-alkyne derivative **3**. The *O*-GlcNAc proteins were immunoprecipitated using an anti-TAMRA antibody to remove non-glycosylated proteins from the lysate, resolved by 1D or 2D gel electrophoresis, and visualized by in-gel fluorescence imaging (**Fig. 4**). Importantly, minimal nonspecific labeling was detected with the TAMRA-alkyne dye (**Fig. 4**, -GalT control lanes), and I observed efficient capture and enrichment of the TAMRA-labeled proteins (+GalT, eluent and flow-through lanes).

To identify *O*-GlcNAc proteins, bands from the gel were excised, proteolytically digested, and subjected to nanoLC-MS/MS analysis. The data acquisition and subsequent database searching methodologies employed are detailed in the methods section. In total, Daniel Mason and I identified 213 proteins, representing 67 previously known and 146 novel, putative *O*-GlcNAc modified proteins (**Table 1**). The majority of the proteins identified participate in neuronal signaling and synaptic function, suggesting important functional roles for *O*-GlcNAc in neuronal communication (**Fig. 5**). Surprisingly, in contrast to previous proteomic analyses of brain tissue<sup>5,15,16</sup>, we identified many proteins involved in metabolism and biosynthesis, consistent with roles for *O*-GlcNAc in nutrient sensing and cell survival observed in other tissues<sup>1-3</sup>. Interestingly, the metabolic proteins included 9 of the 10 enzymes required for glycolysis, suggesting a previously unidentified level of control by *O*-GlcNAc of this pathway. Thus, the approach enables the identification of a large number of unique *O*-GlcNAc modified proteins and has the advantages of ease and accessibility (e.g., short incubation times, simple gel-based detection and separation versus multiple chromatography steps, high-throughput analyses, commercially available reagents).



**Figure 5:** Functional classification of *O*-GlcNAc proteins from rat brain identified by MS

PUGNAc	-	+	-	+
UDP-GalNAz	+	+	-	-



**Figure 6:** Direct detection of changes in *O*-GlcNAc glycosylation levels upon cellular stimulation. Tubulin controls indicate equal loading of protein in each lane.

phenylcarbamate), an inhibitor of the  $\beta$ -*N*-acetylglucosaminidase enzyme that removes *O*-GlcNAc, and the *O*-GlcNAc-modified proteins were labeled and analyzed as before.

Understanding the cellular dynamics of *O*-GlcNAc glycosylation will be critical for elucidating its functional roles in both physiological and diseased states. However, few methods exist for quantifying changes in *O*-GlcNAc glycosylation in response to cellular stimuli. Glycosylation levels are typically monitored by immunoblotting with a general *O*-GlcNAc antibody<sup>17</sup>, which detects only a limited number of *O*-GlcNAc proteins and affords no opportunity to identify proteins undergoing changes in glycosylation. We examined whether our chemoenzymatic approach could overcome such limitations.

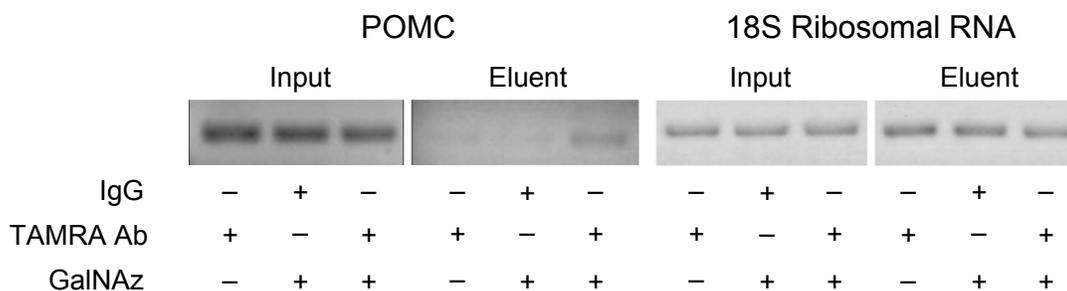
HeLa cells were stimulated with PUGNAc (*O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-

PUGNAc treatment resulted in a  $163 \pm 3\%$  increase in overall *O*-GlcNAc glycosylation levels, and interestingly, ranged from 136–176%, depending on the specific protein (**Fig. 6**). The varying extent to which *O*-GlcNAc is induced upon cellular stimulation may indicate complex regulatory control of the modification. Thus, this approach provides a new method to visualize and quantify dynamic changes in protein *O*-GlcNAc glycosylation which, when coupled with in-gel digestion and MS analyses as described above, will enable the identification of specific proteins undergoing those changes.

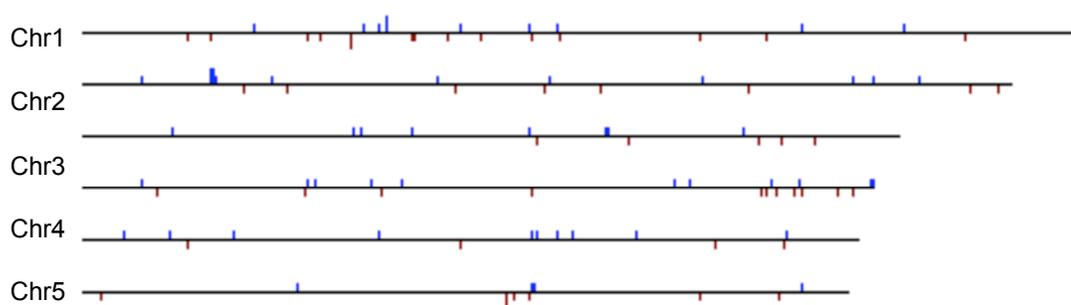
*O*-GlcNAc is known to modify a variety of components of the transcriptional machinery, including RNA polymerase II<sup>18</sup>, CREB<sup>19</sup>, and histone lysine methyltransferase MLL5<sup>20</sup>. Furthermore proteomic studies have demonstrated that transcription factors are over-represented among identified *O*-GlcNAc proteins<sup>1, 5</sup>. Thus I next examined whether our approach could be used to investigate *O*-GlcNAc on chromatin. DNA and proteins were cross-linked with formaldehyde, the chromatin was fragmented, and the lysate was azide-labeled and then reacted with **3**. The *O*-GlcNAc proteins were immunoprecipitated using an anti-TAMRA antibody, the associated DNA was separated from the proteins, and the DNA was amplified by PCR. I observed an enrichment of eluent signal on the POMC promoter specifically in the presence of GalNAz and TAMRA antibody but no enrichment on the control 18S ribosomal RNA promoter (**Fig. 7**). This suggests that our approach can be used to identify specific gene promoters that are enriched in *O*-GlcNAc levels. To more broadly identify such promoters, in collaboration with Jessica Rexach and Rosemarie Tsoa, immunoprecipitated *O*-GlcNAc-associated chromatin and then assayed the results on a promoter array. Using this approach, we identified 154 promoters in which *O*-GlcNAc

levels were specifically enriched (**Table 2**). These enriched promoters are distributed evenly across the mouse chromosomes (**Fig. 8**) and are over-represented in genes important for neural tube development as well as genes important in cell-cell adhesion and alkali metal ion binding. In particular, *O*-GlcNAc levels were enriched on the promoter of four different potassium channels: *Kcnt1*, *Kcnab1*, *Kcne3*, and *Kcnj14*. Thus this approach enables the identification of chromatin regions enriched in *O*-GlcNAc levels and could be used to identify dynamic changes in *O*-GlcNAc levels on the chromatin. Furthermore this approach could be expanded upon with a second immunoprecipitation step to determine colocalization of *O*-GlcNAc and specific proteins or other modifications on a specific region of DNA.

Finally, Jessica Dweck examined whether *O*-GlcNAc modified proteins could be

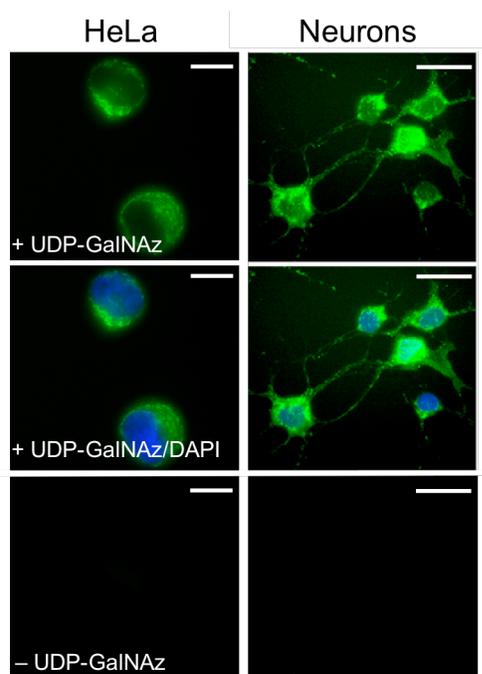


**Figure 7:** Detection of *O*-GlcNAc levels on chromatin



**Figure 8:** *O*-GlcNAc levels are distributed evenly across the mouse chromosomes.

chemoenzymatically tagged and imaged in cells. HeLa cells and cultured cortical neurons were fixed, permeabilized, and labeled with **1** and Y289L GalT, followed by biotin-alkyne **2** or TAMRA-alkyne **3**. The biotin-treated cells were further incubated with a streptavidin-AlexaFluor 488 conjugate. Notably, addition of exogenous GalT and **1** to the cells led to robust labeling of *O*-GlcNAc glycosylated proteins (**Fig. 9**). Although the TAMRA-alkyne **3**



**Figure 9:** Fluorescence imaging of *O*-GlcNAc proteins (green) in HeLa cells (left) or cortical neurons (right). Nuclei were stained with DAPI (blue). Scale bars = 10  $\mu$ m (HeLa) and 25  $\mu$ m

produced background labeling in the absence of **1** (data not shown), strong staining and minimal background labeling were observed using biotin-alkyne **2**. Consistent with the reported localization of *O*-GlcNAc enzymes<sup>1,3</sup>, *O*-GlcNAc glycosylated proteins were found in both the nucleus and cytoplasm. Moreover, Jessica observed robust staining of proteins along neuronal processes, corroborating our mass spectrometric identification of many *O*-GlcNAc proteins involved in synaptic signaling. This is the first example of exploiting chemical tagging methods to image *O*-GlcNAc-modified proteins within cells. The approach affords high labeling sensitivity without perturbing physiological pathways and should be amenable to tissue samples — features that may complicate other strategies such as metabolic labeling.

In summary, we describe an advanced chemoenzymatic labeling approach that exploits [3+2] cycloaddition chemistry to attach fluorescent and biotin tags to *O*-GlcNAc residues. This method enables studies of *O*-GlcNAc glycosylation that were previously inaccessible. The ability to label proteins selectively with a fluorescent reporter group permits rapid and direct in-gel detection of *O*-GlcNAc proteins, facilitating proteomic analyses and providing a new method to quantify dynamic changes in glycosylation. Covalent labeling of proteins allows for cellular imaging of *O*-GlcNAc proteins in their native biological environment. Finally, this approach was developed in conjunction with researchers at Invitrogen with the goal of providing commercially available reagents that are now accessible to the wider research community. We anticipate that this new approach will be a powerful tool for advancing our understanding of the physiological functions and dynamic regulation of *O*-GlcNAc glycosylation within cells.

**Table 1: *O*-GlcNAc glycosylated proteins identified by mass spectrometry.** Proteins are tabulated by function, and the accession number and number of peptides (# Pep.) found for that protein are listed. Previously identified *O*-GlcNAc proteins are indicated. † represents proteins that have been previously identified as *O*-GlcNAc proteins by any method. †† represents proteins that have been previously validated to contain *O*-GlcNAc either by direct identification of the *O*-GlcNAc modification by mass spectrometry or by radioactive GalT labeling.

Protein	Accession Number	# Pep.	Known
<b>Cell Organization / Dynamics</b>			
ACTA2 Actin alpha-2 chain	IPI00008603.1, IPI00021428.1, IPI00023006.1, IPI00025416.3, IPI00110827.1	8	†
ACTB Actin beta chain	IPI00021439.1, IPI00021440.1, IPI00848058.1	27	†
Ank2 Similar to Ankyrin 2 isoform 1	IPI00554111.2	10	
Ank3 Ankyrin 3	IPI00199445.2	19	††
ANXA2 Annexin A2	IPI00797556.1, IPI00848164.1	3	
Anxa6 Annexin A6	IPI00421888.3, IPI00831745.1	7	
ARPC2 Actin-related protein 2/3 complex subunit 2	IPI00005161.3, IPI00661414.2, IPI00764535.2	3	
CAPZB Isoform 1 of F-actin capping protein subunit beta	IPI00026185.5, IPI00191444.3, IPI00218782.2, IPI00269481.7, IPI00365283.1, IPI00406800.4, IPI00474883.2, IPI00642256.1, IPI00776140.1	8	††
Ckap5 Cytoskeleton associated protein 5	IPI00317134.3, IPI00337930.4, IPI00764313.1, IPI00764540.1, IPI00767392.1, IPI00769262.1	4	
Crym Mu-crystallin homolog	IPI00214448.1	14	
Cyln2 CAP-Gly domain-containing linker protein 2	IPI00195929.1	5	
Dnm1 Isoform 1 of Dynamin-1	IPI00272878.6, IPI00331293.3, IPI00413140.3, IPI00657691.2, IPI00816287.2	18	
Dync1h1 Dynein heavy chain, cytosolic	IPI00327630.1	29	†
Epb4.111 Isoform S of Band 4.1-like protein 1	IPI00203237.2, IPI00203239.2, IPI00561718.1	19	
Epb4.113 Type II brain 4.1 minor isoform	IPI00204503.1, IPI00204506.1, IPI00556956.2, IPI00558692.1, IPI00561669.1, IPI00568756.1	14	†
Fscn1 Fascin	IPI00353563.4, IPI00763106.1, IPI00767873.1	13	†
Ina Alpha-internexin	IPI00135965.2, IPI00211936.2, IPI00848753.1	6	†
LOC367171 Microtubule-associated protein 4 isoform 1	IPI00421342.2	15	††
Map1b similar to Microtubule-associated protein 1B	IPI00372009.3	13	††
Mtap1a Microtubule-associated protein 1A	IPI00199693.2	14	
Mtap2 Isoform MAP2x of Microtubule-associated protein 2	IPI00206171.1, IPI00231051.1, IPI00328017.4	42	††

Mtap6 STOP protein	IPI00210119.1, IPI00734617.2	14	
Myh10 Myosin, heavy polypeptide 10	IPI00338604.4, IPI00391300.3, IPI00397526.2, IPI00479307.3, IPI00515398.1, IPI00757312.1, IPI00790503.2	4	
Myo5a Myosin-Va	IPI00118120.1, IPI00214038.1, IPI00390377.2, IPI00776221.1	5	†
NCKAP1 Nck-associated protein 1	IPI00031982.1, IPI00214442.2, IPI00319320.4, IPI00409684.2, IPI00656204.1, IPI00755241.1, IPI00766452.1	7	
Rad23b UV excision repair protein RAD23 homolog B	IPI00008223.3, IPI00108774.1, IPI00210495.1	23	††
RP1-14N1.3 Ifapsoriasin	IPI00397801.4, IPI00787398.1	3	
Snip SNAP25-interacting protein	IPI00190619.3	9	
Spna2 Spectrin alpha chain, brain	IPI00209258.4	4	
Spnb2 Isoform 1 of Spectrin beta chain, brain 1	IPI00319830.7, IPI00555287.2	51	††
SPTAN1 Spectrin alpha, non-erythrocytic 1	IPI00478292.3, IPI00744706.1, IPI00745092.1, IPI00843765.1, IPI00844215.1	3	
TUBA4A Tubulin alpha-4A chain	IPI00007750.1, IPI00794663.1	18	††
TUBB Tubulin beta chain	IPI00011654.2	16	†
TUBB2A Tubulin beta-2A chain	IPI00013475.1	26	†
TUBB2B Tubulin beta-2B chain	IPI00031370.3	155	
TUBB2C Tubulin beta-2C chain	IPI00007752.1	65	†
TUBB3 Tubulin beta-3 chain	IPI00013683.2	10	
TUBB4 Tubulin beta-4 chain	IPI00023598.2	42	†
Wasf1 WAS protein family member 1	IPI00022007.1, IPI00213598.1, IPI00471372.2	6	
Wdr1_predicted WD repeat protein 1	IPI00215349.5	8	
<b>Cellular Communication / Signal Transduction</b>			
Amph1 Amphiphysin	IPI00196508.1	3	
Ap2a1 Isoform A of AP-2 complex subunit alpha-1	IPI00108780.6, IPI00203346.4, IPI00567919.2, IPI00622911.1, IPI00764057.1, IPI00765430.1, IPI00778656.1	10	††
Ap2a2 AP-2 complex subunit alpha-2	IPI00310131.5, IPI00471901.3, IPI00753468.1	8	
Ap2b1 Isoform 1 of AP-2 complex subunit beta-1	IPI00119689.1, IPI00220991.2, IPI00231502.3, IPI00333383.2, IPI00378063.1, IPI00389753.1, IPI00784156.1, IPI00784366.1, IPI00790702.1	6	
Ap3b2_predicted Similar to Adaptor-related protein complex 3 beta 2 subunit	IPI00368200.2	6	††
Bsn Protein bassoon	IPI00212553.3, IPI00556925.1	98	††
Cadps Calcium-dependent secretion activator 1	IPI00199577.5, IPI00199604.4, IPI00297412.4, IPI00330163.3, IPI00374128.3, IPI00384808.2, IPI00478178.4, IPI00668903.1, IPI00670114.1, IPI00747400.1	13	
CAMK2A Isoform A of Calcium/calmodulin-	IPI00215715.3, IPI00550056.1	51	

dependent protein kinase type II alpha chain			
Camkv CaM kinase-like vesicle-associated protein	IPI00205056.1	5	
Coro1a Coronin-1A	IPI00210071.3	13	†
Crmp1 Crmp1 protein	IPI00312527.4, IPI00561065.2	8	
CSNK2A1 Casein kinase 2 alpha 1 polypeptide	IPI00016613.2, IPI00120162.1, IPI00192586.1, IPI00408176.2, IPI00744507.1	14	††
Ctnnd2 Isoform 1 of Catenin delta-2	IPI00136135.1, IPI00228632.1, IPI00553941.3	5	††
Cyfp2 Cytoplasmic FMR1-interacting protein 2	IPI00405625.9, IPI00719600.4, IPI00763802.1, IPI00769269.1, IPI00789699.2	10	
Dclk1 Isoform 1 of Serine/threonine-protein kinase DCLK1	IPI00468380.4, IPI00778626.1	8	
Dctn1 Dynactin subunit 1	IPI00196703.1	8	
Dpysl3 Dihydropyrimidinase-related protein 3	IPI00029111.2, IPI00122349.1, IPI00203250.1, IPI00556970.1	7	
Dpysl4 Similar to Dihydropyrimidinase-related protein 4	IPI00366087.1, IPI00558008.1, IPI00779982.1	9	
Dpysl5 Dihydropyrimidinase-related protein 5	IPI00331981.7	6	
Erc1;LOC100048600 Isoform 1 of ELKS/RAB6-interacting/CAST family member 1	IPI00117731.1, IPI00117733.1, IPI00171230.5, IPI00181684.4, IPI00201791.3, IPI00216719.1, IPI00331792.4, IPI00374976.1, IPI00457547.1, IPI00557326.1, IPI00558224.1	8	
Gdi1 Rab GDP dissociation inhibitor alpha	IPI00324986.1	52	
GIT1 Isoform 1 of ARF GTPase-activating protein GIT1	IPI00384861.3, IPI00470095.1, IPI00649373.1, IPI00795611.1	4	
Gnaq Guanine nucleotide binding protein alpha q polypeptide	IPI00228618.5, IPI00230868.4	4	
GNB1 Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta 1	IPI00026268.3, IPI00120716.3	3	
Gnb2l1 Guanine nucleotide-binding protein subunit beta 2-like 1	IPI00317740.5, IPI00641950.3, IPI00848226.1	5	
Homer1 Isoform 1 of Homer protein homolog 1	IPI00210570.1	6	
Jup Junction plakoglobin	IPI00229475.1, IPI00554711.2, IPI00789324.1	4	†
LOC315676 Similar to Dmx-like 2	IPI00369671.3	9	
LOC681252 Similar to Myristoylated alanine-rich C-kinase substrate	IPI00371946.3, IPI00480687.2	4	
LOC685144;LOC681927 Similar to SEC24 related gene family, member C isoform 3	IPI00365299.2, IPI00388782.2, IPI00763148.1, IPI00767454.1, IPI00769013.1	4	†
Ncdn NORBIN	IPI00205396.1, IPI00331299.9, IPI00549543.1, IPI00555661.1	11	
NSF Vesicle-fusing ATPase	IPI00006451.6, IPI00210635.2, IPI00656325.2	10	
Ogt UDP-N-acetylglucosamine - peptide N-acetylglucosaminyltransferase 110 kDa subunit	IPI00231503.4, IPI00420870.4, IPI00845528.1	10	††
Pacsin1 Protein kinase C and casein kinase substrate in neurons protein 1	IPI00208245.1	7	

Pclo Isoform 1 of Protein piccolo	IPI00203018.1, IPI00231831.1, IPI00758462.1	37	††
Picalm Isoform 2 of Phosphatidylinositol-binding clathrin assembly protein	IPI00194959.5	19	††
Plcb1 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta 1	IPI00192534.1, IPI00468121.1, IPI00558422.1	3	
Ppp1r12a Isoform 1 of Protein phosphatase 1 regulatory subunit 12A	IPI00183002.6, IPI00211695.1, IPI00397730.3, IPI00400680.1, IPI00400681.1, IPI00413191.2, IPI00779684.1	21	
Ppp3ca Isoform 1 of Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	IPI00121545.1, IPI00179415.4, IPI00201410.1, IPI00559849.1, IPI00747748.1, IPI00756703.1	17	
Prkacb Isoform 1 of cAMP-dependent protein kinase beta-catalytic subunit	IPI00263822.7, IPI00560492.1, IPI00742329.1, IPI00742400.1, IPI00742438.1	3	
Prkwnk1 Serine/threonine-protein kinase WNK1	IPI00200557.1, IPI00561348.1	8	††
Ptpn23 Protein tyrosine phosphatase non-receptor type 23	IPI00782007.1	23	
Rap1gds1_predicted Similar to RAP1, GTP-GDP dissociation stimulator 1	IPI00369496.3, IPI00763518.1, IPI00777342.1, IPI00778032.1	8	
Rapgef2_predicted Similar to Rap guanine nucleotide exchange factor 2	IPI00368346.3	7	
RGD1562629_predicted Similar to Protein neurobeachin	IPI00567941.2	5	
RGD1563580_predicted Similar to AP2 associated kinase 1	IPI00556943.2, IPI00559288.2, IPI00786812.1	8	
Rims1 Isoform 1 of Regulating synaptic membrane exocytosis protein 1	IPI00200893.1, IPI00206312.1, IPI00568548.2, IPI00780218.1	3	
Rph3a Rabphilin-3A	IPI00189927.1, IPI00389991.3	3	
Sec23ip Similar to Sec23 interacting protein	IPI00359906.2	7	
Sec311 Isoform 1 of Protein transport protein Sec31A	IPI00210147.2, IPI00515833.1	15	
Shank2 Isoform 2 of SH3 and multiple ankyrin repeat domains protein 2	IPI00231759.3, IPI00231761.1, IPI00400661.2, IPI00470293.3, IPI00475709.1	17	††
SNAP91 Isoform 1 of Clathrin coat assembly protein AP180	IPI00006612.2, IPI00122409.1, IPI00215134.1, IPI00230165.1, IPI00408269.4, IPI00646376.2, IPI00652215.1, IPI00653617.1	41	††
Syn1 Isoform IA of Synapsin-1	IPI00191335.1	4	††
Synj1 Similar to Synaptojanin-1	IPI00210153.3, IPI00229626.7, IPI00231602.2, IPI00850983.1	12	
Ywhab Isoform Long of 14-3-3 protein beta/alpha	IPI00230837.5, IPI00760126.1	4	
YWHAE 14-3-3 protein epsilon	IPI00000816.1	6	
YWHAG 14-3-3 protein gamma	IPI00220642.7	5	
Ywhaq Isoform 1 of 14-3-3 protein theta	IPI00408378.4, IPI00656269.1	7	†
<b>Intracellular Transport</b>			
ATP2A2 Isoform SERCA2A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	IPI00468900.4	9	
ATP6V0A1 Isoform 1 of Vacuolar proton	IPI00465178.5, IPI00743576.1,	15	†

translocating ATPase 116 kDa subunit A isoform 1	IPI00796045.1		
ATP6V1A Vacuolar ATP synthase catalytic subunit A	IPI00007682.2, IPI00373076.1, IPI00407692.3, IPI00844689.1	32	
Atp6v1b2 Vacuolar ATP synthase subunit B brain isoform	IPI00119113.3, IPI00199305.1	39	
Dnm1l Isoform 4 of Dynamin-1-like protein	IPI00193568.3, IPI00208284.3	9	
Gorasp2 Golgi reassembly stacking protein 2	IPI00362488.1	7	††
NAPA Alpha-soluble NSF attachment protein	IPI00009253.2, IPI00189925.1	4	
Nup153 Similar to Nuclear pore complex protein Nup153	IPI00480641.3, IPI00768316.1	5	††
Pacs1 Isoform PACS-1a of Phosphofurin acidic cluster sorting protein 1	IPI00324270.4	4	
SEPT5 Septin-5	IPI00017731.1, IPI00559449.2, IPI00655290.2	4	
Sept6_predicted 49 kDa protein	IPI00363930.4, IPI00420385.4, IPI00454142.5, IPI00454143.3, IPI00780333.1	3	
SEPT7 Isoform 1 of Septin-7	IPI00033025.8, IPI00204899.2, IPI00224626.3, IPI00816201.1	3	
Sept11 Isoform 3 of Septin-11	IPI00420385.4, IPI00454142.5	19	
Slc25a12 Calcium-binding mitochondrial carrier protein Aralar1	IPI00308162.3	15	
Slc25a4 ADP/ATP translocase 1	IPI00115564.5, IPI00231927.11, IPI00676622.1	4	
SLC25A5 ADP/ATP translocase 2	IPI00007188.5, IPI00127841.3, IPI00200466.3, IPI00363182.2, IPI00558425.2, IPI00565507.2	3	
Srprb Signal recognition particle receptor B subunit	IPI00196656.2, IPI00476177.2, IPI00679202.2	6	
VCP Transitional endoplasmic reticulum ATPase	IPI00022774.3, IPI00622235.5, IPI00676914.1	12	†
Vdac2 Voltage-dependent anion-selective channel protein 2	IPI00122547.1, IPI00198327.2	5	
<b>Metabolism / Biosynthesis</b>			
Acot7 Isoform B of Cytosolic acyl coenzyme A thioester hydrolase	IPI00125939.2, IPI00213571.1, IPI00230588.1, IPI00284094.4, IPI00326904.5, IPI00566122.1, IPI00672508.1	3	
Aldoa1 Fructose-bisphosphate aldolase	IPI00195851.1, IPI00221402.7, IPI00231734.5, IPI00465439.5, IPI00796333.1	4	
Aldoc Fructose-bisphosphate aldolase C	IPI00231736.9	15	
Atp5a1 ATP synthase subunit alpha, mitochondrial precursor	IPI00396910.1	11	†
ATP5B ATP synthase subunit beta, mitochondrial precursor	IPI00303476.1, IPI00551812.1	13	
Ctbp1 Isoform 1 of C-terminal-binding protein 1	IPI00128155.2, IPI00392657.1, IPI00754844.1, IPI00780254.1, IPI00845557.1	34	
Dlat Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial precursor	IPI00231714.3, IPI00765153.1	3	

Eno1 Alpha-enolase	IPI00462072.3, IPI00464815.11	40	†
Eno2 Gamma-enolase	IPI00326412.4	7	†
Fasn Fatty acid synthase	IPI00200661.1	4	†
Gda Guanine deaminase	IPI00325884.5, IPI00851130.1	16	
Glud1 Glutamate dehydrogenase 1, mitochondrial precursor	IPI00016801.1, IPI00027146.1, IPI00114209.1, IPI00324633.2, IPI00753095.1	8	
Glul Glutamine synthetase	IPI00324020.6, IPI00626790.2	10	
Got1 Aspartate aminotransferase, cytoplasmic	IPI00421513.8	11	
Got2 Aspartate aminotransferase, mitochondrial precursor	IPI00210920.1	17	
Gpi Glucose-6-phosphate isomerase	IPI00364311.1	3	
Hk1 Hexokinase-1	IPI00202543.1	22	
Hmgcs1 Hydroxymethylglutaryl-CoA synthase, cytoplasmic	IPI00188158.1	4	
IDH3A Isoform 1 of Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial precursor	IPI00030702.1, IPI00198720.1, IPI00459725.2	3	†
LOC316632 NADH dehydrogenase 1 alpha subcomplex 10-like protein	IPI00189759.1, IPI00561513.1	3	
LOC360975 2-oxoglutarate dehydrogenase E1 component, mitochondrial precursor	IPI00215093.1, IPI00390995.2, IPI00782594.1	3	
Ndufs1 NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial precursor	IPI00358033.1	56	
Ndufs2 NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial precursor	IPI00128023.3, IPI00471647.1, IPI00830766.1	3	
Oxr1 Similar to Oxidation resistance 1	IPI00199013.7, IPI00764149.1	3	
Pdha1 Pyruvate dehydrogenase E1 component alpha subunit somatic form, mitochondrial precursor	IPI00191707.4, IPI00337893.2, IPI00393034.3, IPI00764176.1, IPI00768086.2	3	
Pdhb Pyruvate dehydrogenase E1 component subunit beta, mitochondrial precursor	IPI00194324.2	7	
Pfkm 6-phosphofructokinase muscle type	IPI00331541.5	12	
Pfkp 6-phosphofructokinase type C	IPI00231954.5	4	
Pgam1 Phosphoglycerate mutase 1	IPI00421428.9, IPI00453476.2, IPI00457898.3, IPI00549725.6, IPI00740800.1	9	†
Pgk1 Phosphoglycerate kinase 1	IPI00231426.6, IPI00372910.2, IPI00555069.3	9	†
Phgdh D-3-phosphoglycerate dehydrogenase	IPI00225961.5, IPI00475835.3	3	†
Pkm2 Isoform M1 of Pyruvate kinase isozymes M1/M2	IPI00231929.6	11	†
Psat1 Phosphoserine aminotransferase	IPI00331919.5	6	
Pygb Glycogen phosphorylase brain form	IPI00229796.3, IPI00357945.1	6	
Taldo1 Transaldolase	IPI00124692.1, IPI00190377.2	3	
Tpi1 Triosephosphate isomerase	IPI00231767.5, IPI00339162.1	7	†
Tst Thiosulfate sulfurtransferase	IPI00366293.3, IPI00566218.1	3	
<b>mRNA / Protein Processing</b>			
Carm1 Isoform 1 of Histone-arginine methyltransferase CARM1	IPI00125950.2, IPI00279931.1, IPI00366497.3, IPI00412880.2, IPI00568674.2, IPI00639957.2, IPI00650083.2, IPI00655258.2,	5	

	IPI00830611.1		
Cct2 T-complex protein 1 subunit beta	IPI00366218.3	3	
Cct8_predicted Similar to T-complex protein 1 subunit theta	IPI00370815.3	28	†
Fbxo2 F-box only protein 2	IPI00153176.2, IPI00209303.1	3	
Fkbp4 Similar to FK506-binding protein 4	IPI00358443.3, IPI00767393.1	3	
HNRPA1 Isoform A1-B of Heterogeneous nuclear ribonucleoprotein A1	IPI00215965.2, IPI00224251.5, IPI00465365.4, IPI00553777.2, IPI00748262.1, IPI00797148.1	9	†
Hnrpa2b1_predicted Heterogeneous nuclear ribonucleoproteins A2/B1	IPI00212969.2, IPI00358211.3, IPI00396378.3, IPI00405058.6, IPI00414696.1, IPI00622847.2, IPI00828488.1, IPI00853914.1	7	†
Hnrpa3 Isoform 1 of Heterogeneous nuclear ribonucleoprotein A3	IPI00269661.1, IPI00269662.1, IPI00419373.1, IPI00455134.1, IPI00459722.2, IPI00461800.1, IPI00466185.3, IPI00470076.5, IPI00623731.1, IPI00660502.1, IPI00664047.1, IPI00664791.1	12	†
Hnrpc Heterogeneous nuclear ribonucleoprotein C	IPI00130343.2, IPI00187860.3, IPI00216592.2, IPI00223443.1, IPI00223444.1, IPI00477313.3, IPI00759596.1, IPI00759870.1, IPI00759886.1, IPI00781839.1	4	
Hnrpk Hnrpk protein	IPI00194974.2, IPI00216049.1, IPI00216746.1, IPI00223253.1, IPI00224575.1, IPI00514561.1, IPI00777007.1, IPI00780608.1, IPI00807545.1	37	†
Hnrpul2 Heterogeneous nuclear ribonucleoprotein U-like protein 2	IPI00222208.2, IPI00360386.3, IPI00561756.2, IPI00565127.2, IPI00756515.1, IPI00849047.1	4	
Hsp110 Isoform HSP105-alpha of Heat shock protein 105 kDa	IPI00123802.5, IPI00224109.2, IPI00471835.1, IPI00568014.2, IPI00778569.1, IPI00779326.1, IPI00830204.1	3	
Hspa12a_predicted Similar to Heat shock protein 12A	IPI00358537.2	6	
Hspa4 Heat shock 70 kDa protein 4	IPI00387868.2	10	††
Hspd1 Isoform 1 of 60 kDa heat shock protein, mitochondrial precursor	IPI00308885.6, IPI00339148.2, IPI00472102.3, IPI00763910.1, IPI00784154.1, IPI00790763.1	36	††
Hsph1 Heat shock protein 105 kDa	IPI00218993.1, IPI00471835.1, IPI00513743.1, IPI00514983.3	8	
Huwe1 HECT, UBA and WWE domain containing 1	IPI00463909.3, IPI00655012.2	37	
NPEPPS Puromycin-sensitive aminopeptidase	IPI00026216.4, IPI00130000.1, IPI00372700.1, IPI00608097.1, IPI00767572.1, IPI00768609.1	8	†
Otub1;LOC100046081 Ubiquitin thioesterase OTUB1	IPI00154004.1, IPI00371462.3, IPI00755837.1	4	
PABPC1 Isoform 1 of Polyadenylate-binding protein 1	IPI00008524.1, IPI00124287.1, IPI00189074.3, IPI00331552.4,	6	†

	IPI00410017.1, IPI00478522.1, IPI00796945.1		
PCBP2 Poly(rC)-binding protein 2 isoform b	IPI00012066.2, IPI00127707.1, IPI00216689.2, IPI00221796.1, IPI00221799.1, IPI00470509.2, IPI00796337.1	4	†
Pdia3 Protein disulfide-isomerase A3 precursor	IPI00324741.2	9	
Rbm12 Swan	IPI00421433.1, IPI00560597.1	12	
Rbmx Heterogeneous nuclear ribonucleoprotein G	IPI00124979.2, IPI00304692.1, IPI00370207.3, IPI00474144.1, IPI00559910.1, IPI00604873.2, IPI00663587.1, IPI00763272.1, IPI00766882.1, IPI00775821.1, IPI00775899.1	6	
Sf3a1_predicted Similar to Splicing factor 3 subunit 1	IPI00215030.1, IPI00408796.3	5	
SFPQ Isoform Long of Splicing factor, proline- and glutamine-rich	IPI00010740.1, IPI00129430.1, IPI00627068.1, IPI00752791.1, IPI00755611.1, IPI00767277.1, IPI00849080.1	3	†
Thop1 Thimet oligopeptidase 1	IPI00564198.2	4	
Ubqln2_predicted Similar to ubiquilin 2	IPI00362791.3	66	
Uqcrc1 Ubiquinol-cytochrome-c reductase complex core protein 1, mitochondrial precursor	IPI00471577.1	6	
Uqcrc2 Ubiquinol-cytochrome-c reductase complex core protein 2, mitochondrial precursor	IPI00188924.4	3	
USP5 Isoform Long of Ubiquitin carboxyl-terminal hydrolase 5	IPI00024664.1, IPI00207657.1, IPI00375145.1, IPI00767186.1, IPI00768802.1	4	
<b>Transcription / Translation</b>			
Hcfc1_predicted Similar to Host cell factor C1	IPI00367724.3, IPI00765252.1	13	††
CAND1 Isoform 1 of Cullin-associated NEDD8-dissociated protein 1	IPI00100160.3, IPI00205466.1, IPI00420562.5, IPI00746694.1, IPI00753059.1	3	
CNOT1 CCR4-NOT transcription complex subunit 1 isoform A	IPI00166010.6, IPI00359049.4, IPI00673465.1, IPI00674283.1, IPI00752506.1, IPI00757812.1	17	††
DDX17 DEAD box polypeptide 17 isoform 1	IPI00023785.6, IPI00396797.2, IPI00651653.1, IPI00651677.1, IPI00653307.1	4	
DDX5 Probable ATP-dependent RNA helicase DDX5	IPI00017617.1, IPI00420363.2, IPI00464718.1	3	
Eef1a1 Elongation factor 1-alpha 1	IPI00195372.1, IPI00307837.5, IPI00396485.3, IPI00472724.1, IPI00551729.1	5	†
EEF2 Elongation factor 2	IPI00186290.6, IPI00203214.6, IPI00466069.3, IPI00849291.1	4	††
EG268795 Similar to 60S ribosomal protein L7a (Surfeit locus protein 3) isoform 1	IPI00265107.4, IPI00299573.12, IPI00330363.8, IPI00354363.3, IPI00397676.4, IPI00462006.3,	3	

	IPI00462453.4, IPI00478896.2, IPI00479315.2, IPI00622160.3		
Eif4a2 Eukaryotic initiation factor 4A-II	IPI00193595.3, IPI00328328.3, IPI00400432.2, IPI00409717.1, IPI00409918.1	3	
Eif4g3_predicted Similar to Eukaryotic translation initiation factor 4 gamma 3	IPI00365284.3, IPI00767350.1	6	
pur-beta Transcriptional activator protein Pur-beta	IPI00189358.2	3	
RGD1560833_predicted Similar to MKL/myocardin-like 2	IPI00765655.1	2	
Ripx Protein RUFY3	IPI00204065.1, IPI00206350.3	3	
RPS3 40S ribosomal protein S3	IPI00011253.3, IPI00134599.1, IPI00212776.1	4	†
RPS8 40S ribosomal protein S8	IPI00216587.9, IPI00231202.6, IPI00274175.1, IPI00466820.4, IPI00475203.1, IPI00621229.1, IPI00645201.1, IPI00671398.1, IPI00756488.1, IPI00756959.1, IPI00828628.1, IPI00849948.1	3	†
VeZF1_predicted 22 kDa protein	IPI00780927.1	4	
Zfr Similar to Zinc finger RNA binding protein	IPI00367952.3, IPI00765814.1	8	††
<b>Uncharacterized</b>			
HnrpU2 Heterogeneous nuclear ribonucleoprotein U-like protein 2	IPI00222208.2, IPI00360386.3, IPI00561756.2, IPI00565127.2, IPI00756515.1, IPI00849047.1	4	
Immt 82 kDa protein	IPI00364895.4, IPI00566985.1, IPI00777695.1	8	
LOC314432 Similar to Ubiquitin-protein ligase (EC 6.3.2.19) E1-mouse	IPI00368347.2	7	
LOC501546 LOC501546 protein	IPI00201213.3	3	
MGC93707 Mitochondrial antiviral-signaling protein	IPI00364200.1	3	
RGD1562348_predicted Similar to Ankyrin repeat domain protein 17 isoform B	IPI00361795.2, IPI00388314.3	9	
RGD1563977_predicted Similar to Protein 4.1G	IPI00191995.2, IPI00192909.2, IPI00368431.2, IPI00388101.1, IPI00393242.1	3	
RGD1566064_predicted Similar to HBxAg transactivated protein	IPI00363856.3	13	
SH3GLB2 Isoform 1 of SH3 domain GRB2-like protein B2	IPI00024540.3, IPI00153832.1, IPI00398828.1, IPI00626834.2, IPI00756786.1, IPI00776533.1, IPI00779094.1, IPI00828453.1	16	
Ubap2_predicted Similar to Ubiquitin-associated protein 2	IPI00190431.4	13	
Ubap2I Isoform 5 of Ubiquitin-associated protein 2-like	IPI00407835.1, IPI00412535.2, IPI00514856.4, IPI00761937.1	6	

**Table 2:** Promoters enriched in *O*-GlcNAc levels. The corresponding gene name and description for each promoter are listed. The location of the promoter and the log ratio of the chromatin immunoprecipitation eluent over input are also listed.

Gene Name	Description	Location	LogRatio
Rfwd2	Ring finger and WD repeat domain 2	chr1:161066116-161066175	1.12E+00
Arl6ip1	ADP-ribosylation factor-like 6 interacting protein 1	chr7:117911348-117911397	1.05E+00
Gpr21	G protein-coupled receptor 21	chr2:037340628-037340687	1.01E+00
Bhlhb5	Basic helix-loop-helix family, member e22	chr3:018243837-018243896	9.98E-01
Pde6b	Phosphodiesterase 6B, cGMP, rod receptor, beta polypeptide	chr5:108630597-108630649	9.94E-01
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	chr6:125131691-125131742	9.51E-01
Slc4a4	Solute carrier family 4 (anion exchanger), member 4	chr5:090007481-090007540	9.05E-01
Kcnab1	Potassium voltage-gated channel, shaker-related subfamily, beta member 1	chr3:065195669-065195719	8.92E-01
Hspb7	Heat shock protein family, member 7 (cardiovascular)	chr4:140694790-140694840	8.88E-01
Phf13	PHD finger protein 13	chr4:150833827-150833886	8.73E-01
Hivep3	Human immunodeficiency virus type 1 enhancer binding protein 3	chr4:119307364-119307417	8.58E-01
Ifnb1	Interferon beta 1, fibroblast	chr4:087991861-087991916	8.49E-01
Spg20	Spastic paraplegia 20 (Troyer syndrome)	chr3:055222441-055222497	8.46E-01
Pin1	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1	chr9:020401004-020401061	8.46E-01
Tmc7	Transmembrane channel-like gene family 7	chr7:118377551-118377610	8.31E-01
Trp53inp2	Tumor protein p53 inducible nuclear protein 2	chr2:155073250-155073294	8.30E-01
Sgcb	Sarcoglycan, beta (dystrophin-associated glycoprotein)	chr5:073912548-073912601	8.27E-01
NC2_00099332	Unknown	NC2_00099332	8.24E-01
Mmp8	Matrix metalloproteinase 8	chr9:007558381-007558437	8.19E-01
Kcnj14	Potassium inwardly-rectifying channel, subfamily J, member 14	chr7:045690923-045690982	8.11E-01
Pde3a	Phosphodiesterase 3A, cGMP inhibited	chr6:141210944-141210995	8.07E-01
Ccdc106	Coiled-coil domain containing 106	chr7:004654767-004654811	8.00E-01
Gpsm1	G-protein signaling modulator 1 (AGS3-like, <i>C. elegans</i> )	chr2:026137783-026137827	7.86E-01
NC2_00099332	Unknown	NC2_00099332	7.81E-01
Snrp70	Small nuclear ribonucleoprotein 70	chr7:045263625-045263669	7.72E-01
Aspm	Asp (abnormal spindle)-like, microcephaly	chr1:141271216-	7.69E-01

	associated	141271266	
NC2_00099332	Unknown	NC2_00099332	7.58E-01
Ltb4dh	Prostaglandin reductase 1	chr4:059078901-059078954	7.55E-01
NC2_00099332	Unknown	NC2_00099332	7.50E-01
Rexo4-Adamts13	Unknown	chr2:026795163-026795217	7.49E-01
Aox3	Aldehyde oxidase 3	chr1:058058797-058058847	7.48E-01
NC2_00099332	Unknown	NC2_00099332	7.40E-01
Sema3c	Semaphorin 3c	chr5:017086294-017086338	7.31E-01
Gucy2c	Guanylate cyclase 2c	chr6:136750955-136751011	7.28E-01
Krtap5-5	Keratin associated protein 5-5	chr7:142043215-142043274	7.20E-01
Itn2c	Integral membrane protein 2C	chr1:087720489-087720548	7.12E-01
Kcnt1	Potassium channel, subfamily T, member 1	chr2:025700956-025701003	7.10E-01
Fbxl13	F-box and leucine-rich repeat protein 13	chr5:021054005-021054064	7.04E-01
Snrpn	Small nuclear ribonucleoprotein N	chr7:059883275-059883319	6.96E-01
2600010E01Rik	Proline rich 5 like	chr2:101598189-101598234	6.95E-01
Myo3b	Myosin IIIB	chr2:070085726-070085785	6.74E-01
Lrch4	Leucine-rich repeats and calponin homology (CH) domain	chr5:137860009-137860054	6.71E-01
Bai3	Brain-specific angiogenesis inhibitor 3	chr1:025776261-025776316	6.67E-01
Il11	Interleukin 11	chr7:004383274-004383318	6.61E-01
Gm1040	Nucleolar protein with MIF4G domain 1	chr5:029766349-029766408	6.61E-01
1200015F23Rik-Gchfr	Unknown	chr2:118854850-118854901	6.59E-01
Coq10b	Coenzyme Q10 homolog B	chr1:054999769-054999828	6.50E-01
Tas2r139	Taste receptor, type 2, member 139	chr6:042071031-042071090	6.47E-01
Rab23	RAB23, member RAS oncogene family	chr1:033664505-033664560	6.47E-01
Ralb	V-ral simian leukemia viral oncogene homolog B	chr1:121334415-121334464	6.44E-01
BC026590	Unknown	chr4:056900096-056900155	6.43E-01
Ndufs2	NADH dehydrogenase (ubiquinone) Fe-S protein 2	chr1:173071744-173071801	6.43E-01
chr6:054357273-054357332	Unknown	chr6:054357273-054357332	6.40E-01
Klk1b11	Kallikrein 1-related peptidase b11	chr7:043860709-043860757	6.40E-01
Chn1	Chimerin (chimaerin) 1	chr2:073460953-	6.38E-01

		073461012	
Fn1	Fibronectin 1	chr1:071585152-071585211	6.37E-01
Hrh3	Histamine receptor H3	chr2:180037363-180037418	6.37E-01
Arid1a	AT rich interactive domain 1A (SWI-like)	chr4:133027660-133027704	6.37E-01
Seprn1	Selenoprotein N, 1	chr4:133817519-133817563	6.36E-01
4931406C07Rik	Unknown	chr9:015049879-015049938	6.35E-01
Kcne3	Potassium voltage-gated channel, Isk-related subfamily, gene 3	chr7:100047744-100047792	6.34E-01
Prss3-EG436523	Unknown	chr6:041317008-041317067	6.32E-01
Sult2b1	Sulfotransferase family, cytosolic, 2B, member 1	chr7:045626168-045626227	6.30E-01
Nanos2	Nanos homolog 2 (Drosophila)	chr7:018146282-018146328	6.25E-01
Nadk	NAD kinase	chr4:154404524-154404575	6.24E-01
Scly	Selenocysteine lyase	chr1:093150219-093150267	6.20E-01
EG229862	Unknown	chr3:137487520-137487579	6.17E-01
Olf666	Olfactory receptor 666	chr7:104767985-104768044	6.14E-01
D430042O09Rik	Unknown	chr7:125502534-125502585	6.12E-01
Lmo4	LIM domain only 4	chr3:144147353-144147412	6.00E-01
Iqsec3	IQ motif and Sec7 domain 3	chr6:121437055-121437114	5.99E-01
chr7:062089379-062089438	Unknown	chr7:062089379-062089438	5.91E-01
9130221D24Rik	Unknown	chr3:133170194-133170253	5.91E-01
Sh2d2a	SH2 domain protein 2A	chr3:087934416-087934460	5.90E-01
Tmcc2	Transmembrane and coiled-coil domains 2	chr1:134217724-134217783	5.88E-01
Ccdc9	Coiled-coil domain containing 9	chr7:015445399-015445449	5.86E-01
Capn5	Calpain 5	chr7:098018304-098018355	5.80E-01
Rab43	RAB43, member RAS oncogene family	chr6:087781399-087781454	5.73E-01
2610209A20Rik	Lipoyl(octanoyl) transferase 2 (putative)	chr7:100034032-100034090	5.71E-01
Csn1s2b	Casein alpha s2-like B	chr5:088895459-088895518	5.68E-01
Mag	Myelin-associated glycoprotein	chr7:030626049-030626105	5.68E-01
Cort	Cortistatin	chr4:147968893-147968944	5.66E-01

F2r13	Coagulation factor II (thrombin) receptor-like 3	chr8:075653052-075653097	5.64E-01
Tlx2	T-cell leukemia, homeobox 2	chr6:083034418-083034462	5.60E-01
2310038H17Rik	Unknown	chr1:064549300-064549347	5.59E-01
C1qc	Complement component 1, q subcomponent, C chain	chr4:136162907-136162966	5.58E-01
Madd	MAP-kinase activating death domain	chr2:090983205-090983264	5.57E-01
Kdelc1	KDEL (Lys-Asp-Glu-Leu) containing 1	chr1:044046571-044046615	5.55E-01
Inpp1	Inositol polyphosphate-1-phosphatase	chr1:052761859-052761903	5.54E-01
Gm973	Predicted gene 973	chr1:059459724-059459783	5.53E-01
Fhod1	Formin homology 2 domain containing 1	chr8:108237017-108237065	5.51E-01
Gpr109a	G protein-coupled receptor 109A	chr5:124125079-124125129	5.51E-01
Cnr2	Cannabinoid receptor 2 (macrophage)	chr4:135181859-135181915	5.50E-01
Syt6	Synaptotagmin VI	chr3:103706243-103706302	5.47E-01
Rpn1	Ribophorin I	chr6:088050301-088050345	5.47E-01
Stoml3	Stomatin (Epb7.2)-like 3	chr3:053572971-053573030	5.46E-01
Hdac4	Histone deacetylase 4	chr1:093978609-093978663	5.45E-01
Tesk2	Testis-specific kinase 2	chr4:116220557-116220616	5.44E-01
Isg2011	Apoptosis enhancing nuclease	chr7:078761115-078761174	5.42E-01
Il8rb	Interleukin 8 receptor, beta	chr1:074089611-074089670	5.42E-01
Cpsf3l	Cleavage and polyadenylation specific factor 3-like	chr4:154732636-154732695	5.41E-01
Dbnnd2	Dysbindin (dystrobrevin binding protein 1) domain containing 2	chr2:164183228-164183287	5.39E-01
Yipf7-Guf1	Unknown	chr5:069830810-069830869	5.37E-01
2010315L10Rik	Vesicle transport protein USE1 isoform 3	chr8:074298265-074298309	5.37E-01
Ank1	Ankyrin 1, erythroid	chr8:024519526-024519585	5.37E-01
Igsf21	Immunoglobulin superfamily, member 21	chr4:139519948-139519992	5.36E-01
Calcr	Calcitonin receptor	chr6:003719766-003719819	5.35E-01
Siglec1	Sialic acid binding Ig-like lectin 1, sialoadhesin	chr2:130780557-130780616	5.33E-01
Nmur1	Neuromedin U receptor 1	chr1:088217036-088217084	5.32E-01
Shroom3	Shroom family member 3	chr5:093758832-	5.30E-01

		093758876	
Clic3	Chloride intracellular channel 3	chr2:025276329-025276385	5.29E-01
Lim2	Lens intrinsic membrane protein 2	chr7:043296801-043296855	5.29E-01
Otud6b	OTU domain containing 6B	chr4:014753291-014753341	5.29E-01
A930008G19Rik	Family with sequence similarity 53, member B	chr7:132651633-132651677	5.29E-01
Chrm4	Cholinergic receptor, muscarinic 4	chr2:091728313-091728369	5.27E-01
chr4:102789696-102789755	Unknown	chr4:102789696-102789755	5.27E-01
Cblc	Casitas B-lineage lymphoma c	chr7:018955195-018955239	5.24E-01
Gem	GTP binding protein (gene overexpressed in skeletal muscle)	chr4:011628685-011628729	5.24E-01
Cryge	Crystallin, gamma E	chr1:064986389-064986435	5.21E-01
Cyp26b1	Cytochrome P450, family 26, subfamily b, polypeptide 1	chr6:084559974-084560027	5.21E-01
Ceacam9	Carcinoembryonic antigen-related cell adhesion molecule 9	chr7:015875265-015875309	5.20E-01
Mcm3	Minichromosome maintenance complex component 3	chr1:020804983-020805042	5.20E-01
Il15ra	Interleukin 15 receptor, alpha	chr2:011624220-011624274	5.19E-01
Slmo2	Slowmo homolog 2	chr2:174114425-174114484	5.17E-01
Calr3	Calreticulin 3	chr8:075372337-075372395	5.17E-01
Slc6a17	Solute carrier family 6, member 17	chr3:107651018-107651077	5.14E-01
Gatad2a	GATA zinc finger domain containing 2A	chr8:072924369-072924421	5.14E-01
Pax3	Paired box gene 3	chr1:078083580-078083635	5.13E-01
Aldh1b1	Aldehyde dehydrogenase 1 family, member B1	chr4:045817830-045817881	5.12E-01
Col27a1	Collagen, type XXVII, alpha 1	chr4:062702912-062702961	5.12E-01
Casp6	Caspase 6, apoptosis-related cysteine peptidase	chr3:129888382-129888441	5.11E-01
Tspan2	Tetraspanin 2	chr3:102864758-102864809	5.11E-01
Wdr54	WD repeat domain 54	chr6:083121238-083121294	5.10E-01
Defb9	Defensin beta 9	chr8:023352585-023352644	5.10E-01
Slc39a10	Solute carrier family 39 (zinc transporter), member 10	chr1:046797981-046798031	5.09E-01
Fzd7	Frizzled homolog 7	chr1:059426879-059426925	5.09E-01
Cyb5r2	Cytochrome b5 reductase 2	chr7:107550066-107550125	5.08E-01

Lrp1b	Low density lipoprotein receptor-related protein 1B	chr2:042476511-042476570	5.08E-01
Lcn8	Lipocalin 8	chr2:025476940-025476998	5.07E-01
Pcdh7	Protocadherin 7	chr5:058004012-058004063	5.07E-01
Smad1	SMAD family member 1	chr8:082300818-082300874	5.06E-01
Gnb2	Guanine nucleotide binding protein (G protein), beta 2	chr5:137761013-137761064	5.06E-01
Slc25a34	Solute carrier family 25, member 34	chr4:140895902-140895946	5.06E-01
Trim46	Tripartite motif-containing 46	chr3:089329908-089329967	5.06E-01
Melk	Maternal embryonic leucine zipper kinase	chr4:044321192-044321251	5.05E-01
Olf1336	Olfactory receptor 1336	chr7:006059882-006059930	5.04E-01
Olf71	Olfactory receptor 71	chr4:043732351-043732410	5.04E-01
Casc4	Cancer susceptibility candidate 4	chr2:121559691-121559750	5.04E-01
Slc25a40	Solute carrier family 25, member 40	chr5:008432144-008432203	5.03E-01
Gpr175	G protein-coupled receptor 175	chr6:088869789-088869848	5.03E-01
C430003P19Rik	BRISC complex subunit Abro1	chr7:132714914-132714973	5.02E-01
Qrfp	Pyroglutamylated RFamide peptide	chr2:031634928-031634987	5.01E-01
Abca2	ATP-binding cassette, sub-family A (ABC1), member 2	chr2:025251853-025251912	5.00E-01

## Methods

### General Reagents and Methods:

Unless otherwise noted, reagents were purchased from the commercial suppliers Fisher (Fairlawn, NJ) or Sigma-Aldrich (St. Louis, MO). Protease inhibitors were purchased from Roche Applied Sciences (Indianapolis, IN), sequencing grade trypsin was from Promega (Madison, WI), agarose-conjugated protein G was from Pierce (Rockford, IL), and Immobilon-FL PDVF membrane was from Millipore (Billerica, MA). Dulbecco's modified Eagle media (DMEM), B27, fetal bovine serum (FBS) and penicillin/streptomycin were from Gibco (Carlsbad, CA). The anti- $\alpha$ -crystallin (ab5595) and anti- $\beta$ -tubulin antibodies were from Abcam and Sigma, respectively. Click-It™ *O*-GlcNAc Enzymatic Labeling System, Click-It™ Biotin Glycoprotein Detection Kit, Click-It™ Tetramethylrhodamine (TAMRA) Glycoprotein Detection Kit, anti-TAMRA antibody, 4-12% NuPAGE® Bis-Tris Mini gels, pH 4-7 Zoom IPG strips, and lithium dodecyl sulfate (LDS) buffer were from Invitrogen (Carlsbad, CA). Peptide N-glycosidase F (PNGase F) was purchased from New England Biolabs (Beverly, MA). The anti-*O*-GlcNAc antibodies CTD110.6 and RL-2 were from Covance (Princeton, NJ) and Affinity Bioreagents (Golden, CO), respectively. Wheat germ agglutinin (WGA) lectin was from EY laboratories (San Mateo, CA). The secondary goat anti-rabbit antibody conjugated to IRDye800 was from Rockland Immunochemicals (Gilbertsville, PA), and the streptavidin-IR680 conjugate was from Li-COR Biosciences (Lincoln, NE). Sprague-

Dawley and Long Evans rats were from Charles River Laboratories (Wilmington, MA). All protein concentrations were measured using the BCA protein assay (Pierce). Western blots were visualized and quantified using an Odyssey infrared imaging system (LI-COR Biosciences). In-gel fluorescence detection was performed using a FujiFilm FLA-3000 or FLA-5100 scanner, and the fluorescence was displayed in green pseudocolor.

**Chemoenzymatic Labeling of  $\alpha$ -Crystallin.**  $\alpha$ -Crystallin from the Click-It™ *O*-GlcNAc Enzymatic Labeling System (20  $\mu$ g) was labeled with UDP-GalNAz **1** and biotin alkyne **2** as per the Click-It™ *O*-GlcNAc Enzymatic Labeling System and Click-It™ Biotin Glycoprotein Detection Kit instructions. Negative controls were performed under identical conditions, except GalT, **1**, or **2** were left out of the reactions.  $\alpha$ -Crystallin (10 pmol from each reaction) was resolved on a 1.5 mm, 10-well NuPAGE 4-12% Bis-Tris gel and transferred to PDVF. The membrane was blocked with 5% BSA in 50 mM Tris-HCl pH 7.4, 150 mM NaCl containing 0.1% Tween (TBST) for 1 h at RT followed by 1 h incubation with a streptavidin-IR680 conjugate (1:10,000) in TBST. After four washes for 15 min in TBST, the membrane was visualized using an Odyssey imaging system. The same membrane was then blotted with an anti- $\alpha$ -crystallin antibody (1:1000) in 5% nonfat milk / TBST for 1 h at RT. Following three washes in TBST for 5 min, the membrane was incubated with a goat anti-rabbit antibody conjugated to IRDye800 (1:10,000) in the same buffer for 1 h at RT, washed three more times for 10 min, and then visualized using an Odyssey imaging system. To examine the detection sensitivity, 25, 10, 5, 1, 0.5, and 0.25 pmol of labeled  $\alpha$ -crystallin were resolved by SDS-PAGE,

transferred to PDVF, and immunoblotted with a streptavidin-IR680 conjugate (1:10,000). For comparison, unlabeled  $\alpha$ -crystallin (25 pmol) was resolved by SDS-PAGE, transferred to PDVF, and immunoblotted with the CTD110.6 antibody (1:1000), RL-2 antibody (1:1000), or WGA lectin (10  $\mu$ g/mL) for 1 h at RT.

**Chemoenzymatic Labeling of Rat Forebrain Extracts.** The forebrains from three adult rats (~150 g, male Sprague Dawley) were extracted and fractionated using the Qproteome Cell Compartment Kit (Qiagen). Fractions 1 (the cytoplasmic fraction) and 3 (the nuclear fraction) were precipitated with 4 volumes ice-cold acetone followed by overnight incubation at -20 °C and redissolved in 2% SDS plus Complete™ protease inhibitors. Protein from each fraction (1.5 mg for the 2D gel labeling, 0.5 mg for the 1D gel labeling) was precipitated and labeled at 1 mg/mL as per the Click-It™ *O*-GlcNAc Enzymatic Labeling System instructions except that Complete™ protease inhibitors were added during the labeling reaction, followed by labeling with the TAMRA-alkyne dye **3** as per the Click-It™ TAMRA Glycoprotein Detection Kit instructions except that EDTA-free Complete™ protease inhibitors were added during the TAMRA labeling reaction. For the 1D gels, negative controls were performed under identical conditions except that GalT was omitted from the labeling reaction.

**Immunoprecipitation of TAMRA-Labeled *O*-GlcNAc Proteins.** Labeled samples were precipitated using methanol/chloroform/water, brought up to a concentration of 2 mg/mL in 1% SDS plus Complete™ protease inhibitors, and boiled. The SDS was then quenched

with 1 volume of NETFD buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM EDTA, 6% NP-40) plus protease inhibitors, and the lysate was precleared against washed protein G sepharose beads (1 mL/1.5 mg of protein) at 4 °C for 1 h. After centrifugation, the supernatant was collected and incubated with an anti-TAMRA antibody (100 µg/1.5 mg of protein) at 4 °C for 4 h. The samples were then added to pre-washed protein G sepharose beads (1 mL/1.5 mg of protein) at 4 °C for 1.5 h. Following centrifugation, the beads were washed once with 4 column volumes of NETFD buffer and three times with 4 column volumes of NETF buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM EDTA). After washing, the beads were boiled in elution buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 50 µL buffer/100 µL beads). The supernatant was collected after centrifugation and precipitated by adding 4 volumes of ice-cold acetone and incubating at -20 °C for 16 h.

**1D Gel Electrophoresis and Silver Staining.** The precipitated eluents (cytoplasmic, nuclear, and controls) from above, along with the input (before immunoprecipitation) and flow-through fractions (15 µg) were separated on a 1.5 mm, 10-well NuPAGE 4–12% Bis-Tris gel. The gels were imaged using a FujiFilm FLA-3000 or FLA-5100 scanner and silver stained using a protocol adapted from Blum, Shevchenko, and co-workers<sup>21,22</sup>. Briefly, the gels were fixed in an aqueous solution of 50% MeOH, 10% acetic acid for 30 min and then again in 5% MeOH, 1% acetic acid for 15 min. The gels were then washed 3 x 10 min with H<sub>2</sub>O and sensitized for 90 s with Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>•5H<sub>2</sub>O (20 mg/100 mL). After rinsing for 3 x 30 sec with H<sub>2</sub>O, the gels were exposed to AgNO<sub>3</sub> (200 mg/100 mL) for 30

min and rinsed for 3 x 60 s with H<sub>2</sub>O. Finally, the gels were developed for 2.5 min in a solution containing Na<sub>2</sub>CO<sub>3</sub> (6 g/100 mL), 37% formaldehyde (50 µL/100 mL), Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>•5H<sub>2</sub>O (0.4 mg/100 mL). The reaction was stopped with 6% acetic acid. Twelve equally-spaced gel pieces were excised from each of the eluent lanes (cytoplasmic, nuclear, and –GalT controls), spanning the full height of the gel. Individual gel pieces were destained in a solution containing 0.4 g K<sub>3</sub>Fe(CN)<sub>6</sub> in 200 mL of an aqueous sodium thiosulphate solution (0.2 g Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>•5H<sub>2</sub>O in 1L of H<sub>2</sub>O) for 15 min, and washed 4 times for 15 min and 1 time for 16 h with H<sub>2</sub>O.

**2D Gel Electrophoresis.** Precipitated eluents (cytoplasmic, nuclear) were resuspended in 100 mM Tris, pH 8.0, 1% SDS, and then reduced and alkylated with tributyl phosphine (200 mM) and *N,N*-Dimethylacrylamide (0.5%) by heating at 65° C for 10 min, followed by rotation end-over-end at RT for 20 min. The samples were precipitated with methanol/chloroform/water and resuspended in 7 M urea, 2 M thiourea, 2% CHAPS, 2% ASB-14 (Sigma) buffer (173.5 µL), and 2 M DTT (5.5 µL) plus pH 4–7 ampholytes (1 µL) were added. The samples were centrifuged at 20,000 rpm for 3 min, the supernatant was loaded onto pH 4–7 strips, and the sample was rehydrated for 90 min. The strips were focused for 20 min at 200V, 25 min at 450V, 20 min at 700V, and 55 min at 2000V, after which they were incubated in 1x LDS sample buffer plus 50 mM DTT, and resolved on a NuPAGE 4–12% Bis-Tris gel. The gel was imaged using a fluorescence scanner, and the fluorescent spots were excised from the gel and fixed in an aqueous solution of 50% MeOH, 7% acetic acid overnight.

**In-Gel Digestion of Captured *O*-GlcNAc Proteins.** Individual gel pieces (cytoplasmic, nuclear, and –GalT controls) from the 1D and 2D gels were dehydrated with CH<sub>3</sub>CN (2 x 5 min) and then rehydrated with dithiothreitol (1.5 mg/mL in 100 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 8.0) for 30 min. The excess dithiothreitol was removed and iodoacetamide (10 mg/mL in 100 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 8.0) was added in the dark for 30 min. Excess iodoacetamide was removed and the gels were washed twice with 100 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 8.0 and dried with CH<sub>3</sub>CN before being dried using a speed vac. Trypsin (20 ng/μL in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 8.0; 50 μL) was added to each gel piece, and the gel pieces were allowed to swell on ice. After 30 min, excess trypsin was removed, 50 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 8.0 (15 μL for the 2D gel pieces; 30 μL for the 1D gel pieces) was added, and the digestions were incubated at 37 °C. After 16 h, the peptides were extracted with H<sub>2</sub>O (30 μL for the 2D gel pieces; 60 μL for the 1D gel pieces) for 30 min, and the gel pieces were washed twice with an aqueous solution of 5% formic acid containing 50% CH<sub>3</sub>CN (25 μL for the 2D gel pieces; 40 μL for the 1D gel pieces) for 10 min. The combined extract and washes were concentrated using a speed vac for 1 h to remove the CH<sub>3</sub>CN.

**LC-MS Analysis of Captured *O*-GlcNAc Proteins.** Nano LC-MS of in-gel tryptic digests was performed on a Thermo Fisher Surveyor MS plus HPLC and LTQ XL ion trap mass spectrometer using a modified vented column setup and data-dependent scanning<sup>23</sup>. Samples were loaded onto a 360 x 100 μm precolumn (2 cm, 5 μm Monitor C18) and desalted prior to placing the precolumn in-line with the analytical column.

Peptides were then eluted with a linear gradient of 0% to 40% B in 30 min (A, 0.1M aqueous HOAc; B, 0.1M HOAc in CH<sub>3</sub>CN), a flow rate of 250 nL/min and using a 360 x 75 μm self-packed column with integrated electrospray emitter (10 cm of 5 μm Monitor, C18). MS scans were as follows: 1 full scan followed by 5 MS/MS scans of the most intense ions from the full scan using data-dependent analysis with dynamic exclusion. Dynamic exclusion parameters: repeat count — 1; repeat duration — 15s; exclusion duration — 30s.

MS/MS spectra were searched against a human, rat and mouse subset of the European Bioinformatics Institute — International Protein Index (EBI-IPI) database (downloaded 08-01-2007), with an appended reversed database and using Sequest 3.0. A fixed modification of Cys (+57), a variable modification of Met (+16) and trypsin cleavage were specified. Search results were compiled and filtered in Scaffold 2.0 (Proteome Software, Inc, Portland, OR). For analysis of 2D gel bands, a protein identification was accepted if it was established with a 99% probability of a correct identification and a minimum of 2 peptides (90% probability of a correct identification) were matched to the protein. For analysis of 1D gel bands, a protein identification was accepted if a minimum of 3 peptides were matched to the protein and peptide identifications satisfied XCorr versus m/z thresholds of +1/1.8, +2/2.5, and +3/3.5, and a DeltaCn threshold of 0.1. Proteins published as putative *O*-GlcNAc proteins were chosen by taking the list of proteins identified in the experimental eluent lane and subtracting out those proteins found in the corresponding control lane, as well as any extracellular proteins that contaminated the protein fractionations.

**In-Gel Fluorescence Detection of *O*-GlcNAc Dynamics.** HeLa cells were grown to 80–

90% confluence in DMEM containing 10% FBS and penicillin/streptomycin (100 U/mL) and harvested. Cells were incubated in DMEM with PUGNAc (100  $\mu$ M) or H<sub>2</sub>O for 9 h at 37 deg and 5% CO<sub>2</sub>. The cells were lysed in boiling 1% SDS, sonicated, and boiled for 5 min. The resulting lysate (200  $\mu$ g) was chemoenzymatically labeled with **1**, followed by **3**, as described above. A negative control was performed under identical conditions, except that **1** was omitted from the reaction mixture. After TAMRA-labeling, protein (21  $\mu$ g) was resolved on a 1.0 mm, 12-well NuPAGE 4-12% Bis-Tris Gel. The gel was imaged using a FLA-5100 scanner. Western blotting was done as described for  $\alpha$ -crystallin above but using an anti-tubulin antibody (1:10,000).

Total changes in *O*-GlcNAc glycosylation levels with PUGNAc were quantified using Multi Gauge software (Fujifilm). Quantification was determined by taking the ratio of the total fluorescent signal of the PUGNAc lane to the total fluorescent signal of the control lane, corrected to tubulin levels. Quantification represents the mean  $\pm$  standard deviation for n=2 experiments. The range over which PUGNAc changed *O*-GlcNAc glycosylation levels was determined by taking the ratio of the fluorescent signal of the PUGNAc lane to the fluorescent signal of the control lane for the 18 strongest bands, corrected for tubulin levels.

**Chromatin Immunoprecipitation (ChIP).** Cortical neuronal cultures were prepared from embryonic day 15 C57BL/6 mice as described<sup>24</sup>. Neurons were cultured for four days in Neurobasal media (Invitrogen), 2 mM Glutamax (Invitrogen), penicillin/streptomycin (Invitrogen, 100 U/ml), 2% B-27 (Invitrogen) for five days, after which the media was removed and the cells were fixed with 1% formaldehyde / PBS

(0.01 M phosphate, 0.138 M NaCl, 0.0027 M KCl, pH 7.4) for 20 min at RT. The formaldehyde was quenched with 50  $\mu$ l / ml of 2.5 M glycine, the cells were washed three times with 1X PBS, collected in 1X PBS, and then pelleted by centrifugation at 23,500 x g for 5 min at 4 °C. Cells were resuspended in cell lysis buffer (25 mM HEPES pH 7.8, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.1% NP-40, 1 mM DTT; 200  $\mu$ l / 10 cm dish of cells) and incubated on ice for 10 min. The nuclei were pelleted by centrifugation at 10,000 x g for 5 min at 4 °C. The supernatant was removed and the nuclear pellet was resuspended in nuclear lysis buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Sodium deoxycholate, 0.1% SDS, 200  $\mu$ l / 10 cm dish) and incubated on ice for 10 min. The samples were sonicated on ice 5 x 30 sec at 40% amplitude using a Sonics Vibra Cell sonicator, centrifuged for 10 min at 22,000 x g, and the supernatant was retained as the nuclear lysate.

Protein concentration from the nuclear lysate was determined by BCA protein concentration assay. Nuclear lysate (200  $\mu$ g) was supplemented with nuclear lysis buffer to 165  $\mu$ l and 4  $\mu$ l of Complete™ protease inhibitors (50X), 11  $\mu$ l of MnCl<sub>2</sub> (100 mM), 10  $\mu$ l UDP-GalNAz (0.5 mM), 10  $\mu$ l GalT. GalT was left out of the control reaction as indicated. The samples were incubated end-over-end overnight at 4°C and then the samples were precipitated by addition of 20  $\mu$ l NaOAc (5 M, pH 5.2) followed by 1750  $\mu$ l of ice-cold EtOH (100%). The samples were quickly vortexed, placed at -20°C for 1 hr, and then centrifuged at 23,500xg for 15 minutes. The supernatant was discarded and the pellets were resuspended in 50  $\mu$ l of 1% SDS, 50 mM Tris pH 8. Labeling with the TAMRA-alkyne dye **3** was performed per the Click-It™ TAMRA Glycoprotein

Detection Kit instructions except for the samples were EtOH precipitated as described above.

The samples were resuspended in 20  $\mu$ l of 0.5% SDS with Complete™ protease inhibitors followed by addition of 173  $\mu$ l of nuclear lysis buffer and 1  $\mu$ l of 100% Triton X-100. Protein A Sepharose beads (20  $\mu$ l), salmon sperm DNA (4  $\mu$ l, 2 mg/ml), and normal rabbit IgG (5  $\mu$ l, 0.4 mg/ml) were added to lysate and the samples were rotated end-over-end for 1 h at 4°C. The beads were spun down on a benchtop centrifuge for 30 sec, the supernatant was transferred to a new tube, 10% was saved for input, 2  $\mu$ g of TAMRA antibody or normal rabbit IgG were added to the remaining sample as indicated, and the samples were rotated end-over-end overnight at 4 °C.

After overnight incubation, 20  $\mu$ l of Protein A Sepharose beads was added to each sample, and the samples were rotated end-over-end for 1 h at 4 °C. The beads were then spun down on a benchtop centrifuge for 30 sec and wash successively with 1 ml of nuclear lysis buffer, nuclear lysis buffer supplemented with 360  $\mu$ l NaCl, wash buffer (20 mM Tris pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Sodium deoxycholate), and TE (20 mM Tris pH 8.0, 1 mM EDTA). The final wash was removed, 100  $\mu$ l of TE, 1% SDS was added the beads, and the beads were rotated end-over-end for 10 min at 65 °C. The beads were centrifuged, the first eluent was saved, 160  $\mu$ l TE, 0.67% SDS was added to the sample, and the beads were again rotated end-over-end for 10 min at 65 °C. The beads were centrifuged and the second eluent was combined with the first. TE and SDS was added to the input such that the final input SDS concentration in the inputs was 0.81% and the eluents and inputs were decrosslinked by incubating them at 65 °C for 4

hours. After decrosslinking, 250  $\mu$ l TE and 10  $\mu$ l Proteinase K (10 mg/ml) were added and the reactions were incubated for 2 h at 37°C shaking at 235 RPM.

56  $\mu$ l of 4 M LiCl in TE was added to each sample and the DNA was extracted with 560  $\mu$ l of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) followed by 560  $\mu$ l chloroform. 56  $\mu$ l of NaOAc (5 M, pH 5.2), 1.5 ml ethanol (100%, ice-cold), and 1  $\mu$ l glycogen (20 mg/mL) was added to each sample, the sample were vortexed and then incubated at -20°C overnight. After overnight incubation, the samples were centrifuged for 30 minutes at 22,000 x g, the supernatant was discarded, the pellets were allowed to air dry, and the pellets were redissolved in 60  $\mu$ L DNase-free H<sub>2</sub>O. These samples were either taken on to PCR or given to Rosemary Tao in the Sun lab (UCLA) for analysis on a promoter microarray.

**PCR.** 2  $\mu$ l of DNA from the ChIP experiments was combined with 14.5  $\mu$ l DNase-free H<sub>2</sub>O, 0.5  $\mu$ l PCR Nucleotide Mix (10 mM), 0.75  $\mu$ l MgCl<sub>2</sub> (50 mM), 2.5  $\mu$ l 10X Enhancer, 2.5  $\mu$ l 10X Amplifier, 2  $\mu$ l primers (10  $\mu$ M, forward and reverse mix), and 0.25  $\mu$ l Taq polymerase. The samples were then heated in a thermocycler using the following heating profile: 95 °C, 1 min (1x), 95 °C, 30 sec, annealing temperature, 1 min, 72 °C, 1 min, 30 sec (36x), 72 °C, 7 min (1x), 4 °C, unlimited hold. 5  $\mu$ l of 6x gel loading dye was added to each sample, and the DNA was resolved on a 2% agarose / TEA gel. The following primers were used for ChIP:

Gene	Forward Primer	Reverse Primer
Pomc	TACCTCCAAATGCCAGGAAG	CGCTGGTGGT TAGGAAGAAC
18S Ribosomal RNA	CGCGGTTCTATTTTGTGGT	AGTCGGCATCGTTATGGTC

**Chemoenzymatic Labeling and Fluorescence Imaging of *O*-GlcNAc Proteins in Cells.** HeLa cells were counted, diluted into DMEM containing 10% FBS and penicillin/streptomycin (100 U/mL) and seeded on poly-D-lysine-coated (0.1 mg/mL poly-D-lysine in 50 mM sodium borate, pH 10, 100  $\mu$ L/cover slip for 30 min at 37 °C) 15 mm glass coverslips (Carolina Biologicals) at a density of 75 cells/mm<sup>2</sup> (100  $\mu$ L/cover slip). After 30 min, 400  $\mu$ L of media was added to each coverslip, and the cultures were incubated at 5% CO<sub>2</sub> at 37 °C for 6 h.

Cortical neuronal cultures were prepared from embryonic day 18 Long Evans rats as described<sup>24</sup>. Neurons were counted, diluted into supplemented Basal Media Eagle (BME, Sigma; 450 mL media, 10 mL L-glutamine (200 mM), 5 mL penicillin/streptomycin (10,000 U/mL), 10 mL B-27 serum-free supplement (50X stock), 25 mL FBS) and seeded on poly-DL-ornithine-coated 18-mm glass coverslips (Carolina Biologicals) at a density of 100 cells/mm<sup>2</sup> (150  $\mu$ L/cover slip). After 30 min, 350  $\mu$ L of supplemented BME media was added to each coverslip. The cultures were incubated in 5% CO<sub>2</sub> at 37 °C for 7 days.

To image *O*-GlcNAc glycosylated proteins, the media was removed, and the coverslips were rinsed one time with PBS, fixed in 4% paraformaldehyde for 20 min at RT, washed twice with PBS, permeabilized in 0.3% Triton X-100 for 5 min at RT, and washed twice with enzymatic labeling buffer (50 mM HEPES, 125 mM NaCl, pH 7.9). Reaction mixtures and negative controls without UDP-GalNAz **1** were prepared as described in the Click-It™ *O*-GlcNAc Enzymatic Labeling System instructions except that Component C, the enzymatic labeling buffer, was replaced with a buffer containing

125 mM NaCl, 50 mM HEPES, pH 7.9. These mixtures were added to each coverslip (50  $\mu$ L), and the coverslips were incubated at 4 °C for 14–20 h. For the HeLa cells, PNGase F (2500 U/mL) was added to the enzymatic labeling reaction mixture; no difference in staining was observed in the presence or absence of PNGaseF. Coverslips were washed one time with 125 mM NaCl, 50 mM HEPES, pH 7.9 and twice with 50 mM Tris, pH 8.0. Biotin labeling reaction mixtures were prepared as per the Click-It™ Biotin Glycoprotein Detection Kit instructions using 50 mM Tris, pH 8.0 without SDS, added to each coverslip (50  $\mu$ L), and the reaction allowed to proceed for 1 h at RT. For TAMRA labeling, TAMRA-alkyne **3** was substituted above for biotin-alkyne **2**. The TAMRA-alkyne **3** produced high background labeling in the absence of GalT, likely due to noncovalently sticking of **3** to hydrophobic regions of membranes and proteins. The background could be reduced by washing the cells with organic solvents (similar to the precipitation steps after TAMRA labeling on lysates), but these solvents also distorted and destroyed the fixed cells.

After the reaction was finished, the coverslips were washed once with PBS, three times with 0.1% Triton-X100 in PBS, and once with PBS. Following the PBS wash, nonspecific binding was blocked by incubating with 3% BSA in PBS for 1 h at RT and then rinsing once with PBS. Cells were then incubated with streptavidin-AlexaFluor 488 (1:800; Molecular Probes) in 3% BSA in PBS for 1 h at 37 °C. Coverslips were rinsed three times with 0.2% Triton-X100 in PBS and once with PBS. The coverslips were mounted onto glass slides using Vectashield mounting medium with DAPI (2  $\mu$ L; Vector Labs) and sealed with clear nail polish. Cells were imaged using a Nikon Eclipse

TE2000-S inverted microscope, and images were captured with Metamorph software using a 40x plan fluor oil objective.

## References

1. Rexach, J.E., Clark, P.M. & Hsieh-Wilson, L.C. Chemical approaches to understanding O-GlcNAc glycosylation in the brain. *Nat Chem Biol* **4**, 97-106 (2008).
2. Hart, G.W., Housley, M.P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022 (2007).
3. Love, D.C. & Hanover, J.A. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci STKE* **2005**, re13 (2005).
4. Khidekel, N. *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J Am Chem Soc* **125**, 16162-16163 (2003).
5. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci U S A* **101**, 13132-13137 (2004).
6. Vocadlo, D.J., Hang, H.C., Kim, E.J., Hanover, J.A. & Bertozzi, C.R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci U S A* **100**, 9116-9121 (2003).
7. Gross, B.J., Kraybill, B.C. & Walker, S. Discovery of O-GlcNAc transferase inhibitors. *J Am Chem Soc* **127**, 14588-14589 (2005).
8. 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).
9. Kim, E.J. *et al.* Distinctive inhibition of O-GlcNAcase isoforms by an alpha-GlcNAc thiol-sulfonate. *J Am Chem Soc* **129**, 14854-14855 (2007).
10. Carrillo, L.D., Krishnamoorthy, L. & Mahal, L.K. A cellular FRET-based sensor for beta-O-GlcNAc, a dynamic carbohydrate modification involved in signaling. *J Am Chem Soc* **128**, 14768-14769 (2006).
11. Tai, H.C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel identification of O-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc* **126**, 10500-10501 (2004).
12. Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* **41**, 2596-2599 (2002).
13. Prescher, J.A. & Bertozzi, C.R. Chemistry in living systems. *Nat Chem Biol* **1**, 13-21 (2005).
14. Speers, A.E. & Cravatt, B.F. Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol* **11**, 535-546 (2004).

15. 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**, 923-934 (2006).
16. Khidekel, N. *et al.* Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* **3**, 339-348 (2007).
17. 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**, 30133-30142 (2004).
18. Kelly, W.G., Dahmus, M.E. & Hart, G.W. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* **268**, 10416-10424 (1993).
19. Lamarre-Vincent, N. & Hsieh-Wilson, L.C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J Am Chem Soc* **125**, 6612-6613 (2003).
20. Fujiki, R. *et al.* GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. *Nature* **459**, 455-459 (2009).
21. Blum, H., Beier, H. & Gross, H.J. Improved Silver Staining of Plant-Proteins, Rna and DNA in Polyacrylamide Gels. *Electrophoresis* **8**, 93-99 (1987).
22. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal Chem* **68**, 850-858 (1996).
23. Licklider, L.J., Thoreen, C.C., Peng, J.M. & Gygi, S.P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal Chem* **74**, 3076-3083 (2002).
24. Gama, C.I. *et al.* Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Bio* **2**, 467-473 (2006).

**Chapter 5: CREB — A Key Transcription Factor in the  
Nervous System**

**cAMP response element binding protein (CREB) is a complex transcription factor that integrates various cellular signals and outputs specific transcriptional programs. In response to extracellular stimuli, CREB is post-translationally modified, most notably by phosphorylation. This, in turn, alters CREB-dependent transcription and affects such processes as neuronal development and survival, drug addiction, and learning and memory. Here, we review the literature on CREB regulation and function in the nervous system.**

Cyclic-AMP response element binding protein (CREB) is a key transcription factor that translates extracellular stimuli into specific gene programs<sup>1-3</sup>. CREB is expressed in all cells in the brain<sup>3</sup> and complete knockout of CREB in mice is lethal<sup>4</sup>. CREB is a member of a family of transcription factors that includes CREB, CREM, and ATF-1 and binds as a homo- or heterodimer to cyclic-AMP response elements (CREs) containing the DNA sequence TGACGTCA<sup>1</sup>. CREB consists of four domains: glutamine rich Q1 and Q2 domains, a central kinase inducible domain (KID) and a bZIP domain. The KID domain contains the majority of known CREB phosphorylation sites<sup>2</sup>, including the major phosphorylation site at Ser133, and is necessary for phosphorylation-induced CREB activity. The bZIP domain is necessary for CREB to bind the DNA. The Q2 domain, primarily, and the Q1 domain, secondarily, are important for interactions with the basal transcriptional machinery and have been shown to mediate CREB transactivation independent on stimulus<sup>5</sup>.

### *Post-Translational Modifications of CREB*

*Phosphorylation.* The best-studied post-translational modification of CREB is phosphorylation of CREB at Ser133. Many different kinases, including CaMK, PKA, PCK, PKG, Rsk, and MEK, and many different stimuli, including growth factors, neurosignalling molecules, cytokines, and environmental stress<sup>2</sup> all phosphorylate CREB at Ser133. This phosphorylation event promotes the interaction between CREB and the coactivators CREB binding protein (CBP) and its paralogue p300, dramatically enhances CREB activity, and drives the expression on different CREB-mediated gene programs. Although phosphorylation at this site has been extensively studied and is often associated with CREB activation, it appears to be neither necessary<sup>6</sup> nor sufficient<sup>2</sup> for CREB activation.

CREB is phosphorylated at Ser142 by CaMKII<sup>7</sup> and this phosphorylation is modulated by a number of different stimuli, including formalin injection into the spinal cord, a model for peripheral noxious stimulation<sup>8</sup>, circadian rhythm, light, and glutamate in the suprachiasmatic nucleus (SCN)<sup>9</sup>, and calcium influx in cultured cortical neurons (which also leads to phosphorylation of Ser143)<sup>10</sup>. Ser142 phosphorylation is necessary for complete light-induced expression of *c-fos* and *mPer1* and for light-induced phase shifts of the circadian clock in the SCN<sup>9</sup>. Ser142Ala CREB, mutants show enhanced, while Ser142Ala / Ser143Ala double mutants show repressed, CREB activity in response to neuronal depolarization but no effect on CREB activity in response to increased cAMP levels compared to WT CREB<sup>10</sup>. Mechanistically, Ser142 phosphorylation has been shown to cause the dissociation of CREB dimers<sup>7</sup> as well as to block the interaction between CREB and the KIX domain of CBP<sup>10</sup>.

Hypoxia induces hyperphosphorylation of CREB at sites other than Ser133<sup>11</sup>, although the exact sites are unknown. Ionizing radiation, which causes DNA damage, promotes the phosphorylation of a cascade of CREB phosphorylation sites by ATM, CK1, and CK2 starting with Ser111 followed by Ser108, Ser114, Ser117, and Ser121, which finally blocks the interaction between CREB and CBP<sup>12</sup>. Alternatively, CREB undergoes cell-cycle-dependent phosphorylation at Ser108, Ser111, and Ser114, and CREB with Ser111 and Ser114 mutated to glutamic acid to mimic phosphorylation show enhanced transcription on non-induced CREB<sup>13</sup>. Furthermore, DNA damage leads to CREB phosphorylation at Ser271 by HIPK2, which enhances CREB transactivation by recruiting CBP<sup>14</sup>. Finally GSK-3 phosphorylates CREB at Ser129 subsequent to phosphorylation of Ser133, and this phosphorylation event is required for CREB transactivation by forskolin in PC12 cells<sup>15</sup>.

*Other Post-Translational Modifications.* Along with phosphorylation, CREB is also modified by acetylation, glycosylation, ubiquitination, and SUMO-ylation. CREB is acetylated by CBP at Lys91, Lys96, and Lys136, and mutations that replace these lysines with alanines or arginines enhance both basal and PKA-induced CREB activity<sup>16</sup>. CREB is glycosylated *in vivo* within amino acids 259–261<sup>17</sup>. Finally, following onset of hypoxia, CREB is phosphorylated, ubiquitinated, and degraded<sup>11</sup>, but after prolonged hypoxia, CREB is SUMO-ylated at Lys285 and Lys304, which stabilizes CREB, and in the case of Lys304 SUMO-ylation, may lead to increased nuclear localization<sup>18</sup>. Taken together, the abundance of post-translational modifications on CREB suggests that CREB is tightly controlled in response to stimuli and that, although Ser133 phosphorylation may

be a predominant post-translational modification for activating CREB, these other modifications may be important for the finely tuned modulation of CREB activity.

### *CREB-Interacting Proteins*

CREB interacts directly with a variety of different proteins that either activate or repress CREB activity. Two of the most widely studied CREB-interacting proteins are CBP / p300 and transducer of regulated CREB activity (TORCs). CBP / p300 are transcriptional transactivators that interact with multiple transcription factors, including CREB. The interaction between CREB and CBP is modulated by phosphorylation of CREB at Ser133, which is necessary but not sufficient for CBP to bind CREB<sup>19</sup>. CBP / p300 promote the recruitment of RNA polymerase II through RNA helicase A and have histone acetyltransferase activity, which facilitates the opening of chromatin to allow access to the DNA<sup>20</sup>.

TORCs are a recently discovered family of proteins that coactivate CREB transcription independent of CREB Ser133 phosphorylation<sup>6,21</sup> and have been shown to contribute to important physiological and disease processes such as gluconeogenesis<sup>22</sup> and diabetes<sup>23</sup>. TORCs bind to the bZIP domain of CREB, likely through ionic interactions<sup>24</sup>, and have been proposed to enhance CREB transactivation by promoting the interaction between CREB and TAF<sub>II</sub>130, a component of the basal transcriptional machinery<sup>6</sup>. TORCs are actively shuttled out of the nucleus, and thus, in untreated cells, TORCs are found predominantly although not exclusively in the cytoplasm<sup>25</sup>. TORCs are modified by phosphorylation and *O*-GlcNAc glycosylation. In response to low-energy signals such as low ATP levels or low glucose levels, AMPK in hepatic cells and

the hypothalamus and SIK in hepatic cells phosphorylate TORC2 at Ser171<sup>22,26</sup>, and in response to low glucose in islet cells, MARK2 phosphorylates TORC2 at Ser275<sup>27</sup>. Both of these phosphorylation events recruit the 14-3-3 protein, which causes TORC to be sequestered in the cytoplasm. Alternatively, in response to high glucose, TORC2 is *O*-GlcNAc glycosylated at Ser70 and Ser171<sup>23</sup> in hepatic cells or dephosphorylated by calcineurin in islet cells<sup>24,27</sup>, which blocks the interaction with 14-3-3, and allows TORC2 to accumulate in the nucleus and enhance CREB activity. In neurons, TORC1 and TORC2 translocate into the nucleus downstream of calcineurin in response to neuronal depolarization<sup>28</sup>, where they are required for activity-dependent gene expression of SIK, activity-dependent dendritic growth<sup>29</sup>, stress sensitivity<sup>30</sup>, and maintenance of L-LTP in the Schaffer collateral–CA1 pathway<sup>28</sup>.

CREB can also interact with other proteins both to alter CREB activity as well as to alter the activity of the other protein. For example, CREB binds TAF<sub>II</sub>130/135, which interacts with the Q2 domain of CREB and activates CREB<sup>31</sup>, as well as YY-1, which inhibits CREB activity by bending the DNA around CREB and thus blocking the interaction between CREB and the basal transcriptional machinery<sup>32</sup>. Alternatively, CREB interacts with MeCP2, which modulates MeCP2 function such that it activates rather than represses genes<sup>33</sup>, and CREB bridges p53 and CBP, thereby enhancing p53 transcriptional activation<sup>34</sup>.

### *CREB Targets*

The complement of confirmed and predicted CREB target genes is extensive and continues to grow. 1349 mouse and 1663 human putative CREB binding sites have been

identified<sup>35</sup> and 6302 CREB loci have been mapped from PC12 cells using a serial analysis of chromatin occupancy (SACO) approach<sup>36</sup>. CREB has been shown to regulate the expression of many different classes of proteins, including proteins important in neurotransmitter release, cell structure, signal transduction, and metabolism<sup>1</sup>. Furthermore, microarray analysis of CREB knockdown in myeloid leukemia cells<sup>37</sup>, CREB overexpression in the nucleus accumbens, and S133A CREB overexpression in the nucleus accumbens<sup>38</sup> identified many transcripts whose expression was modified in response to modified CREB activity.

### *Specificity of CREB Signal*

CREB integrates a diversity of extracellular signals and translates them into unique gene programs. Yet how CREB differentiates each of these signals and transcribes the correct set of genes for the correct time period and for a given signal within a given cell type remains an open and pressing question in the field. For example, tyrosine hydroxylase, a CREB-dependent gene, is expressed in only specific cell types throughout the brain<sup>39</sup> and c-fos expression has a distinct time-course that returns to basal levels within an hour, independent of stimulation time, in 3T3 cells<sup>40</sup>. The post-translational modifications of CREB (discussed above), which can be induced or repressed in response to different stimuli, may account for some of these effects. Nevertheless other signals and events are also likely required to differentiate the set of targets transcribed following CREB activation.

The exact duration of CREB Ser133 phosphorylation may in part contribute to CREB transactivation. H<sub>2</sub>O<sub>2</sub> induces transient (15 min) CREB phosphorylation without

CREB transactivation, whereas estradiol induces prolonged (5 hr) CREB phosphorylation and CREB activation<sup>41</sup>. Nevertheless the kinetics of CREB phosphorylation are not the sole mediator of CREB activity as forskolin, which activates PKA, and phorbol 12-tetradecanoate 13-acetate (TPA), which activates PKC; both phosphorylate CREB equally in NIH3T3 cells but only forskolin induces CREB activity<sup>42</sup>, and, similarly, forskolin and TPA both produce comparable levels of CREB phosphorylation in PC12 cells, but only forskolin induces expression of the CREB target gene *Icer*<sup>43</sup>. Interestingly, in the latter case, only forskolin was found to promote the formation of a nuclear CREB–CBP complex although both forskolin and TPA promoted the formation of cytoplasmic CREB–CBP complexes<sup>43</sup>.

Chromatin modification and accessibility may also contribute to stimuli-specific differences in CREB-mediated transcription. CREB forms a stable complex with HDAC1 and PP1, which is disrupted by forskolin<sup>44</sup>. Furthermore, treatment with Trichostatin A (TSA), an HDAC inhibitor, enhances the expression of a subset of CREB-mediated genes, including *c-fos* and *NUR77*, following forskolin treatment while blocking the activation of a different subset of CREB-mediated genes, including *ICER* and *NOR-1*<sup>45</sup>. Similarly, *BDNF*, a CREB-target gene, has exon-specific changes in chromatin modifications on its promoter following NMDA treatment<sup>46</sup>. Finally, CREB occupancy of different promoters, including the *somatostatin* promoter, differs depending on the cell type, possibly in response to different availability of the promoter for protein binding<sup>47</sup>.

Finally, post-translational modifications of CREB coactivators or corepressors may specify which stimuli activate CREB transcription. CBP is phosphorylated<sup>48</sup>,

methylated<sup>49</sup>, SUMO-ylated<sup>50</sup>, and glycosylated<sup>51</sup>. CBP Ser301 phosphorylation is induced by NMDA in hippocampal neurons and is required for full CREB transactivation<sup>48</sup>. Furthermore, methylation at Arg300 blocks the interaction between CREB and CBP and represses CREB transcription. Finally, CBP can be SUMO-ylated at Lys999, Lys1034, and Lys1057, which recruits the transcriptional corepressor Daxx and represses CBP transcriptional activity<sup>50</sup>. Thus, signals that modify the duration of CREB phosphorylation, the chromatin modifications around CREB target genes, and the activity of CREB coactivator and corepressor, in addition to activating CREB, may be important in differentiating specific stimuli and activating unique CREB-dependent gene programs.

#### *Pathways that Activate CREB Following Neuronal Activity*

Three major kinase pathways regulate CREB activity in neurons in response to neuronal activity — PKA, CaMK, and Ras/ERK. PKA is regulated by cAMP levels, which are themselves regulated by adenylate cyclase activity. Adenylate cyclase activity can be regulated downstream of G-protein-coupled receptors (GPCRs) for neurotransmitters as well as directly through Ca<sup>2+</sup> flux<sup>1,52</sup>. Ca<sup>2+</sup> flux also activates both MAPK and calmodulin. MAPK and calmodulin further activate Rsk kinases and CaMK, respectively, which directly phosphorylate CREB<sup>53</sup>. Once phosphorylated, two phosphatases are able to directly dephosphorylate CREB, PP1 and PP2A<sup>1</sup>. Inhibition of calcineurin is also known to enhance CREB phosphorylation but calcineurin has yet to be shown to directly dephosphorylate CREB *in vivo*<sup>54</sup>.

### *Functions of CREB in the Nervous System*

CREB performs a number of different functions in the nervous system. Among the best studied include its role in neuronal development and survival, in drug addiction, and in learning and memory.

*CREB and Neuronal Development and Survival.* Inhibition of CREB family members *in vitro* and CREB knockout studies *in vivo* substantiate a critical role for CREB in neuronal survival<sup>55-58</sup>. *In vitro* transfection with a dominant-negative CREB construct blocks BDNF-mediated cell survival in cerebellar granular neurons<sup>55</sup> and NGF-mediated cell survival in sympathetic neurons<sup>56</sup>. Furthermore, overexpression of a constitutively-active CREB construct was sufficient to promote sympathetic neuron survival in the absence of NGF<sup>56</sup>. Finally, CREB null mice have a reduction in the size of the corpus callosum and the anterior commissures, as well as enhanced apoptosis of sensory and sympathetic neurons<sup>57</sup>, and dentate gyrus neuronal survival following ischemic insult depends on CREB-dependent gene expression<sup>58</sup>.

CREB has also been shown to contribute to neuronal development. Studies have shown that CREB, through the expression of Wnt-2, in hippocampal neurons<sup>59</sup> and CREB and TORC1 in cortical neurons<sup>29,60</sup> are necessary for activity-dependent dendrite growth both *in vitro*. Furthermore, CREB knockout studies indicate axonal growth defects in DRG neurons<sup>57</sup>.

*CREB and Drug Addiction.* CREB activation, expression, and activity appear intricately linked to the process of drug addiction. Multiple regions of the brain demonstrate up-

regulation of the cyclic AMP pathway following opiate addiction<sup>61</sup>. In the locus coeruleus, morphine administration inhibits, whereas morphine withdrawal activates, CREB phosphorylation<sup>62</sup>. Alternatively morphine and cocaine administration induces CREB activity in the nucleus accumbens and striatum<sup>63,64</sup>. Furthermore, mice lacking CREB in the entire brain show decreased development of morphine dependence<sup>65</sup>. More specifically, overexpression of a dominant-negative form of CREB in the nucleus accumbens or the caudal ventral tegmental area (VTA) increases the rewarding effects of morphine and cocaine whereas overexpression of WT CREB decreases them<sup>66-68</sup>. Alternatively, overexpression of WT CREB in the rostral VTA makes cocaine more rewarding and overexpression of a dominant-negative form of CREB makes cocaine less rewarding<sup>68</sup>.

*CREB and Memory.* CREB is a key transcription factor for regulating long-term memory across many different species. Long-term facilitation, a model of memory, can be blocked or induced in *Aplysia* by injecting CRE oligonucleotides or phosphorylated CREB, respectively, demonstrating that CREB is necessary and sufficient for this process<sup>69</sup>. Additionally, activation of long-term facilitation activates CREB-mediated transcription in *Aplysia*<sup>70</sup>. Induction of a dominant negative CREB transgene blocks while induction of a CREB activator isoform enhances long-term memory in *Drosophila*<sup>71,72</sup>. In mice, learning tasks are associated with an increase in CREB phosphorylation at Ser133<sup>73,74</sup>. Furthermore, injection of CREB antisense oligos into the hippocampus of WT CREB causes deficits in spatial memory<sup>75,76</sup>. Moreover, CREB  $\alpha\Delta$  knockout mice show a deficit in long-term but not short-term memory following

contextual and cued fear conditioning<sup>76</sup> and social transmission of food preferences<sup>77</sup> as well as a deficit in the Morris Water maze test, which could be caused by learning or memory impairment<sup>76</sup>. Although follow-up experiments have shown that these effects depend, in part, on the specific mouse strain used for the study and may require knockout of additional CREB isoforms<sup>78,79</sup>, they nevertheless suggest a significant role for CREB in long term memory formation. Similarly, induction of a dominant-negative CREB blocks consolidation of long-term memory whereas overexpression of WT CREB enhances this consolidation<sup>80,81</sup>. Taken together, these studies demonstrate that CREB contributes to critical and complex functions in the brain.

### *Conclusions*

CREB is a complex transcription factor that is activated by multiple, distinct signaling pathways in neurons and that is essential for many critical neuronal functions. Although CREB has been well studied, a number of open questions remain: How does CREB distinguish between the different pathways that induce phosphorylation at Ser133? Do other CREB post-translational modifications contribute, and if so, what are these other modifications, how are they regulated, and how do they affect CREB function? Finally, given a variety of inputs, how does CREB activate transcriptional programs specific for the context? Addressing these questions will provide in-depth information on the function of CREB in neurons. Moreover, insofar as CREB is a model for other transcription factors, additional data of CREB will more broadly inform our understanding of transcription.

## References

1. Lonze, B.E. & Ginty, D.D. Function and regulation of CREB family transcription factors in the nervous system. *Neuron* 35, 605-23 (2002).
2. Johannessen, M., Delghandi, M.P. & Moens, U. What turns CREB on? *Cell Signal* 16, 1211-27 (2004).
3. Carlezon, W.A., Duman, R.S. & Nestler, E.J. The many faces of CREB. *Trends Neurosci.* 28, 436-45 (2005).
4. Rudolph, D. et al. Impaired fetal T cell development and perinatal lethality in mice lacking the cAMP response element binding protein. *Proc Natl Acad Sci USA* 95, 4481-6 (1998).
5. Shaywitz, A.J. & Greenberg, M.E. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem* 68, 821-61 (1999).
6. Conkright, M.D. et al. TORCs: transducers of regulated CREB activity. *Mol Cell* 12, 413-23 (2003).
7. Wu, X. & McMurray, C.T. Calmodulin kinase II attenuation of gene transcription by preventing cAMP response element-binding protein (CREB) dimerization and binding of the CREB-binding protein. *J Biol Chem* 276, 1735-41 (2001).
8. Niederberger, E., Ehnert, C., Gao, W. & Coste, O. The impact of CREB and its phosphorylation at Ser142 on inflammatory nociception. *Biochem Biophys Res Commun* (2007).
9. Gau, D. et al. Phosphorylation of CREB Ser142 regulates light-induced phase shifts of the circadian clock. *Neuron* 34, 245-53 (2002).
10. Kornhauser, J.M. et al. CREB transcriptional activity in neurons is regulated by multiple, calcium-specific phosphorylation events. *Neuron* 34, 221-33 (2002).
11. Taylor, C.T., Furuta, G.T., Synnestvedt, K. & Colgan, S.P. Phosphorylation-dependent targeting of cAMP response element binding protein to the ubiquitin/proteasome pathway in hypoxia. *Proc Natl Acad Sci USA* 97, 12091-6 (2000).
12. Shanware, N.P., Trinh, A.T., Williams, L.M. & Tibbetts, R.S. Coregulated ataxia telangiectasia-mutated and casein kinase sites modulate cAMP-response element-binding protein-coactivator interactions in response to DNA damage. *J Biol Chem* 282, 6283-91 (2007).
13. Saeki, K., Yuo, A. & Takaku, F. Cell-cycle-regulated phosphorylation of cAMP response element-binding protein: identification of novel phosphorylation sites. *Biochem J* 338 ( Pt 1), 49-54 (1999).
14. Sakamoto, K. et al. Regulation of Genotoxic Stress Response by HIPK2 through Phosphorylation of CREB at Serine 271. *Molecular biology of the cell* (2010).
15. Fiol, C.J. et al. A secondary phosphorylation of CREB341 at Ser129 is required for the cAMP-mediated control of gene expression. A role for glycogen synthase kinase-3 in the control of gene expression. *J Biol Chem* 269, 32187-93 (1994).
16. Lu, Q., Hutchins, A.E., Doyle, C.M., Lundblad, J.R. & Kwok, R.P.S. Acetylation of cAMP-responsive element-binding protein (CREB) by CREB-binding protein enhances CREB-dependent transcription. *J Biol Chem* 278, 15727-34 (2003).

17. Lamarre-Vincent, N. & Hsieh-Wilson, L.C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J Am Chem Soc* 125, 6612-3 (2003).
18. Comerford, K.M. et al. Small ubiquitin-related modifier-1 modification mediates resolution of CREB-dependent responses to hypoxia. *Proc Natl Acad Sci USA* 100, 986-91 (2003).
19. Brindle, P., Nakajima, T. & Montminy, M. Multiple protein kinase A-regulated events are required for transcriptional induction by cAMP. *Proc Natl Acad Sci USA* 92, 10521-5 (1995).
20. Nakajima, T. et al. RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* 90, 1107-12 (1997).
21. Iourgenko, V. et al. Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc Natl Acad Sci USA* 100, 12147-52 (2003).
22. Koo, S.-H. et al. The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. *Nature* 437, 1109-11 (2005).
23. Dentin, R., Hedrick, S., Xie, J., Yates, J. & Montminy, M. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* 319, 1402-5 (2008).
24. Screaton, R.A. et al. The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector. *Cell* 119, 61-74 (2004).
25. Bittinger, M.A. et al. Activation of cAMP response element-mediated gene expression by regulated nuclear transport of TORC proteins. *Curr Biol* 14, 2156-61 (2004).
26. Lerner, R.G., Depatie, C., Rutter, G.A., Screaton, R.A. & Balthasar, N. A role for the CREB co-activator CRTC2 in the hypothalamic mechanisms linking glucose sensing with gene regulation. *EMBO reports* 10, 1175-81 (2009).
27. Jansson, D. et al. Glucose controls CREB activity in islet cells via regulated phosphorylation of TORC2. *Proc Natl Acad Sci USA* 105, 10161-6 (2008).
28. Kovács, K.A. et al. TORC1 is a calcium- and cAMP-sensitive coincidence detector involved in hippocampal long-term synaptic plasticity. *Proc Natl Acad Sci USA* 104, 4700-5 (2007).
29. Li, S., Zhang, C., Takemori, H., Zhou, Y. & Xiong, Z.-Q. TORC1 regulates activity-dependent CREB-target gene transcription and dendritic growth of developing cortical neurons. *J Neurosci* 29, 2334-43 (2009).
30. Wang, B. et al. The insulin-regulated CREB coactivator TORC promotes stress resistance in *Drosophila*. *Cell Metab* 7, 434-44 (2008).
31. Tanese, N., Saluja, D., Vassallo, M.F., Chen, J.L. & Admon, A. Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100. *Proc Natl Acad Sci USA* 93, 13611-6 (1996).
32. Zhou, Q., Gedrich, R.W. & Engel, D.A. Transcriptional repression of the c-fos gene by YY1 is mediated by a direct interaction with ATF/CREB. *J Virol* 69, 4323-30 (1995).
33. Chahrour, M. et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224-9 (2008).

34. Giebler, H.A., Lemasson, I. & Nyborg, J.K. p53 recruitment of CREB binding protein mediated through phosphorylated CREB: a novel pathway of tumor suppressor regulation. *Mol Cell Biol* 20, 4849-58 (2000).
35. Conkright, M.D. et al. Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness. *Mol Cell* 11, 1101-8 (2003).
36. Impey, S. et al. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* 119, 1041-54 (2004).
37. Pellegrini, M. et al. Expression Profile of CREB knockdown in Myeloid Leukemia Cells. *BMC Cancer* 8, 264 (2008).
38. McClung, C.A. & Nestler, E.J. Regulation of gene expression and cocaine reward by CREB and DeltaFosB. *Nat Neurosci* 6, 1208-15 (2003).
39. Cambi, F., Fung, B. & Chikaraishi, D. 5' flanking DNA sequences direct cell-specific expression of rat tyrosine hydroxylase. *Journal of Neurochemistry* 53, 1656-9 (1989).
40. Greenberg, M.E. & Ziff, E.B. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* 311, 433-8 (1984).
41. Kanda, N. & Watanabe, S. 17beta-estradiol inhibits oxidative stress-induced apoptosis in keratinocytes by promoting Bcl-2 expression. *J Invest Dermatol* 121, 1500-9 (2003).
42. Seternes, O.M., Johansen, B. & Moens, U. A dominant role for the Raf-MEK pathway in forskolin, 12-O-tetradecanoyl-phorbol acetate, and platelet-derived growth factor-induced CREB (cAMP-responsive element-binding protein) activation, uncoupled from serine 133 phosphorylation in NIH 3T3 cells. *Mol Endocrinol* 13, 1071-83 (1999).
43. Mayr, B.M., Canettieri, G. & Montminy, M.R. Distinct effects of cAMP and mitogenic signals on CREB-binding protein recruitment impart specificity to target gene activation via CREB. *Proc Natl Acad Sci USA* 98, 10936-41 (2001).
44. Canettieri, G. et al. Attenuation of a phosphorylation-dependent activator by an HDAC-PP1 complex. *Nat Struct Biol* 10, 175-81 (2003).
45. Fass, D.M., Butler, J.E.F. & Goodman, R.H. Deacetylase activity is required for cAMP activation of a subset of CREB target genes. *J Biol Chem* 278, 43014-9 (2003).
46. Tian, F. et al. Dynamic chromatin remodeling events in hippocampal neurons are associated with NMDA receptor-mediated activation of Bdnf gene promoter 1. *Journal of Neurochemistry* 109, 1375-88 (2009).
47. Cha-Molstad, H., Keller, D.M., Yochum, G.S., Impey, S. & Goodman, R.H. Cell-type-specific binding of the transcription factor CREB to the cAMP-response element. *Proc Natl Acad Sci USA* 101, 13572-7 (2004).
48. Impey, S. et al. Phosphorylation of CBP mediates transcriptional activation by neural activity and CaM kinase IV. *Neuron* 34, 235-44 (2002).
49. Xu, W. et al. A transcriptional switch mediated by cofactor methylation. *Science* 294, 2507-11 (2001).
50. Kuo, H.-Y. et al. SUMO modification negatively modulates the transcriptional activity of CREB-binding protein via the recruitment of Daxx. *Proc Natl Acad Sci USA* 102, 16973-8 (2005).

51. Tai, H.-C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel identification of O-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc* 126, 10500-1 (2004).
52. Poser, S. & Storm, D.R. Role of Ca<sup>2+</sup>-stimulated adenylyl cyclases in LTP and memory formation. *Int J Dev Neurosci* 19, 387-94 (2001).
53. West, A.E. et al. Calcium regulation of neuronal gene expression. *Proc Natl Acad Sci USA* 98, 11024-31 (2001).
54. Christie-Fougere, M.M., Darby-King, A., Harley, C.W. & Mclean, J.H. Calcineurin inhibition eliminates the normal inverted U curve, enhances acquisition and prolongs memory in a mammalian 3'-5'-cyclic AMP-dependent learning paradigm. *NSC* 158, 1277-83 (2009).
55. Bonni, A. et al. Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science* 286, 1358-62 (1999).
56. Riccio, A., Ahn, S., Davenport, C.M., Blendy, J.A. & Ginty, D.D. Mediation by a CREB family transcription factor of NGF-dependent survival of sympathetic neurons. *Science* 286, 2358-61 (1999).
57. Lonze, B.E., Riccio, A., Cohen, S. & Ginty, D.D. Apoptosis, axonal growth defects, and degeneration of peripheral neurons in mice lacking CREB. *Neuron* 34, 371-85 (2002).
58. Mabuchi, T. et al. Phosphorylation of cAMP response element-binding protein in hippocampal neurons as a protective response after exposure to glutamate in vitro and ischemia in vivo. *J Neurosci* 21, 9204-13 (2001).
59. Wayman, G.A. et al. Activity-dependent dendritic arborization mediated by CaM-kinase I activation and enhanced CREB-dependent transcription of Wnt-2. *Neuron* 50, 897-909 (2006).
60. Redmond, L., Kashani, A.H. & Ghosh, A. Calcium regulation of dendritic growth via CaM kinase IV and CREB-mediated transcription. *Neuron* 34, 999-1010 (2002).
61. Nestler, E.J. Molecular basis of long-term plasticity underlying addiction. *Nat Rev Neurosci* 2, 119-28 (2001).
62. Guitart, X., Thompson, M.A., Mirante, C.K., Greenberg, M.E. & Nestler, E.J. Regulation of cyclic AMP response element-binding protein (CREB) phosphorylation by acute and chronic morphine in the rat locus coeruleus. *Journal of Neurochemistry* 58, 1168-71 (1992).
63. Cole, R.L., Konradi, C., Douglass, J. & Hyman, S.E. Neuronal adaptation to amphetamine and dopamine: molecular mechanisms of prodynorphin gene regulation in rat striatum. *Neuron* 14, 813-23 (1995).
64. Kano, T., Suzuki, Y., Shibuya, M., Kiuchi, K. & Hagiwara, M. Cocaine-induced CREB phosphorylation and c-Fos expression are suppressed in Parkinsonism model mice. *Neuroreport* 6, 2197-200 (1995).
65. Maldonado, R. et al. Reduction of morphine abstinence in mice with a mutation in the gene encoding CREB. *Science* 273, 657-9 (1996).
66. Carlezon, W.A. et al. Regulation of cocaine reward by CREB. *Science* 282, 2272-5 (1998).

67. Barrot, M. et al. CREB activity in the nucleus accumbens shell controls gating of behavioral responses to emotional stimuli. *Proc Natl Acad Sci USA* 99, 11435-40 (2002).
68. Olson, V.G. et al. Regulation of drug reward by cAMP response element-binding protein: evidence for two functionally distinct subregions of the ventral tegmental area. *J Neurosci* 25, 5553-62 (2005).
69. Kandel, E.R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 294, 1030-8 (2001).
70. Kaang, B.K., Kandel, E.R. & Grant, S.G. Activation of cAMP-responsive genes by stimuli that produce long-term facilitation in *Aplysia* sensory neurons. *Neuron* 10, 427-35 (1993).
71. Yin, J.C. et al. Induction of a dominant negative CREB transgene specifically blocks long-term memory in *Drosophila*. *Cell* 79, 49-58 (1994).
72. Yin, J.C., Del Vecchio, M., Zhou, H. & Tully, T. CREB as a memory modulator: induced expression of a dCREB2 activator isoform enhances long-term memory in *Drosophila*. *Cell* 81, 107-15 (1995).
73. Viola, H. et al. Phosphorylated cAMP response element-binding protein as a molecular marker of memory processing in rat hippocampus: effect of novelty. *J Neurosci* 20, RC112 (2000).
74. Stanciu, M., Radulovic, J. & Spiess, J. Phosphorylated cAMP response element binding protein in the mouse brain after fear conditioning: relationship to Fos production. *Brain Res Mol Brain Res* 94, 15-24 (2001).
75. Guzowski, J.F. & McGaugh, J.L. Antisense oligodeoxynucleotide-mediated disruption of hippocampal cAMP response element binding protein levels impairs consolidation of memory for water maze training. *Proc Natl Acad Sci USA* 94, 2693-8 (1997).
76. Bourtchuladze, R. et al. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* 79, 59-68 (1994).
77. Kogan, J.H. et al. Spaced training induces normal long-term memory in CREB mutant mice. *Curr Biol* 7, 1-11 (1997).
78. Gass, P. et al. Deficits in memory tasks of mice with CREB mutations depend on gene dosage. *Learn Mem* 5, 274-88 (1998).
79. Graves, L., Dalvi, A., Lucki, I., Blendy, J.A. & Abel, T. Behavioral analysis of CREB alphasdelta mutation on a B6/129 F1 hybrid background. *Hippocampus* 12, 18-26 (2002).
80. Kida, S. et al. CREB required for the stability of new and reactivated fear memories. *Nat Neurosci* 5, 348-55 (2002).
81. Han, J.-H. et al. Neuronal competition and selection during memory formation. *Science* 316, 457-60 (2007).

## **Chapter 6: Dynamic *O*-Glycosylation Regulates CREB-Mediated Neuronal Gene Expression and Memory Formation**

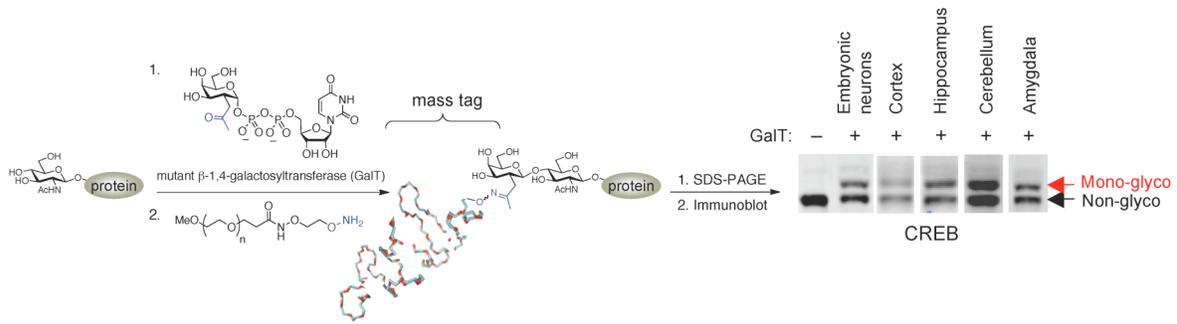
Portions of this chapter are from Clark, P.M., Rexach, J.E., Mason, D.E., Neve, R.L., Peters, E.C. & Hsieh-Wilson, L.C. Regulation of Neuronal Gene Expression and Memory Formation by Dynamic Glycosylation. *Submitted* (2010).

***O*-Glycosylation of proteins with *N*-acetyl-D-glucosamine (*O*-GlcNAc) is an abundant post-translational modification that shares key features with protein phosphorylation; however, its precise functions in the brain are not well understood. We show that *O*-GlcNAc glycosylation regulates cAMP-response element binding protein (CREB), a transcription factor critical for neuronal activity-dependent gene expression, neuronal development, and long-term memory storage. Glycosylation of CREB was dynamically induced by membrane depolarization and repressed CREB-dependent transcription by impairing the association of CREB with the CREB-regulated transcriptional co-activator (CRTC/TORC). Blocking glycosylation of CREB at a single amino acid site promoted axonal and dendritic growth and enhanced long-term memory consolidation. Our studies reveal that *O*-GlcNAc glycosylation plays a major role in essential neuronal processes and higher-order brain functions.**

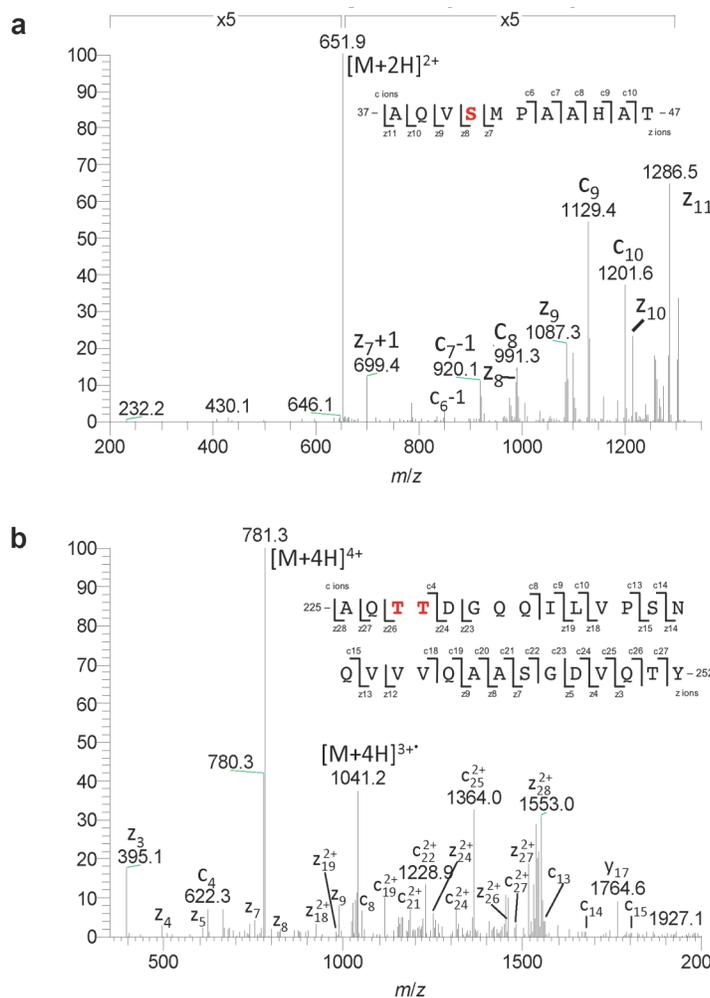
Dynamic *O*-GlcNAc glycosylation of intracellular proteins is emerging as a crucial regulatory post-translational modification<sup>1-5</sup>. Attachment of this simple glycan, *N*-acetylglucosamine (GlcNAc), to serine or threonine residues occurs on more than 1,000 proteins, including transcription factors, translational regulators, cytoskeletal components, and other nucleocytoplasmic proteins<sup>1-3,6</sup>. The fact that *O*-GlcNAc glycosylation shares key features with protein phosphorylation, which regulates neuronal processes such as cell signaling, synaptic plasticity, and learning and memory<sup>7</sup>, suggests similar, critical roles for *O*-GlcNAc in the brain. Moreover, several studies have linked *O*-GlcNAc to various neuropathologies, particularly Alzheimer's disease<sup>1-3,8,9</sup>. Despite

this intriguing body of evidence, little is known about the specific contributions of *O*-GlcNAc to fundamental neuronal functions. Thus, there is a strong rationale to explore how site-specific protein *O*-glycosylation may serve as a critical regulator of higher-order brain function.

To determine specific roles for *O*-GlcNAc in the nervous system, we examined glycosylation of CREB, a transcription factor important for neuronal development and survival, circadian rhythms, drug addiction, and long-term memory consolidation<sup>10-14</sup>. Jessica Rexach chemoenzymatically labeled proteins with terminal GlcNAc sugars with a 2,000-Da polyethylene glycol (PEG) mass tag and immunoblotted with an anti-CREB antibody to visualize the glycosylated species (**Fig. 1**). We found that a large fraction of CREB (44–48%) was mono-glycosylated in both cultured cortical neurons and various brain regions of adult mice. To map the glycosylation sites, I transiently expressed CREB in neuro2a cells, immunoprecipitated it, and subjected it to electron transfer dissociation mass spectrometry (ETD-MS) analysis. In addition to the three sites initially identified (Thr259, Ser260, Thr261)<sup>15</sup>, *O*-GlcNAc glycosylation was mapped to Ser40 and Thr227 or Thr228 within the Q1 and Q2 domains (**Figs. 2, 3**). Jessica found that expression of a mutant form of CREB, in which Ser40 was mutated to alanine (S40A), led to a large reduction in CREB glycosylation levels in cortical neurons (56%), whereas mutation of both Thr227 and Thr228 (TT227-8AA) led to a smaller decrease in glycosylation (36%; **Fig. 4**). CREB was also glycosylated within Thr259, Ser260, and Thr261 at low levels (TST259-61AAA; 13%), and simultaneous mutation of all six sites (A6) abolished the glycosylation of CREB. These results demonstrate that CREB is

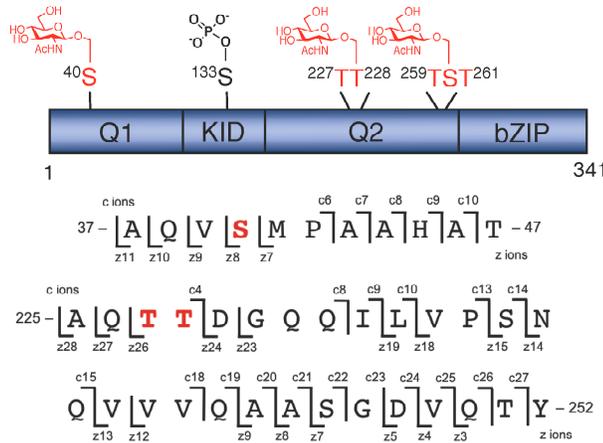


**Figure 1:** CREB is *O*-GlcNAc glycosylated in neurons. Detection of *O*-GlcNAc glycosylated CREB in neurons by chemoenzymatic labeling with a 2000-Da mass tag

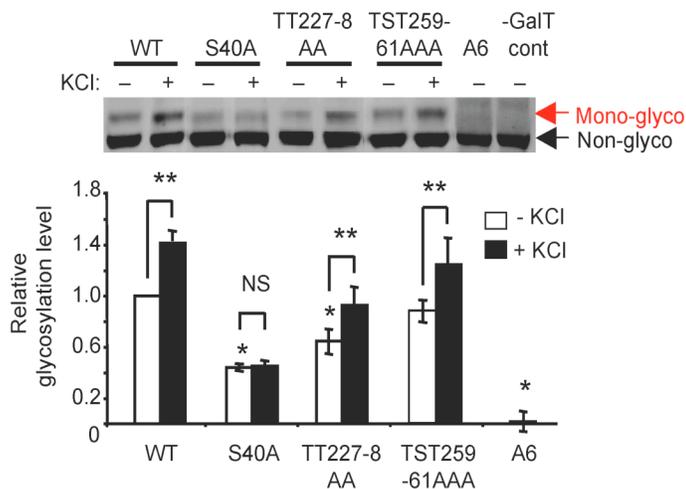


**Figure 2:** CREB is *O*-GlcNAc glycosylated at Ser40 and Thr227 or Thr228. Mass spectrum of chymotrypsin digested FLAG-CREB from Neuro2a cells. Electron transfer dissociation mass spectrometry (ETD-MS) was performed on the  $m/z$  651.9 ion (a) and the  $m/z$  781.3 ion (b). Shown are annotated spectra for two peptides identified to contain the *O*-GlcNAc modification. The c and z fragment ions observed were used to map the glycosylation sites to the residues indicated

highly glycosylated in neurons, identify all major glycosylation sites on neuronal CREB, and establish Ser40 as the predominant site of *O*-GlcNAc glycosylation.



**Figure 3:** CREB glycosylation sites. Glycosylation sites on CREB mapped by ETD-MS



**Figure 4:** CREB is *O*-GlcNAc glycosylated at Ser40 in response to neuronal activity. Glycosylation levels of Flag-tagged WT CREB and various alanine mutants after expression in cultured cortical neurons. Neurons were depolarized with KCl where indicated. ( $n = 7$  for WT and S40A CREB;  $n = 3-5$  for other mutants;  $*P < 0.01$  compared to WT, unstimulated cells;  $**P < 0.05$ ; NS, not significant). Error bars, means, and standard errors of the mean in this and subsequent figures.

Next, Jessica determined if CREB glycosylation is dynamically modulated by neuronal activity. Membrane depolarization of cortical neurons by treatment with KCl induced glycosylation of CREB. CREB

glycosylation levels increased steadily by  $42.0 \pm 4.8\%$  over the

course of 6 h, in contrast to the rapid induction of CREB phosphorylation at Ser133 (Fig. 5).

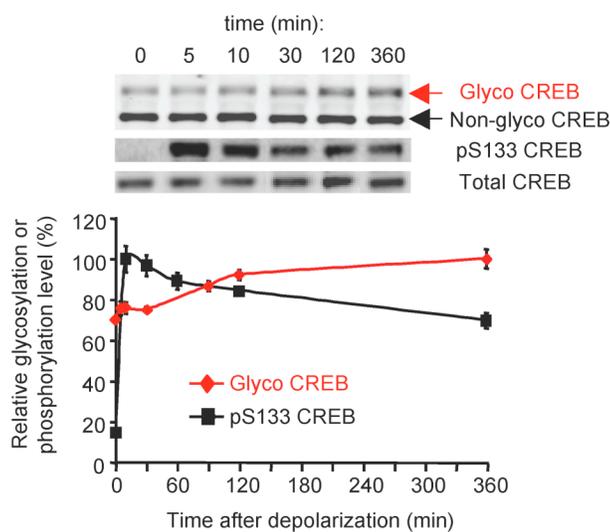
Mutation of Ser40 to alanine blocked depolarization-induced CREB glycosylation, while mutation of the other glycosylation sites had no effect (Fig. 4).

Given the slow kinetics of glycosylation, I examined whether CREB glycosylation is dependent on

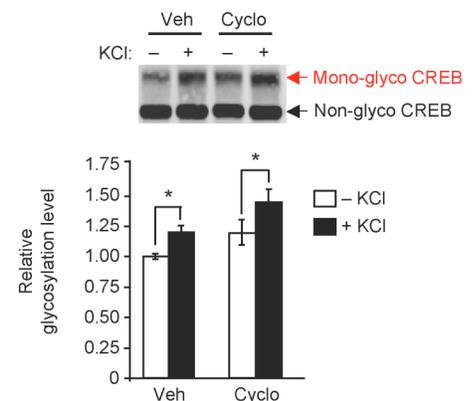
new protein synthesis. Treatment with the protein synthesis inhibitor

cycloheximide did not block the increase in glycosylation (**Fig. 6**), suggesting that glycosylation is triggered directly by signal transduction pathways. Jessica showed that inhibition of L-type calcium channels with nimodipine abolished the depolarization-induced glycosylation of CREB, indicating a requirement for voltage-sensitive calcium influx (**Fig. 7**). Moreover, inhibition of  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinases (CaMKs) or mitogen-activated protein kinase (MAPK) blocked the increase in CREB glycosylation, while inhibitors of protein kinase C or protein phosphatases PP-2B or PP-1/2A had no effect (**Fig. 7**). Together, these results show for the first time that neuronal activity stimulates *O*-GlcNAc glycosylation and specifically induces CREB glycosylation at Ser40 in a calcium- and kinase-dependent manner.

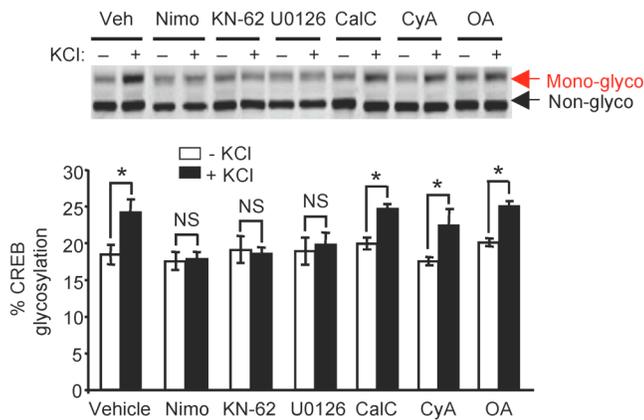
CaMKs and MAPK are known to phosphorylate CREB at Ser133, which leads to



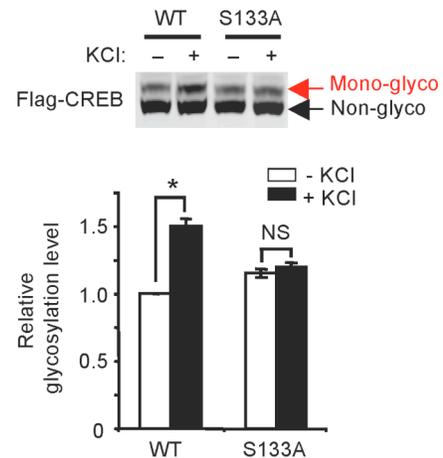
**Figure 5:** Kinetics of endogenous CREB glycosylation and Ser133 phosphorylation upon depolarization of cortical neurons. Levels of glycosylation or phosphorylation are plotted relative to the maximum signal for each modification. ( $n = 4-6$ ).



**Figure 6:** The protein synthesis inhibitor cycloheximide does not block depolarization-induced glycosylation of CREB. Glycosylation levels of endogenous CREB. Neurons were pretreated with cycloheximide (Cyclo) or DMSO vehicle (Veh) and then incubated in the presence or absence of KCl (55 mM) ( $n = 3$ ,  $*P < 0.01$ ).



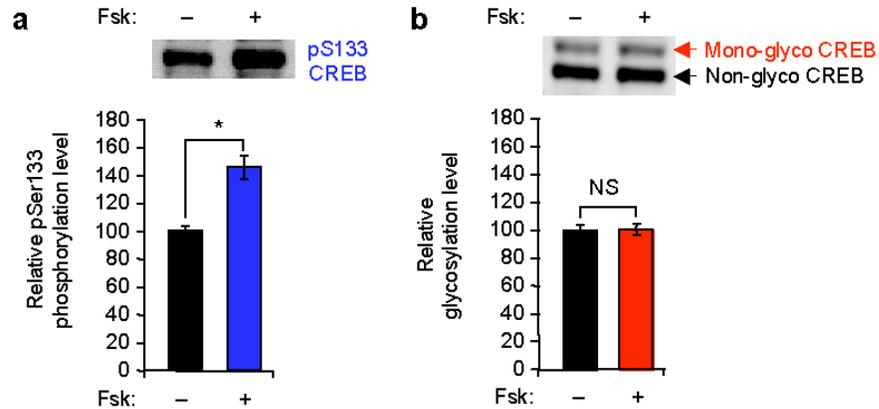
**Figure 7:** CREB glycosylation is modulated by specific kinase pathways. Glycosylation levels of endogenous CREB in unstimulated or KCl-stimulated cortical neurons upon treatment with inhibitors of L-type calcium channels (nimodipine; nimo), CaMKs (KN-62), MAPK (U0126), protein kinase C (calphostin C; CalC), PP-2B (cyclosporin A; CyA), or PP-1/2A (okadaic acid; OA). ( $n = 3-6$ ,  $*P < 0.02$ ).



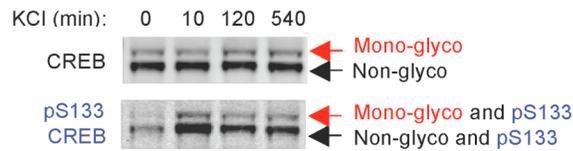
**Figure 8:** Glycosylation is not induced on S133A CREB. Glycosylation levels were analyzed on Flag-tagged WT or S133A CREB expressed in cortical neurons. ( $n = 22$ ,  $*P < 0.001$ ).

recruitment of the coactivator CREB-binding protein (CBP) and activation of CREB-mediated transcription<sup>16,17</sup>. As these same kinases are necessary for activity-dependent glycosylation of CREB, Jessica determined whether Ser133 phosphorylation is required for CREB glycosylation. Mutation of Ser133 to Ala (S133A) blocked the KCl-induced increase in CREB glycosylation (**Fig. 8**). However, I showed that forskolin-mediated stimulation of Ser133 phosphorylation via the cAMP pathway failed to induce CREB glycosylation (**Fig. 9**), suggesting that phosphorylation at Ser133 may be required but is not sufficient to activate CREB glycosylation. We recently reported a rapid, chemoenzymatic strategy to probe the interplay between phosphorylation and *O*-GlcNAc glycosylation on target proteins<sup>18</sup>. Using this approach, Jessica examined the interdependence of Ser133 phosphorylation and Ser40 glycosylation on CREB. Cortical neuronal lysates were labeled with a 2,000-Da mass tag and immunoblotted with a

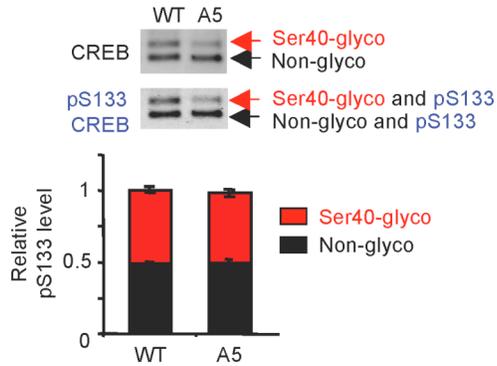
phospho-Ser133-specific or total CREB antibody to enable visualization of four distinct subpopulations: (1) mono-glycosylated, (2) nonglycosylated, (3) mono-glycosylated and Ser133-phosphorylated, and (4) nonglycosylated and Ser133-phosphorylated CREB (**Fig. 10**). A significant subpopulation of endogenous CREB was simultaneously phosphorylated and glycosylated in both unstimulated and depolarized neurons (**Fig. 10**), consistent with the notion that phosphorylation and glycosylation cooperatively regulate CREB activity. Moreover, I showed that WT CREB and a mutant form (A5), in which all glycosylation sites except Ser40 were mutated to alanine, showed comparable levels of phospho-Ser133 induction when expressed in neurons, and both Ser40 glycosylation and Ser133 phosphorylation occurred concomitantly on the same protein molecule (**Fig. 11**). Notably, Jessica found that the kinetics of Ser133 phosphorylation upon KCl depolarization was similar for both the glycosylated and nonglycosylated subpopulations of endogenous CREB (**Fig. 10, 12**), indicating that Ser133 phosphorylation occurs independent of the glycosylation state. However, glycosylation was more rapidly induced on the Ser133-phosphorylated subpopulation compared to the total population of endogenous CREB (**Fig. 10, 13**). Collectively, these results strongly suggest that neuronal activity-dependent glycosylation of CREB at Ser40 is induced preferentially on the Ser133-phosphorylated subpopulation.



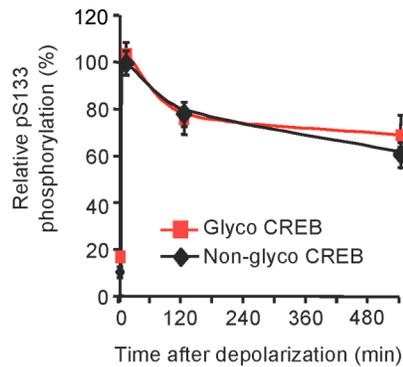
**Figure 9:** Forskolin induces CREB phosphorylation (a) but not CREB glycosylation (b) in cortical neurons. Neurons were treated with forskolin or DMSO vehicle and lysates were either immunoblotted for pSer133 CREB or chemoenzymatically labeled with a polyethylene glycol mass tag and immunoblotted for CREB to visualize the glycosylated CREB subpopulation. ( $n = 3$ ,  $*P < 0.01$ ; NS, not significant).



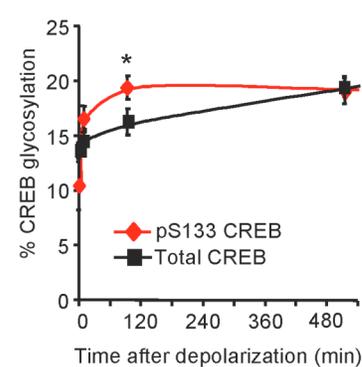
**Figure 10:** Chemoenzymatic labeling of endogenous CREB for visualizing phosphorylation and glycosylation within the same protein molecule and for quantifying the levels of each modification within distinct post-translationally modified subpopulations



**Figure 11:** CREB is simultaneously phosphorylated at Ser133 and glycosylated at Ser40. Quantification of pSer133 levels on Flag-tagged WT or A5 mutant CREB following 10-min depolarization. ( $n = 3$ ).



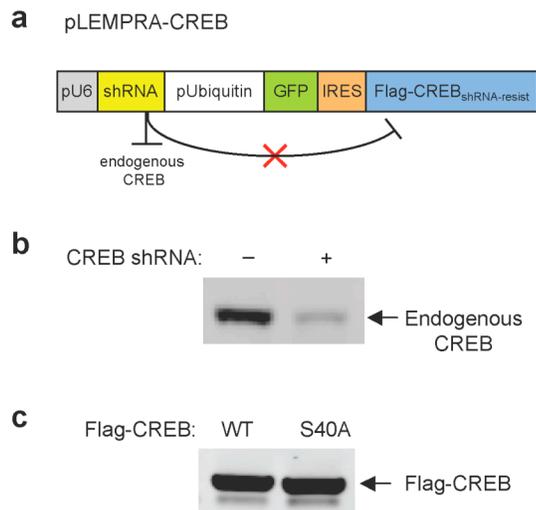
**Figure 12:** Kinetics of Ser133 phosphorylation for specific post-translationally modified subpopulations of endogenous CREB. ( $n = 4$ ,  $*P < 0.03$ ).



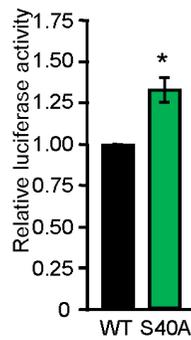
**Figure 13:** Kinetics of glycosylation for specific post-translationally modified subpopulations of endogenous CREB. ( $n = 4$ ,  $*P < 0.03$ ).

To determine whether glycosylation modulates CREB activity, I compared the ability of wild-type (WT) and S40A mutant CREB to regulate CRE-dependent gene expression. A short hairpin RNA (shRNA) was used to knockdown endogenous CREB in neuro2a neuroblastoma cells, and shRNA-resistant WT or S40A mutant CREB was overexpressed (**Fig. 14**). Replacement of endogenous CREB with the S40A mutant resulted in increased CRE-luciferase activity (**Fig. 15**), suggesting that glycosylation represses the transcriptional activity of CREB. The S40A substitution also upregulated expression of endogenous CREB target genes, including *CDKN1A*, *NR4A2*, and *OPA3*

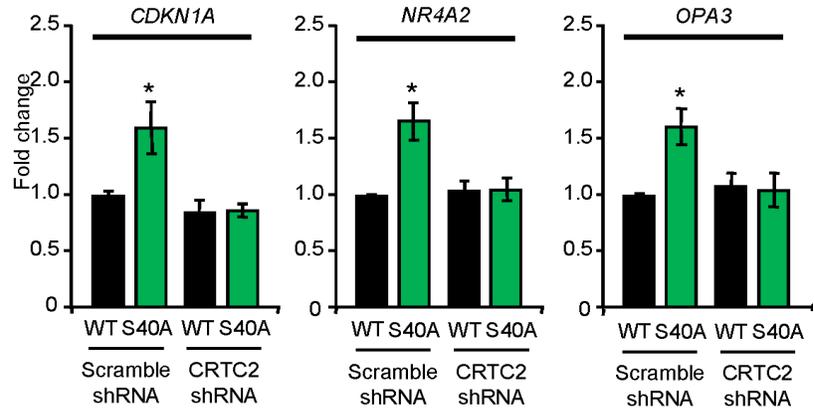
(**Fig. 16**). To investigate the mechanism, I evaluated whether glycosylation affects the ability of CREB to associate with DNA or transcriptional co-activators. Binding of CREB to the CRE promoter was unaffected by the S40A mutation in an electrophoretic mobility shift assay (**Fig. 17**). However, binding of CREB to CRTC2, a co-activator that stimulates both basal and induced CREB transcription in neurons<sup>19</sup>, was significantly enhanced by the S40A mutation in reciprocal co-immunoprecipitation assays (**Fig. 18**).



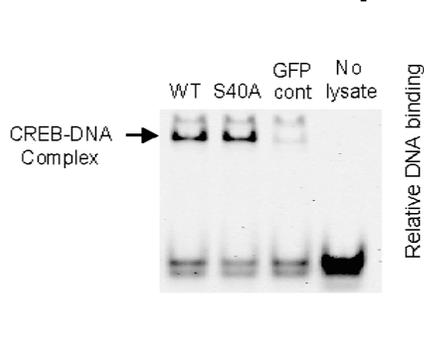
**Figure 14:** Knockdown of endogenous CREB and overexpression of shRNA-resistant, Flag-tagged WT or S40A CREB in neuro2a cells. **a**, pLEMPRA-CREB construct used to knockdown endogenous CREB and express Flag-tagged CREB. **b**, Knockdown of endogenous CREB using a vector containing the shRNA sequence. **c**, Overexpression of shRNA-resistant CREB from the pLEMPRA construct. Western blots are representative of five independent experiments.



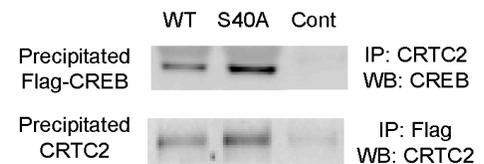
**Figure 15:** Glycosylation at Ser40 represses CREB activity. CRE-luciferase activity in neuro2a cells expressing WT or S40A CREB. ( $n = 11$ ,  $*P < 0.01$ ).



**Figure 16:** Glycosylation at Ser40 represses CREB activity. Quantitative polymerase chain reaction (qPCR) analysis of *CDKN1A*, *NR4A2*, and *OPA3* expression in cells transfected with the indicated shRNAs or expression vectors using *RPL3* as an internal control ( $n = 8$ ,  $*P < 0.01$ ).



**Figure 17:** WT and S40A CREB show similar binding to the CRE promoter *in vitro*. Neuro2a cells were transfected with WT or S40A CREB, lysed, and an electrophoretic mobility shift assay was performed using IRDye 700-labeled oligos containing a CRE promoter sequence. ( $n = 3$ ; NS, not significant).

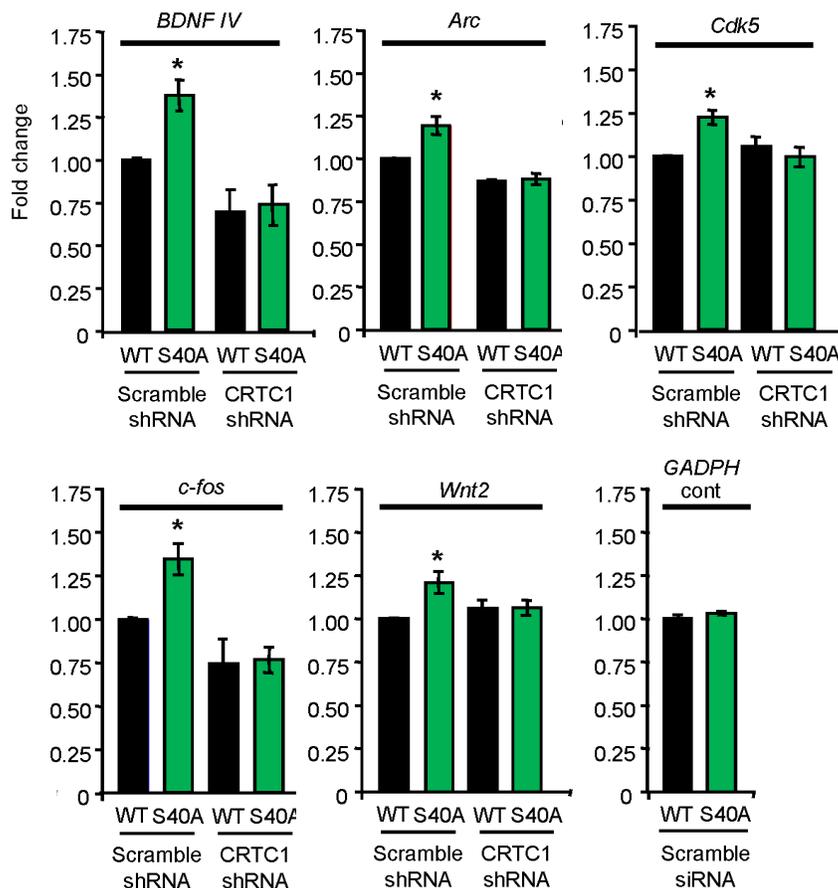


**Figure 18:** Glycosylation at Ser40 blocks the interaction between CREB and CRTC2. Co-immunoprecipitation of the CREB-CRTC2 complex from neuro2a cells expressing WT or S40A CREB. ( $n = 4$ ).

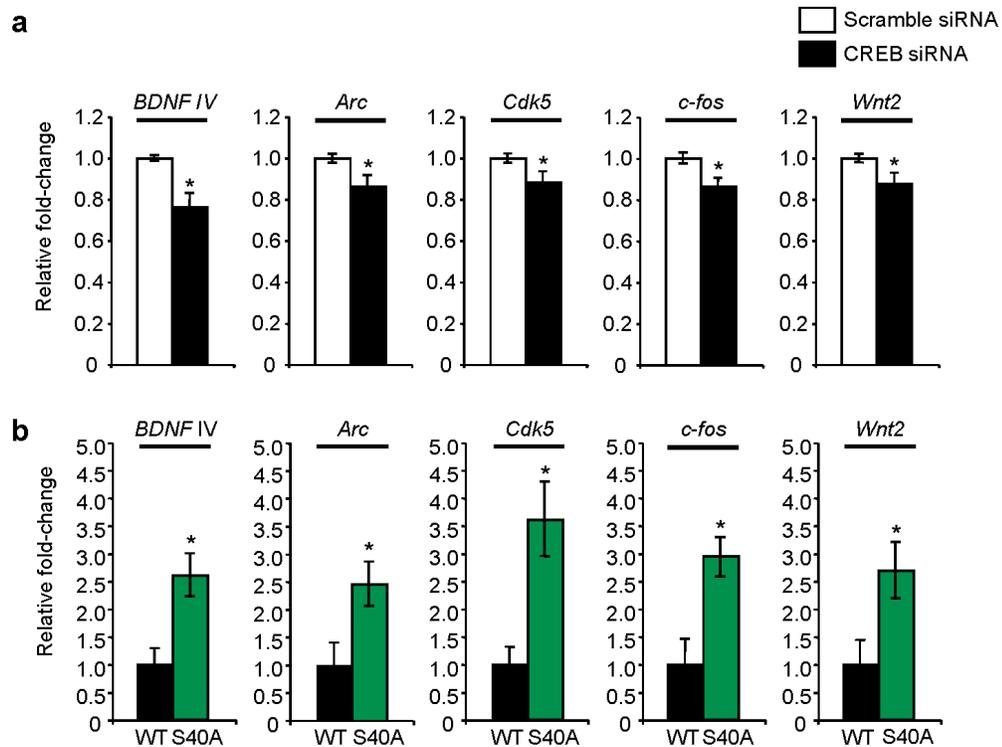
Furthermore, knockdown of CRTC2 expression in neuro2a cells abolished the observed increases in *CDKN1A*, *NR4A2*, and *OPA3* transcript levels for S40A CREB compared to

WT CREB (**Fig. 16**). Together, these findings indicate that glycosylation impairs the ability of CREB to activate transcription by disrupting the CREB-CRTC interaction.

I next determined whether glycosylation at Ser40 regulates neuronal gene expression, focusing on well-characterized genes involved in brain development and memory consolidation<sup>20-24</sup>. Relative to WT CREB, expression of S40A CREB in cortical neurons increased the levels of *BDNF* exon IV, *Arc*, *Cdk5*, *c-fos*, and *Wnt2* transcripts (**Fig. 19**). Taking into account the contribution of CREB to the expression of each gene, as measured using CREB siRNA, the observed increases correspond approximately to a 2.5–3.6-fold induction in CREB-dependent transcription (**Fig. 20**). Consistent with a mechanism involving direct regulation of these genes through modulation of the CREB-

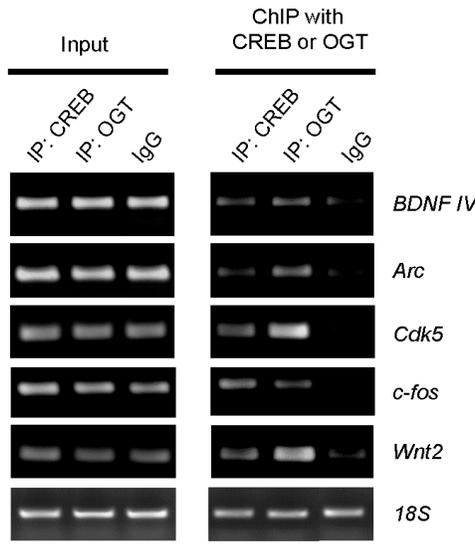


**Figure 19:** Glycosylation at Ser40 represses CREB activity via a CRTC-dependent mechanism in neurons. qPCR analysis of *BDNF* exon IV, *Arc*, *Cdk5*, *c-fos*, and *Wnt2* expression in cultured cortical neurons electroporated with the indicated siRNAs or expression vectors using *RPL3* as an internal control. ( $n = 4-9$ ,  $*P < 0.01$ ).

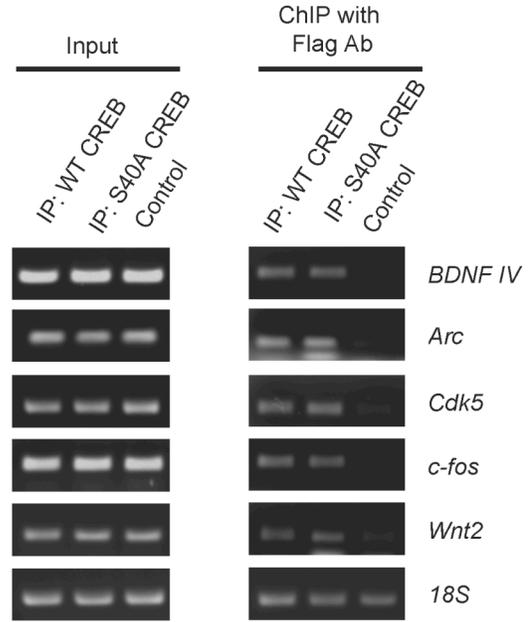


**Figure 20:** S40A CREB produces approximately a 2.5–3.6-fold increase in mRNA expression relative to WT CREB. **a**, Cortical neurons were electroporated with scramble or CREB siRNA and the mRNA levels of each gene were measured by quantitative RT-PCR. ( $n = 14$ ,  $*P < 0.05$ ). **b**, Cortical neurons were electroporated with WT CREB or S40A CREB and the mRNA levels of each gene were measured by quantitative RT-PCR. The fold-change was calculated by subtracting the transcript level of each gene in neurons transfected with CREB siRNA from the transcript levels of each gene in WT or S40A CREB-expressing neurons and then normalizing the WT CREB levels to 1. ( $n = 9$ ,  $*P < 0.01$ ).

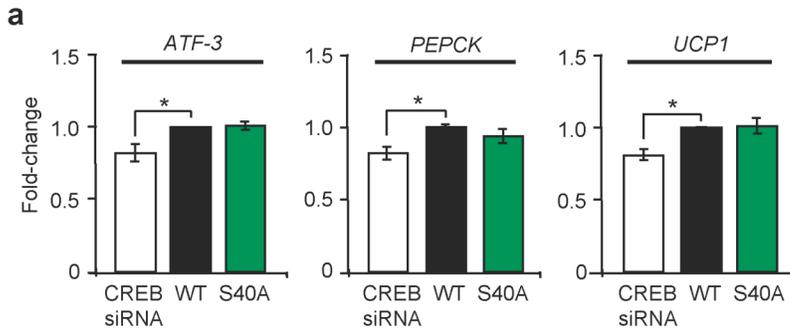
CRTC interaction, both CREB and *O*-GlcNAc transferase (OGT) were bound to the promoters of each gene, and WT and S40A CREB showed comparable levels of promoter occupancy in chromatin immunoprecipitation assays (Fig. 21, 22). Moreover, siRNA-mediated knockdown of CRTC1 reversed the effects of S40A CREB on neuronal gene expression (Fig. 19). Interestingly, no increases in the transcript levels of *ATF3*, *PEPCK*, or *UCP1* were detected even though CREB was bound to their promoters (Fig. 23), suggesting that *O*-glycosylation at Ser40 may confer specificity in the regulation of a subset of CREB-dependent genes. Finally, I investigated the effects of Ser40



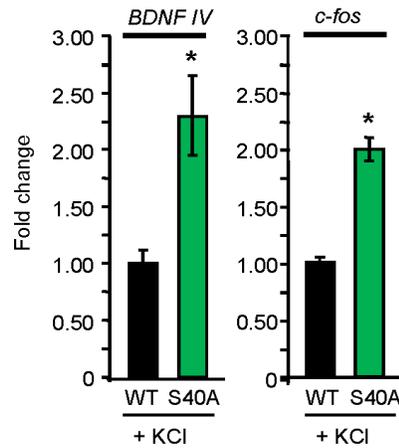
**Figure 21:** Both CREB and OGT occupy the *BDNF* exon IV, *Arc*, *Cdk5*, *c-fos*, and *Wnt2* promoters. Chromatin immunoprecipitation with an anti-CREB, anti-OGT, or IgG antibody was followed by PCR for the indicated promoters. ( $n = 3$ ).



**Figure 22:** WT and S40A CREB show comparable levels of promoter occupancy on *BDNF* exon IV, *Arc*, *Cdk5*, *c-fos*, and *Wnt2* promoters. Cortical neurons were electroporated with WT or S40A Flag-tagged CREB. Chromatin immunoprecipitation with an anti-FLAG or IgG antibody was followed by PCR for the indicated promoters. ( $n = 3$ ).



**Figure 23:** S40A CREB regulates a subset of CREB-mediated genes. **a**, Cortical neurons were electroporated with CREB siRNA, WT CREB, or S40A CREB as noted. Quantitative PCR was performed on the indicated genes. ( $n = 4$ ,  $*P < 0.05$ ). **b**, Chromatin immunoprecipitation was performed with an anti-CREB or IgG control antibody. PCR was carried out using primers specific for the promoters of the indicated genes. Representative gels from three independent experiments are shown.



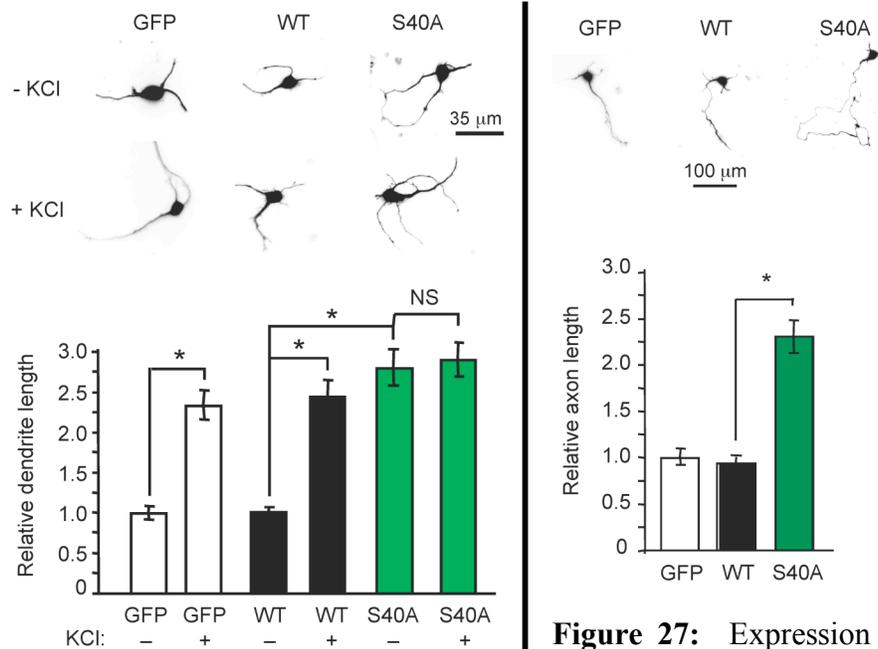
**Figure 24:** Blocking glycosylation enhances *BDNF IV* and *c-fos* expression even more after depolarization. qPCR analysis of *BDNF* exon IV and *c-fos* expression after membrane depolarization of cultured cortical neurons expressing WT or S40A CREB. ( $n = 10$ ,  $*P < 0.01$ ).

glycosylation on activity-dependent gene expression. Consistent with the observation that neuronal activity enhances CREB glycosylation, blocking glycosylation of CREB at Ser40 increased the levels of *BDNF* exon IV and *c-fos* transcripts to a greater extent in membrane-depolarized neurons compared to unstimulated neurons (**Fig. 24**). Taken together, the results indicate that CREB glycosylation at Ser40 modulates both basal and activity-dependent gene expression, thereby regulating genes important for neuronal development, survival, and synaptic plasticity.



**Figure 25:** WT and S40A Flag-tagged CREB are expressed at similar levels in cortical neurons. Cortical neurons were electroporated with WT or S40A CREB. After 3 DIV, the lysate was immunoblotted for Flag-tagged CREB.

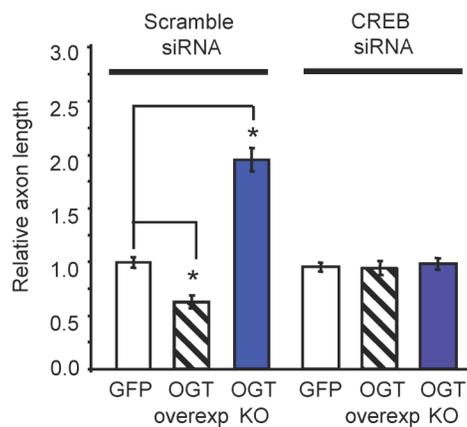
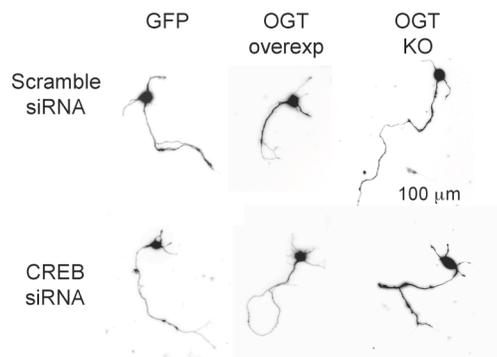
CREB has critical roles in several aspects of neuronal development, including axon growth, activity-dependent dendrite development, and synaptogenesis<sup>10,22,25</sup>. To assess the functional consequences of Ser40 glycosylation on neuronal growth, Jessica assayed axonal and dendritic extension in cortical neurons expressing WT or



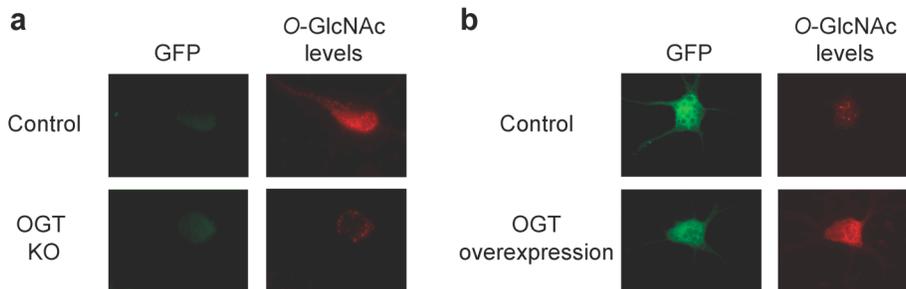
**Figure 26:** CREB glycosylation at Ser40 represses dendritic growth. Relative total dendrite lengths of cortical neurons expressing GFP, WT CREB or S40A CREB. ( $n = 90$  from three independent experiments,  $*P < 0.001$ ; NS, not significant).

**Figure 27:** Expression of S40A CREB enhances axon outgrowth. Relative axon lengths of cortical neurons expressing, GFP, WT CREB or S40A CREB ( $n = 90$  from three independent experiments,  $*P < 0.001$ ).

S40A mutant CREB (**Fig. 25**). Dendrites of neurons expressing WT CREB or a GFP control exhibited similar lengths, and as expected, their growth was stimulated by membrane depolarization (**Fig. 26**). In contrast, neurons expressing S40A CREB had significantly longer dendrites than WT CREB-expressing neurons (2.77-fold increase), displaying lengths comparable to those of depolarized neurons, and their dendrites showed no further elongation upon membrane depolarization (**Fig. 26**). Additionally, neurons expressing S40A CREB had significantly longer axons compared to WT CREB- and GFP-expressing controls (**Fig. 27**). Therefore, preventing CREB glycosylation at Ser40 enhances the outgrowth of both axons and dendrites *in vitro*, indicating that glycosylation has a large, chronic repressive effect on multiple developmental pathways.



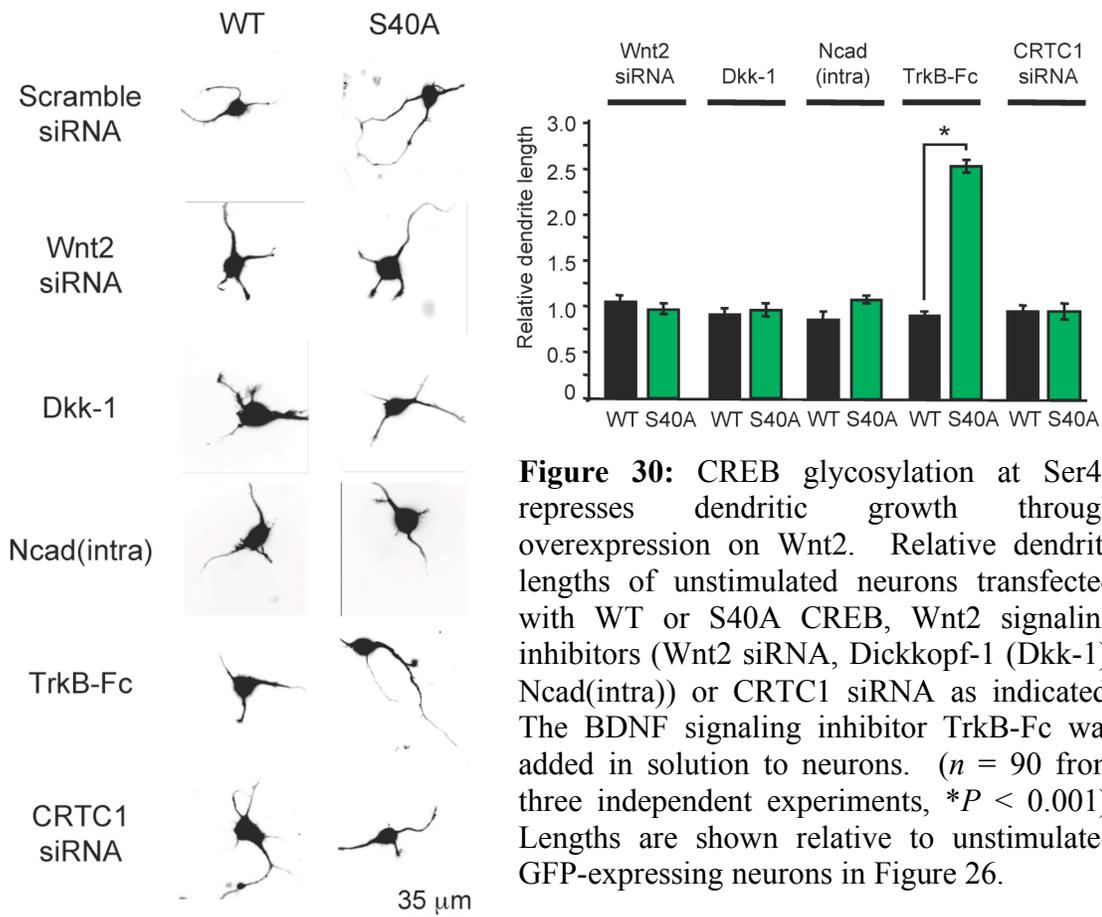
**Figure 28:** OGT expression represses while OGT knockout enhances axon length. Relative axon lengths of OGT floxed cortical neurons expressing GFP, Flag-tagged OGT, or Cre recombinase (OGT KO). Neurons were also transfected with scramble or CREB siRNA as indicated. ( $n = 90$  from three independent experiments,  $*P < 0.001$ ).



**Figure 29:** OGT knockout decreases *O*-GlcNAc levels, and OGT overexpression increases *O*-GlcNAc levels. **a**, Neurons from OGT floxed mice electroporated with a GFP construct (control) or with GFP and CRE recombinase constructs (OGT KO). **b**, Neurons from WT mice electroporated with a GFP construct (control) or with an OGT construct containing GFP (OGT overexpression). Cells were immunostained for overall *O*-GlcNAc levels using the pan-specific *O*-GlcNAc antibodies CTD110.6 and RL2, respectively.

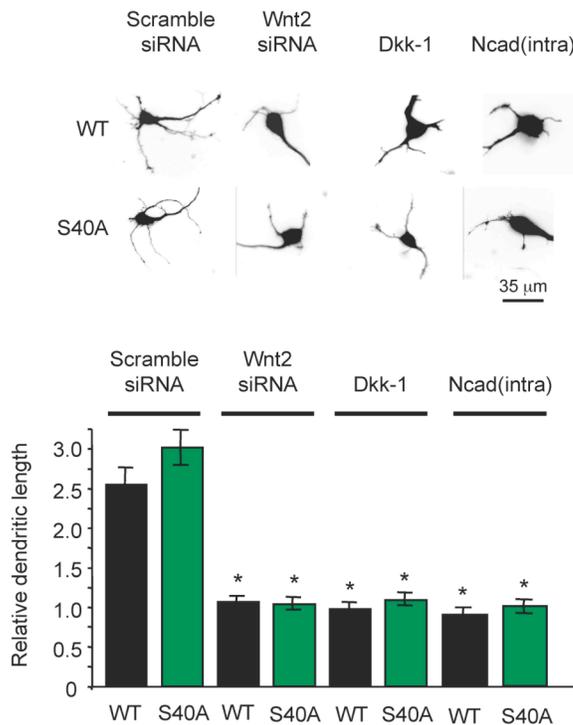
To confirm further that these effects are occurring directly through *O*-glycosylation of CREB, Jessica performed OGT gain- and loss-of-function experiments. OGT-null neurons were generated from OGT floxed mice<sup>26</sup> by expression of CRE

recombinase in cultured cortical neurons. Knocking out OGT reduced overall *O*-GlcNAc glycosylation levels and stimulated axonal growth, whereas OGT overexpression enhanced *O*-GlcNAc glycosylation levels and attenuated axonal growth (**Fig. 28, 29**). In both cases, siRNA-mediated knockdown of endogenous CREB reversed the effects of OGT knockout or overexpression and restored axon lengths to those of GFP-expressing neurons (**Fig. 28**), indicating that *O*-GlcNAc glycosylation modulates axonal growth through a CREB-dependent mechanism.

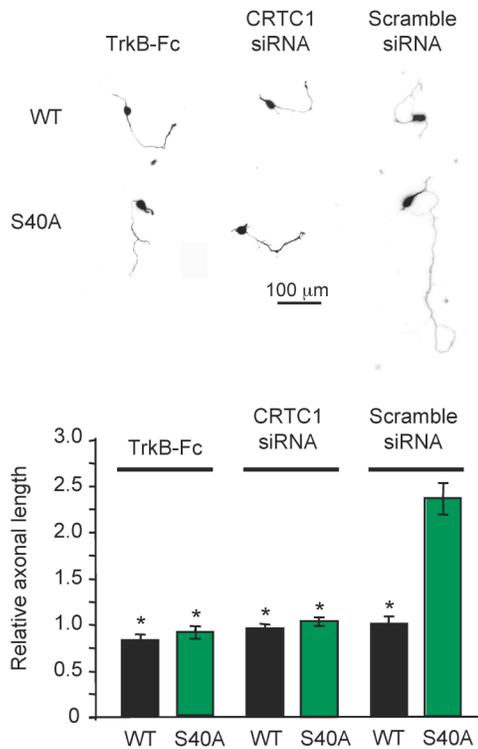


**Figure 30:** CREB glycosylation at Ser40 represses dendritic growth through overexpression on Wnt2. Relative dendrite lengths of unstimulated neurons transfected with WT or S40A CREB, Wnt2 signaling inhibitors (Wnt2 siRNA, Dickkopf-1 (Dkk-1), Ncad(intra)) or CRTIC1 siRNA as indicated. The BDNF signaling inhibitor TrkB-Fc was added in solution to neurons. ( $n = 90$  from three independent experiments,  $*P < 0.001$ ). Lengths are shown relative to unstimulated GFP-expressing neurons in Figure 26.

To investigate the molecular mechanisms further, Jessica and I considered known mediators of CREB-dependent dendrite and axon elongation. Activation of CREB drives the expression of the secreted mitogen *Wnt2* to regulate activity-dependent dendritic growth, whereas application of the neurotrophin BDNF leads to axon elongation<sup>22,27</sup>. Both *Wnt2* and *BDNF* transcript levels were significantly increased in cortical neurons expressing S40A CREB as compared to WT CREB (**Fig. 19**). I showed that knockdown of *Wnt2*, overexpression of the *Wnt2* antagonist Dickkopf-1, or overexpression of the  $\beta$ -catenin sequestrant *Ncad*(intra) reversed the stimulatory effects of both S40A CREB (**Fig. 30**) and neuronal depolarization (**Fig. 31**) on dendritic growth. Alternatively, Jessica showed that treatment with the BDNF/NT-4/5 scavenger TrkB-Fc blocked the effects of S40A CREB specifically on axon growth, but not dendrite growth (**Fig. 30, 32**). Moreover, knockdown of *CRTC1* abolished the S40A CREB-dependent increases in *Wnt2/BDNF* gene expression (**Fig. 19**), dendritic growth (**Fig. 30**), and axonal growth



**Figure 31:** Depolarization-induced dendrite growth is blocked by *Wnt2* siRNA, *Dkk-1*, and *Ncad*(intra). Relative axon lengths of cortical neurons were electroporated with Flag-tagged WT or S40A CREB constructs containing GFP, along with scramble siRNA, *Wnt2* siRNA, *Dkk-1*, or *Ncad*(intra) constructs and treated with KCl as indicated. Lengths relative to neurons expressing WT CREB in the absence of KCl stimulation are shown. (\* $P < 0.001$  relative to scramble siRNA for each respective genotype).



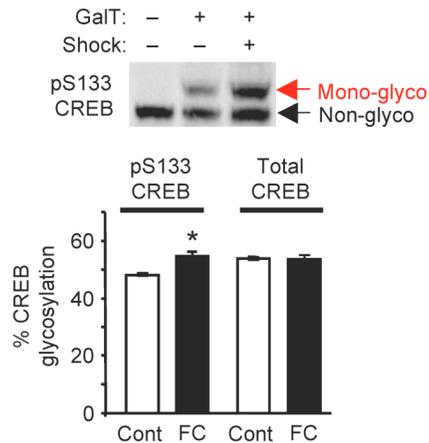
**Figure 32:** The enhanced axon outgrowth caused by expression of S40A CREB is blocked by TrkB-Fc or CRTC1 knockdown. Cortical neurons were electroporated with Flag-tagged WT or S40A CREB constructs containing GFP, along with CRTC1 or scramble siRNA as indicated. TrkB-Fc (0.7  $\mu$ g/ml) or vehicle was added after 1 DIV where indicated. ( $n = 90$  from three independent experiments,  $*P < 0.0001$ ).

(**Fig. 32**). Thus, CREB glycosylation modulates dendrite and axon elongation via the CRTC-dependent downregulation of Wnt2 and BDNF signaling, respectively. These findings strongly suggest that *O*-glycosylation of CREB functions as a critical regulator of neuronal growth. By controlling the basal threshold levels of key genes, glycosylation of CREB exerts a chronic repressive effect on neuronal growth and enables appropriate stimuli-induced growth responses.

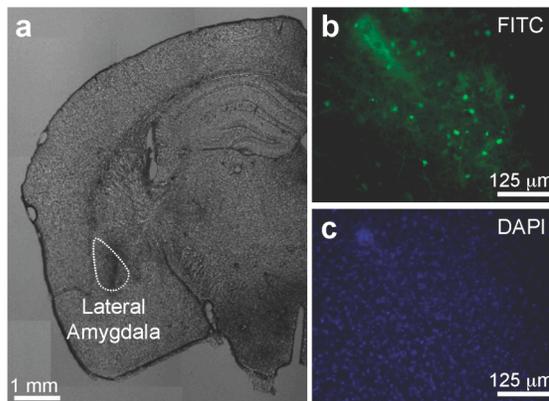
Having shown that *O*-GlcNAc glycosylation modulates important CREB-dependent cellular processes, we next examined the role of CREB glycosylation in higher-order brain functions *in vivo*. CREB-dependent transcription is essential for the consolidation of long-term conditioned fear memories<sup>14,28</sup>. Jessica and I first examined whether glycosylation is induced on endogenous CREB in response to auditory fear conditioning. Specifically, the glycosylation levels of CREB were compared in the lateral amygdala of fear-conditioned mice and tone-only trained controls. An increase in glycosylation ( $13.6 \pm 0.3\%$ ) was detected specifically within the activated CREB subpopulation (phosphorylated at Ser133; **Fig. 33**), indicating that glycosylation

of CREB is induced following activation of amygdala neurons *in vivo*. To determine whether CREB glycosylation affects memory formation *in vivo*, Jessica bilaterally microinjected replication-defective herpes simplex viral (HSV) vectors expressing WT CREB and GFP, S40A CREB and GFP, or GFP alone into the lateral amygdala of mice before fear conditioning (**Fig. 34**), and assessed memory 30 min, 2 h, and 24 h after training. Similar to previous experiments<sup>29,30</sup>, mice infused with WT CREB vector had enhanced memory compared to GFP vector-infused mice after 24 h, but not after 30 min or 2 h (**Fig. 35**;  $F_{1,25} = 4.34$ ,  $P = 0.048$ ), indicating that CREB overexpression increases long-term fear memory. Notably, Jessica found that mice infused with S40A CREB vector exhibited significant memory enhancement 2 h after training compared to mice infused with WT CREB or GFP (**Fig. 35**;  $F_{2,45} = 9.70$ ,  $P = 0.0003$ ). To test whether this effect represents enhanced long-term memory, a CREB-dependent process that requires *de novo* mRNA and protein synthesis<sup>14,28</sup>, Jessica injected the mice with the protein synthesis inhibitor, anisomycin, at various points after training and then assessed memory. Inhibiting protein synthesis immediately after training blocked the memory enhancement of S40A CREB at 2 h (**Fig. 36**;  $F_{1,12} = 24.57$ ,  $P = 0.0003$ ), while inhibiting protein synthesis 2 h after training failed to block the memory enhancement at 24 h (**Fig. 37**;  $F_{1,13} = 0.23$ ,  $P = 0.64$ ). These results indicate that mice expressing S40A CREB have enhanced, consolidated long-term memory at 2 h. As with the neurite outgrowth studies, we observed an accelerated response upon removing the repressive effects of glycosylation. Limited information is known about the genes and mechanisms that control the rate of memory consolidation. However, our results are consistent with the requirement for *de novo* gene expression and suggest that blocking CREB glycosylation

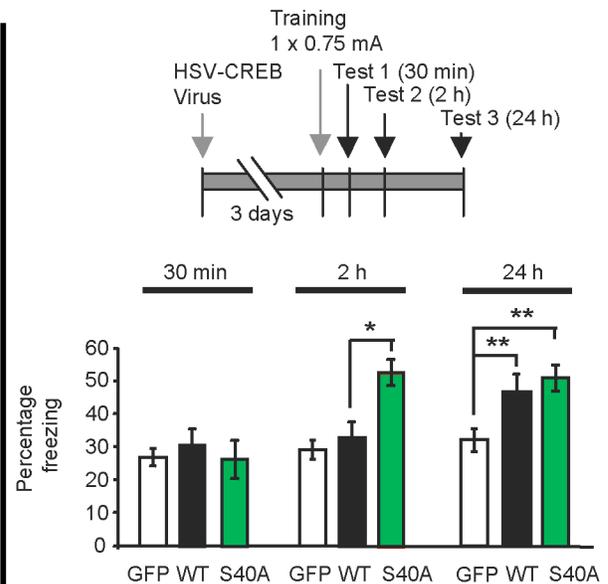
leads to the accumulation of plasticity-related transcripts and the facilitation of rapid long-term memory consolidation. More broadly, these findings provide the first direct demonstration that *O*-GlcNAc glycosylation plays a role in higher-order brain functions.



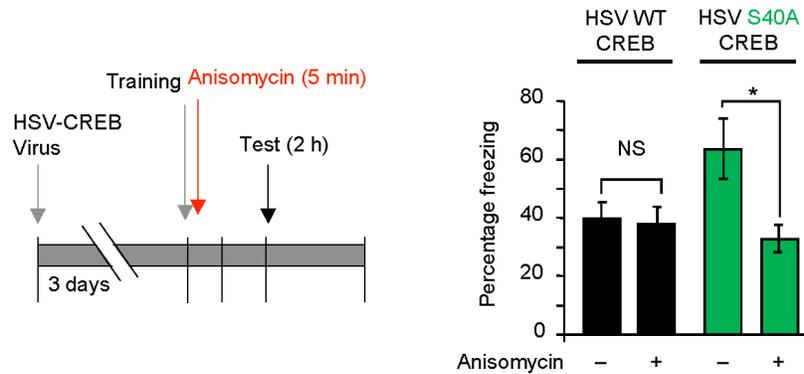
**Figure 33:** CREB glycosylation is induced following activation of neurons *in vivo*. Glycosylation levels of activated Ser133-phosphorylated CREB and total CREB in the amygdala 15 min after auditory fear conditioning. ( $n = 3$ ,  $*P < 0.01$ ).



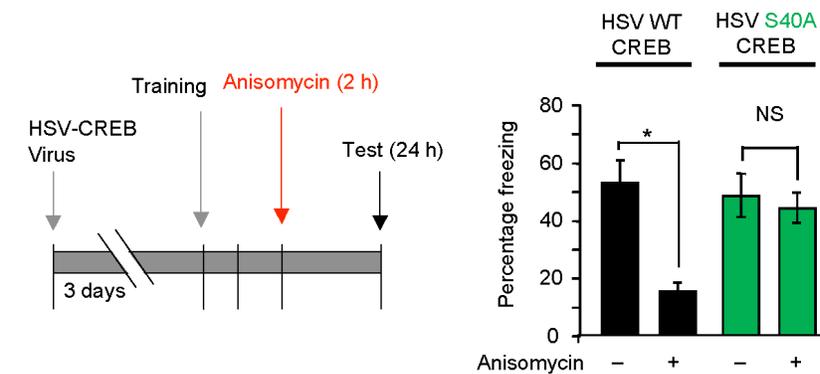
**Figure 34:** Herpes simplex virus (HSV) infection of the lateral amygdala of mice. The lateral amygdala [AP = -1.3, ML = +/-3.3, V = -4.8 from bregma] of 9–10 week old male C57BL/6 mice was injected with WT CREB, S40A CREB, or GFP HSV (1.5 ml) for 25 min. **a**, Bright field image of a representative brain section infused with WT CREB vector showing an outline of the lateral amygdala. **b**, FITC image of injection site. **c**, DAPI image of (b)



**Figure 35:** CREB glycosylation at Ser40 modulates long-term conditioned fear memory. Freezing behavior after auditory fear conditioning of mice infused with HSV vectors expressing GFP, WT or S40A CREB. ( $n = 11$  for GFP,  $n = 16$  for WT, and  $n = 20$  for S40A at 2 h and 24 h,  $n = 6$  for all vectors at 30 min.  $*P < 0.05$  compared to GFP after 24 h,  $**P < 0.0005$  compared to WT and GFP at 2 h).



**Figure 36:** Inhibition of protein synthesis immediately after training blocks the memory enhancement of S40A CREB at 2 h. Replication-defective HSV vectors expressing WT CREB and GFP, S40A CREB and GFP, or GFP alone were bilaterally microinjected into the lateral amygdala of mice 3 days prior to auditory fear conditioning training. Mice were injected with anisomycin or saline 5 min after fear conditioning, and freezing behavior was assessed after 2 h. ( $n = 6$  for WT,  $n = 7$  for S40A,  $*P < 0.0005$ ; NS, not significant)



**Figure 37:** Inhibition of protein synthesis 2 h after training blocks the memory enhancement of WT CREB at 24 h, but not that of S40A CREB. Replication-defective HSV vectors expressing WT CREB and GFP, S40A CREB and GFP, or GFP alone were bilaterally microinjected into the lateral amygdala of mice 3 days prior to auditory fear conditioning training. Mice were injected with anisomycin or saline 2 h after fear conditioning, and freezing behavior was assessed after 24 h. The memory enhancement of S40A CREB at 24 h was anisomycin-resistant. ( $n = 6$  for WT,  $n = 8$  for S40A anisomycin,  $n = 7$  for S40A saline,  $*P < 0.005$ ; NS, not significant).

The importance of dynamic *O*-GlcNAc glycosylation in the regulation of glucose homeostasis and insulin signaling is well appreciated<sup>1,3-5</sup>. Our study expands the scope of

cellular regulation by *O*-glycosylation to the brain and demonstrates that it serves functions in the brain comparable to other major posttranslational modifications such as phosphorylation. We show that *O*-glycosylation is dynamically modulated by neuronal signaling pathways and works cooperatively with Ser133 phosphorylation to allow for graded suppression of the activated CREB subpopulation. *O*-Glycosylation also suppresses the constitutive transcriptional activity of CREB in quiescent neurons and limits basal gene expression levels, allowing for a larger dynamic range of induction and proper activity-induced responses such as neuronal growth and memory consolidation. In these ways, glycosylation enables the fine-tuned, exquisite coupling of extracellular stimuli to transcriptional regulation in ways not achieved by phosphorylation or transcriptional repressors. Collectively, our results demonstrate how site-specific protein *O*-glycosylation contributes to complex neuronal processes and reveal its potential as a critical regulator of higher-order brain function.

## **Methods**

**Construction of expression plasmids and viruses.** All constructs were generated using standard molecular biology methods and verified by DNA sequencing.

*pLEMPRA CREB.* Rat CREB cDNA, containing an N-terminal Flag tag, was cloned into the lentiviral expression vector pLEMPRA-GOI (provided by M. Greenberg) immediately following the GFP-IRES sequence. Subsequently a CREB shRNA cassette (described below) was subcloned into the pLEMPRA vector. The CREB sequence was made shRNA-resistant by introducing the following five silent mutations (lower case

letters): GGAGagcGTGGATAGcGTg. Various CREB alanine mutations were created from the resulting vector using the QuikChange Lightning Mutagenesis kit (Stratagene).

*pLenti CREB.* The pLenti WT and alanine mutant CREB plasmids were created by subcloning Flag-tagged WT and alanine mutant CREB sequences from the pLEMPRA CREB vectors in place of the H2B sequence in the lentiviral expression plasmid pLenti PGK:H2B:mCherry (provided by R. Lansford). A T2A sequence was inserted between the CREB and mCherry sequences.

*shRNA.* shRNA sequences targeting mouse CREB (5'-GGAGTCTGTGGATAGTGTA-3'<sup>31</sup>), mouse CRT2 (5'-GATGCTAAAGTCCCTGCTATT-3') or no mouse transcript (5'-CAACAAGATGAAGAGCACC-3'; scramble) were inserted into the lentiviral expression vector pLLX-shRNA<sup>32</sup>.

*HSV CREB.* The S40A HSV CREB plasmid was constructed by cloning S40A rat CREB cDNA in place of WT CREB into the bicistronic HSV amplicon p1005+:CREB (provided by S. Josselyn), in which CREB and eGFP expression are driven by IE4/5 and T7 promoters, respectively. The amplicons were packaged as previously described<sup>30</sup>.

*pA2UCOE-OGT.* Rat OGT cDNA was first cloned into the pLEMPRA-GOI vector immediately following the GFP-IRES-Flag sequence. Subsequently the entire GFP-IRES-Flag-OGT sequence was subcloned in place of the EGFP sequence in the lentiviral expression vector pA2UCOE-EGFP<sup>33</sup>.

**Cell cultures and transfection.** Neuro2a cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin (100 U ml<sup>-1</sup>). Transfections were performed using Lipofectamine LTX and PLUS reagents according to the manufacturer's protocol (Invitrogen).

Cortical neuronal cultures were prepared as previously described<sup>18</sup>, except that neurons from E15-16 timed-pregnant C57BL/6 mice were plated onto poly-DL-lysine coated plates or coverslips. For pharmacological treatment and ChIP experiments, neurons were cultured in Neurobasal medium (NBM) supplemented with 2 mM Glutamax-I, penicillin/streptomycin (100 U ml<sup>-1</sup>), and 2% B-27 (Invitrogen). For reverse transcription PCR (RT-PCR) and neurite outgrowth experiments, neurons were cultured in NBM supplemented with 10% FBS and 2 mM Glutamax-I. Neurons were electroporated with vectors and siRNA using the program K-09 on the Nucleofector Device (Lonza) according to the manufacturer's instructions.

**Quantification of *O*-GlcNAc glycosylation and Ser133 phosphorylation levels on CREB.** For those experiments using exogenously expressed CREB mutants (**Figs. 4, 8, 11**), WT or mutant pLenti CREB constructs were electroporated into neurons. Neurons were treated with the following after 4-6 DIV: KCl (55 mM, 2 h for **Figs. 4, 6, 7, 8**; 10 min for **Fig. 11**; 10 min–9 h for **Figs. 5, 10, 12, 13**), forskolin (10 μM, 2 h). Prior to KCl treatments, both treated and control neurons were silenced overnight with tetrodotoxin (TTX, 1 μM; Tocris Biosciences). Where indicated, cells were treated with the following

inhibitors for 30 min prior to the addition of KCl: nimodipine (5  $\mu$ M), KN-62 (5  $\mu$ M), U0126 (10  $\mu$ M), calphostin C (2.5  $\mu$ M), cyclosporin A (5  $\mu$ M), okadaic acid (50 nM), cycloheximide (0.3 mg ml<sup>-1</sup>) or vehicle (water, EtOH, or DMSO). All drugs except KCl and TTX were from Axxora Alexis.

Cultured neurons or dissected brain tissues were lysed and chemoenzymatically labeled with a PEG mass tag as previously described<sup>18</sup>. The lysates were subjected to 4–12% SDS-PAGE (Invitrogen) and immunoblotted. Anti-CREB (Chemicon) and anti-phospho-Ser133 CREB (Affinity BioReagents) antibodies were used to quantify the percentage of glycosylation on endogenous CREB (**Figs. 1, 5, 6, 7, 9, 10, 12, 13**). An anti-Flag (Sigma) antibody was used to quantify the percentage of glycosylation on exogenously expressed CREB mutants (**Figs. 4, 8, 11**). Relative levels of S133 phosphorylation were measured by normalizing phospho-Ser133 levels to total CREB levels. Western blots were visualized and quantified using an Odyssey Infrared Imaging System and software (Li-Cor, Version 2.1).

**Identification of *O*-GlcNAc glycosylation sites on CREB.** Neuro2a cells transfected with WT pLEMPRA CREB were treated with *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino *N*-phenylcarbamate (PUGNAc, 100  $\mu$ M, 6 h; Toronto Research Chemicals) to inhibit  $\beta$ -*N*-acetylglucosaminidase and lysed in 1.5% SDS-containing protease inhibitor cocktail (PIC; Roche) and 5  $\mu$ M PUGNAc. The lysate (7.5 mg) was diluted 1.5-fold, quenched with one volume of NETFS buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM EDTA, PIC, 5  $\mu$ M PUGNAc) containing 6% (v/v) NP-

40 and then was further diluted to 2 mg ml<sup>-1</sup> with NETFS buffer. The sample was passed over 400 µl of anti-Flag M2 affinity gel three times, washed three times with 10 ml of NETFS containing 1% (v/v) NP-40, washed twice with 10 ml of NETFS, eluted in 400 µl of 4% SDS, 100 mM Tris pH 7.9, and concentrated to a volume of 20 µl. After SDS-PAGE (4–12% Bis-Tris gels), the CREB band was excised and manually digested in-gel with chymotrypsin as previously described<sup>34</sup>. Nano LC-MS of peptides was performed as previously described using a 60-min linear gradient on an LTQ XL (Thermo Fisher)<sup>35</sup>. MS/MS spectra were collected in both CID and ETD modes using separate analyses. MS/MS were searched using Mascot 2.2 against a custom database containing Flag-CREB and 200 other proteins. ETD spectra were first converted to \*.DTA files using RawXtract (Version 1.9.1) and allowing charge states up to +5 prior to conversion to the MGF file format using Bioworks (Version 3.3.1). Searches were performed with an enzyme specificity of chymotrypsin at one terminus only, fixed modification of carbamidomethyl (C), and variable modifications of oxidation (M) and GlcNAc (S,T). The search results were evaluated by applying a Mascot Ion Cutoff score of 20 and then manually evaluating each putative GlcNAc modified peptide. Raw CID data was further evaluated for the prominent neutral loss of GlcNAc using SALSA<sup>35</sup>.

**Luciferase reporter assays.** Neuro2a cells were transfected with WT or S40A pLEMPRA CREB, pCRE-Luc (Stratagene), pRL-TK (Promega), and CREB shRNA vectors as indicated. The cells were cultured for 72 h, and luciferase activities were measured with the Dual-Glo Luciferase Assay System (Promega) on a Victor 3 plate reader (Perkin Elmer). pRL-TK containing *Renilla* luciferase was used to normalize for

transfection efficiency. Lysates from neuro2a cells transfected with CREB or scramble shRNA, and WT or S40A pLEMPRA CREB were immunoblotted for CREB to confirm endogenous CREB knockdown and equal levels of WT and S40A CREB expression.

**Electrophoretic mobility shift assay (EMSA).** Neuro2a cells were transfected with WT or S40A pLEMPRA CREB and CREB shRNA vectors or the pMaxGFP (Lonza) control vector alone. The nuclear fractions were isolated<sup>36</sup>, and the DNA-binding reaction was performed for 30 min at room temperature in the dark. The 20- $\mu$ l reaction consisted of protein (5  $\mu$ g), 10 mM Tris pH 7.5, 150 mM KCl, 10 mM DTT, 0.25 mM EDTA, Poly(dIdC) (2  $\mu$ g; Pierce), 20  $\mu$ M PUGNAc, 2% glycerol, and IRDye 700-labeled EMSA oligos containing a CRE sequence (1  $\mu$ l; Li-COR). The samples were resolved on a 10% polyacrylamide gel, and bands were quantified using an Odyssey Infrared Imaging System.

**Co-immunoprecipitation of the CREB-CRTC complex.** Neuro2a cells were transfected with WT or S40A pLEMPRA CREB and CREB shRNA vectors. After 72 h, the cells were lysed in 25 mM Tris pH 7.8, 1 mM EDTA, 150 mM NaCl, 0.5% NP-40, 5  $\mu$ M PUGNAc, and protease inhibitors. After pre-clearing the lysate with protein A sepharose beads (GE Healthcare), the lysates were immunoprecipitated with anti-Flag M2 affinity gel or an anti-CRTC2 antibody (Calbiochem). The co-immunoprecipitated complexes were resolved by 4–12% SDS-PAGE and immunoblotted for Flag-CREB and CRTC2. For the Flag IP control, nontransfected samples were immunoprecipitated with

anti-Flag M2 affinity gel. For the CRTC2 IP control, transfected samples were immunoprecipitated with a rabbit IgG antibody (Santa Cruz).

**Quantitative RT-PCR (qPCR).** For neuro2a gene expression experiments, the cells were transfected with WT or S40A pLEMPRA CREB, CREB shRNA, and scramble or CRTC2 shRNA. For neuronal gene expression experiments, cortical neurons were electroporated with WT or S40A pLEMPRA CREB and scramble (UUCUCCGAACGUGUCACGUdTdT) or CRTC1 (CGAACAAUCCGCGGAAAUdTdT) siRNA<sup>37</sup>. To measure the contribution of CREB to the expression of each gene, neurons were electroporated with scramble or CREB siRNA (UACACUAUCCACAGACUCCdTdT)<sup>31</sup>. After 3–4 days, neurons were pretreated overnight with TTX and depolarized with KCl for 6 h (where indicated), and the mRNA was extracted and purified using an RNeasy kit (Qiagen) and reverse-transcribed with SuperScript III and random primers (Invitrogen) according to the manufacturer's protocol. Quantitative PCR was performed with FastStart Universal SYBR Green Master (Rox; Roche) using an ABI 7300 real-time instrument, version 1.2. Relative quantities of mRNA were normalized to the ribosomal protein L3 (RPL3) mRNA content. PCR primers are described in **Table 1**.

**Chromatin Immunoprecipitation (ChIP).** For experiments with exogenously expressed CREB mutants (**Fig. 23**), WT or S40A pLenti CREB constructs were electroporated into neurons. ChIP was performed as previously described<sup>38</sup>, except that the neurons were fixed for 20 min, the samples were treated with proteinase K (10 mg

ml<sup>-1</sup>) for 2 h at 37 °C after elution, and the DNA was isolated by phenol/chloroform extraction followed by ethanol precipitation. Samples were immunoprecipitated using protein A sepharose beads and anti-OGT (provided by G. Hart), anti-CREB (Upstate), anti-Flag M2, and anti-rabbit IgG (Santa Cruz) antibodies. In the case of the Flag IP, a control IP of nontransfected lysate with anti-Flag antibody was performed. Purified DNA samples were subjected to PCR for 36 cycles. PCR primers are described in Table 2.

**Table 1. Primers used for RT-PCR**

Gene name	Forward primer 5' – 3'	Reverse primer 5' – 3'
<i>Arc</i>	TGGAGCAGCTTATCCAGAGG	TATTCAGGCTGGGTCTGTG
<i>ATF3</i>	CCAAGAGCCGTTGGGGCAGG	TCACTCGGGGGCAGAGTGGG
<i>BDNF IV</i>	CAGAGCAGCTGCCTTGATGTT	GCCTTGCCGTGGACGTTTA
<i>Cdk5</i>	TTTCCCTCCCTCCGTG	TGGGAAAGGAGCCAATTTATG
<i>CDKN1A</i>	GTTCCGCACAGGAGCAAAG	GAGTGCAAGACAGCGACAAG
<i>c-fos</i>	CCGACTCCTTCTCCAGCAT	TCACCGTGGGGATAAAGTTG
<i>GADPH</i>	CTGAGTATGTCGTGGAGTCTACTGG	GTCATATTTCTCGTGGTTCACACC
<i>NR4A2</i>	GCATACAGGTCCAACCCAGT	AATGCAGGAGAAGGCAGAAA
<i>OPA3</i>	GCAAAGGCAAAAGATGGAAC	GTGTTCAACCAAGGAAGGAG
<i>PEPCK</i>	GGGCCTGCAACCCTGAGCTG	GGCGATCCGCAACGCAAAGC
<i>RPL3</i>	TCATTGACACCACCTCCAAA	GCACAAAGTGGTCCTGGAAT
<i>UCP1</i>	TTGAGCTGCTCCACAGCGCC	CCGCGACTTCGGACTCCTGC
<i>Wnt2</i>	CATAGCCCCCACCCTGT	AGTTCCTTCGCTATGTGATGTTTCT

**Table 2. Primers used for ChIP PCR**

Gene name	Forward primer 5' – 3'	Reverse primer 5' – 3'
<i>Arc</i>	GGCTGGCTCTGGGAGGTATTTA	CCCCCAGAGCTGAGAGTTCAGA
<i>ATF3</i>	CCAGTTCTCCCTGGAAGCTA	CGTTGCATCACCCCTTTTAA
<i>BDNF exon IV</i>	TGGACTCCCACCCACTTT	GTGGCCGATATGTACTCC
<i>Cdk5</i>	GCTGAAGCTGTCAGGAGGTC	GTGCCCCGCTCTTGTATTA
<i>c-fos</i>	CCTCCCTCCTTTACACAG	GTCTTGGCATAACATCTTTC
<i>PEPCK</i>	GGCCTCCCAACATTCATTAAC	GTAGCCCCGCCCTCCTTGCTTTAA
<i>UCP1</i>	TCCTCTGGGCATAATCAGGAACT	CAGGTCTCAAAGAGCTGCTAGT
<i>Wnt2</i>	CCCGCACACGGAGTCTGACC	AATCCATCAGCACCGCGCCC
<i>18S Ribosomal RNA</i>	CGCGTTCTATTTTGTGGT	AGTCGGCATCGTTATGGTC

**Neurite Outgrowth.** Neurons were electroporated with WT or S40A pLEMPRA, pMaxGFP, pcDNA3-Dkk-1-Flag (provided by X. Yu), or Ncad(intra) (provided by X. Yu) vectors and scramble, CRTCl, or Wnt-2 siRNA (Santa Cruz) as indicated and then plated at a density of 25,000 neurons  $\text{cm}^{-2}$ . Neurons from B6.129-*Ogt*<sup>tm1Gwh</sup>/J mice (Jackson Laboratories) were electroporated with pMaxGFP, the CRE recombinase pBOB-CAG-iCRE-SD (Addgene), or pA2UCOE-OGT vector, along with scramble or CREB siRNA as indicated. One day prior to imaging for dendrites, neurons were depolarized with KCl (50 mM) where indicated. After 1 DIV, neurons were treated with TrkB-Fc (R&D Biosystems; 0.7  $\mu\text{g ml}^{-1}$ ) where indicated. After 4–5 DIV, all neurons were fixed with 4% paraformaldehyde (PFA) for 20 min at room temperature, washed twice with PBS, once with H<sub>2</sub>O, and mounted onto glass slides. Transfected GFP-expressing cells were imaged using a Nikon Eclipse TE2000-S inverted microscope equipped with Metamorph software. Neurite lengths were quantified with NeuronStudio (Version 0.9.92)<sup>39</sup>. Lysates from neurons electroporated with WT or S40A CREB were immunoblotted for Flag-CREB to confirm equal levels of CREB expression. Neurons from B6.129-*Ogt*<sup>tm1Gwh</sup>/J or C57BL/6 mice were immunostained with an anti-OGT (Sigma) or anti-*O*-GlcNAc antibody (CTD110.6, Covance or RL2, Pierce) to confirm the effects of OGT knockdown and overexpression on OGT and *O*-GlcNAc levels, respectively.

**Auditory fear conditioning.** *Surgery:* 9–10 week old male C57BL/6 mice (Charles River) were anesthetized with isoflurane and placed in a stereotaxis frame. Holes were

drilled in the skull above the lateral amygdala [AP = -1.3, ML = +/-3.3, V = -4.8 from bregma according to previous methods<sup>30</sup>]. Bilateral injections of WT, S40A, or control (p1005+ vector without CREB) HSV (1.5  $\mu$ l) were delivered through a Hamilton syringe over 25 min. The syringe was left in place for an additional 10 min prior to retraction. Mice were trained 3 days after the injections.

*Training:* Mice were placed into a conditioning chamber and, after 2 min, a tone (85 dB, 2000 Hz) was played for 30 s and co-terminating with a footshock (2 s, 0.75 mA). This protocol was chosen because it afforded robust yet non-ceiling levels of freezing. Immediately following or 2 h after training, mice were administered 150 mg kg<sup>-1</sup> intraperitoneal anisomycin (Sigma) or saline where indicated.

*Conditioning:* After 30 min, 2 h, or 24 h, mice were placed into a new cage. After 3 min, a tone (85 dB, 2000 Hz) was played continuously for 3 min. The mice were recorded and monitored for freezing (defined as no movement except breathing) every 5 s during the first 3 min (pre-tone freezing) and the last 3 min (post-tone freezing) by two independent observers, one of whom was blind to the experimental conditions. The percentage freezing was calculated as the mean from both observers divided by the total number of observations. No significant differences were measured in pre-tone freezing scores across all experiments. Following conditioning, mice were perfused with 4% PFA. The brains were embedded in 2% agarose, cut into 50  $\mu$ m sections with a vibratome (Leica VT1000s), and imaged for GFP to confirm the correct injection site.

**Quantification of CREB glycosylation levels after auditory fear conditioning.** Mice were placed into a conditioning chamber. After 2 min, a tone (85 dB, 2000 Hz) was

played for 30 s and co-terminated with a footshock (2 s, 0.75 m). After an additional 2 min, the above sequence was repeated. Control mice were subjected to the above sequence without the shock. Mice were placed back in their transport cage for 15 min, after which they were quickly euthanized with an overdose of isoflurane and decapitated. Brains were removed, the amygdala dissected on ice, and the samples processed for chemoenzymatic labeling as described above.

**Statistics.** *P* values were calculated from Student's paired t-test when comparing within groups and from Student's unpaired t-test when comparing between groups. ANOVA was used to analyze *in vivo* data. All calculations were performed using the program Excel.

## References

1. Hart, G.W., Housley, M.P. & Slawson, C. Cycling of *O*-linked beta-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-22 (2007).
2. Rexach, J.E., Clark, P.M. & Hsieh-Wilson, L.C. Chemical approaches to understanding *O*-GlcNAc glycosylation in the brain. *Nat. Chem. Biol.* **4**, 97-106 (2008).
3. Love, D.C. & Hanover, J.A. The hexosamine signaling pathway: deciphering the "*O*-GlcNAc code". *Sci. STKE* **2005**, re13 (2005).
4. Yang, X. et al. Phosphoinositide signalling links *O*-GlcNAc transferase to insulin resistance. *Nature* **451**, 964-9 (2008).
5. Dentin, R., Hedrick, S., Xie, J., Yates, J., 3rd & Montminy, M. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* **319**, 1402-5 (2008).
6. 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).
7. Greengard, P. The neurobiology of slow synaptic transmission. *Science* **294**, 1024-30 (2001).
8. 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. USA* **101**, 10804-9 (2004).

9. 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).
10. Lonze, B.E., Riccio, A., Cohen, S. & Ginty, D.D. Apoptosis, axonal growth defects, and degeneration of peripheral neurons in mice lacking CREB. *Neuron* **34**, 371-85 (2002).
11. Gau, D. et al. Phosphorylation of CREB Ser142 regulates light-induced phase shifts of the circadian clock. *Neuron* **34**, 245-53 (2002).
12. Carlezon, W.A., Jr. et al. Regulation of cocaine reward by CREB. *Science* **282**, 2272-5 (1998).
13. Pittenger, C. et al. Reversible inhibition of CREB/ATF transcription factors in region CA1 of the dorsal hippocampus disrupts hippocampus-dependent spatial memory. *Neuron* **34**, 447-62 (2002).
14. Kida, S. et al. CREB required for the stability of new and reactivated fear memories. *Nat. Neurosci.* **5**, 348-55 (2002).
15. Lamarre-Vincent, N. & Hsieh-Wilson, L.C. Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J. Am. Chem. Soc.* **125**, 6612-3 (2003).
16. Finkbeiner, S. et al. CREB: a major mediator of neuronal neurotrophin responses. *Neuron* **19**, 1031-47 (1997).
17. Chrivia, J.C. et al. Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855-9 (1993).
18. Rexach, J.E. et al. Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. *Nat. Chem. Biol.* **6**, 645-51 (2010).
19. Zhou, Y. et al. Requirement of TORC1 for late-phase long-term potentiation in the hippocampus. *PLoS One* **1**, e16 (2006).
20. Plath, N. et al. Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories. *Neuron* **52**, 437-44 (2006).
21. Egan, M.F. et al. The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* **112**, 257-69 (2003).
22. Wayman, G.A. et al. Activity-dependent dendritic arborization mediated by CaM-kinase I activation and enhanced CREB-dependent transcription of Wnt-2. *Neuron* **50**, 897-909 (2006).
23. Ohshima, T. et al. Targeted disruption of the cyclin-dependent kinase 5 gene results in abnormal corticogenesis, neuronal pathology and perinatal death. *Proc. Natl. Acad. Sci. USA* **93**, 11173-8 (1996).
24. Fleischmann, A. et al. Impaired long-term memory and NR2A-type NMDA receptor-dependent synaptic plasticity in mice lacking c-Fos in the CNS. *J. Neurosci.* **23**, 9116-22 (2003).
25. Aguado, F. et al. The CREB/CREM transcription factors negatively regulate early synaptogenesis and spontaneous network activity. *J. Neurosci.* **29**, 328-33 (2009).
26. 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. USA* **97**, 5735-9 (2000).
27. Tucker, K.L., Meyer, M. & Barde, Y.A. Neurotrophins are required for nerve growth during development. *Nat. Neurosci.* **4**, 29-37 (2001).

28. Viosca, J., Lopez de Armentia, M., Jancic, D. & Barco, A. Enhanced CREB-dependent gene expression increases the excitability of neurons in the basal amygdala and primes the consolidation of contextual and cued fear memory. *Learn. Mem.* **16**, 193-7 (2009).
29. Zhou, Y. et al. CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nat. Neurosci.* **12**, 1438-43 (2009).
30. Han, J.H. et al. Neuronal competition and selection during memory formation. *Science* **316**, 457-60 (2007).
31. Cox, L.J., Hengst, U., Gurskaya, N.G., Lukyanov, K.A. & Jaffrey, S.R. Intra-axonal translation and retrograde trafficking of CREB promotes neuronal survival. *Nat. Cell. Biol.* **10**, 149-59 (2008).
32. Zhou, Z. et al. Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* **52**, 255-69 (2006).
33. Zhang, F. et al. Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. *Blood* **110**, 1448-57 (2007).
34. Clark, P.M. et al. Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J. Am. Chem. Soc.* **130**, 11576-7 (2008).
35. Hansen, B.T., Jones, J.A., Mason, D.E. & Liebler, D.C. SALSA: a pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal. Chem.* **73**, 1676-83 (2001).
36. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. USA* **101**, 13132-7 (2004).
37. Li, S., Zhang, C., Takemori, H., Zhou, Y. & Xiong, Z.Q. TORC1 regulates activity-dependent CREB-target gene transcription and dendritic growth of developing cortical neurons. *J. Neurosci.* **29**, 2334-43 (2009).
38. Fan, X., Lamarre-Vincent, N., Wang, Q. & Struhl, K. Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res.* **36**, e125 (2008).
39. Wearne, S.L. et al. New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience* **136**, 661-80 (2005).

## **Chapter 7: Computational Modeling of Glycosaminoglycans**

Parts of this chapter are from Clark, P.M., Rogers, C.J., Tully, S.E., Garcia, K.C., Abrol, R., Goddard, W.A. & Hsieh-Wilson, L.C. Microarray and Computational Approaches to Understanding Glycosaminoglycan-Growth Factor Interactions. *In preparation* (2010).

**Glycosaminoglycans (GAGs) are sulfated linear polysaccharides that are important in neuronal development, viral invasion, and cancer. Recent work in our lab has shown that chondroitin sulfate (CS) tetrasaccharides, a type of GAG, are able to promote neuronal outgrowth in a manner that depends on the pattern of their sulfation. Here, we use computational approaches to better understand how the CS sulfation patterns affect their activity. We modeled the solution structure of CS-A, CS-C, CS-E, and CS-R and found that each CS tetrasaccharide favors a distinct set of torsion angles and presents a unique electrostatic surface. We further employed computational docking algorithms to determine the CS-E binding sites on a variety of proteins, including BDNF, NGF, and TNF. We found that CS-E binds to a general CS-E binding site characterized by two closely placed basic amino acids and a more distant third basic amino acid. Based on the modeled CS-E binding sites, we predict that CS-E stabilizes the interaction between certain NGF family of neurotrophins and their Trk receptors, results that were supported using carbohydrate microarrays.**

## **Introduction**

Glycosaminoglycans (GAGs) are a set of diverse sulfated carbohydrates that are important in numerous biological processes including neuronal development<sup>1,2</sup>, angiogenesis<sup>3</sup>, and viral invasion<sup>4</sup>. They have also been implicated in a number of diseases including cancer<sup>5</sup>, spinal cord injury<sup>6,7</sup>, and Alzheimer's disease<sup>8</sup>. GAGs are linear polysaccharides that are composed of repeating disaccharide units. These linear polysaccharides are usually found on the cell surface or in the extracellular matrix and are

attached to a protein core as a post-translational modification<sup>9</sup>. Different GAGs have been shown to bind to a wide variety of proteins, and it is through these interactions that the GAGs have their effects<sup>10</sup>.

There are multiple classes of GAGs, with the two most heavily studied being heparin sulfate (HS) or heparin and chondroitin sulfate (CS). HS and heparin consists of repeating units of D-glucosamine and D-glucuronic acid (GlcA) or L-iduronic acid whereas CS consists of repeating units of N-acetylglucosamine (GlcN) and GlcA. Within the linear GAG polysaccharide, any free hydroxyl group can be sulfated<sup>11</sup>, leading to a diverse combination of sulfate motifs, and the pattern of sulfation uniquely identifies each disaccharide unit<sup>9</sup>. For example, Chondroitin Sulfate A (CS-A) consists of a CS disaccharide with a sulfate group on C-4 position of the GlcN residue whereas Chondroitin Sulfate C (CS-C) consists of a CS disaccharide with a sulfate group on the C-6 position of the GlcN residue. Recent work in our lab has begun to show that the molecular level activity of these GAGs depends intimately on their sulfation pattern. Tetrasaccharides of CS-E have been shown to produce neurite outgrowth whereas tetrasaccharides of CS-C or CS-A do not<sup>12</sup>. Similarly CS-E tetrasaccharides have been shown to inhibit the interaction between Tumor Necrosis Factor  $\alpha$  (TNF-  $\alpha$ ) and Tumor Necrosis Factor Receptor 1 (TNF-R1)<sup>13</sup>.

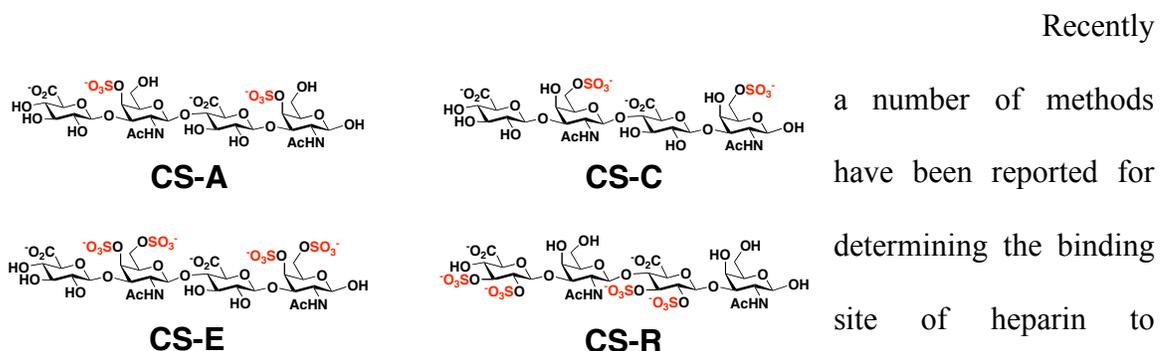
Because GAGs control their activity through intricately positioned molecular interactions, they have interested structural biologists who want to study how these sulfate groups affect CS or HS structure and their interactions with different protein targets<sup>14-17</sup>. Crystallographers performed some of the early studies on the molecular level interactions between GAGs and their binding partners. Doug Rees and his group

discovered the crystal structures of heparin bound to fibroblast growth factor 1 and 2 (FGF-1 and FGF-2)<sup>18,19</sup>. More recent studies have focused on the ternary complex between FGF-1, heparin, and its receptor<sup>20</sup>, and FGF-2, heparin, and its receptor<sup>21</sup> as well as heparin bound to other proteins including antithrombin III<sup>22</sup>, annexin V<sup>23</sup>, and annexin II<sup>24</sup>.

These crystallography studies point to some common themes in GAG–protein interactions<sup>25</sup>. Confirming experimental evidence, the structures indicate that the positions of the GAG sulfate groups are critical for specific interactions with key lysine and arginine residues on the protein surface. Yet in those cases where the GAGs bind to a protein monomer, all of the sulfate groups on the GAG are not positioned to interact with the protein, but rather many of the GAG sulfate groups are positioned out into solution where they interact with salts and water. Furthermore, unlike small molecule–protein interactions, which usually occur in a deep pocket of the protein, GAG–protein interactions occur much closer to the surface of the protein and in shallow pockets.

Computational chemists have also studied GAGs and their interactions with proteins. Computational approaches are particularly useful in this field as they are able to elucidate molecular details that would be otherwise difficult to obtain experimentally. Some of the first computational studies looked at the structure of GAGs in solution. Perez et al.<sup>26</sup> used molecular dynamics to look at the lowest energy torsion angles between the GlcA monomer and the GlcN monomer for CS-A and CS-C. Similarly Mulloy et al. employed molecular modeling, in combination with limited NMR data, to investigate the conformation of heparin in solution<sup>27</sup>. These molecular models indicated

that CS and heparin form repeating helical chains in solution, and that the degree of rotation of these chains depends on the sulfation pattern.



**Figure 1:** CS-A, CS-C, CS-E, and CS-R tetrasaccharides

Most of these methods work by identifying potential binding sites around the protein and sampling each of these sites with a heparin saccharide to assay which site affords the lowest complex energy or highest surface complementarities. However, it is not known whether similar methods could be used to model the relationship of less highly charged CS with proteins. Additionally how these methods could be used to investigate larger GAG protein ternary complexes remains unknown. New validated methods for determining the interactions between CS and individual proteins or larger protein complexes are needed.

Here, I describe work I have done to elucidate the solution structures of CS-A, CS-C, CS-E, and an unnatural CS motif, CS-R tetrasaccharides (**Fig. 1**). These tetrasaccharides have been previously synthesized in our laboratory, and they have been shown to have different biological activity from one another. Next, I describe a computational method I developed for determining chondroitin sulfate binding sites on proteins and show that it correctly predicts the binding sites of heparin on FGF-1 and FGF-2, as well as the likely binding site of CS-E on midkine. Finally, I use this method

to predict the CS-E binding sites on a number of proteins. These binding sites demonstrate a common CS-E binding motif and predict a role for CS-E in stabilizing complexes between neurotrophins and their receptors.

## **Results and Discussion**

*Solution Structures of CS-A, CS-C, CS-E, and CS-R.* To understand how the different chondroitin sulfate molecules exert their unique effects, we chose to first model the solution structure of the CS-A, CS-C, CS-E, and CS-R tetrasaccharides. By examining the CS solution structures, we reasoned that we could see how the sulfate groups could affect the structure. We build each of the tetrasaccharides (CS-A, CS-C, CS-E, and CS-R) into the Cerius2 program (Accelrys Inc.), charged them, and minimized them in vacuum within the confines of the Dreiding force field, which had to be first modified to accommodate the sulfate groups. Although each tetrasaccharide carries a formal charge of -4 or -6, we chose to afford each atom of the tetrasaccharide with partial charges but to keep the tetrasaccharide overall neutral. This was done to account for the fact that actual tetrasaccharides are in polarizable water with counter-ions that would dampen the charges of the sulfate and carboxyl groups<sup>32</sup>.

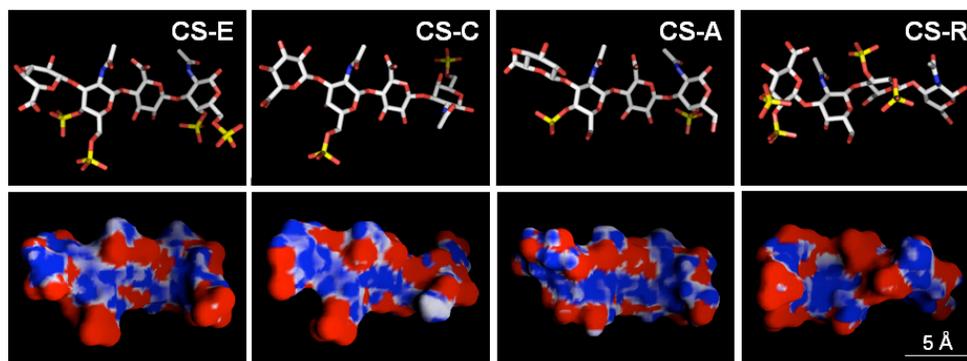
Once the tetrasaccharides had been built and minimized, each was subject to a Boltzmann jump algorithm to create 1000 different conformations. A Boltzmann jump works by taking the original conformation of a molecule and rotating a specific set of torsions to create a new conformation. If the new conformation is lower in energy than the original conformation, the Boltzmann jump algorithm accepts the new conformation. If the new conformation has a higher energy than the original conformation, then the

Boltzmann jump algorithm accepts the new conformation at a probability of  $\exp(-\Delta E/RT)$  where  $\Delta E$  is the difference in energy between the old and the new conformation. The Boltzmann jump algorithm continues to follow this method until it reaches a predefined number of conformations. In our case that limit was 1000 conformations.

Once the 1000 conformations of a given tetrasaccharide had been created, these conformations were sorted into five groups based upon the RMSD to each other. The five groups were chosen in such a way as to maximize the average RMSD between each group. This was done to ensure that over the next steps, a diverse set of structures would be explored to enhance the probability that the global minimum energy structure was indeed reached. We scored the energy of all of the structures in each of the five groups and chose the top two lowest-energy structures from each of the groups to bring on to the next step. We then took each of these ten structures, immersed them in a water box, and ran 300 ps of molecular dynamics on each of the structures to allow them to find their minimum energy structure. The structures from the last 100 ps of a given run were averaged and the energy of this average structure was calculated. We used the lowest-energy structure among these ten to represent the predicted solution structure for each tetrasaccharide (**Fig. 2**).

We found that each CS tetrasaccharide favors a distinct set of torsion angles and presents a unique electrostatic and van der Waals surface for interaction with proteins. Whereas the negatively charged sulfate and carboxylate groups on CS-C point toward either the top or bottom face of the molecule, as oriented in Figure 2a, the same charges on CS-A point in several different directions. Similarly, although CS-E and CS-R have the same number of sulfate groups, the relative orientation of these groups along the

carbohydrate backbone leads to distinctly different predicted solution structures. Whereas the CS-R tetrasaccharide has the sulfate groups distributed along several faces of the molecule, the CS-E tetrasaccharide presents all four sulfate groups along a single face, which may position the groups to interact with basic residues characteristic of glycosaminoglycan binding sites on proteins<sup>25</sup>.



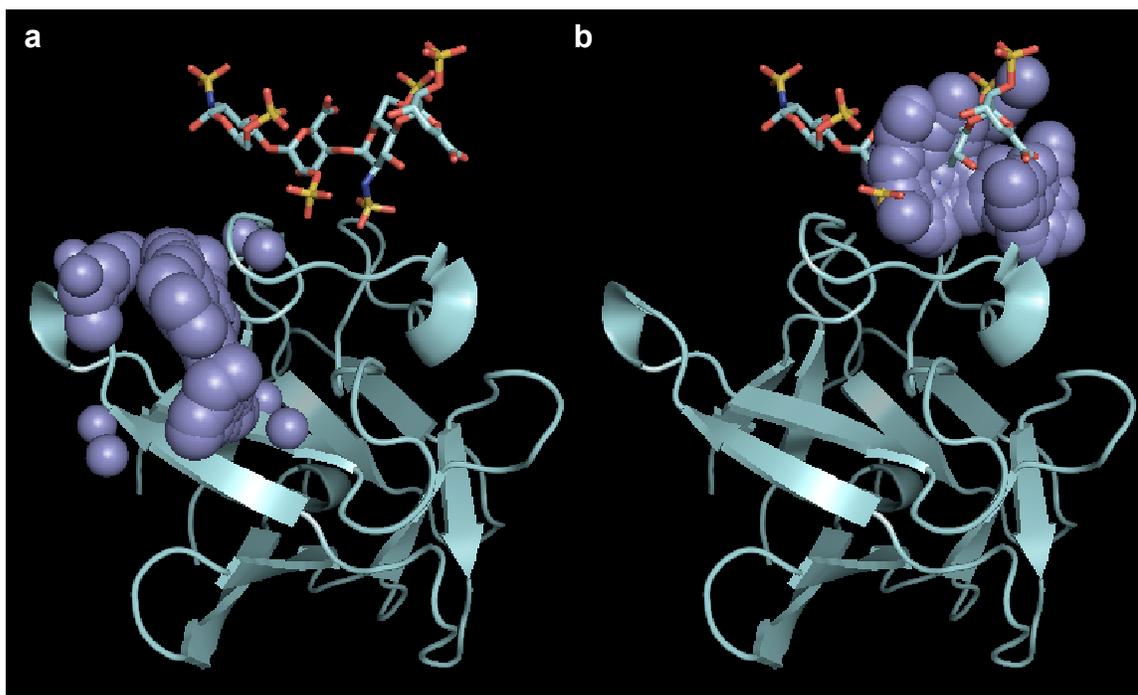
**Figure 2:** *Top:* The lowest energy structures of CS-A, CS-C, CS-E, and CS-R tetrasaccharides. *Bottom:* Electrostatic representations of these structures

*Predicting CS Binding Sites on Proteins.* CS interacts with a variety of different proteins, including VAR2CSA<sup>33</sup>, TNF- $\alpha$ <sup>13</sup>, BDNF<sup>12</sup>, and NGF<sup>34</sup>. Thus having determined the solution structure of CS molecules, we next wanted to investigate the CS binding sites on these and other proteins.

*Method Development and Confirmation.* The program ScanBindSite has previously been employed to correctly predict the binding sites of small molecules into proteins<sup>35,36</sup>. Thus we chose to use this program to investigate CS binding sites on proteins. The input for ScanBindSite is a protein and a ligand file. ScanBindSite calculates the molecular surface of the protein and then determines the surface innervations from the negative image of the molecular surface. These innervations are represented by spheres and are

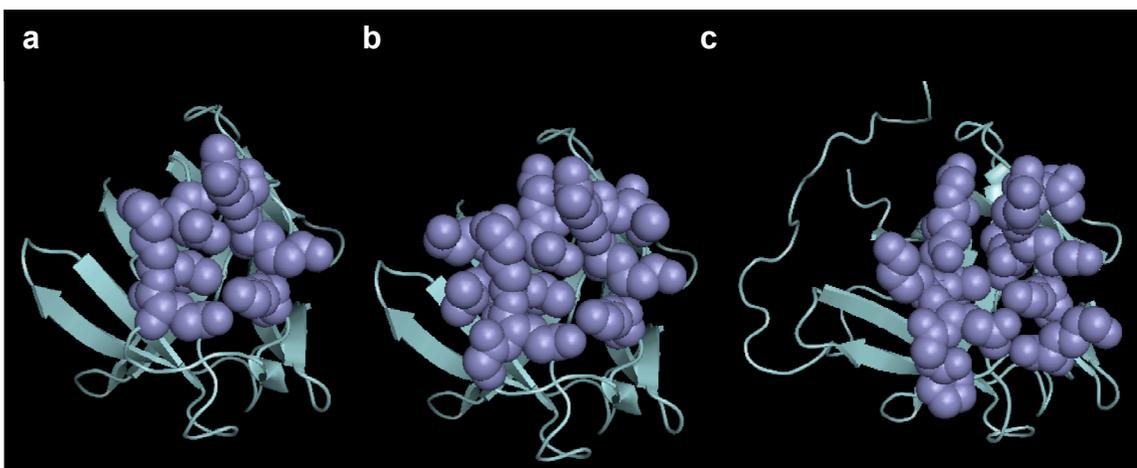
grouped into potential binding sites. ScanBindSite then maps the ligand atoms onto the spheres in each of these binding sites and calculates the energy of that docked conformation. At the end these energies are tabulated and can be used to determine which of the potential regions is the likely site for the ligand to bind.

To test whether we could use ScanBindSite to predict the binding sites of GAG tetrasaccharides, we first predicted the GAG binding sites on structures for which the GAG binding sites are known. We improved and then validated the method using two heparin–protein co-crystal structures, heparin binding to the FGF-2 monomer and heparin binding to the FGF-1 dimer<sup>37,38</sup>, and two domains of the protein VAR2CSA, DBL3X and DBL6ε, for which the CS-A binding site has been proposed by mutagenesis<sup>33</sup>. We reasoned that these would be good structures to test our approach as they represent the interaction between charged GAG polysaccharides and proteins, which is similar to the system we were interested in exploring.



**Figure 3:** **a**, Predicted heparin binding site of FGF-2 from 1BFB using default ScanBindSite parameters. **b**, Predicted heparin binding site of FGF-2 from 1BFB using modified ScanBindSite parameters

ScanBindSite requires a number of input parameters that can be optimized for a given application. To determine the correct parameters to use for predicting CS binding sites on proteins, we tested which set of parameters best predicted the binding site of the heparin tetrasaccharide onto the surface of FGF-2. Initially, we found that the default ScanBindSite parameters failed to predict the binding site of the heparin tetrasaccharide onto the surface of FGF-2. Initially, we found that the default ScanBindSite parameters failed to predict the heparin binding site on FGF-2 (**Fig. 3a**). Further analysis indicated that ScanBindSite failed to even identify the heparin binding site as a potential binding site, instead favoring more innervated regions of the protein. We also found that the potential binding sites determined by the program were much smaller than the size of a tetrasaccharide. To expand the potential binding sites determined by the ScanBindSite program, we changed the *radmax* parameter from 4.0 to

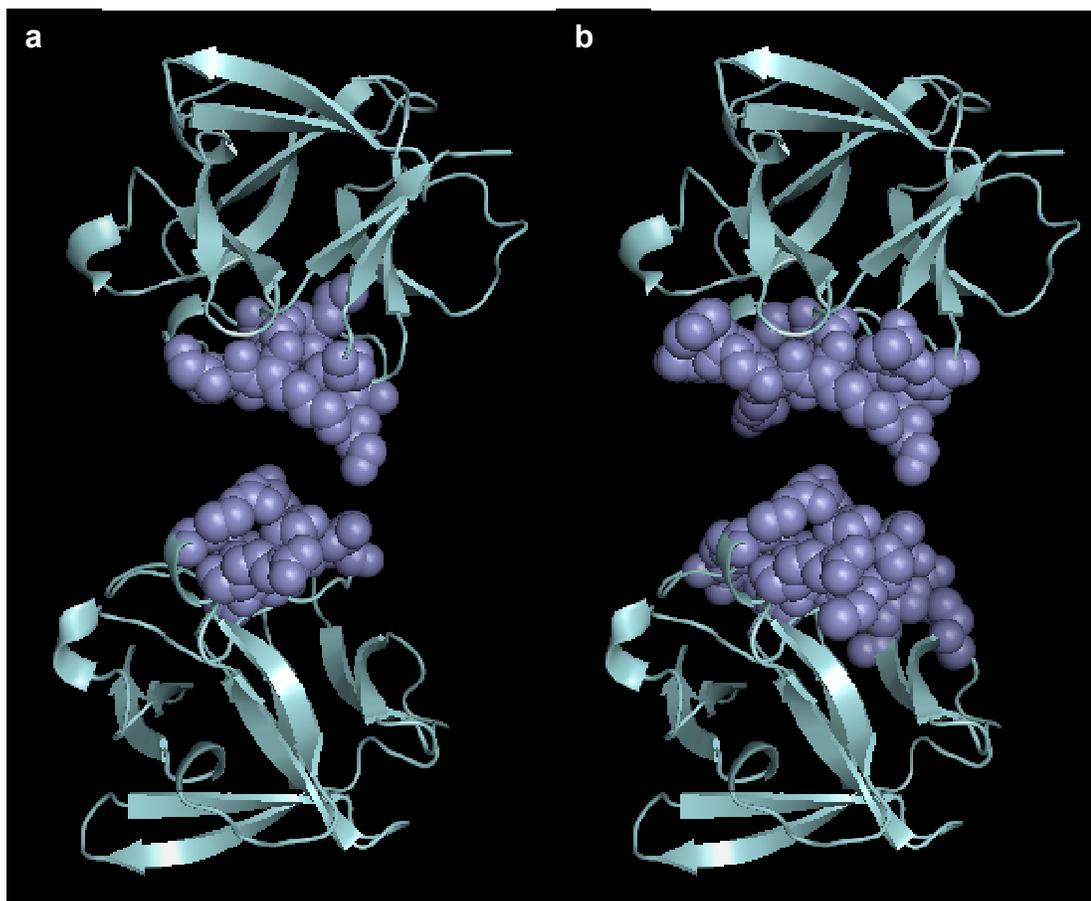


**Figure 4:** **a**, Heparin binding site on FGF-2 (from 1BFB). **b**, Predicted heparin binding site (from 1BFB crystal structure). **c**, Predicted heparin binding site (from 1BLA crystal structure).

5.0, which favors the formation of larger binding regions and the *dotlim* parameter from 0.0 to -0.5, which favors the formation of potential binding sites within the flatter surface regions of the protein that characterize GAG binding sites. Using these new parameters on the same heparin–FGF-2 systems, we correctly determined the heparin binding site of FGF-2 (**Fig. 3b**). Further optimization of these parameters by changing the *radmax* to 6.0 and the *dotlim* to -0.75 and -1.0 gave worse results.

Residues within four angstroms of heparin in the 1BFB crystal structure	Residues within four angstroms of heparin in the computationally determined binding site from 1BFB	Residues within four angstroms of heparin in the computationally determined site from 1BLA*
	Lys 27	
Asn 28	Asn 28	Asn 28
	Gly 29	Gly 29
		Leu 119
Lys 120	Lys 120	Lys 120
Arg 121	Arg 121	Arg 121
	Thr 122	Thr 122
Lys 126	Lys 126	Lys 126
		Lys 130
		Pro 133
	Gly 134	Gly 134
Gln 135	Gln 135	Gln 135
	Lys 136	Lys 136
Ala 137	Ala 137	Ala 137
	Leu 139	

**Table 1:** Residues that interact with heparin from the 1BFB crystal structure and in the predicted heparin binding site from FGF-2 in the 1BFB and 1BLA crystal structure. \* Eight was subtracted from each residue number to make the 1BLA residue numbers align with the 1BFB residue numbers.



**Figure 5: a,** Heparin binding site on FGF-1 (from 2AXM). **b,** Predicted heparin binding site

We next wanted to investigate whether the protein residues with which the GAG interacts could be extracted from the calculated binding sites. To do this, we examined the five lowest-energy heparin–FGF-2 structures within the calculated binding site and determined which residues interact with the docked heparin molecules in these structures. We found that our models predicted all of the residues that heparin interacts with in the crystal structure while predicting a limited number of extra residues (**Fig. 4a, b, Table 1**). Furthermore to determine whether the GAG binding site could be correctly predicted from the apoprotein as well as the co-crystal structure, the binding site of heparin was predicted from a crystal structure of the FGF-2 apoprotein, which differs from the heparin

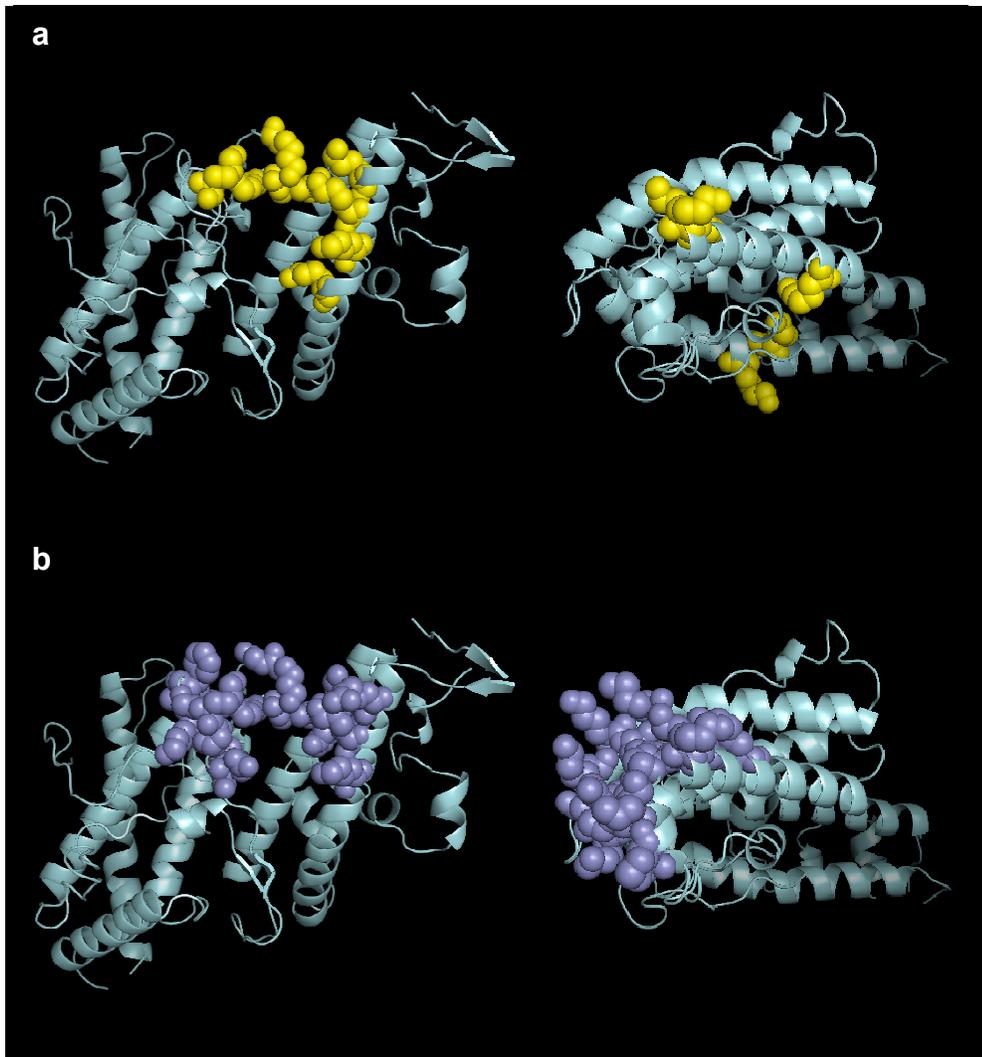
-FGF-2 co-crystal structure by an RMSD of 1.172 angstroms. We found that the calculated binding site of heparin on the FGF-2 apoprotein structure was similar to the binding site calculated from the co-crystal structure (**Fig. 4c**). Furthermore this binding site correctly predicted all of the residues that interact with heparin in the co-crystal structure while predicting few extra residues with which heparin does not interact (**Table 1**).

To test whether these modified parameters worked more generally for predicting glycosaminoglycan binding sites, we predicted the heparin binding site on FGF-1. No crystal structure of a heparin tetrasaccharide bound to FGF-1 exists, so we extracted the central heparin tetrasaccharide from the FGF-1–heparin hexasaccharide crystal structure. We then used this tetrasaccharide and the FGF-1 structure to predict the heparin binding site on FGF-1 (**Fig. 5a, b**). Again we successfully identified the heparin binding site on FGF-1 and predicted all but one of the amino acids with which the heparin tetrasaccharide interacts in the crystal structure (**Table 2**). This suggests that the modified ScanBindSite program could successfully predict GAG binding site on proteins.

Finally, we wanted to determine whether our approach could successfully predict the binding sites of CS on proteins. CS-A binds to two domains on the protein VAR2CSA, DBL3X and DBL6 $\epsilon$ <sup>33</sup>. Crystal structures of these domains are available and previous work has proposed the CS-A binding sites by mutagenesis<sup>33</sup>. We predicted the CS-A binding site on DBL3X and DBL6 $\epsilon$  using ScanBindSite (**Fig. 6, Table 3, 4**). Excitingly, the predicted CS-A binding site on DBL3X contains seven of the eight basic

<b>Residues within four angstroms of heparin in the 2AXM crystal structure</b>	<b>Residues within four angstroms of heparin in the predicted binding site from 2AXM</b>
	Ser 17, Chain A
Asn 18, Chain A	Asn 18, Chain A
	Gly 19, Chain A
Leu 111, Chain A	
Lys 112, Chain A	Lys 112, Chain A
Lys 113, Chain A	Lys 113, Chain A
	Asn 114, Chain A
Lys 118, Chain A	Lys 118, Chain A
	Arg 119, Chain A
	Arg 122, Chain A
	His 124, Chain A
	Tyr 125, Chain A
Gly 126, Chain A	Gly 126, Chain A
Gln 127, Chain A	Gln 127, Chain A
Lys 128, Chain A	Lys 128, Chain A
Ala 129, Chain A	Ala 129, Chain A
Asn 18, Chain B	Asn 18, Chain B
	Gly 19, Chain B
	His 21, Chain B
	Arg 35, Chain B
Lys 112, Chain B	Lys 112, Chain B
Lys 113, Chain B	Lys 113, Chain B
Asn 114, Chain B	Asn 114, Chain B
	Ser 116, Chain B
	Cys 117, Chain B
Lys 118, Chain B	Lys 118, Chain B
Arg 119, Chain B	Arg 119, Chain B
Arg 122, Chain B	Arg 122, Chain B
	Gln 127, Chain B
	Lys 128, Chain B
Ala 129, Chain B	Ala 129, Chain B

**Table 2:** Residues that interact with heparin from the 2AXM crystal structure and in the predicted heparin binding site from FGF-1 in the 2AXM crystal structure



**Figure 6:** **a**, Residues important for CS-A binding to DLB3X (left) and DLB6 (right), as previously determined by mutagenesis experiments (Khunrae et al. 2009). **b**, Predicted CS-A binding site on DLB3X (left) and DLB6 (right)

residues shown previously to be important for CS-A binding. Furthermore, the predicted CS-A binding site on DBL6 $\epsilon$  contains the two basic amino acids shown to be most important for CS-A binding, K2392 and K2395. Although the predicted CS-A binding site did not span all of the residues shown to be important for CS-A binding in the DBL6 $\epsilon$  mutagenesis experiments, the extra residues not predicted to be part of the binding site by our program were found by mutagenesis studies using endogenous CS,

Predicted CS-A Binding Site	Mutagenesis <sup>a</sup>	Kd
	WT	33 mM
Asp 1236		
Gly 1237		
Lys 1238		
Phe 1240		
Gly 1242		
Lys 1243	Lys1243Ala	367 mM
Gly 1244		
Glu 1246		
Thr 1317		
Gly 1318		
Thr 1319		
Lys 1324	K1324A	122 mM
Lys 1328	Lys1328Ala	89 mM
Gly 1329		
Arg 1467	Arg1467Ala	122 mM
Tyr 1468		
Arg 1503		
Lys 1504	Lys1504Ala	172 mM
Lys 1507	Lys1510Ala	193 mM
Lys 1510		
	Lys1515Ala	488 mM

<sup>a</sup> Mutagenesis values from Khunrae *et al*, 2009

**Table 3:** Predicted CS-A binding site on DBL3X and residues experimentally determined important for CS-A binding

Predicted CS-A Binding Site	Mutagenesis <sup>a</sup>	Kd
	WT	80 mM
	Lys2346Ala <sup>b</sup>	190 mM
Ile 2384		
Cys 2385		
Lys 2388		
Arg 2389		
Pro 2391	Lys2392Ala	ND <sup>c</sup>
Lys 2392	Lys2395Ala	ND <sup>c</sup>
Lys 2395		
Tyr 2399		
	Arg2408Ala	151 mM
Lys 2451		
Ile 2452		
Leu 2453		
Gly 2454		
Lys 2462		
Lys 2465		
Trp 2466		
Met 2469		
	Lys2565Ala	215 mM
	Lys2567Ala	440 mM

<sup>a</sup> Mutagenesis values from Khunrae *et al*, 2009

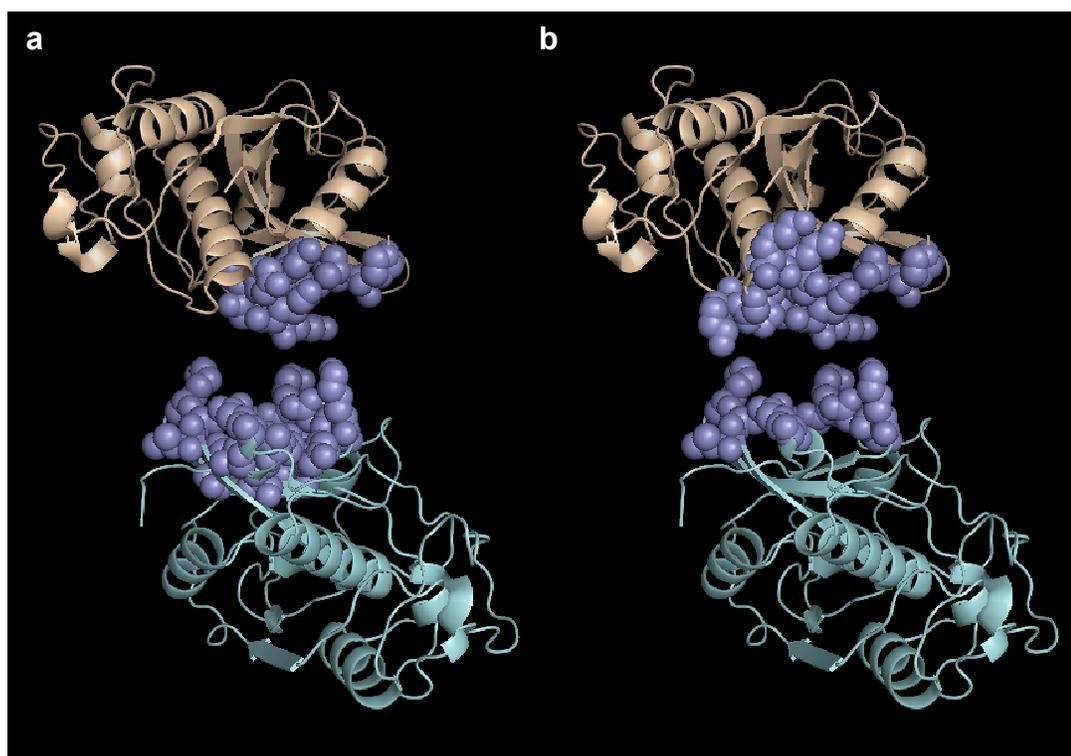
<sup>b</sup> Residue is not resolved in the crystal structure

<sup>c</sup> CS-A bound these mutants too weakly to accurately determine a Kd value

**Table 4:** Predicted CS-A binding site on DBL6 and residues experimentally determine important for CS-A binding

which is likely to be longer than a tetrasaccharide, and indeed the residues not found by our predictions are 19.4 angstroms and 26.9 angstroms from the major CS-E binding site and thus are unlikely to interact with a CS-A tetrasaccharide. Furthermore, based on other experimental evidence, Khunrae and coworkers propose that K2392 and K2395 represent the true CS-A binding site whereas the other amino acids determined by mutagenesis are likely an artifact of using only the DBL6ε domain for binding studies<sup>33</sup>.

Finally CS-A has been co-crystallized with cathepsin K<sup>39</sup>. Again using ScanBindSite, we successfully predicted the CS-A binding site on cathepsin K (**Fig. 7**) as well as many of the CS-A interacting residues (**Table 5**). This suggests that our method for predicting GAG binding sites is successful at predicting known binding sites and can be used to predict GAG binding sites on protein where the site is unknown.

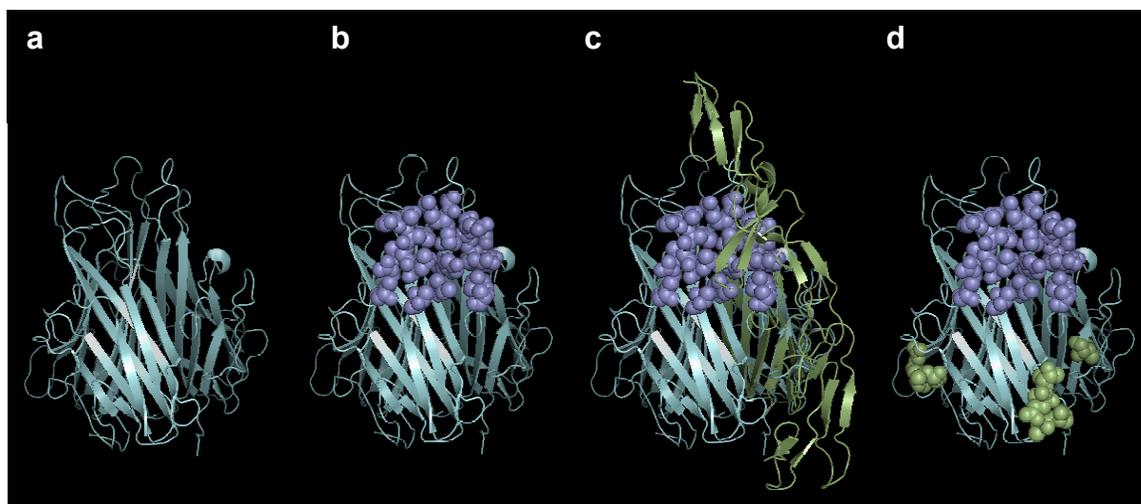


**Figure 7:** **a**, Crystal structure CS-A binding site on Cathepsin K (blue). **b**, Computationally-predicted CS-A binding site on Cathepsin K (blue)

Residues within four angstroms of CS-A in the 3C9E crystal structure	Residues within four angstroms of CS-A in the predicted binding site from 3C9E
Pro 2, Chain A	Pro 2, Chain A
Ser 4, Chain A	Ser 4, Chain A
Val 5, Chain A	Val 5, Chain A
Asp 6, Chain A	Asp 6, Chain A
Tyr 7, Chain A	
Lys 9, Chain A	Lys 9, Chain A
Lys 10, Chain A	Lys 10, Chain A
Gly 11, Chain A	Gly 11, Chain A
Tyr 12, Chain A	Tyr 12, Chain A
	Lys 39, Chain A
	Lys 40, Chain A
	Lys 41, Chain A
	Gly 43, Chain A
	Lys 44, Chain A
Asp 6, Chain B	
Arg 8, Chain B	
Lys 9, Chain B	Lys 9, Chain B
Tyr 145, Chain B	
Ser 146, Chain B	
Lys 147, Chain B	Lys 147, Chain B
Gly 148, Chain B	Gly 148, Chain B
Ile 171, Chain B	Ile 171, Chain B
Gln 172, Chain B	Gln 172, Chain B
Lys 173, Chain B	Lys 173, Chain B
His 177, Chain B	
Ile 179, Chain B	
Gly 189, Chain B	Gly 189, Chain B
Asn 190, Chain B	Asn 190, Chain B
Lys 191, Chain B	Lys 191, Chain B
Tyr 193, Chain B	
Ile 194, Chain B	
Leu 195, Chain B	

**Table 5:** Residues that interact with CS-A from the 3C9E crystal structure and in the predicted CS-A binding site from Cathepsin K in the 3C9E crystal structure

Given our success at predicting the binding sites of heparin of FGF-1 and FGF-2 and CS-A on DBL3 and DBL6 $\epsilon$ , we decided to employ our approach to predict the binding site of CS-E on a number of proteins with which it is known to interact, including TNF- $\alpha$ , BDNF, NGF, NT-3, NT-4/5, TrkA, TrkB, TrkC, midkine, GDNF receptor alpha 1, Nogo-66, and Nogo receptor (S. Tully, C. Rogers, unpublished data).

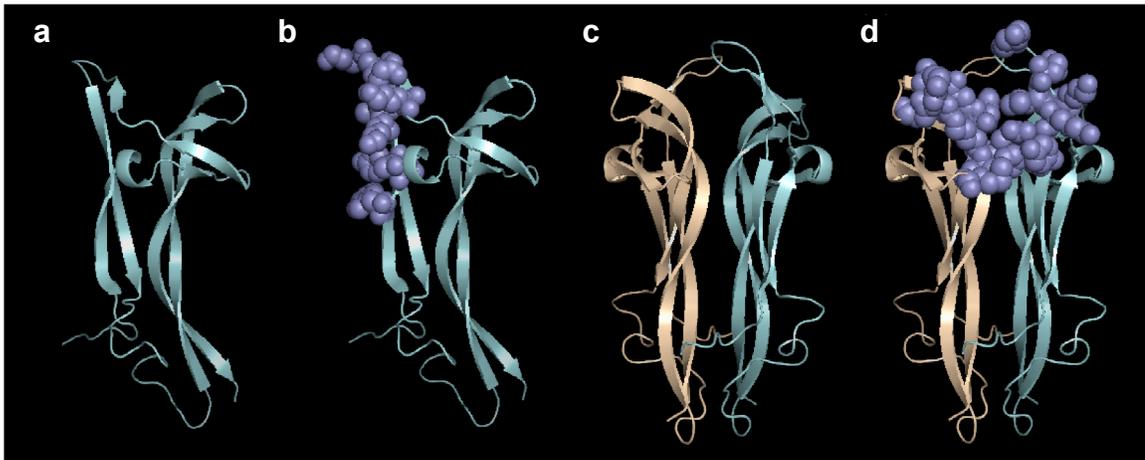


**Figure 8:** **a**, Crystal structure of TNF trimer (from 1TNF). **b**, Predicted CS-E binding site on TNF trimer. **c**, Overlay between predicted CS-E binding site (slate) and predicted TNF-R1 (green, from 1TNR) complex. **d**, Overlay between predicted CS-E binding site (slate) and residues important for TNF-R2 (green) binding

*TNF- $\alpha$* . *TNF- $\alpha$*  is a molecule important in the pathogenesis of such diseases as rheumatoid arthritis, Crohn's disease, and psoriasis<sup>40</sup>. *TNF- $\alpha$*  interacts with two receptors TNF-R1 (p55) or TNF-R2 (p75) that modulate its biological functions. We predicted the CS-E binding site from the *TNF- $\alpha$*  structure in 1TNF, which consists of amino acids 6 through 157 of human *TNF- $\alpha$*  (**Fig. 8a, b**). The CS-E binding site on the TNF trimer spans two of the three monomers and includes basic amino acids on the first monomer — Arg103 and Arg138 — as well as basic amino acids on the second monomer — Lys65 and Lys112.

We next wanted to know how the interaction between CS-E and *TNF- $\alpha$*  might affect the interaction between *TNF- $\alpha$*  and its receptors. To determine how *TNF- $\alpha$*  interacts with TNF-R1, we constructed a homology model of this complex based on the crystal structure of TNF- $\beta$  and TNF-R1<sup>41</sup>. This model, along with other mutagenesis

studies<sup>42</sup>, indicates that TNF-R1 interacts with TNF- $\alpha$  at the same interface as CS-E interacts with TNF- $\alpha$  (**Fig. 8c**). Thus one would predict from the CS-E binding site that the CS-E might block the interaction between TNF- $\alpha$  and TNF-R1. Alternatively, mutagenesis studies indicate that TNF-R2 interacts with TNF- $\alpha$  on a different part of the protein from the predicted CS-E binding site<sup>42</sup> suggesting that CS-E should not block the interaction between TNF- $\alpha$  and TNF-R2 (**Fig. 8d**). Excitingly, previous work by Tully and coworkers<sup>13,34</sup> has demonstrated that CS-E blocks the interaction between TNF- $\alpha$  and TNF-R1 but not TNF-R2, confirming the computational predictions.

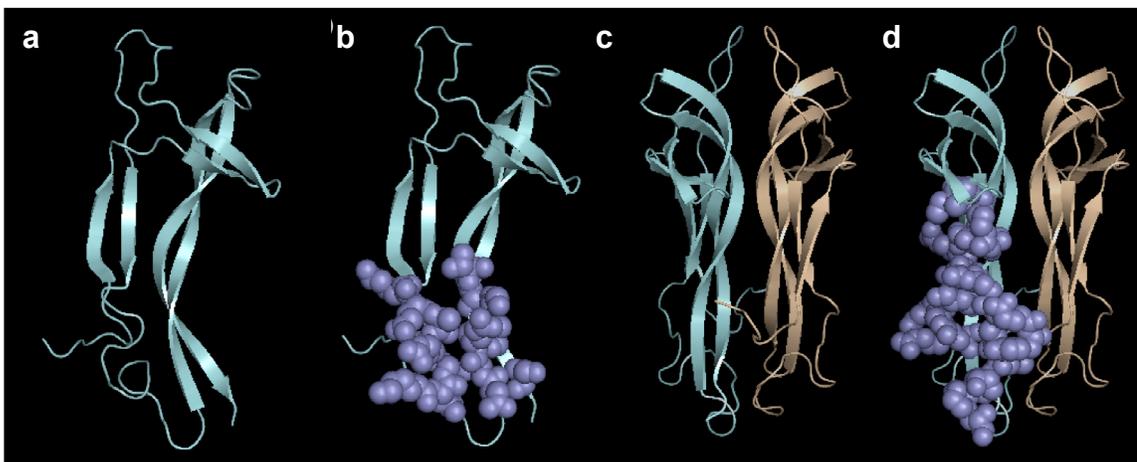


**Figure 9:** **a**, BDNF monomer crystal structure (from 1BND). **b**, CS-E binding site (slate). **c**, Homology model of BDNF dimer crystal structure (wheat and cyan). **d**, CS-E binding site (slate)

*Neurotrophins and Trk Receptors.* The NGF family of neurotrophins contains BDNF, NGF, NT-3, and NT-4/5 and shows a high degree of structural homology between these structures (average RMSD for C $\alpha$  atoms = 0.926 angstroms). Neurotrophins function generally to regulate growth, survival, and differentiation of neurons<sup>43</sup> and have been shown to signal predominantly through the common p75 neurotrophin receptor and through distinct Trk receptors, TrkA, TrkB, and TrkC. CS-E has been shown to bind to

all four members of this family<sup>44</sup>.

*BDNF*. BDNF is an important molecule for synaptic plasticity and learning and memory and functions, in part, through its interaction with the high-affinity TrkB receptor<sup>45</sup>. We predicted the CS-E binding site from the BDNF structure in 1BND, which contains amino acids 8 through 116 of human BDNF. The CS-E binding site on the BDNF monomer is predominantly across a beta-sheet and within a charged loop region that contains three basic residues — Lys41, Lys46, and Lys50 (**Fig. 9a, b**). The CS-E binding site was also predicted from a homology model of the BDNF dimer (**Fig. 9c, d, Table 6**). The CS-E binding site on the dimer structure is within a similar region to the BDNF monomer but also contains amino acids from the second dimer molecule. This includes basic residues Arg88, Arg97, and Arg101 on the second BDNF molecule.



**Figure 10:** **a**, NT-3 monomer crystal structure (from 1BND). **b**, CS-E binding site (slate). **c**, NT-3 dimer crystal structure (from 1NT3). **d**, CS-E binding site (slate)

*NT-3*. NT-3 contributes to neuronal survival, neurotransmission, and synaptic plasticity<sup>46</sup>. NT-3 interacts preferentially with TrkC although it has also been shown to signal through TrkA and TrkB in certain cellular contexts<sup>47</sup>. We predicted the CS-E

<b>BDNF</b>	<b>NGF</b>	<b>NT-4/5</b>	<b>NT-3</b>
Chain A, Lys41	Chain A, Asn46	Chain A, Asp32	Chain A, Ile28
Chain A, Lys46	Chain A, Ser47	Chain A, Leu33	Chain A, Arg56
Chain A, Gln48	Chain A, Val48	Chain A, Arg34	Chain A, Cys57
Chain A, Leu49	Chain A, Phe49	Chain A, Arg36	Chain A, Glu59
Chain A, Lys50	Chain A, Lys50	Chain A, Arg98	Chain A, Ala60
Chain A, Tyr52	Chain A, Tyr52	Chain A, Asp103	Chain A, Arg61
Chain B, Met31	Chain B, Lys32	Chain A, Gln105	Chain A, Asn76
Chain B, Arg88	Chain B, Lys34	Chain A, Arg107	Chain A, Gln78
Chain B, Asp93	Chain B, Lys88	Chain A, Val108	Chain A, Lys80
Chain B, Arg97	Chain B, Asp93	Chain A, Gly109	Chain A, Thr81
Chain B, Ile98	Chain B, Gly94	Chain A, Trp110	Chain A, Gln83
Chain B, Gly99	Chain B, Lys95	Chain A, Arg111	Chain A, Arg103
Chain B, Trp100	Chain B, Gln96	Chain A, Trp112	Chain A, Asp105
Chain B, Arg101	Chain B, Ala98	Chain B, Trp23	Chain A, Ala111
Chain B, Phe102	Chain B, Trp99	Chain B, Ala47	Chain A, Leu112
	Chain B, Arg100	Chain B, Leu52	Chain A, Ser113
	Chain B, Phe101	Chain B, Arg53	Chain A, Lys115
		Chain B, Tyr55	Chain B, Arg8
			Chain B, Glu10
			Chain B, Tyr11

**Table 6:** Predicted CS-E binding sites on BDNF, NGF, NT-4/5, and NT-3

binding site on NT-3 from NT-3 in 1BND, which contains amino acids 8 through 116 of human NT-3. The predicted CS-E binding site on the NT-3 monomer is predominantly within a loop region between the fourth and fifth beta sheet of the structure and contains four basic amino acids — Lys58, Arg61, Lys64, and Lys80 (**Fig. 10a, b**). Interestingly, although NT-3 has a high degree of structural homology to BDNF (RMSD 0.967 angstroms), the CS-E binding site on NT-3 is different from the CS-E binding site on BDNF. The preference for one binding site over the other is likely due to changes in basic residues in the loop regions of BDNF and NT-3 (**Fig. 11**). In particular, the loop that contains the CS-E binding site in BDNF is very different from the homologous loop within NT-3, with Lys41 and Lys46 in BDNF being homologous to Glu40 and Asn45 in NT-3. Similarly the CS-E binding site in NT-3 contains Lys62, which is homologous to

Gly63 in BDNF. The CS-E binding site on the NT-3 dimer is similar to the CS-E binding site on the NT-3 monomer but fails to contain Lys58 and Arg64, although it does contain other basic residues such as Arg103 and Lys115 on the first NT-3 monomer and Arg8 on the second one (**Fig. 10c, d, Table 6**).

```

BDNF      ----HSDPARRGQLSVCDSISEWVTAADKKTAVDMSGGTVTVLEKVPVSKGQ-LKQYFYE 55
NGF       --SSSHPIFHRGEFVCDSVSVVVG--DKTTATDIKGKEVMVLGEVNINNSV-FKQYFFE 55
NT-4/5    GVSETAPASRRGELAVCDAVSGWVT--DRRTAVDLRGREVEVLGEVPAAGGSPLRQYFFE 58
NT-3      ---YAEHKSHRGEVSVCDSESLWVT--DKSSAIDIRGHQVTVLGEIKTQNSP-VKQYFYE 54

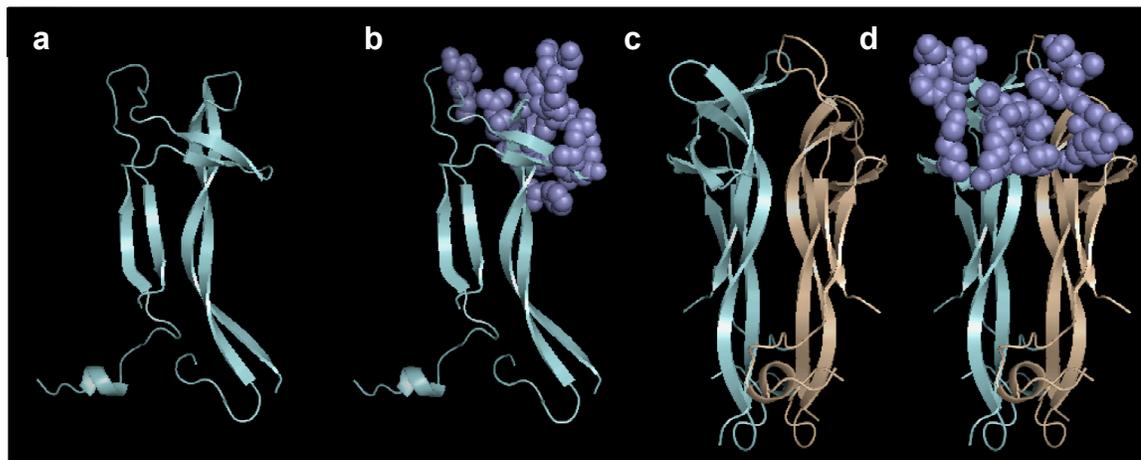
BDNF      TKCNPMGYTKEG-----CRGIDKRHWNSQCRTTQSYVVRALTMSKKRIGWRFIRIDTS 108
NGF       TKCRDPNPVDSG-----CRGIDSKHWNSYCTTHTTFVKALTMDG-KOAAWRFIRIDTA 107
NT-4/5    TRCKADNAEEGGPGAGGGGCRGVDRRHWVSECKAKQSYVVRALTADAQGRVGWWRIRIDTA 118
NT-3      TRCKEARVPVKN-----CRGIDDKHWNSQCKTSQTYVVRALTSENNKLVGWRWIRIDTS 107

BDNF      CVCTLTIKRGR-- 119
NGF       CVCVLSRKAVRRA 120
NT-4/5    CVCTLLSRTGRA- 130
NT-3      CVCALSRKIGRT- 119

```

#### Beta Sheets

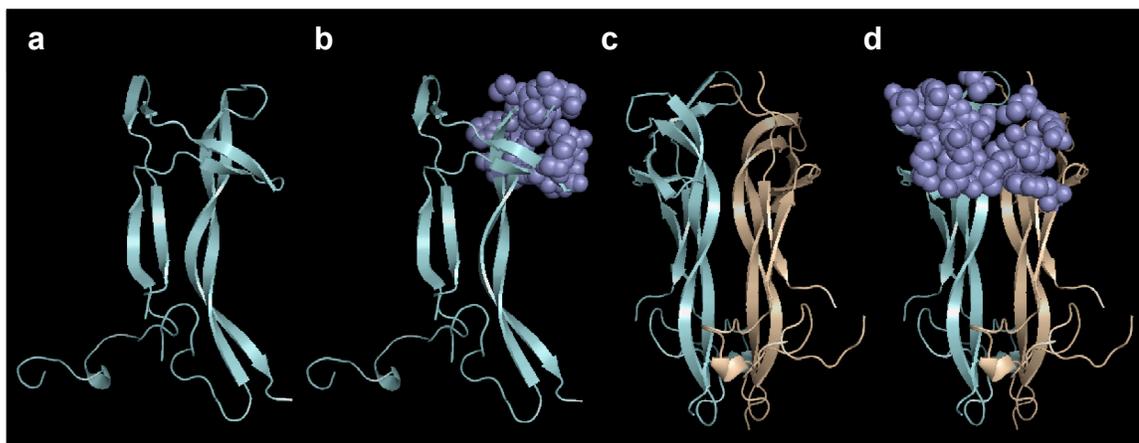
**Figure 11:** Cationic amino acids (yellow) in the respective CS-E binding sites.



**Figure 12:** **a**, NGF monomer crystal structure (from 2IFG). **b**, CS-E binding site (slate). **c**, NGF dimer crystal (from 2IFG). **d**, CS-E binding site (slate)

*NGF*. NGF is also a neurotrophin that is involved in maintenance and survival of peripheral and sensory neurons<sup>48</sup> and has been shown to signal through TrkA. We predicted the CS-E binding site on the NGF monomer from NGF in chain E of 2IFG,

which contains amino acids 2 through 115 of human NGF. CS-E binds to two hairpin loops and the adjacent beta sheets and contains five basic amino acids — Lys32, Lys34, Lys88, Lys95, and Arg100 (**Fig. 12a, b**). The CS-E binding site on NGF is distinct from the CS-E binding site both on BDNF as well as NT-3. Again this can be attributed to differences in the amino acid sequences in the loops with which CS-E interacts (**Fig. 11**). For example, the homologous amino acids to Lys41 and Lys46 in the CS-E binding site of BDNF are Glu41 and Asp46 on NGF, and the homologous amino acids to Arg61 and Lys64 in the CS-E binding site of NT-3 are Asn62 and Asp64 in NGF. Correspondingly, Lys32 and Lys34 that make up the CS-E binding site on NGF are homologous to Arg31 and His33 in NT-3 and homologous to Ser32 and Gly34 in BDNF, respectively. Although CS-E interacts with different loops of the NGF and BDNF monomers, CS-E interacts with the same loops of the NGF and BDNF dimers (**Fig. 12c, d, Table 6**). This



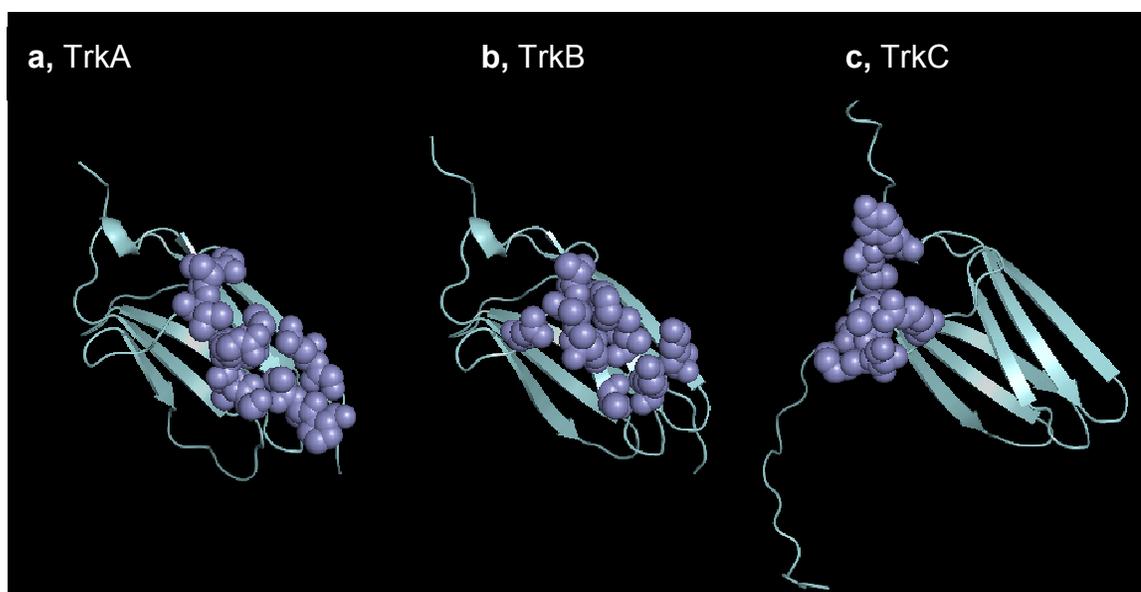
**Figure 13:** **a**, NT-4/5 monomer crystal structure (from 1HCF). **b**, CS-E binding site (slate). **c**, NT-4/5 dimer crystal (from 1HCF). **d**, CS-E binding site (slate)

is due, in part, to the fact that when the NGF and BDNF dimers form, loops that are distant from each other in the monomer structures and now close to each other in the dimer structure.

*NT-4/5*. *NT-4/5* is a neurotrophin that has been shown to promote peripheral sensory and sympathetic neuronal survival, and, like BDNF, signals through TrkB<sup>49</sup>. We predicted the CS-E binding site on *NT-4/5* from the structure of the *NT-4/5* monomer in 1HCF, which contains amino acids 1 through 127 of human *NT-4/5*. CS-E binds predominantly in two loop regions between the second and third beta-sheet and the seventh and eighth beta-sheet of *NT-4/5* (**Fig. 13a, b**). The CS-E binding site contains five basic amino acids — Arg34, Arg36, Arg98, Arg107, and Arg111 — and is very similar to the CS-E binding site on NGF. Indeed, four of the five basic residues are homologous between the CS-E binding sites on NGF and *NT-4/5* (**Fig. 11**). The CS-E binding site on the *NT-4/5* dimer is also similar to the CS-E binding site on the NGF dimer and BDNF dimer and includes further interaction with the second *NT-4/5* molecule including with Arg53 (**Fig. 13c, d, Table 6**).

The predicted neurotrophin CS-E binding sites share a number of common features. Although different in its exact binding site, CS-E predominantly binds in the loops that connect that beta sheets. In the case of *NT-3*, this corresponds to loop 3, while in the remaining neurotrophins, this corresponds to loops 1, 2, and 4. Furthermore, none of the binding sites fall within the dimerization interface or the actual or predicted Trk or p75 receptor interface, suggesting that CS-E would not block these interactions. In the dimer structures, the CS-E binding site is always across the face of the two neurotrophins with potential electrostatic interactions between CS-E and both monomers in the complex. For example, in the NGF monomer, CS-E is predicted to interact with Lys32, 34, 88, and 95, and Arg100 while in the NGF dimer, CS-E is predicted to interact with these amino

acids on the first protein in the monomer as well as Lys 50 on the second protein. Similarly in the NT-3 monomer, CS-E is predicted to interact with Arg 61 and 103 and Lys 80 and 115 while in the NT-3 dimer, CS-E is predicted to also interact with Arg 8 on the second NT-3 molecule. Since neurotrophins are suggested to exist predominantly as dimers in nature<sup>50</sup>, this data suggests that CS-E may primarily interact with the neurotrophin dimers rather than contribute to dimer formation.



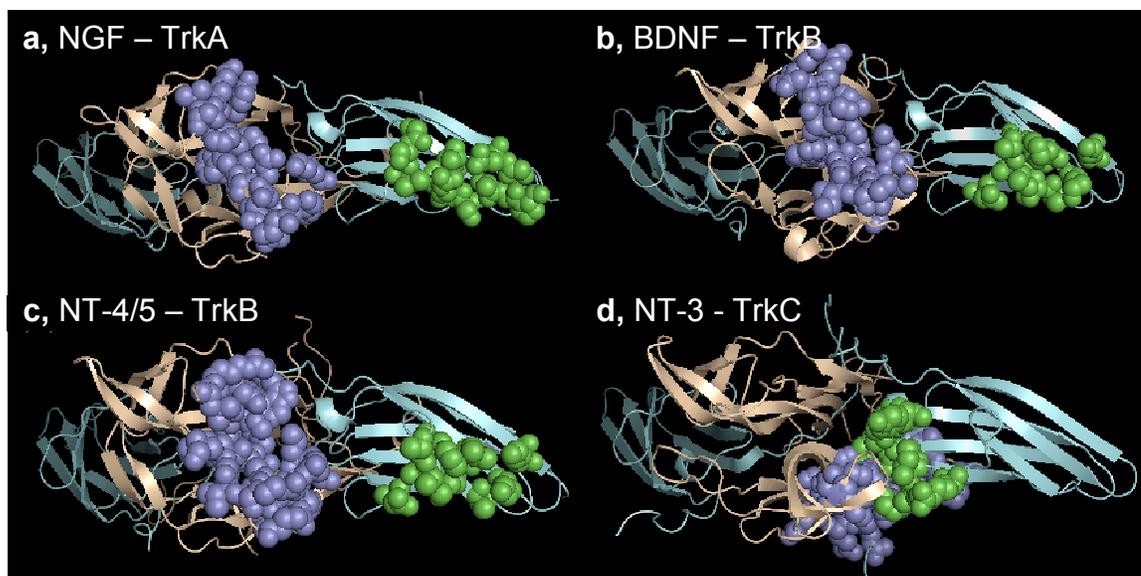
**Figure 14:** CS-E binding sites on the Trk family of receptors. **a**, TrkA **b**, TrkB **c**, TrkC

*Trk Receptors.* The Trk family of proteins, which includes TrkA, TrkB, and TrkC, are one set of receptors for the NGF family of neurotrophins and interact weakly with CS-E. We modeled the CS-E binding sites on TrkA, TrkB, and TrkC proteins from their crystal structures in 1WWW, which contain amino acids on 282 through 382 of human TrkA, 1HCF, which contain amino acids 286 through 383 of human TrkB, and 1WWC, which contain amino acids 300 through 404 of human TrkC (**Fig. 14, Table 7**). TrkA and TrkB

<b>TrkA</b>	<b>TrkB</b>	<b>TrkC</b>
Ser312	Ala314	Arg343
Leu313	Gln316	Ser345
Arg314	Phe318	Lys346
Gly319	Ala322	Ile347
Ser320	Ile323	Asn366
Val321	Asn325	Lys367
Leu362	Ile362	Pro368
Ala364	Lys364	Thr369
Asn365	Lys369	Tyr371
Pro366		
Gly368		
Gln369		

**Table 7:** Predicted CS-E binding sites on TrkA, TrkB, and TrkC

have similar CS-E binding sites, which reside mostly across the face of three beta sheets. The TrkA binding site contains only one basic residue — Arg314 — whereas the TrkB binding site contains two basic residues — Lys364 and 369. The CS-E binding site on TrkC, however, is in a different region of the protein from the homologous sites on TrkA and TrkB. The CS-E binding site on TrkC is near the top of the  $\beta$ -barrel structure and contains three basic residues – Arg343, Lys346, and Lys367.

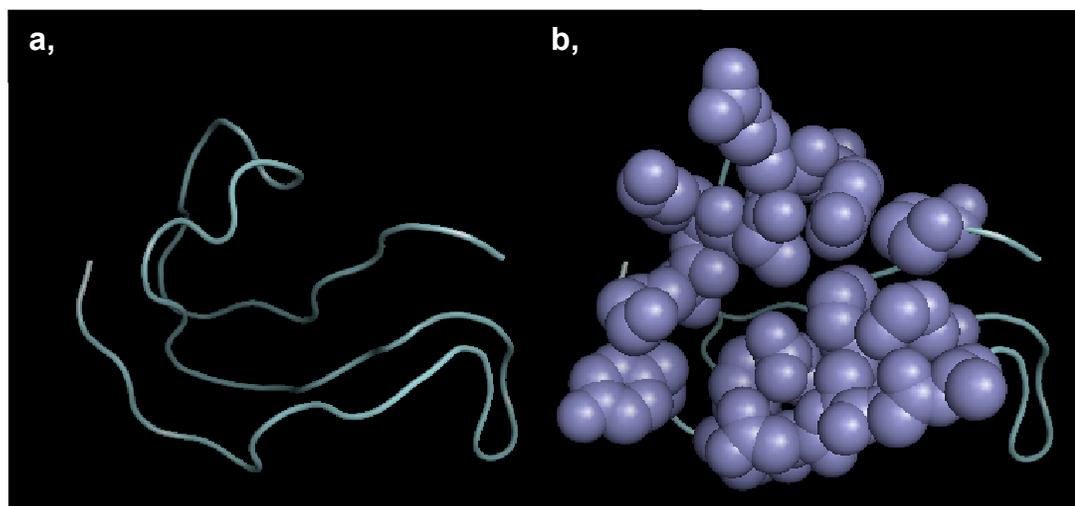


**Figure 15:** CS-E binding sites for the neurotrophins (slate) and Trk receptors (green), projected onto the neurotrophin (wheat) – receptor (cyan) complexes. **a**, NGF – TrkA, **b**, BDNF – TrkB, **c**, NT-4/5 – TrkB, **d**, NT-3 – TrkC

*Complex Formation.* Heparin polysaccharides interact with multiple proteins in a protein complex<sup>20,21</sup> and facilitate in signaling<sup>51</sup>. CS-E binds to neurotrophins such as NT-4/5 and NGF as well as weakly to their receptors, including TrkA and TrkB<sup>34,44</sup>. To investigate whether CS-E might facilitate or stabilize these neurotrophin - receptor complexes, we plotted the CS-E binding sites for the neurotrophin dimers and the Trk proteins onto predicted or actual crystal structures of the neurotrophin - Trk complexes, including NT-3 / TrkC, NT-4/5 / TrkB, BDNF / TrkB, and NGF / TrkA (**Fig. 15**). In the case of the predicted NT-3 / TrkC complex, the CS-E binding site for TrkC falls within the NT-3 / TrkC protein-protein interaction interface, thus making it unlikely that CS-E would facilitate the complex formation (although at the same time, the interaction between CS-E and TrkC would appear too weak to block the formation of the NT-3 / TrkC complex). In every case besides the NT-3 / TrkC complex, however, the CS-E binding sites on the neurotrophin dimer and the Trk protein occur on the same face of the protein complex. Thus, one long CS molecule, with CS-E motifs spaced at the correct

distance, could potentially span both the CS-E binding site on the Trk protein and the CS-E binding site on the neurotrophin dimer. Furthermore, the distance between the basic amino acids on the neurotrophin dimer binding sites and those of the Trk binding sites is such that these amino acids would be correctly positioned to interact with the sulfate groups on repeating CS units. For example, the average distance between the exposed basic amino acids within the CS-E binding site on NGF and TrkA is 25.9 angstroms, which is approximately twice the distance between the sulfate groups on the four position of CS-E or approximately the distance between the furthestmost four sulfate groups on a CS-E hexasaccharide. This suggests that CS-E might facilitate the formation or stabilization of the neurotrophin / Trk protein complex. Indeed, previous studies have reported that mutations of residues within the predicted CS-E binding site on NGF — in particular, Lys32, Lys34, and Glu35 to alanine or Lys32, and Arg95 to alanine — decreased binding of NGF to a fibroblast cell line that expresses only TrkA by 45 and 60% even though none of these residues make direct side-chain contacts with TrkA<sup>52</sup>. Nevertheless, the CS-E binding sites on the Trk molecules do not contain a high density of basic amino acids that is characteristic of traditional CS-E binding sites and thus are more likely secondary binding sites of CS-E molecules, suggesting that CS-E may not necessarily assist in bringing the neurotrophin and the Trk receptor together but rather may stabilize the preformed neurotrophin-Trk complex.

## Other Proteins



**Figure 16:** a, Midkine (from 1MKN). b, Predicted CS-E binding site (slate)

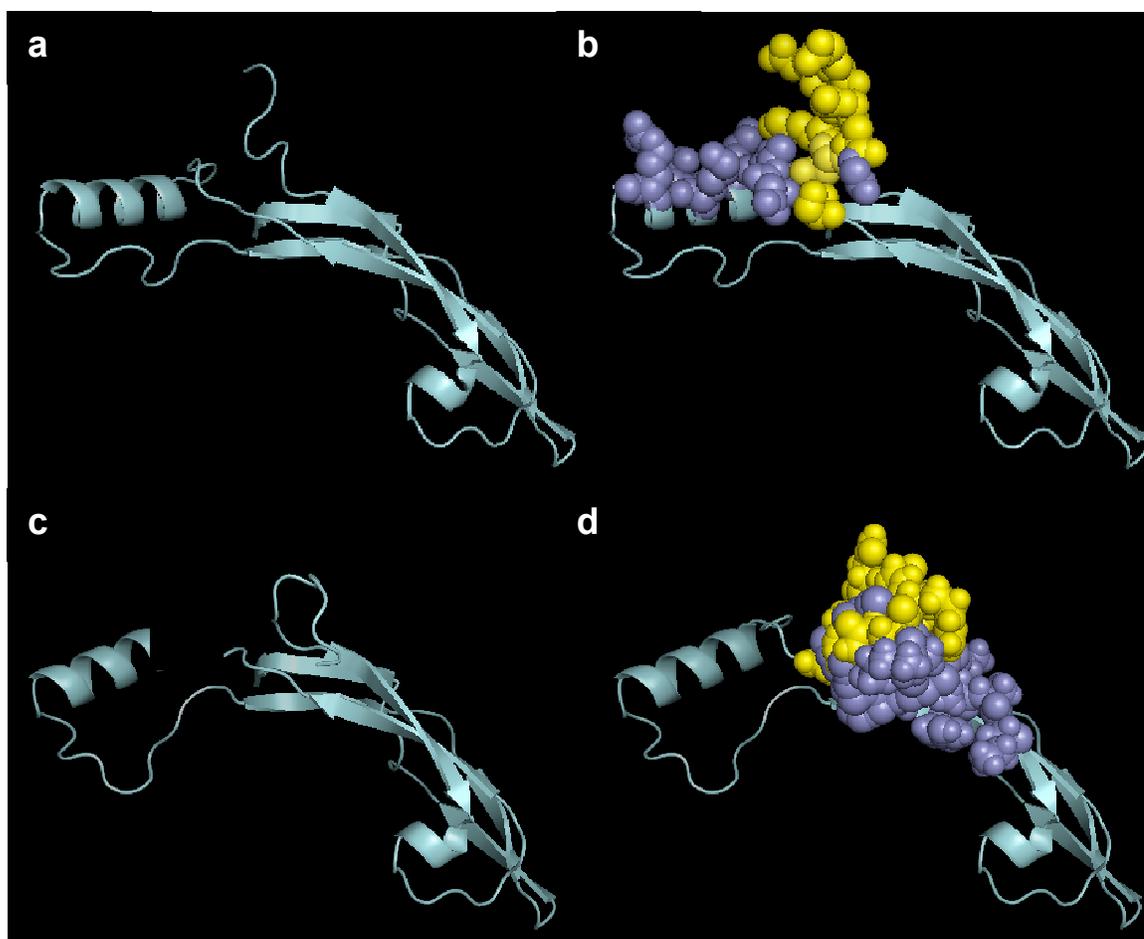
Residues whose NMR chemical shift changes with addition of Heparin 12-mer	Predicted CS-E Binding Site
	Tyr 64
	Phe 66
Glu 67	
Asn 68	Asn 68
Trp 69	Trp 69
	Gly 70
	Ala 71
Lys 79	Lys 79
Val 80	
Arg 81	Arg 81
Leu 85	
Lys 86	Lys 86
Lys 87	Lys 87
Ala 88	Ala 88
Arg 89	Arg 89
Tyr 90	Tyr 90
Asn 91	
Cys 94	
	Lys 102

**Table 8:** Residues that interact with heparin as previously determined by NMR (Iwasaki 1997) and the predicted CS-E binding site

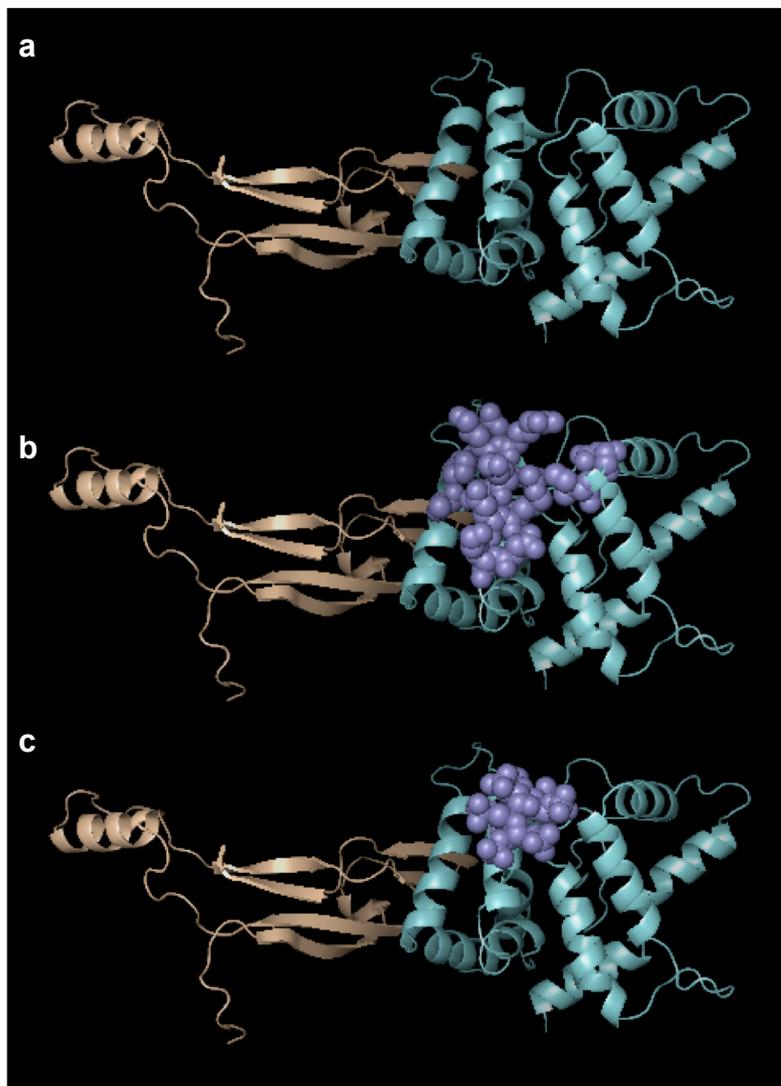
*Midkine*. Midkine is a 13 kDa protein whose expression is regulated by retinoic acid and which has been shown to enhanced neurite outgrowth and survival<sup>53</sup>. We predicted the CS-E binding site using the structure of midkine from 1MKN, which contains amino acids 23 through 81 of human midkine. The predicted CS-E binding site is within the N-terminal region of the protein and spans most of one face of the protein (**Fig. 16**). The binding site contains five basic amino acids and contains nine of the fourteen amino acids whose NMR chemical shifts<sup>53</sup> were affected upon the addition of a heparin 12mer (**Table 8**).

*GDNF*. GDNF is a neuronal survival factor that is structurally distinct from the NGF family of neurotrophins. GDNF has been implicated in neuronal differentiation, survival, and protection<sup>54</sup>. We predicted the binding site on CS-E based on the structure of GDNF from 2V5E, which contains amino acids 34 through 134 of human GDNF and chain D of 3FUB, which contains amino acids 32 through 134 of human GDNF. The CS-E binding site on GDNF is near the N-terminus of the protein and within a long stretch of basic amino acids, RRGQRGKNR, that is devoid of traditional secondary structure (**Fig. 17, Table 9**). The two GDNF structures are from different crystallization of GDNF, one as a monomer and the other as a dimer, and although they differ significantly (RMSD = 1.480 angstroms), they have similar CS-E binding sites, with the largest difference in the binding site being a consequence of differences in the placement of the alpha helix in the protein structure. This suggests that our method is robust to differences in protein structure such as might occur during different crystallization processes. Furthermore since the GDNF structure appears very mobile around the CS-E binding site, this may indicate a role for CS-E in stabilizing specific structural conformations.

*GDNF Receptor.* The GDNF-family receptor  $\alpha$  1 binds to GDNF and signals through the receptor tyrosine kinase RET<sup>55</sup>. We predicted the CS-E binding site on the GDNF receptor from the structure of the GDNF receptor from Chain C in 3FUB, which contains amino acids 150 through 384 of the rat GDNF receptor. The predicted CS-E binding site on the isolated GDNF receptor is on the surface on two alpha helices and is close to but separate from the binding interface between GDNF and the GDNF receptor (**Fig. 18a, b,**



**Figure 17:** **a**, GDNF crystal structure (from 2V5E). **b**, Predicted CS-E binding site on GDNF from (**a**) (slate and yellow). **c**, GDNF crystal structure (from Chain D 3FUB). **d**, Predicted CS-E binding site on GDNF (slate and yellow) from (**c**). Residues predicted to be in the CS-E binding site for both GDNF crystal structures are yellow; those predicted to be in the CS-E binding site for only one GDNF crystal structure are colored slate.



**Figure 18:** **a**, GDNF protein–receptor complex (wheat and cyan, from 3FUB). **b**, Predicted CS-E binding site (slate). **c**, Binding site of heparin mimic sucrose octasulfate (slate) as determined from the co-crystal structure (2V5E)

**Table 9).** The CS-E binding site consists of four charged residues — Lys169, Lys191, Lys194, and Lys202. The CS-E binding site on the GDNF protein - receptor complex is shifted slightly toward GDNF compared with the binding site on isolated GDNF receptor and overlaps with the binding site of the heparin mimic sucrose octasulfate found in the GDNF - GDNF receptor crystal structure (**Fig. 18c**).

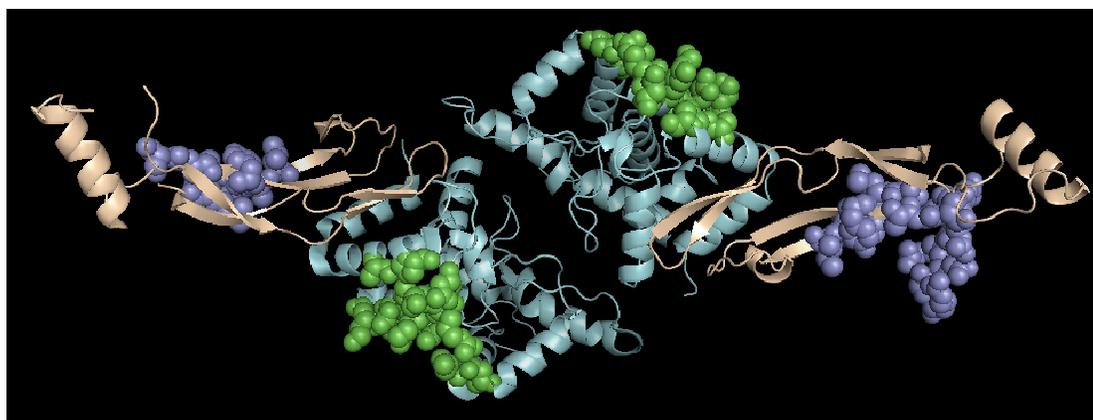
Unlike the neurotrophins and their receptors, the CS-E binding site on GDNF and

<b>GDNF Monomer (3FUB)</b>	<b>GDNF Monomer (2V5E)</b>	<b>GDNF Receptor</b>	<b>Nogo</b>
Arg32	Gln34	Lys169	Thr243
Gly33	Arg35	Tyr170	Leu246
Gln34	Gly36	Ala173	Ala247
Arg35	Lys37	Thr176	Pro248
Gly36	Asn38	Pro177	Leu249
Lys37	Arg39	Asn188	Arg250
Asn38	Gly40	Arg190	Ala251
Arg39	Ser71	Lys191	Gln253
Val42	Asp73	Lys194	Arg269
Leu43	Ala74	Ala195	Pro270
Thr44	Ala75	Gln198	Ala273
Ala45	Lys81	Lys202	Trp274
Ile46	Lys84		Lys277
Tyr67	Asn85		Phe278
Ser69	Arg88		Arg279
Gly70			
Ser71			

**Table 8:** CS-E binding site on GDNF monomer (3FUB, 2V5E crystal structure), GDNF receptor, and Nogo

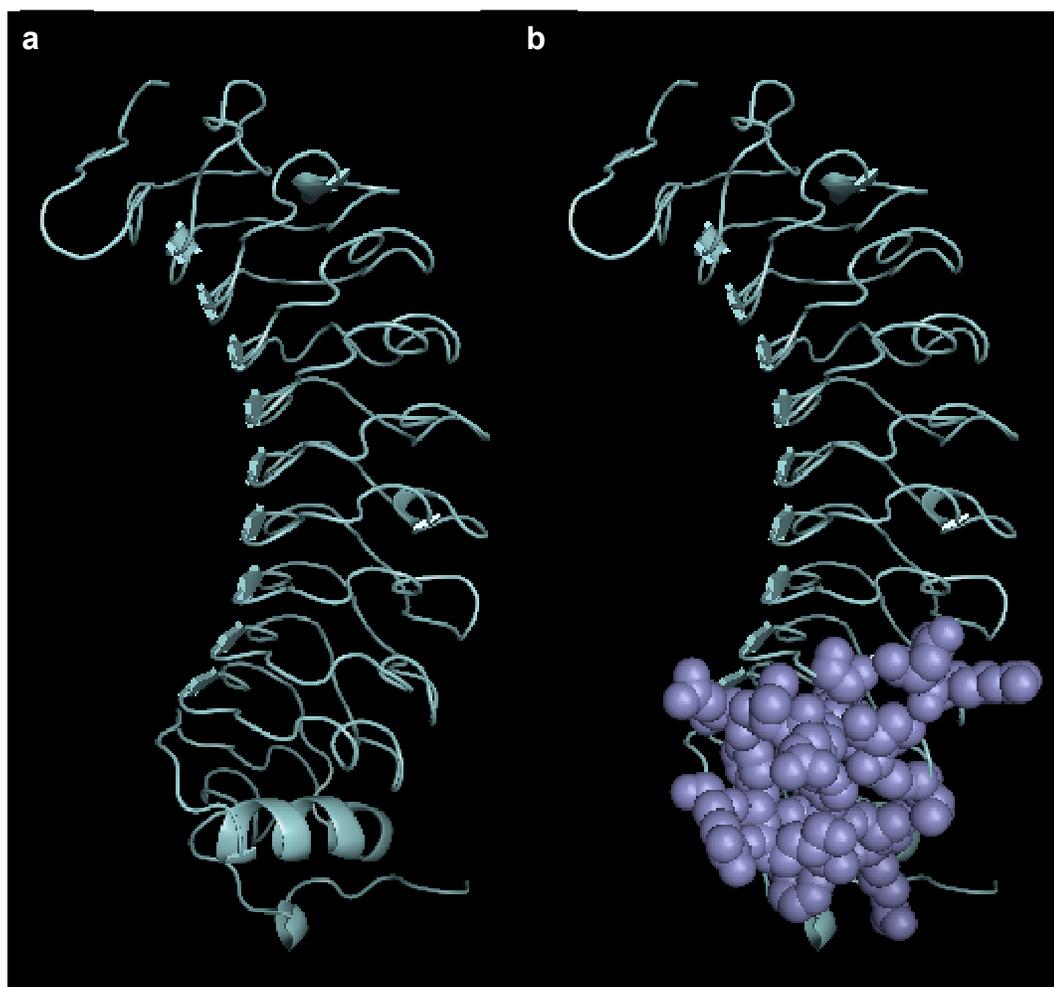
its receptor are not close to each other but rather on opposite sides of the protein, and even the binding site on the adjacent receptor appears to be in the wrong orientation to allow one CS-E molecule to span both binding sites (**Fig. 19**). Thus it does not look likely that CS-E would facilitate or necessarily stabilize a complex between GDNF and its receptor. Nevertheless since the CS-E binding site on GDNF is within an unstructured region of the protein that is structurally different in different crystallizations, it is possible that the CS-E binding site on GDNF could orient in such a way as to interact with the CS-E binding site on the GDNF receptor.

*Nogo Receptor.* The Nogo receptor interacts with Nogo and is important for axonal regeneration in the adult vertebrae central nervous system<sup>56</sup>. The Nogo receptor consists of a signal peptide followed by eight leucine-rich repeats (LRR), a leucine-rich region C-terminal domain (LRRCT), and a predicted transmembrane / glycosylphosphatidylinositol linkage<sup>57</sup>. We predicted the CS-E binding site from the Nogo receptor structure in 1P8T, which contains amino acids 27–311 of the human Nogo receptor. The CS-E binding site is at the end of the final LRR and within the LRRCT and contains four basic residues — Arg250, Arg269, Lys277, and Arg279 (**Fig. 20, Table 9**). This binding site is also separate from the predicted ligand-binding regions on the protein<sup>56</sup>, suggesting that CS-E is not likely to block the interaction between the Nogo receptor and its ligands.



**Figure 19:** CS-E binding sites (slate) on GDNF (wheat) and CS-E binding sites (yellow) on the GDNF receptor (cyan) as mapped onto the GDNF – GDNF receptor complex (3FUB)

*CS-E Binding Site Characteristics.* Although CS-E binds to a variety of different proteins with distinct structural motifs, we found that certain general features characterize many of the CS-E binding sites. CS-E binding sites are enriched in basic residues, as might be expected given the six acidic groups on the CS-E tetrasaccharide. Of the proteins modeled that strongly interact with CS-E, the median number of basic residues



**Figure 20:** **a**, Nogo receptor crystal structure (from 1P8T). **b**, CS-E binding site (slate)

in the CS-E binding sites is four. Furthermore, some of the CS-E binding sites are characterized in primary sequence by two basic amino acids that are usually within a few

residues from one another and a third basic amino acid that is more distant from the first two. One example of this type of binding site is the NT-4/5 dimer. The CS-E binding site on the NT-4/5 dimer consists of two arginine residues at amino acid positions 34 and 36 on the first protein and then a third arginine at amino acid position 53 on the second. Nevertheless, this binding site also contains additional basic residues that do not fit within this simplified tetrasaccharide binding site motif and may be important in further stabilizing the CS-E tetrasaccharide or for making extra contacts with longer CS-E chains.

Further analysis of the secondary structure of the CS-E binding sites reveals more similarities between proteins. The secondary structure of the CS-E binding sites are characterized by two basic amino acids that are approximately 5 Å from each other and a third basic amino acid approximately 15 Å away from the first two but that can be connected by a line to the first two without bisecting the protein. Thus, in the NT-4/5 dimer, the average distance between the terminal nitrogen of the guanidinium groups of Arg34 and Arg36 is 4.1 Å while the terminal nitrogen of the guanidinium group of Arg53 are an average of 16.9 Å from the terminal nitrogen of the guanidinium group of Lys34. Similarly in the BDNF dimer, the terminal nitrogen of the guanidinium groups of Arg97 and Arg101 are an average of 7.0 Å from each other and the epsilon nitrogen of Lys46 is an average 14.7 Å from the terminal nitrogen of the guanidinium group of Arg97. Interestingly, these distances correspond quite closely to the distances between the sulfate groups on the CS-E tetrasaccharide. The distance between the sulfur atoms on the four and the six position of the same sugar is on average 5.5 Å while the furthest distance between the sulfate atoms on the four position of the first sugar and the six position of the

third sugar is 15.0 Å and the average distance between the sulfate atoms on the two sugars is 12.9 Å. Indeed, these characteristic distances between the positively charged side chains more accurately characterize the CS-E binding site than does the primary sequence characterization. For example, the CS-E binding site on the TNF trimer structure contains four positively charged residues, Arg103 and Arg138 on one monomer of the trimer and Lys65 and Lys112 on a second monomer of the trimer, which does not fit well into the proposed primary sequence identification of a CS-E binding site. Yet, the distance between the epsilon nitrogen of Lys65 and Lys112 is 4.9 Å and the average distance between epsilon nitrogen on Lys112 in the first monomer and the terminal guanidinium nitrogen on Arg138 in the second monomer is 15.5 Å, again corresponding well to the distances between sulfate groups on CS-E.

Yet, surprisingly the third basic amino acid that lies approximately 15 Å away from the first two is almost exclusively found at a minimum of 10 Å away from any other basic amino acids in the CS-E binding site. This suggests that CS-E may need a strong positively charged region, consisting of two or more basic residues, to anchor it and then other, less positively charged regions, to orient it. Indeed the mutations that distinguish the CS-E binding sites on BDNF, NGF, NT-3, and NT-4/5 occur within one or both of the two closely positioned basic amino acids, whereas the third more distantly placed basic amino acids is more conserved between the proteins. Such a situation could explain the preference of these proteins for CS-E molecules over other chondroitin sulfate molecules with less concentrated charge. In particular, it would suggest that CS-E is able to most strongly bind to these proteins due to the ability of the negatively-charged sulfates on the four and six positions of the GlcN sugar to make strong salt bridges with

two correctly positioned positively-charged amino acids in the CS-E binding site for these proteins. This strong salt bridge may be required, in part, because unlike other small molecules, GAGs do not bind into a deep binding pocket and thus fail to make the full set of molecular interactions that they would otherwise make in a deep pocket. This would imply that a full CS-E tetrasaccharide may not be necessary for binding to these molecules but rather a tetrasaccharide consisting of one CS-E disaccharide unit and another singly charged CS-motif may be sufficient.

*Non-traditional binding sites.* The predicted CS-E binding sites for the Nogo receptor and the GDNF receptor do not immediately fit within this CS-E binding site rubric. The distance between any two of the positively-charged amino acids in these binding sites is greater than 10 Å, making them too far away to make meaningful contacts with the sulfate groups on the four and six position of a CS-E sugar. Nevertheless these binding sites each have amino acids that could rotate upon CS-E binding to afford a more characteristic CS-E binding site that would provide strong interactions between the CS-E sulfate groups and basic residues.

*Limitations.* Since the predicted CS-E binding sites are based on calculations, they suffer from a number of limitations. Beyond the inherent limitations in trying to model a complex system with a limited number of equations and variables, other limitations exist in this system. In particular, for most of the proteins, a greater region of the protein was used for the CS-E binding experiments than was resolved in the crystal structure. That is, CS-E could be binding to parts of the protein not found in the crystal structures and thus the predicted CS-E binding site based on the crystal structure would be incorrect. An

example of where this almost becomes a problem is in the case of GDNF. The CS-E binding studies were performed of amino acids 1 – 134 of GDNF, whereas the crystal structures of GDNF begin at amino acids 40, 32, and 34 and end at amino acids 134, 134, and 134 for chain B of 3FUB, chain D of 3FUB, and 2V5E, respectively. The predicted CS-E binding site based on the structures that start at amino acids 32 and 34 predict that CS-E interacts heavily with the amino acids 32 through 40 thus making this region a key part of the predicted CS-E binding site. Indeed, calculating the CS-E binding site based on the GDNF structure that starts at amino acid 40 predicts that the CS-E binding site is on the opposite side of the protein to that predicted from the structures that start at amino acids 32 and 34. Thus, if only the structure for GDNF that starts at amino acid 40 were available, then the predicted binding site would likely be incorrect.

## Methods

*Forcefields:* The Dreiding force field<sup>58</sup>, adapted to include sulfate groups, was used throughout the calculations. The force field was modified by optimizing the bond lengths and angles of a model  $\text{CH}_3\text{OSO}_3\text{Na}$  system through quantum mechanics (Jaguar)<sup>59</sup> and adjusting the Dreiding force field parameters based on this optimum geometry. All charges for the ligands were calculated using the charge equilibrium (QEq)<sup>60</sup> method. CHARMM22<sup>61</sup> charges were used for the protein.

*Molecular dynamics simulations:* For each tetrasaccharide, charge equilibrium (QEq)<sup>60</sup> charges were assigned and 1,000 conformations were generated using a Boltzmann jump method with rotation around the glycosidic bonds followed by structural minimization. The resulting conformations were sorted into five groups by root mean square deviation

in coordinates and ranked by their potential energies. 300 ps of explicit water molecular dynamics was run at 300 K on the two lowest-energy conformations from each of the five groups. For each of the ten molecular dynamics runs, the tetrasaccharide conformations were averaged from the last 100 ps and their potential energies were calculated with explicit solvation. The lowest-energy structure among these ten was used to represent the predicted solution structure. All Boltzmann jumps and the molecular dynamics calculations were performed using Cerius2 (Accelrys Inc.)<sup>62</sup>.

*Preparation of the Proteins:* The pdb file for each protein was downloaded from the RCSB Protein Data Bank ([www.pdb.org](http://www.pdb.org)). Water and other non-protein molecules were removed, missing residues were added using Swiss PDB Viewer, and hydrogen were added using the WhatIf program. CHARM22 charges were added, NaCl atoms were added to neutralize the protein, and the protein was minimized using the MPSim program in SGB implicit solvation.

*Binding Site Calculations:* We determined the binding sites as per previously described<sup>63</sup> with the following changes: We used a buried surface criteria of 10%, the parameter 'Grow' rather than 'Pass', the  $radmax = 5.0$ , and the  $dotlim = -0.5$ . Once the potential binding sites were identified and the top docked conformations and corresponding binding sites were ranked by energy, the predicted GAG binding site was identified by the following. The top twenty-five docked conformations and corresponding binding sites were tabulated and the sum of the inverse energy ranks for each binding site was determined. Any binding site in which this value was greater than zero was considered a GAG binding site. To determine which residues contributed to the predicted binding site,

the five lowest-energy dock conformations for GAG binding site were determined and those residues within 4 Å of any of those conformations were taken to be part of the potential GAG binding site. Similarly to determine the heparin binding site from the heparin containing crystal structures, residues within 4 angstroms of heparin were determined and were considered to contribute to the heparin binding site.

## References

1. Carulli, D., Laabs, T., Geller, H.M. & Fawcett, J.W. Chondroitin sulfate proteoglycans in neural development and regeneration. *Curr Opin Neurobiol* **15**, 116-20 (2005).
2. Bulow, H.E. & Hobert, O. Differential sulfations and epimerization define heparan sulfate specificity in nervous system development. *Neuron* **41**, 723-36 (2004).
3. Goretzki, L., Lombardo, C.R. & Stallcup, W.B. Binding of the NG2 proteoglycan to kringle domains modulates the functional properties of angiostatin and plasmin(ogen). *J Biol Chem* **275**, 28625-33 (2000).
4. Trybala, E. et al. Structural and functional features of the polycationic peptide required for inhibition of herpes simplex virus invasion of cells. *Antiviral Res* **62**, 125-34 (2004).
5. Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat Rev Cancer* **2**, 521-8 (2002).
6. Huang, W.C. et al. Chondroitinase ABC promotes axonal re-growth and behavior recovery in spinal cord injury. *Biochem Biophys Res Commun* **349**, 963-968 (2006).
7. Bradbury, E.J. et al. Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature* **416**, 636-40 (2002).
8. Pangalos, M.N., Shioi, J., Efthimiopoulos, S., Wu, A. & Robakis, N.K. Characterization of appican, the chondroitin sulfate proteoglycan form of the Alzheimer amyloid precursor protein. *Neurodegeneration* **5**, 445-51 (1996).
9. Gama, C.I. & Hsieh-Wilson, L.C. Chemical approaches to deciphering the glycosaminoglycan code. *Curr Opin Chem Biol* **9**, 609-19 (2005).
10. Capila, I. & Linhardt, R.J. Heparin-protein interactions. *Angew Chem Int Ed Engl* **41**, 391-412 (2002).
11. Kinoshita-Toyoda, A. et al. Structural determination of five novel tetrasaccharides containing 3-O-sulfated D-glucuronic acid and two rare oligosaccharides containing a beta-D-glucose branch isolated from squid cartilage chondroitin sulfate E. *Biochemistry* **43**, 11063-74 (2004).
12. Gama, C.I. et al. Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* **2**, 467-73 (2006).

13. Tully, S.E., Rawat, M. & Hsieh-Wilson, L.C. Discovery of a TNF-alpha antagonist using chondroitin sulfate microarrays. *J Am Chem Soc* **128**, 7740-1 (2006).
14. Yates, E.A. et al. Protein-GAG interactions: new surface-based techniques, spectroscopies and nanotechnology probes. *Biochem Soc Trans* **34**, 427-30 (2006).
15. Shaya, D. et al. Crystal structure of heparinase II from *Pedobacter heparinus* and its complex with a disaccharide product. *J Biol Chem* **281**, 15525-35 (2006).
16. Tan, K. et al. The structures of the thrombospondin-1 N-terminal domain and its complex with a synthetic pentameric heparin. *Structure* **14**, 33-42 (2006).
17. Mohammadi, M., Olsen, S.K. & Goetz, R. A protein canyon in the FGF-FGF receptor dimer selects from an a la carte menu of heparan sulfate motifs. *Curr Opin Struct Biol* **15**, 506-16 (2005).
18. Zhu, X. et al. Three-dimensional structures of acidic and basic fibroblast growth factors. *Science* **251**, 90-3 (1991).
19. Zhu, X., Hsu, B.T. & Rees, D.C. Structural studies of the binding of the anti-ulcer drug sucrose octasulfate to acidic fibroblast growth factor. *Structure* **1**, 27-34 (1993).
20. Pellegrini, L., Burke, D.F., von Delft, F., Mulloy, B. & Blundell, T.L. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **407**, 1029-34 (2000).
21. Schlessinger, J. et al. Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol Cell* **6**, 743-50 (2000).
22. Mourey, L. et al. Crystal structure of cleaved bovine antithrombin III at 3.2 Å resolution. *J Mol Biol* **232**, 223-41 (1993).
23. Bewley, M.C., Boustead, C.M., Walker, J.H., Waller, D.A. & Huber, R. Structure of chicken annexin V at 2.25-Å resolution. *Biochem* **32**, 3923-9 (1993).
24. Burger, A. et al. The crystal structure and ion channel activity of human annexin II, a peripheral membrane protein. *J Mol Biol* **257**, 839-47 (1996).
25. Faham, S., Linhardt, R.J. & Rees, D.C. Diversity does make a difference: fibroblast growth factor-heparin interactions. *Curr Opin Struct Biol* **8**, 578-86 (1998).
26. Rodriguez-Carvajal, M.A., Imberty, A. & Perez, S. Conformational behavior of chondroitin and chondroitin sulfate in relation to their physical properties as inferred by molecular modeling. *Biopolymers* **69**, 15-28 (2003).
27. Mulloy, B., Forster, M.J., Jones, C. & Davies, D.B. N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem J* **293** ( Pt 3), 849-58 (1993).
28. Bitomsky, W. & Wade, R.C. Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8. *J. Am. Chem. Soc.* **121**, 3004-3013 (1999).
29. Forster, M. & Mulloy, B. Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochem Soc Trans* **34**, 431-434 (2006).

30. Sachchidanand et al. Mapping the heparin-binding site on the 13-14F3 fragment of fibronectin. *J Biol Chem* **277**, 50629-35 (2002).
31. Gandhi, N.S., Coombe, D.R. & Mancera, R.L. Platelet endothelial cell adhesion molecule 1 (PECAM-1) and its interactions with glycosaminoglycans: 1. Molecular modeling studies. *Biochemistry* **47**, 4851-62 (2008).
32. Millane, R.P., Mitra, A.K. & Arnott, S. Chondroitin 4-sulfate: comparison of the structures of the potassium and sodium salts. *J Mol Biol* **169**, 903-20 (1983).
33. Khunrae, P., Philip, J.M.D., Bull, D.R. & Higgins, M.K. Structural comparison of two CSPG-binding DBL domains from the VAR2CSA protein important in malaria during pregnancy. *J Mol Biol* **393**, 202-13 (2009).
34. Tully, S.E. & Hsieh-Wilson, L.C. Unpublished data. (2010).
35. Floriano, W.B. et al. Modeling the human PTC bitter-taste receptor interactions with bitter tastants. *J Mol Model* **12**, 931-41 (2006).
36. Floriano, W.B., Vaidehi, N. & Goddard, W.A. Making sense of olfaction through predictions of the 3-D structure and function of olfactory receptors. *Chem Senses* **29**, 269-90 (2004).
37. Faham, S., Hileman, R.E., Fromm, J.R., Linhardt, R.J. & Rees, D.C. Heparin structure and interactions with basic fibroblast growth factor. *Science* **271**, 1116-20 (1996).
38. DiGabriele, A.D. et al. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **393**, 812-7 (1998).
39. Li, Z., Kienetz, M., Cherney, M.M., James, M.N.G. & Brömme, D. The crystal and molecular structures of a cathepsin K:chondroitin sulfate complex. *J Mol Biol* **383**, 78-91 (2008).
40. Tracey, D., Klareskog, L., Sasso, E.H., Salfeld, J.G. & Tak, P.P. Tumor necrosis factor antagonist mechanisms of action: a comprehensive review. *Pharmacol Ther* **117**, 244-79 (2008).
41. Banner, D.W. et al. Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation. *Cell* **73**, 431-45 (1993).
42. Mukai, Y. et al. Structure-function relationship of tumor necrosis factor (TNF) and its receptor interaction based on 3D structural analysis of a fully active TNFR1-selective TNF mutant. *J Mol Biol* **385**, 1221-9 (2009).
43. Kaplan, D.R. & Miller, F.D. Neurotrophin signal transduction in the nervous system. *Curr Opin Neurobiol* **10**, 381-91 (2000).
44. Rogers, C. & Hsieh-Wilson, L.C. Unpublished data. (2010).
45. Numakawa, T. et al. BDNF function and intracellular signaling in neurons. *Histol Histopathol* **25**, 237-58 (2010).
46. Pae, C.U., Marks, D.M., Han, C., Patkar, A.A. & Steffens, D. Does neurotrophin-3 have a therapeutic implication in major depression? *Int J Neurosci* **118**, 1515-22 (2008).
47. Patapoutian, A. & Reichardt, L.F. Trk receptors: mediators of neurotrophin action. *Curr Opin Neurobiol* **11**, 272-80 (2001).
48. Sofroniew, M.V., Howe, C.L. & Mobley, W.C. Nerve growth factor signaling, neuroprotection, and neural repair. *Annu Rev Neurosci* **24**, 1217-81 (2001).

49. Berkemeier, L.R. et al. Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB. *Neuron* **7**, 857-66 (1991).
50. Kolbeck, R., Jungbluth, S. & Barde, Y.A. Characterisation of neurotrophin dimers and monomers. *Eur J Biochem* **225**, 995-1003 (1994).
51. Irie, F., Okuno, M., Matsumoto, K., Pasquale, E.B. & Yamaguchi, Y. Heparan sulfate regulates ephrin-A3/EphA receptor signaling. *Proc Natl Acad Sci USA* **105**, 12307-12312 (2008).
52. Ibáñez, C.F. et al. Disruption of the low affinity receptor-binding site in NGF allows neuronal survival and differentiation by binding to the trk gene product. *Cell* **69**, 329-41 (1992).
53. Iwasaki, W. et al. Solution structure of midkine, a new heparin-binding growth factor. *The EMBO Journal* **16**, 6936-46 (1997).
54. Parkash, V. et al. The structure of the glial cell line-derived neurotrophic factor-coreceptor complex: insights into RET signaling and heparin binding. *J Biol Chem* **283**, 35164-72 (2008).
55. Baloh, R.H., Enomoto, H., Johnson, E.M. & Milbrandt, J. The GDNF family ligands and receptors - implications for neural development. *Curr Opin Neurobiol* **10**, 103-10 (2000).
56. He, X.L. et al. Structure of the Nogo receptor ectodomain: a recognition module implicated in myelin inhibition. *Neuron* **38**, 177-85 (2003).
57. Fournier, A.E., GrandPre, T. & Strittmatter, S.M. Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. *Nature* **409**, 341-6 (2001).
58. Mayo, S.L., Olafson, B.D. & Goddard, W.A. Dreiding - a Generic Force-Field for Molecular Simulations. *Journal of Physical Chemistry* **94**, 8897-8909 (1990).
59. Greeley, B.H. et al. New Pseudospectral Algorithms for Electronic-Structure Calculations - Length Scale Separation and Analytical 2-Electron Integral Corrections. *Journal of Chemical Physics* **101**, 4028-4041 (1994).
60. Rappe, A.K. & Goddard, W.A. Charge Equilibration for Molecular-Dynamics Simulations. *Journal of Physical Chemistry* **95**, 3358-3363 (1991).
61. MacKerell, A.D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586-3616 (1998).
62. Lim, K.T. et al. Molecular dynamics for very large systems on massively parallel computers: The MPSim program. *Journal of Computational Chemistry* **18**, 501-521 (1997).
63. Kalani, M.Y. et al. The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists. *Proc Natl Acad Sci U S A* **101**, 3815-20 (2004).

**Appendix I: *O*-GlcNAc Glycosylation Regulates CREB  
Activity in Pancreatic Beta Cells**

This work was done in collaboration with Dr. Nathan Lamarre-Vincent.

**CREB and *O*-GlcNAc glycosylation have both independently been shown to be critical for beta cell function and dysfunction. Yet little is known about the role of *O*-GlcNAc glycosylation on CREB in beta cell function. Here we show that CREB is highly glycosylated at Ser40 and that CREB glycosylation can be induced by high glucose, glucosamine, and PUGNAc in pancreatic beta cells. We furthermore show that CREB glycosylation represses induced CREB activity on a CRE luciferase construct and on IRS-2, a protein known to be critical for beta cell survival. This data suggests a mechanism wherein hyperglycemia induces CREB glycosylation, which leads to enhanced beta cell death.**

## **Introduction**

CREB is a key regulator of both beta cell function and dysfunction. CREB mediates the levels of anti-apoptotic genes *bcl-2* and *IRS-2* in response to prosurvival signals such as GIP and GLP<sup>1, 2</sup>, and CREB overexpression in cultured beta cell lines protects these cells against cytokine-induced cell death<sup>3</sup>. Alternatively, suppression of CREB-dependent *IRS-2* gene expression has been implicated in beta cell death<sup>1</sup>. Furthermore overexpression of a dominant negative form of CREB in human islets leads to increased apoptosis<sup>4</sup>, and transgenic animals with a beta-cell-specific expression of either a repressive isoform of CREB or a dominant negative form of CREB develop severe diabetes and exhibit decreased beta cell mass<sup>1, 5</sup>.

Previous studies have indicated a role for *O*-GlcNAc in diabetes. Approximately 2–5% of all cellular glucose is metabolized through the hexosamine biosynthesis pathway (HBP) to generate UDP-GlcNAc, and OGT activity is particularly sensitive to UDP-GlcNAc concentrations<sup>6</sup>. *O*-GlcNAc modifies many of the proteins important for insulin

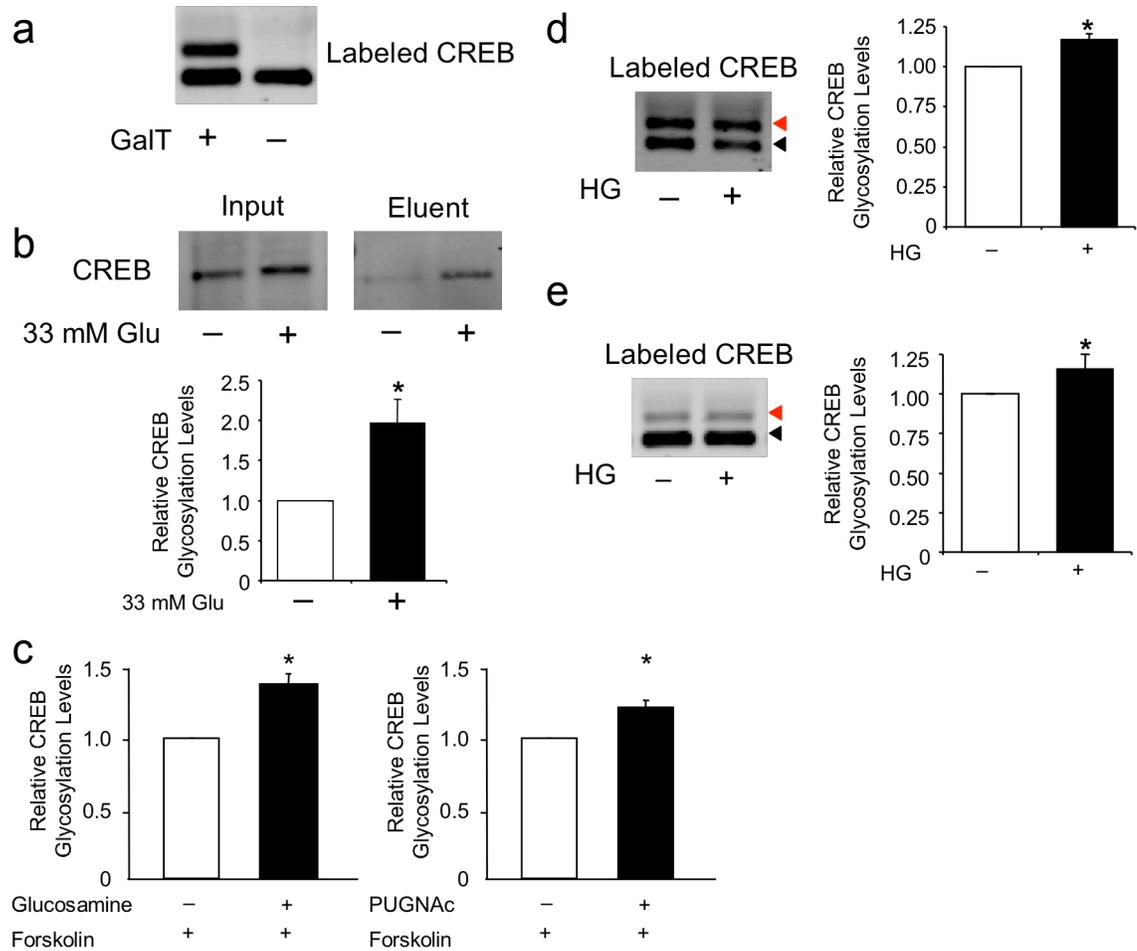
signaling, including IRS-1, IRS-2, and GLUT 4<sup>7-9</sup>, and increased *O*-GlcNAc levels have been shown to cause insulin resistance in a number of cell types including adipocytes and skeletal muscle<sup>10, 11</sup>. Studies have also shown that *in vivo* glucosamine (GlcN) injections lead to insulin resistance in normoglycemic rats but not in diabetic rats<sup>12</sup>. Furthermore overexpression of OGT *in vivo* in skeletal muscle and fat or in hepatic cells leads to insulin resistance<sup>13,14</sup> whereas overexpression of OGA in the liver of diabetic mice improves glucose homeostasis<sup>15</sup>. Similarly *O*-GlcNAc has been shown to regulate transcription factors important for glucose homeostasis, including FoxO1, NeuroD1, MafA, and PDX-1<sup>16, 17</sup>.

More recently, flux through the HBP which leads to increased *O*-GlcNAc levels has been shown to contribute to high glucose-induced beta cell death. OGT expression is highest in beta cells, and beta cell *O*-GlcNAc levels respond rapidly to changes in glucose levels *in vivo*<sup>18</sup>. OGT and *O*-GlcNAc levels are significantly induced in the islets of diabetic rats as compared to control rats<sup>19</sup>. Furthermore, inhibiting glucose flux through the HBP or inhibiting *O*-glycosylation blocks high glucose-induced apoptosis of human islets and cultured rat insulinoma cells<sup>8</sup>. Similarly, blocking the removal of *O*-GlcNAc from proteins in beta cells *in vivo* leads to beta cell death after exposure to high glucose, an affect that is abrogated by inhibiting flux through the HBP<sup>18</sup>.

Given the important roles of both CREB and *O*-GlcNAc in beta cell function, we sought to determine how *O*-GlcNAc glycosylation downstream of hyperglycemia may regulate CREB activity.

## Results and Discussion

We first sought to determine whether CREB was *O*-GlcNAc glycosylated in pancreatic cells. Pancreases from WT euglycemic rats were isolated and the proteins were chemoenzymatically labeled followed by reaction with aminoxy-PEG to specifically label the *O*-GlcNAc-modified proteins. Although the standard protocol for this technique is to immediately resolve the PEG-labeled lysate by SDS-PAGE, we found that immunoblotting for CREB on such a gel gave numerous background bands, making it difficult to identify the PEG-shifted *O*-GlcNAc-modified CREB population. Thus we immunoprecipitated CREB to better isolate it from the surrounding protein before resolving the sample by SDS-PAGE. Immunoblotting for CREB showed two clear bands, representing nonglycosylated and glycosylated CREB (**Fig. 1a**). From this gel, we could see that CREB is highly glycosylated ( $36.5 \pm 4.90\%$ ) in the pancreas. Furthermore only one PEG-shifted band could be detected for CREB, indicating that CREB is predominantly monoglycosylated in the pancreas. Nevertheless the background antibody signal, near where a second, doubly PEG-shifted CREB signal would reside, makes it difficult to rule out a very low abundant doubly-glycosylated CREB subpopulation such as seen in neurons (see Chapter 5).

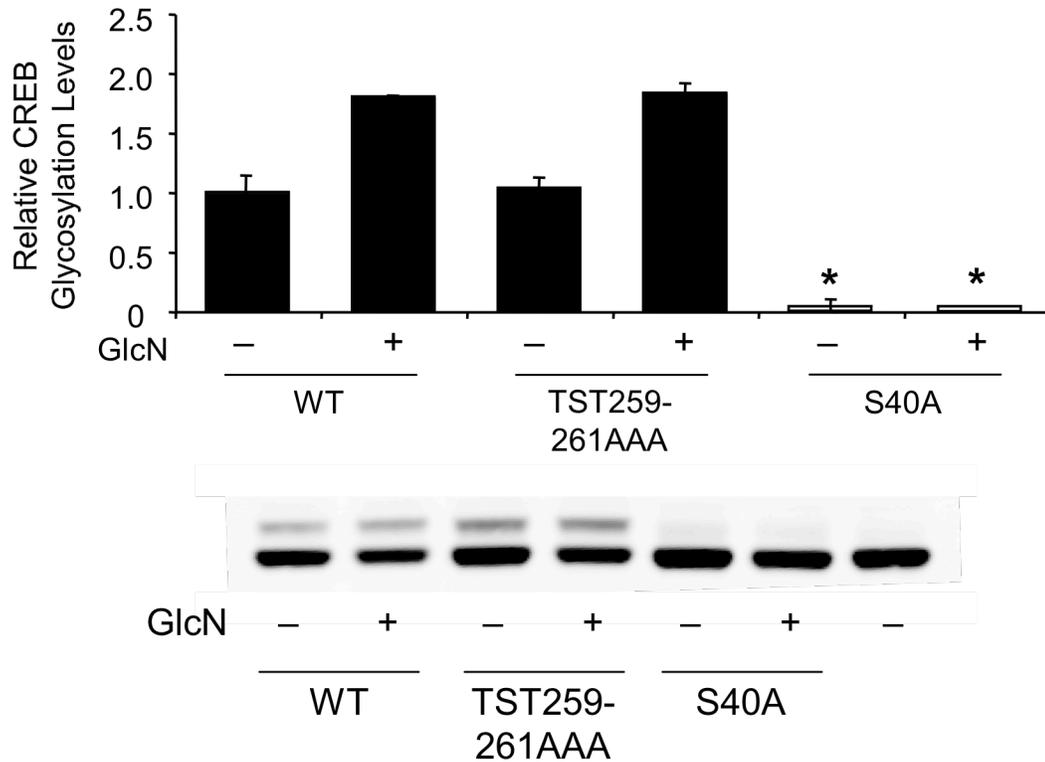


**Figure 1:** Hyperglycemia induces CREB glycosylation. (a) Rat pancreases were removed and the lysate was chemoenzymatically labeled followed by reaction with aminoxy-PEG and immunoprecipitation of CREB. Eluents were resolved by SDS-PAGE and immunoblotted for CREB. (b) MIN6 cells were cultured in the presence of 5.5 or 33 mM glucose for 96 hours. Cell lysates were chemoenzymatically labeled followed by reaction with aminoxy-biotin and capture by streptavidin. Input and eluent to the streptavidin capture was resolved by SDS-PAGE and immunoblotted for CREB.  $n = 3$ , \*  $P < 0.0002$ . (c) HIT-T15 cells were treated with GlcN, PUGNAc, or fsk as indicated. Cell lysates were treated as in (b) except that the labeled lysate was immunoprecipitated for CREB.  $n = 3$ , \*  $P < 0.01$ . (d) Euglycemic and hyperglycemic GK rat pancreases were processed as in (a).  $n = 3$ , \*  $P < 0.01$ . (e) Euglycemic and hyperglycemic STZ-treated rat pancreases were processed as in (a).  $n = 3$ , \*  $P < 0.1$ .

General *O*-GlcNAc levels fluctuate with glucose concentration in beta cells, so we wanted to determine whether CREB glycosylation was affected by changes in glucose concentration. We cultured the pancreatic cell line MIN6 in the presence of 5.5 mM or 33 mM glucose for 96 hours and then assayed the levels of *O*-GlcNAc on CREB by

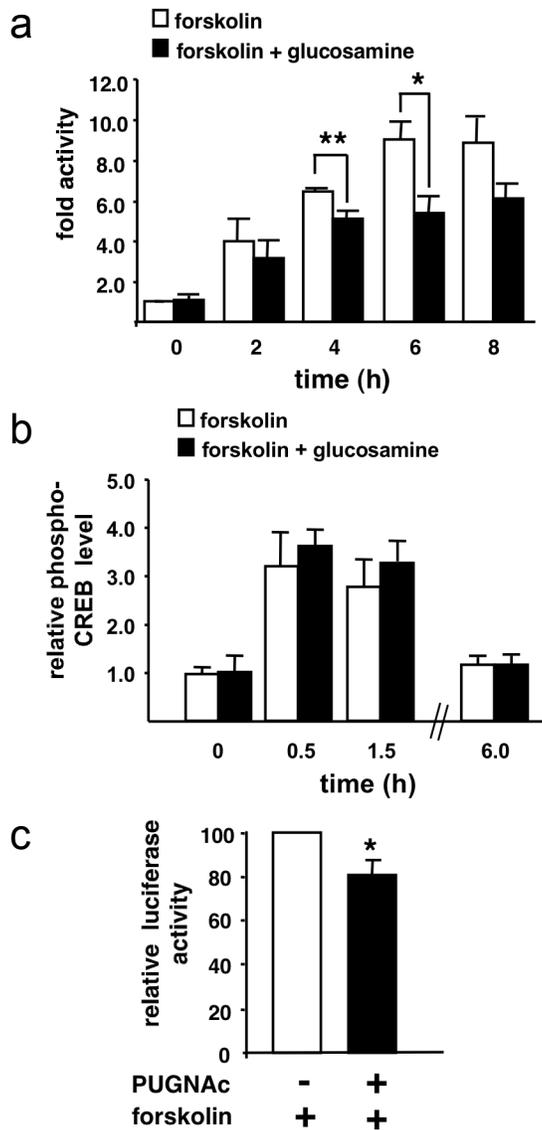
chemoenzymatic labeling followed by biotinylation and streptavidin pulldown (**Fig. 1b**). We found that high glucose induced CREB levels and further induced CREB glycosylation levels. Correcting for the increase in CREB levels, high glucose induced *O*-GlcNAc glycosylation on CREB by  $1.97 \pm 0.30$ -fold compared to low glucose. Previously we have shown that neuronal depolarization induces CREB glycosylation downstream of  $\text{Ca}^{2+}$  influx (see Chapter 5), and glucose is known to depolarize beta cells leading to  $\text{Ca}^{2+}$  influx<sup>20</sup>, so it may be that similar pathways are activated in both cell types. Furthermore, GlcN as well as PUGNAc induced CREB glycosylation in the pancreatic cell line HIT-T15 (**Fig. 1c**). Taken together, this data indicates that *in vitro*, glucose, GlcN, and PUGNAc induce CREB glycosylation levels.

To determine whether hyperglycemia induces CREB glycosylation *in vivo*, we assayed CREB glycosylation levels in a number of different animal models of diabetes, including streptozocin-induced hyperglycemia as well as Goto-Kakizaki (GK) and Zucker Diabetic Fatty (ZDF) rat genetic models of diabetes. In each of these models, the blood glucose level of the experimental, diabetic rats were all over 300 mg / dL compared with control rats, which had blood glucose levels around 100 mg / dL. Again here we PEG labeled the *O*-GlcNAc-modified population of the protein before immunoprecipitating and Western blotting for CREB. We found in the case of the GK and STZ rat models that hyperglycemia induced CREB glycosylation levels  $16.9 \pm 0.01\%$  and  $15.4 \pm 0.09\%$  respectively (**Fig. 1d, e**). Hyperglycemia did not induce CREB glycosylation in the ZDF rats ( $1.03 \pm 0.12\%$ ). These studies suggest that at least in certain animal models, hyperglycemia is linked to increased CREB glycosylation.



**Figure 2:** CREB is dynamically glycosylated at Ser40 in HIT-T15 cells. WT, TST259-261AAA, and S40A FLAG-CREB were transfected into HIT-T15 cells and the cells were treated with glucosamine (GlcN) as indicated. Lysates were chemoenzymatically labeled, reacted with aminoxy PEG, and immunoblotted for FLAG.  $n = 3$ , \*  $p < 0.01$  compared to WT CREB for each condition.

We have previously identified Ser40 as the major dynamic site of glycosylation in Neuro2A cells and embryonic neurons (see Chapter 5), so we next investigated whether Ser40 glycosylation was also the major dynamic site in beta cells. We transfected HIT-T15 cells with WT, TST259-261AAA, and S40A FLAG-CREB and assayed their glycosylation levels before and after GlcN treatment. The S40A mutant blocked the vast majority ( $95 \pm 7\%$ ) of FLAG-CREB glycosylation while the TST259-261AAA FLAG-CREB mutant had no effect on FLAG-CREB glycosylation levels in HIT-T15 cells (**Fig. 2**). Furthermore, GlcN induced CREB glycosylation on both WT and TST259-261AAA FLAG-CREB ( $81\% \pm 14\%$ ,  $83\% \pm 9\%$ , respectively) but had no effect on CREB



**Figure 3:** *O*-GlcNAc glycosylation represses CREB transcriptional activity. (a) HIT-T15 cells were transfected with CRE-luciferase and incubated with or without GlcN. Fsk was added to the medium for the times indicated, the harvested cells were lysed, and luciferase activity was quantified.  $n = 3-7$ , \*  $P < 0.05$ . (b) HIT-T15 cells were incubated with GlcN prior to the addition of Fsk. Cells were harvested and the lysate was analyzed by immunoblotting with an anti-phospho-S133 CREB antibody. (c) Cells were prepared as in (a) except that they were treated with PUGNAc rather than GlcN.  $n = 3$ , \*  $P < 0.0005$ .

glycosylation of S40A FLAG-CREB ( $-0.008\% \pm 0.05\%$ ), indicating that Ser40 is the only CREB glycosylation site dynamic to GlcN.

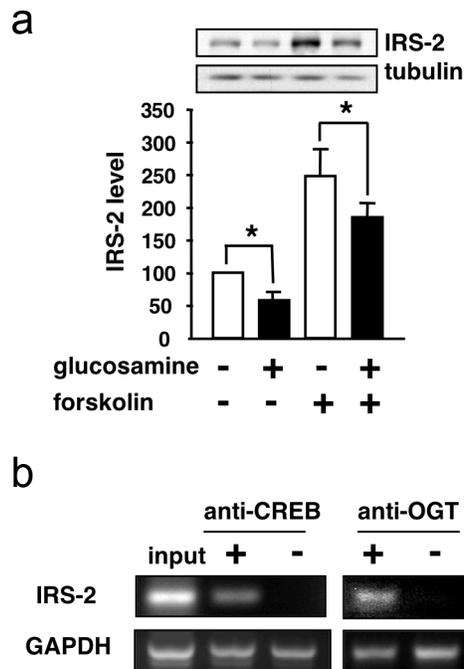
Having shown that CREB glycosylation could be dramatically modulated in response to glucose, GlcN, and PUGNAc, we next examined the effect of glycosylation on CREB activity. We transfected HIT-T15 cells with a CRE-luciferase reporter and then tested CREB activity in the presence or absence of GlcN, to enhance CREB glycosylation, or forskolin (fsk), to enhance CREB phosphorylation. Interestingly, we found that GlcN had no effect on CREB activity in non-induced cells, which is different than in

neurons, where removal of glycosylation increased non-induced CREB activity (see Chapter 5). Non-induced nuclear TORC levels differ depending on the cell type<sup>21</sup>, and we have previously shown that CREB glycosylation blocks the interaction with the TORC coactivator (see Chapter 5), so one explanation for this difference could be that the amount of nuclear TORC in HIT-T15 cells is lower than in neurons. Indeed although the absolute amount of TORC in the nucleus was not measured, previous studies have shown that only approximately 15% of non-induced HIT-T15 cells have TORC2 localized to the nucleus<sup>22</sup> whereas approximately 40% of TORC1 is localized to the nucleus in non-induced neurons<sup>23</sup>. Although glycosylation had little effect on non-induced CREB levels, glycosylation depressed induced CREB activity in a time-dependent manner, reducing CREB activity  $39 \pm 10\%$  after 6 hours (**Fig. 3a**). Similar effects were seen following 9 hour PUGNAC treatment (**Fig. 3b**). As seen in neurons, changes in CREB glycosylation had no effect on phospho-Ser133 levels (**Fig. 3c**). This data suggests that induced CREB glycosylation blocks induced CREB activity independent on changes in phospho-Ser133 levels.

CREB has been shown to be critical for regulating *IRS-2* levels that support beta cell survival<sup>1</sup>, thus we wanted to know whether CREB glycosylation could effect *IRS-2* mRNA levels in beta cells. Beta cells were treated with GlcN for 9 hours and forskolin for 3 hours. This time GlcN had a significant affect on both non-induced as well as induced levels of *IRS-2* (**Fig. 4a**). To determine whether this could be a direct effect of CREB and OGT on the *IRS-2* promoter, we performed a chromatin immunoprecipitation assay. We found that both CREB and OGT were enriched on the *IRS-2* promoter but not

on the GAPDH promoter (**Fig. 4b**). This indicates that CREB glycosylation may have an important effect on IRS-2 levels.

Taken together, this data suggests a new model for the role of CREB glycosylation in hyperglycemia-induced beta cell death. In this model, hyperglycemia causes increased CREB glycosylation at Ser40. This increased glycosylation leads to decreased IRS-2 levels in both non-induced and induced CREB. These lower IRS-2 levels, then, enhance the susceptibility of beta cells to apoptosis.



**Figure 4:** The *IRS-2* gene is a direct target for regulation by CREB and OGT. (a) Cells were treated with or without GlcN. Fsk was added to the medium 3 h after addition of GlcN, and the cells were incubated for an additional 6 h. IRS-2 expression was analyzed by immunoblotting of cell lysates using an anti-IRS-2 antibody and corrected to protein concentration as measured by an anti- $\alpha$ -tubulin antibody.  $n = 4$ ,  $*P < 0.03$ . (b) Chromatin immunoprecipitation was performed on HIT-T15 cells using antibodies against CREB and OGT. PCR was performed on the IRS-2 or GAPDH promoter.

## Methods

**Materials.** All reagents were purchased from Fisher Scientific unless otherwise specified. RPMI-1640, DMEM, Hank's Buffered Saline Solution (HBSS), fetal bovine serum (FBS) and penicillin/streptomycin were purchased from Invitrogen (Carlsbad, CA). CREB and insulin receptor substrate 2 (IRS-2) antibodies were purchased from Upstate (Charlottesville, VA). OGT and  $\alpha$ -tubulin antibodies were purchased from

Sigma-Aldrich (St. Louis, MO). Anti-OGT (AL28) ascites was a generous gift of G.W. Hart (The Johns Hopkins University School of Medicine, Baltimore, MD). *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate (PUGNAc) was purchased from Toronto Research Chemicals (Toronto, Canada). Forskolin (Fsk) was purchased from Axxora (San Diego, CA), and glucosamine (GlcN) was purchased from Fluka (St. Louis, MO).

**Cell Culture.** HIT-T15 cells (American Type Culture Collection) were grown in RPMI-1640 supplemented with 10% FBS, 100 U/ml penicillin, 0.1 mg/ml streptomycin (RPMI-1640 complete). Cell passages 66–79 were used for experiments, and cells were subcultured every 6–8 days. MIN6 cells (a generous gift from Dr. Marc Montminy, Salk Institute, La Jolla, CA) were grown in DMEM supplemented with 10% FBS, 100 U/ml penicillin, 0.1mg/ml streptomycin (DMEM complete) and cell passages 25–35 were used for experiments.

**Pancreatic Isolation.** 150–200 g male Sprague-Dawley rats (Charles River) were euthanized with CO<sub>2</sub> in accordance with IACUC guidelines at the California Institute of Technology. The pancreases were removed and immediately lysed in 2% SDS with 2x protease inhibitor cocktail (Roche) with sonication.

**Diabetic Rat Models.** Homozygous Goto-Kakizaki rats (11 weeks old) and Wistar Hanover GALAS control rats (11 weeks old) were obtained from Taconic and maintained on a NIH-31 diet formulation (Taconic) and the pancreases were harvested after one

week. Homozygous obese diabetic Zucker diabetic fatty rats (10 weeks old) and heterozygous lean Zucker rats (10 weeks) were obtained from Charles River and the pancreases were harvested after one week. 150–200 g male Sprague-Dawley rats (Charles River) were i.p. injected with 50 mg / kg streptozocin dissolved at 10 mg / mL in 100 mM sodium citrate, pH 4.5. The pancreases were harvested after 72 hours.

**Plasmids.** pFC6a-CREB, pFC6a-CREB(AAA), pFC6a-CREB(S40A) were generated by subcloning wild-type CREB, TST259-261AAA, and S40A mutant CREB from the respective pET23b+ vectors into the pFLAG-CMV-6a *E. coli* and mammalian expression vector (Sigma-Aldrich) using primers with 5' EcoRI and 3' BamHI restriction sites.

**Drug Treatments.** GlcN and Fsk treatments were performed as follows. HIT-T15 and MIN-6 cells were used at 50–75% confluence. Cells were pretreated with RPMI-1640 complete (HIT-T15 cells) or DMEM complete (MIN6 cells) supplemented with 10 mM GlcN in 2 mM HEPES, pH 7.5 for 3–12 h as indicated before treatment with the appropriate media supplemented with 10 mM Fsk or vehicle (DMSO) with or without 10 mM GlcN in 2 mM HEPES, pH 7.5 for indicated times.

PUGNAC treatments were performed as follows. Cells were pretreated for 3–12 h as indicated with the appropriate media supplemented with 100 mM PUGNAC before the addition of 10 mM Fsk for indicated times.

#### **Chemoenzymatic Labeling and Detection of O-GlcNAc Glycosylation.**

Chemoenzymatic and PEG labeling was performed as previously described (see Chapter

5) except that 2X Complete™ (Roche) and 2X Halt™ (Pierce) protease inhibitors were added at each step. CREB was immunoprecipitated from the PEG reaction as described below. Chemoenzymatic and biotin labeling followed by streptavidin capture and elution was performed as previously described (see Chapter 2). Chemoenzymatic and biotin labeling followed by CREB immunoprecipitation is described here: After drug treatment, cells were harvested and the cell pellet was lysed in boiling lysis buffer (1% SDS supplemented with protease inhibitors (Complete Protease Inhibitor Cocktail Tablets, Roche, Indianapolis, IN), sonicated for 5 s, and boiled for 8 min. After centrifugation at 21,500 x g for 5 min, the supernatant was collected, and the protein concentration was measured using the BCA assay (Pierce, Rockford, IL). One volume of denatured cell extract was diluted into four volumes of dilution buffer (10 mM HEPES pH 7.9, 1.8% Triton X-100, 100 mM NaCl, 10 mM MnCl<sub>2</sub>, containing protease inhibitors (5 mg/ml pepstatin, 5 mg/ml chymostatin, 20 mg/ml leupeptin, 20 mg/ml aprotinin and 20 mg/ml antipain) and 1 mM phenylmethylsulfonyl fluoride (PMSF). Diluted extract was then supplemented with 1.25 mM adenosine 5'-diphosphate, 0.5 mM analogue **1** and 20 µg/mL mutant Y289L GalT (64). Control reactions were prepared lacking enzyme or analogue **1** to measure any nonspecific reactivity of streptavidin-HRP (see **Fig. 1**). The reaction mixture was incubated at ~5–7° C for 10–12 h, and dialyzed into 10 mM HEPES, pH 7.9 containing 5 M urea at ~5–7° C (3 x 3 h). Following dialysis the sample was acidified to pH 4.8 by adding NaOAc pH 3.9 (50 mM final concentration). The reaction was initiated with the addition of *N*-(aminoxycetyl)-*N'*-(D-biotinoyl) hydrazine (2.5 mM final concentration, Dojindo, Gaithersburg, MD) and incubated at room temperature for 16–20 h, the sample was then dialyzed into CREB immunoprecipitation

(IP) buffer (10 mM HEPES pH 7.9, 100 mM KCl, .2% Triton X-100 and 1mM EDTA) (1x overnight and 2 x 2 h). After dialysis, the sample was centrifuged for 5 min at 21,500 x g, and protein concentration was measured. Lysate was supplemented with protease inhibitors (5 mg/ml pepstatin, 5 mg/ml chymostatin, 20 mg/ml leupeptin, 20 mg/ml aprotinin, and 20 mg/ml antipain) and volumes were normalized such that equivalent amounts and concentrations of lysate were used for immunoprecipitation. Lysate was first precleared for 1 h at  $\sim 5-7^{\circ}$  C with protein A sepharose and the supernatant was incubated with rabbit anti-CREB antibody (Upstate) for 3-4 h at  $\sim 5-7^{\circ}$  C. Protein A sepharose was then added and incubated with the lysate for 1 h at  $\sim 5-7^{\circ}$  C. The supernatant was collected as flowthrough, and the protein A sepharose was washed with CREB IP buffer (3 X), PBS (3 X) and 50 mM  $\text{Na}_2\text{HPO}_4$  (1 X). The immunoprecipitated protein was eluted using SDS-PAGE loading buffer, resolved by SDS-PAGE, and analyzed by Western blotting.

**Luciferase Assay.** HIT-T15 cells transfected using Targefect F2 transfection reagent (3 mg/ml, Targeting Systems, Santee CA) following the manufacturers instructions. Briefly, HIT-T15 cells were grown to  $\sim 75\%$  confluence in 60 mm dishes. In studies of endogenous CREB activity, cells were co-transfected with 1  $\mu\text{g/ml}$  pCRE-Luc (Stratagene) and 0.5  $\mu\text{g/ml}$  pSV- $\beta\text{Gal}$  using the Targefect F2 transfection reagent. In control Gal4 reactions, cells were co-transfected with 1  $\mu\text{g/ml}$  pFR-Luc (Stratagene), 0.5  $\mu\text{g/ml}$  pcDNA3.1-Gal4, and 0.5  $\mu\text{g/ml}$  pSV- $\beta\text{Gal}$ . Cells were treated with Fsk/GlcN or Fsk/PUGNAc  $\sim 24$  h post-transfection.

Harvested cells were lysed in 1 x Reporter Lysis Buffer (Promega) with brief sonication on ice. Samples were centrifuged for 5 min at 21,000 x g. Supernatant was used for the measurement of luciferase and  $\beta$ -galactosidase and Western blot analysis. Luciferase activity was measured using Bright-Glo<sup>TM</sup> luciferase assay system (Promega) on an Opticom 1 luminometer (MGM instruments, Hamden, CT). Luciferase activity and transfection efficiency were corrected by measurement of  $\beta$ -galactosidase activity.

**Western Blotting Analysis.** Total cell lysates were prepared by cell lysis in boiling 1% SDS with protease (5  $\mu$ g/ml pepstatin, 5  $\mu$ g/ml chymostatin, 20  $\mu$ g/ml leupeptin, 20  $\mu$ g/ml aprotinin and 20  $\mu$ g/ml antipain) and phosphatase inhibitors (20 mM NaF, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 0.05 mM Na<sub>2</sub>MO<sub>4</sub>) by sonication. Samples were resolved by 10% SDS-PAGE or by 4–12% Bis-Tris PAGE and transferred to nitrocellulose. The following antibodies were used for Western blot analysis: anti-CREB antibody (1  $\mu$ g/ml), anti-phospho-CREB(Ser133) antibody (1  $\mu$ g/ml), anti-IRS-2 antibody (1  $\mu$ g/ml), and anti-OGT (1  $\mu$ g/ml), anti- $\alpha$ -tubulin (0.2  $\mu$ g/ml). Blots were visualized by chemiluminescence (Supersignal West Pico, Pierce) on Hyperfilm ECL chemiluminescent film (GE Healthcare Bio-Sciences). Relative band intensity was quantified by analysis of scanned images using NIH Image 1.52 software.

**Chromatin Immunoprecipitation (ChIP).** ChIP was performed as previously described (see Chapter 5). The following primers were used. IRS-2 (NM\_001081212) primers 5'-CCCGCCAGCACTCGCTC-3' and 5'-CGGACGTCATCAGAGCC-3' amplify a 174 bp product corresponding to bp -343 to -169.

**RNA isolation and RT-PCR.** HIT-T15 cells were grown in 60 mm dishes to a confluence of 50–75% ( $\sim 5\text{--}7 \times 10^5$  cells). Cells were pretreated with either RPMI-1640 complete supplemented with 10 mM GlcN in 2 mM HEPES pH 7.5 or RPMI-1640 complete for 3 h before treatment with RPMI-1640 complete supplemented with 10  $\mu$ M Fsk and 10 mM GlcN in 2 mM HEPES pH 7.5 or 10  $\mu$ M Fsk for 1 h, respectively.

RNA was isolated using Qiagen RNeasy mini-columns following the manufacturer's procedure for the isolation of cytoplasmic RNA from animal cells (Qiagen). cDNA was prepared using oligo dT<sub>12-18</sub> primers (Invitrogen) and Transcriptor reverse transcriptase (Roche). IRS-2 (NM\_001081212) primers 5'-GAGCATGGATAGACCCTGA-3' and 5'-GCAGAGGCGACCTGAACTAC-3' amplify a 211 bp product corresponding to bp +1617 to +1817. Mouse  $\beta$ -actin (NM\_007393.3) primers 5'-TGTTACCAACTGGGACGACA-3' and 5'-GGGGTGTGAAGGTCTCAA-3' amplify a 165 bp product corresponding to bp +225 to +390. Samples were analyzed with 25–38 cycles of semi-quantitative PCR using Taq PCR<sub>X</sub> DNA polymerase (Invitrogen). PCR products were loaded onto 2% agarose/ethidium bromide gels and visualized on an AlphaImager 3400 (Alpha Innotech Corp., San Leandro CA). Band intensity was quantified using software AlphaEaseFC software version 4.0.1 (Alpha Innotech Corp.).

**Statistical Analysis.** All experiments were repeated a minimum of three times. Results are presented as the mean value  $\pm$  one standard error of the mean (SEM). Statistical significance was calculated using the Student's T-test.

## References

1. Jhala, U.S. *et al.* cAMP promotes pancreatic beta-cell survival via CREB-mediated induction of IRS2. *Genes Dev* **17**, 1575-1580 (2003).
2. Kim, S.-J., Nian, C., Widenmaier, S. & McIntosh, C.H.S. Glucose-dependent insulinotropic polypeptide-mediated up-regulation of beta-cell antiapoptotic Bcl-2 gene expression is coordinated by cyclic AMP (cAMP) response element binding protein (CREB) and cAMP-responsive CREB coactivator 2. *Mol Cell Biol* **28**, 1644-1656 (2008).
3. Jambal, P. *et al.* Cytokine-mediated down-regulation of the transcription factor cAMP-response element-binding protein in pancreatic beta-cells. *J Biol Chem* **278**, 23055-23065 (2003).
4. Sarkar, S.A. *et al.* Dominant negative mutant forms of the cAMP response element binding protein induce apoptosis and decrease the anti-apoptotic action of growth factors in human islets. *Diabetologia* **50**, 1649-1659 (2007).
5. Inada, A. *et al.* Overexpression of inducible cyclic AMP early repressor inhibits transactivation of genes and cell proliferation in pancreatic beta cells. *Mol Cell Biol* **24**, 2831-2841 (2004).
6. Hart, G.W., Housley, M.P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022 (2007).
7. Ball, L.E., Berkaw, M.N. & Buse, M.G. Identification of the major site of O-linked beta-N-acetylglucosamine modification in the C terminus of insulin receptor substrate-1. *Mol Cell Proteomics* **5**, 313-323 (2006).
8. D'Alessandris, C. *et al.* Increased O-glycosylation of insulin signaling proteins results in their impaired activation and enhanced susceptibility to apoptosis in pancreatic beta-cells. *FASEB J* **18**, 959-961 (2004).
9. Buse, M.G., Robinson, K.A., Marshall, B.A., Hresko, R.C. & Mueckler, M.M. Enhanced O-GlcNAc protein modification is associated with insulin resistance in GLUT1-overexpressing muscles. *Am J Physiol Endocrinol Metab* **283**, E241-250 (2002).
10. Teo, C.F., Wollaston-Hayden, E.E. & Wells, L. Hexosamine flux, the O-GlcNAc modification, and the development of insulin resistance in adipocytes. *Mol Cell Endocrinol* **318**, 44-53 (2010).
11. Arias, E.B., Kim, J. & Cartee, G.D. Prolonged incubation in PUGNAc results in increased protein O-Linked glycosylation and insulin resistance in rat skeletal muscle. *Diabetes* **53**, 921-930 (2004).
12. Rossetti, L., Hawkins, M., Chen, W., Gindi, J. & Barzilai, N. In vivo glucosamine infusion induces insulin resistance in normoglycemic but not in hyperglycemic conscious rats. *J Clin Invest* **96**, 132-140 (1995).
13. McClain, D.A. *et al.* Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *Proc Natl Acad Sci USA* **99**, 10695-10699 (2002).

14. Yang, X. *et al.* Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* **451**, 964-969 (2008).
15. Dentin, R., Hedrick, S., Xie, J., Yates, J. & Montminy, M. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* **319**, 1402-1405 (2008).
16. Issad, T. & Kuo, M. O-GlcNAc modification of transcription factors, glucose sensing and glucotoxicity. *Trends Endocrinol Metab* **19**, 380-389 (2008).
17. Vanderford, N.L., Andrali, S.S. & Ozcan, S. Glucose induces MafA expression in pancreatic beta cell lines via the hexosamine biosynthetic pathway. *J Biol Chem* **282**, 1577-1584 (2007).
18. 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 USA* **97**, 2820-2825 (2000).
19. Akimoto, Y., Kreppel, L.K., Hirano, H. & Hart, G.W. Increased O-GlcNAc transferase in pancreas of rats with streptozotocin-induced diabetes. *Diabetologia* **43**, 1239-1247 (2000).
20. Henquin, J.C. Triggering and amplifying pathways of regulation of insulin secretion by glucose. *Diabetes* **49**, 1751-1760 (2000).
21. Bittinger, M.A. *et al.* Activation of cAMP response element-mediated gene expression by regulated nuclear transport of TORC proteins. *Curr Biol* **14**, 2156-2161 (2004).
22. Jansson, D. *et al.* Glucose controls CREB activity in islet cells via regulated phosphorylation of TORC2. *Proc Natl Acad Sci USA* **105**, 10161-10166 (2008).
23. Li, S., Zhang, C., Takemori, H., Zhou, Y. & Xiong, Z.-Q. TORC1 regulates activity-dependent CREB-target gene transcription and dendritic growth of developing cortical neurons. *J Neurosci* **29**, 2334-2343 (2009).