A Chemoenzymatic Strategy toward Understanding *O*-GlcNAc Glycosylation in the Brain

Thesis by

Nelly Khidekel

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended July 28 2006)

…for my parents Raisa and Roman…

Acknowledgments

I would like to thank Linda Hsieh-Wilson for her intellectual guidance and support over the last six years. Working in a young lab was scientifically challenging and invigorating and Linda had a fantastic ability to bring in the "30,000 ft" perspective when it was needed most.

I would also like to thank my committee, Dennis Dougherty, Doug Rees, and Erin Schuman. Dennis and the entire Dougherty group consistently gave of their time for the young students of the Hsieh-Wilson lab, and I will always be grateful for their support and encouragement. I did not quite make it to a rotation in Doug Rees' lab, but I had the opportunity to TA for him and it was a real joy to be able to think about biophysics, in light of my very mathless life at Caltech. I always appreciated Erin's neuroscience insight during our meetings and I had the opportunity to learn from and interact with several of her students, which was likewise terrific.

I owe so much of my scientific development and the success of our work to Eric Peters and Scott Ficarro at GNF. Working with them was not only scientifically rewarding, but also a heck of a lot of fun. They were my second mentors and I will always be indebted to them for their boundless wit and wisdom.

Of course, I am tremendously indebted to all the amazing members of the Hsieh-Wilson lab over the years, who always gave of their time and energy and who supported me on the best days and were there for me on the worst. In particular, I must thank Nathan Lamarre-Vincent who has been my cohort on the *O*-GlcNAc project from the very beginning. We had some colorful discussions about science and life and I hope to have many more as we move onward…and eastward. I would also like to thank Bruce Tai for the collaborative effort in the development of our methodology and Sabine Arndt for the synthesis of our probe molecule. I owe much to Katherine Poulin-Kerstien, who was my scientific partner in the beginning, and a tremendous friend throughout and to Peter Clark, who has been a great collaborator in the last couple of years. I am also indebted to Helen Cheng, with whom I shared many lunchtime conversations about the meaning of *O*-GlcNAc and with whom I became fast friends. I owe my brief foray into synthetic chemistry to Sarah Tully and Claude Rogers and have appreciated Sarah's generosity and camaraderie from the very beginning.

v

Ultimately, this thesis would not have been possible without the companionship and support of my wonderful friends at Caltech, Sarah Miller, Katherine Poulin-Kerstien, Sarah Tully, Adam Poulin-Kerstien, Nathan Lamarre-Vincent, Ray Doss, Nathan Lundblad and many others. In particular, I must thank Sherry Tsai—our conversations were bottomless, our analyses endless, and her friendship deeply meaningful to me. I hope it will stay bright for many years to come.

Finally, I thank my family Roman, Raisa, Ann…and Mr. Fluff, the fattest cat the world has ever known. My parents are some of the most thoughtful, courageous, passionate people imaginable and they have been my life-long mentors and friends. They taught Ann and me the real meaning of what it is to be a 'scientist' from the time we were young, when they supported our every inquiry and discovery. Ann's spirit, wit and kindness have only bettered with age, and I thank her for her friendship and love. Finally, to Ian Swanson—my best friend and the most indelibly charming man I have ever met. People of *his* caliber are rare and I am deeply grateful to be able to share my life with him.

So here's to Club 202 and Pre-Prufrock, sugars and 'juicy' mass spectra, Wendy's runs and book club, Dim Sum and Vegas, and all the gifted, warm and wonderful people who have made the Caltech experience nothing short of magical.

Abstract

Posttranslational modification to proteins represents a fundamental mechanism by which protein function is extended and elaborated. In the brain, modifications such as phosphorylation play critical roles in mediating neuronal communication and development. Unique among carbohydrate modifications is the addition of a single monosaccharide, *N*-acetyl-D-glucosamine, to serine and threonine residues of proteins (*O*-GlcNAc glycosylation). The modification shares intriguing features with phosphorylation, including its intracellular and dynamic nature. The enzyme responsible for adding the modification to proteins is necessary for life at the single cell level and *O*-GlcNAc glycosylation has been linked to nutrient sensing, gene expression, and in the brain, to neurodegeneration. Despite tantalizing evidence for the modification's importance, understanding *O*-GlcNAc glycosylation has been hampered by insufficient strategies to study it at single-protein level as well as across the proteome. Here we describe the development of a new, chemoenzymatic strategy to facilitate the discovery of *O*-GlcNAc proteins, as well as the first studies aimed at understanding *O*-GlcNAc proteome-wide, in the brain.

Our approach capitalizes on an engineered enzyme and synthetic unnatural substrate to specifically 'tag' *O*-GlcNAc-modified proteins for rapid and sensitive detection. We applied the methodology to the discovery of low-abundance, endogenous *O*-GlcNAc proteins from cells. We also combined the approach with mass spectrometry for the isolation of *O*-GlcNAc peptides and the mapping of glycosylation sites, the first step toward functional analysis of the modification. Overall, our efforts led to the identification of nearly fifty new *O*-GlcNAc proteins, several of which serve as targets for mechanistic study. Many of the proteins function in the control of transcription and translation, highlighting the proposed role for *O*-GlcNAc in regulating gene expression. Additionally, we provide evidence that *O*-GlcNAc glycosylation is particularly prevalent on proteins at the nerve terminal, or synaptosome, where it may function to control vesicle cycling and neurotransmitter release. Finally, our work has also led to the first bioanalytical, quantitative assays for *O*-GlcNAc dynamics in both cells and tissue. We

vii

have shown that *O*-GlcNAc is reversible in neuronal tissue and can respond rapidly and robustly to neuronal stimulation *in vivo*.

Table of Contents

Figures

Chapter 3

Chapter 4

Chapter 5

Schemes

Tables

Appendix I

Chapter 1

Introduction: *O*-GlcNAc Glycosylation in the Brain

1.1 Post-Translational Modifications as Critical Regulators of Neuronal Function

One of the most intriguing revelations of the human genome sequencing project, completed in 2003,[1] was the discovery of the relative paucity of genes encoding all the proteins necessary for human development and functioning. That number, now estimated at 20-25,000 genes,[1] is barely larger than the genome of far simpler organisms such as the nematode *Caenorhabditis elegans* (19,000 genes).[2] While genomics has revealed the genes in various organisms, it tells us little about the actual gene products, the proteins, produced in physiological settings. Because these proteins are so essential for cellular function, 'decoding' their interrelationship will be as critical as sequencing the genome to understanding normal and pathological organismal function. One of the fields that has emerged to tackle this challenge is that of proteomics.[3] As a counterpart to genomics, proteomics asks which proteins are expressed in specific cellular states, how these proteins interact in networks to govern cell function, and how these expressions/interactions change in response to physiological stimulation. The number of proteins far outnumbers genes due to alternative transcript formation, proteolytic processing, and the addition of chemical groups to protein side chains, referred to as post-translational modifications (PTMs). As such, the study of protein regulatory networks represents a formidable task.[3]

Perhaps nowhere is complexity more clearly displayed than in the human brain, which is composed of over 10^{11} neurons and over 10-fold more supporting glial cells.[3] The proteins within each of these cells are responsible for dictating neuronal

communication, growth during development, and degradation. Moreover, aberrations in protein function are hallmarks of debilitating neurodegenerative diseases such as Alzheimer's and Huntington's disease. One mechanism that has evolved to orchestrate protein behavior is the post-translational modification (PTM) of protein side chains. PTMs extend and elaborate the standard 20 amino acids,[4,5] and can serve as molecular 'signals' to modulate protein activity and interactions. One of the best characterized PTMs is the addition of a phosphoryl group to serine, threonine, and tyrosine residues of proteins. Over 500 members of the human genome are kinases, highlighting the necessity for this type of modification.[6] Moreover, fifty years of research has demonstrated that phosphorylation represents a central mechanism by which proteins transmit intracellular signals in response to external cues, assemble in macromolecular complexes, undergo degradation and drive a host of processes such as gene expression.[6,7]

In the neuronal junctions that mediate communication, known as synapses, phosphorylation signaling events underlie the proposed molecular basis of learning and memory (Figure 1.1). For example, presynaptically, phosphorylation regulates synaptic vesicle exocytosis and the release of neurotransmitter.[8] Postsynaptically, phosphorylation modulates glutamate receptor conductance, receptor trafficking to and from the membrane, as well as the complexation and degradation of cytoskeletal-remodeling proteins necessary for proper synapse function.[9-11] In fact, one of the most abundant components of neuronal synapses is the protein kinase Ca2+/calmodulin-dependent protein kinase II (CaMKII). CaMKII is a well-characterized molecular 'switch' whose activity is necessary for the activity-induced strengthening of synapses (long-term potentiation (LTP)). At synapses, CaMKII mediates the induction of LTP

both by directly phosphorylating α-amino-3-hydroxy-5-methyl-4-isoxazole propionic

acid receptors (AMPA-R), thus increasing their conductance, and modulating AMPA

receptor trafficking to the cell surface by phosphorylating their associated proteins.[12,13]
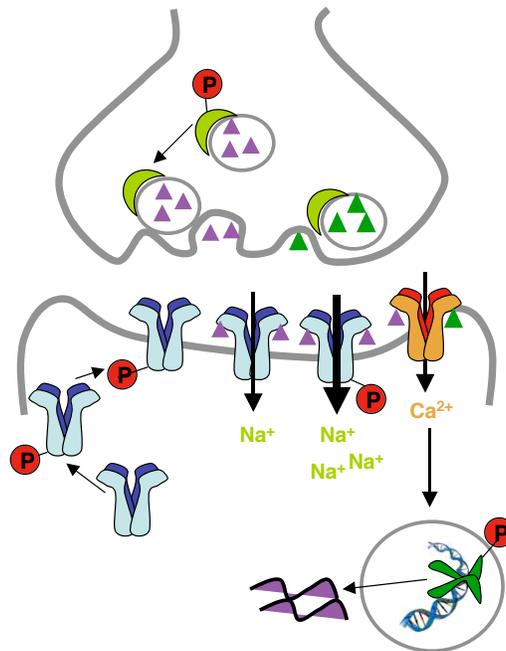
Figure 1.1. Protein phosphorylation regulates numerous facets of neuronal communication, such as the release of neurotransmitter presynaptically, trafficking of ionotropic glutamate receptors into the membrane, and ion channel conductance. Signal transduction via phosphorylation is also necessary for gene transcription and *de novo* protein synthesis, which are required for long-term memory formation.

Among its varied roles, phosphorylation mediates the molecular changes

characteristic of long-term memory formation. Elegant work by Eric Kandel and others,

beginning in the nineteen-sixties, demonstrated that long-term memory formation was

dependent on transcription and translation of new proteins. In particular, they showed

that long-term facilitation of the gill-withdrawal reflex in *Aplysia* was dependent on

transcription mediated by cAMP-response element binding protein (CREB).[14] Related

work demonstrated that LTP, a molecular model for long-term memory, was dependent

on phosphorylation pathways mediated by the cAMP-dependent kinase PKA. As PKA is one of the major kinases acting on CREB, regulated phosphorylation of CREB has been implicated as a molecular marker of long-term memory storage.[15-17] As such, CREB and the phosphorylation pathways that act on it have become targets for drug development to improve cognitive deficits.[14]

In addition to phosphorylation, numerous other PTMs play critical roles in protein function in the brain and other tissue. Currently, over 200 distinct PTMs have been identified, comprising diverse groups of molecules such as lipids, acetates, and polypeptides (Figure 1.2).[4,5]
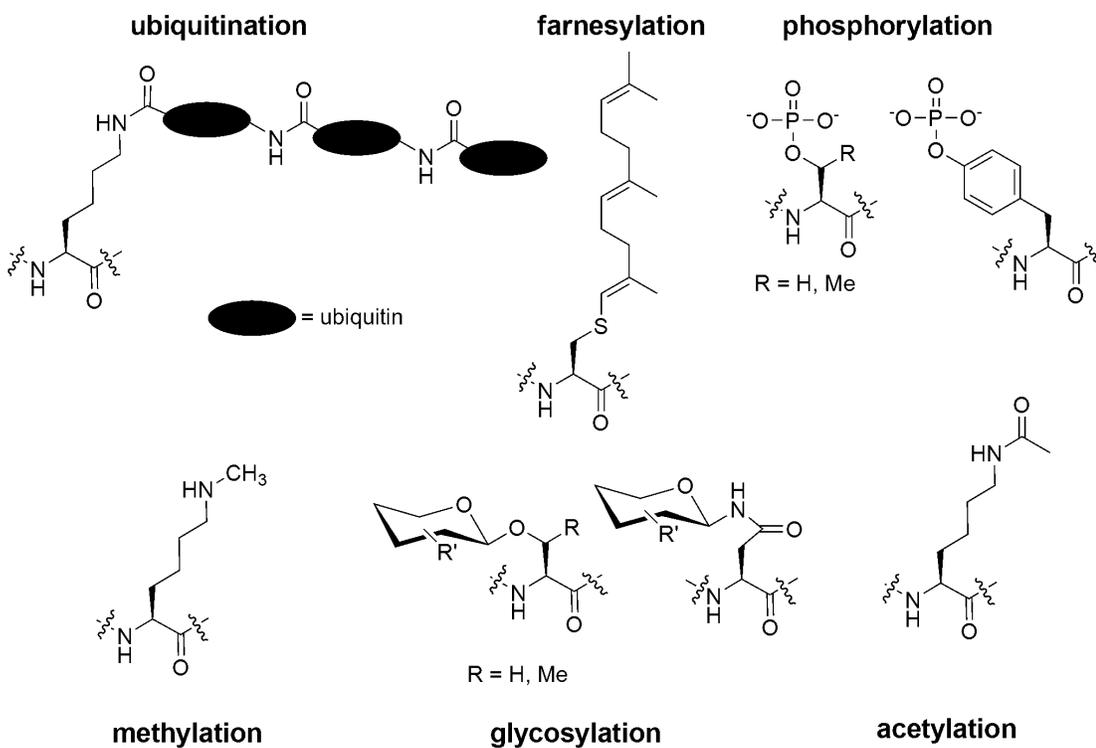
Figure 1.2. Diverse molecules serve to modify proteins posttranslationally. PTMs are critical for all aspects of cellular function including protein turnover and localization, gene expression, and cell-cell communication.

Lipid modifications, such as farnesylation, drive protein translocation and association with the membrane. In the brain, farnesylation has been shown to affect localization of the protein transducin, in the retina, which is critical for light adaptation.[18] At the post-synaptic density (PSD) of the synapse, polyubiquitination (the addition of multiple 76-amino acid ubiquitins to lysine residues) has been shown to promote degradation of critical scaffolding proteins in a manner dependent on neuronal activity.[19,20] Recent efforts to understand PTMs have focused on understanding not just individual modifications, but how networks of modifications drive particular cellular function. One example of such PTM networks exists along the N-terminal 'tails' of histones, the proteins that coil around DNA, and serve as important determinants of transcriptional activation. The distinct network of histone modifications, ranging from lysine acetylation to phosphorylation to methylation, has been proposed to serve as a 'code' that marks dynamic transitions between transcriptionally active or repressed chromatin states.[21]

While some PTMs, like phosphorylation and acetylation, are dynamically regulated, others function stably over the course of a protein's lifetime. One of the most classical of these modifications is the addition of carbohydrate moieties to proteins (glycosylation). Estimates suggest that nearly half of all proteins are glycosylated, making this the most common form of protein modification *in vivo*.[22] *N*-linked glycosylation, via the consensus sequence Asn-X-Ser/Thr, (where X can not be proline) is characterized by extraordinarily long chains of carbohydrates.[5,23] *O*-linked glycosylation, which lacks a consensus motif, is often restricted to just a few carbohydrate units but may be elaborated to great lengths and structural diversity as exemplified by the glycosaminoglyan proteoglycans.[24,25] The enzymes of glycosylation

are largely restricted to the endoplasmic reticulum and Golgi apparatus, where they modify proteins destined for membrane insertion or secretion.[5] As with kinases, the great number and diversity of glycosyltransferases (~250 in the human genome) highlights the essential role of this PTM for cell function.[26]

Studies have implicated carbohydrate modifications in a number of processes, including proper protein folding, protein targeting to subcellular compartments and receptor-ligand binding.[23,24,27] On a cellular level, these modifications are important for cell-cell communication and cell growth and division.[28] One of the most well-characterized roles of glycosylation is in the immune response and inflammation.[29] In humans, red blood cells display one of three glycosphingolipids on the cell surface that differ by a single monosaccharide residue, forming the basis of the so-called A, B or O blood groups. The antibodies generated against the nonnative carbohydrate antigen prohibit blood transfusion among those with different blood types.[24]

In the nervous system, glycosylation plays a vital role in cytokine and growth factor-receptor interactions.[24,28] For example, the cell surface heparan sulfate (HS) proteoglycans are necessary to mediate the interaction of the secreted slit protein and its cell-surface receptor robo. Slit functions as a repulsive guidance cue for growing axons and removing the carbohydrates from HS proteoglycans ablates its chemorepulsive effect.[30] Glycosylation also mediates neurite and axonal outgrowth in the spinal cord, where studies have shown that chondroitin sulfate proteoglycans specifically prevent neuronal regeneration after spinal cord injury.[31] Interestingly, deglycosylation of the proteoglycans 'disinhibits' neurite outgrowth,[32] highlighting the potential benefits of targeting protein glycosylation for therapeutic gain.

1.2 *O*-GlcNAc Glycosylation

Unique to carbohydrate protein modifications is the addition of a monosaccharide β-*N*-acetylglucosamine to serine and threonine residues of proteins (*O*-GlcNAc glycosylation) (Figure 1.3).[33-35]
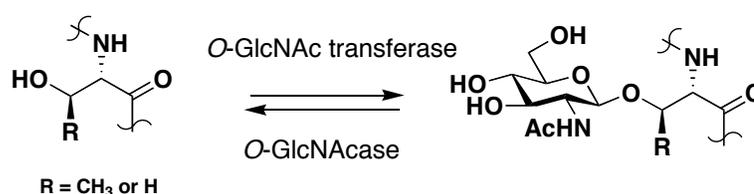
Figure 1.3. *O*-GlcNAc glycosylation is a unique carbohydrate modification to proteins. Mediated by two soluble, intracellular enzymes, *O*-GlcNAc is dynamic and ubiquitous, sharing more in common with phosphorylation than traditional forms of glycosylation.

First identified by Hart and Torres in 1984 in lymphocyte cells,[36] *O*-GlcNAc is distinguished from other carbohydrate modifications by occurring predominantly on intracellular proteins rather than those sequestered to membrane compartments.[37] Moreover, unlike classical forms of glycosylation, *O*-GlcNAc appears to be dynamically regulated. Pulse-chase experiments including [^{3}H]-GlcNAc have shown that for certain protein substrates, the half-life of the *O*-GlcNAc modification is on the order of hours compared to days for the protein backbone.[38] Moreover, recent studies have shown that *O*-GlcNAc levels are strongly upregulated in response to a number of cell stress stimuli.[39] In neutrophils (white blood cells), *O*-GlcNAc can respond directly to ligand binding within minutes.[40]

In the twenty years since its discovery, *O*-GlcNAc has been identified in all higher eukaryotic organisms, from *C. elegans* to humans[33] and found to modify ~100 proteins.[39] The enzymes responsible for catalyzing the modification have been cloned and characterized.[41] Demonstrating that there is no compensatory mechanism for

O-GlcNAc in mammals, an *O*-GlcNAc transferase (OGT) knockout is lethal at the embryonic stem cell level.[42,43] OGT shares little to no homology to traditional glycosyltransferases. Moreover, unlike these enzymes, which are restricted to the ER and Golgi, OGT is a soluble, alternatively spliced, multimeric protein found predominantly in the nucleus and cytoplasm, with additional limited distribution in the mitochondria.[41] Interestingly, unlike enzymes that govern well-characterized modifications, such as phosphorylation, there is only one known OGT gene. While there is no consensus sequence for glycosylation, OGT may modulate its substrate specificity through protein-protein interactions with its tetratricopeptide N-terminal repeat domain (TPR).[44] Recently crystallized, the OGT TPR bears significant similarity to that of importin-α, in which the TPR serves as a protein-interaction domain.[45] Indeed, several binding partners of OGT have been identified,[46,47] and may function to govern its subcellular localization or substrate targeting in a manner analogous to the A-kinase anchoring proteins (AKAPs) associated with the cAMP pathway and PKA.[48]

Interestingly, the β-*N*-acetylglucosaminidase (*O*-GlcNAcase) that removes *O*-GlcNAc has also been cloned and is distributed to the cytoplasm and nucleus.[49,50] Unlike classical hexosaminidases that reside in the lysozome and function largely to degrade membrane glycoproteins, *O*-GlcNAcase has a pH optimum near neutral, coincident with its non-lysosomal activity.[49] The *O*-GlcNAcase appears to be a bifunctional enzyme, as it shares sequence homology with acetyltransferases, and recent studies have shown that it can behave as an acetyltransferase *in vitro*.[51] In dividing cells, targeted pharmacological disruption of the *O*-GlcNAcase induces growth arrest, whereas overexpression induces early mitotic exit and severe perturbations in mitotic

phosphorylation.[52]

The intriguing features of O-GlcNAc led many groups to posit that this modification may function similarly to phosphorylation in the cell.[53] Like phosphorylation, O-GlcNAc is intracellular, dynamic and targets serine and threonine residues. Moreover, recent studies have identified proteins with lectin-like properties for O-GlcNAc, (such as the heat shock protein 70). These lectins may behave analogously to known proteins baring phosphorylation-binding domains such as the WW superfamily.[54] Additionally, several studies highlight the apparent antagonism between O-GlcNAc glycosylation and phosphorylation. For example, O-GlcNAc modifies the carboxy-terminal domain (CTD) of RNA polymerase II (RNAP II) *in vivo*. However, this form of RNAP II lacks the phosphorylation necessary for transcriptional initiation, elongation and termination and thus is transcriptionally inactive.[55,56] Moreover, synthetic peptides that are glycosylated cannot be phosphorylated by the CTD $TF_{II}H$ kinase CDK7,[56] while phosphorylated peptides are no longer substrates for OGT, even though the two primary sites of modification are distinct. In other examples, O-GlcNAc and phosphate compete for the same site of modification. This is the case with the oncogene c-myc, in which the lymphomal hotspot Thr58 can be interchangeably modified depending on extracellular conditions,[57] as well as the transcription factor estrogen receptor β (ER-β), in which glycosylation and phosphorylation may have opposing effects on protein stability.[58] These studies support what has been characterized as a 'yin-O-yang' between O-GlcNAc and phosphorylation. Nonetheless, in the case of certain proteins, such as the signal transducer and activator of transcription 5 (Stat5), O-GlcNAc and phosphorylation are not mutually exclusive[59] and may behave cooperatively to regulate function. Moreover,

establishing the interplay between the two modifications on complex networks of proteins, if it exists, has proved challenging, in part because the function of *O*-GlcNAc was, for many years, not well understood.

Initial efforts to understand the role of *O*-GlcNAc glycosylation centered on identifying modified proteins. Nearly one hundred *O*-GlcNAc proteins were individually identified, during a twenty-year period.[39] These proteins include transcription factors such as Sp1and CREB, the 26S cap of the proteasome, cytoskeletal proteins such as the neurofilament proteins H, M, L, and proteins involved in glucose metabolism, such as glycogen synthase.[60-63]

In part due to the challenges in identifying *O*-GlcNAc targets,[64] the functional significance of *O*-GlcNAc modification remains elusive. Nonetheless, in recent years, several intriguing hypotheses have emerged as to the role of *O*-GlcNAc, including nutrient sensing and transcriptional repression (Figure 1.4). [33,35]
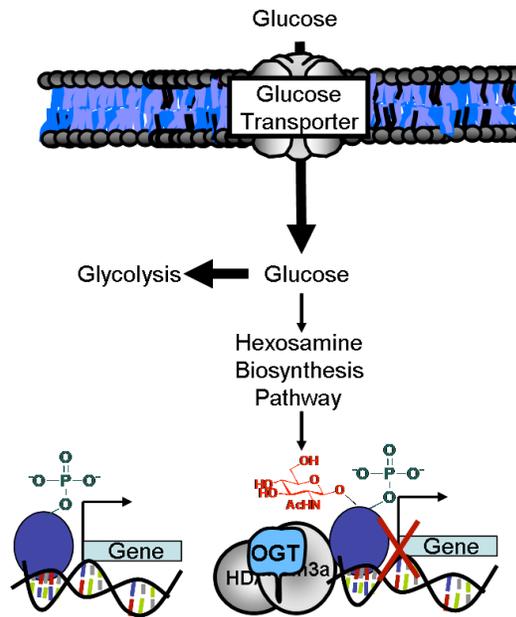
Figure 1.4. *O*-GlcNAc glycosylation plays a role in many aspects of cell function including transcriptional repression and nutrient sensing. The substrate for the *O*-GlcNAc transferase (OGT), UDP-GlcNAc, is a product of the hexosamine biosynthesis pathway and OGT's unique sensitivity to it can tune *O*-GlcNAc levels in response to extracellular glucose. Upregulation of *O*-GlcNAc on transcriptional machinery can inhibit gene transcription in response to changes in glucose or other stimuli.

Roughly 3% of all glucose that enters cells is shuttled into the hexosamine biosynthesis pathway[62] of which the primary product is uridine-diphospho-*N*-actetyl-D-glucosamine (UDP-GlcNAc), the substrate of OGT. Interestingly, UDP-GlcNAc levels are known to fluctuate significantly, depending on cell state.[65] Moreover, OGT has 3 apparent K_ms for UDP-GlcNAc (6, 35 and 217 μM) toward peptide acceptor substrates, and its activity toward those substrates changes with increasing UDP-GlcNAc concentration.[66] These and other findings suggest that glucose levels may be able to fine-tune the *O*-GlcNAc modification in a highly sensitive manner,[67] which prompted many groups to propose that *O*-GlcNAc functions as a nutrient sensor in the cell. Several studies support this hypothesis, demonstrating that GlcNAc levels, both globally and on individual proteins, are responsive to varying glucose concentrations in a number of cell

types.[68-70] Because aberrantly high levels of glucose can affect O-GlcNAc levels, it has been proposed that the modification contributes to type II diabetes, a hallmark of which is hyperglycemia-induced insulin resistance. In support of such a role, Hart and colleagues demonstrated that treatment of 3T3-L1 adipocyte cell lines with the O-GlcNAcase inhibitor O-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-N-phenylcarbamate (PUGNAc) decreased insulin-stimulated glucose uptake and interfered with insulin-activated phosphorylation signaling cascades.[71] A similar finding was reported for skeletal muscle.[72] Others have demonstrated that overexpression of OGT induces insulin resistance in both adipocyte and skeletal tissue.[73] Finally, a $C.$ $elegans$ OGT knockout, while not lethal, induces significant changes in metabolite storage and suppresses the conversion to the dauer form of the nematode, suggesting that O-GlcNAc plays a significant role in $C.$ $elegans$ nutrient storage and longevity.[74]

In addition to nutrient sensing, O-GlcNAc also appears to figure prominently in the control of transcription. These two roles for O-GlcNAc are likely interrelated as it is well known that nutrient cues from the extracellular environment can initiate particular patterns of gene expression in numerous cell types.[35] Recent efforts in the field have focused on the role of O-GlcNAc in transcriptional repression, although it should be noted that for several molecular targets such as nuclear factor kappa B (NFκB), O-GlcNAc serves to upregulate transcriptional activity.[35] Consistent with a role in repression, OGT was found to stably interact with the corepressor mSin3A, a member of the histone deacetylase complex, where it was shown to be involved in synergistic transcriptional repression along with histone deacetylation.[75] Recently, mSIN3A and OGT interactions were tied to yet another PTM, the nonenzymatic addition of

methylglyoxal, a product of glycolysis, to arginine residues of proteins. Methylglyoxal modification is associated with aging, inflammation, and a host of pathologies including diabetes. Methylglyoxal modification of mSin3A in response to high glucose led to enhanced recruitment of OGT and increased *O*-GlcNAc modification of the transcription factor Sp3. The resultant elevation in *O*-GlcNAc led to an increase in expression of the angiopoeitin-2 gene, one of several genes whose proliferative function is associated with late stages of various diseases, including diabetic retinopathy.[76]

In addition to its interaction with mSin3A, *O*-GlcNAc glycosylation has been shown to be associated with transcriptionally silenced areas of euchromatin.[77] Moreover, Kudlow and colleagues demonstrated a negative effect on Sp1 transcription factor activity in response to glycosylation[78]. Our group identified CREB as a new target of *O*-GlcNAc modification, and showed that CREB-mediated transcription is likewise downregulated in response to glycosylation.[60] Finally, OGT colocalizes with RNAP II in cells,[46] and as described earlier, is associated with transcriptionally inactive forms of the RNAP II holonenzyme.[55]

1.3 *O*-GlcNAc Glycosylation in the Brain

Given the similarities and potential interplay between phosphorylation and *O*-GlcNAc glycosylation and the essential role phosphorylation plays in brain function, we became interested in the role of *O*-GlcNAc in the nervous system. In particular, we were interested in the molecular targets of *O*-GlcNAc glycosylation and the ensuing effect of *O*-GlcNAc on brain function. In addition, we were intrigued by the dynamic nature of the modification and sought to understand *O*-GlcNAc regulation in neuronal tissue.

A number of studies suggested that *O*-GlcNAc plays an important role in the

brain. Both OGT and O-GlcNAcase mRNA and protein expression are highest in brain and pancreas.[79] A neuron-specific OGT conditional knockout in the mouse produced viable pups who failed to nurse, had significant motor deficiencies and died within several days.[42] Further supporting a 'yin-O-yang' between phosphorylation and O-GlcNAc, OGT was found to form stable complexes with protein phosphatase 1 in the brain.[80]

Prior to our work, only a handful of O-GlcNAc-modified proteins from the brain had been identified. Among these proteins were the synaptic vesicle associated protein synapsin, β-synuclein, neurofilament proteins H, L, M, the collapsin response mediator protein-2 (CRMP-2), the amyloid precursor protein (APP) and the microtubule associated protein tau.[81] Hyperphosphorylated tau and aberrantly processed APP are hallmarks of Alzheimer's disease, which has been linked to insufficient glucose uptake in the aging brain,[82] suggesting that O-GlcNAc may play a role in neurodegeneration. Indeed, the OGT conditional mouse knockout showed hyperphosphorylated tau,[42] and studies on hippocampal slices from adult rat have shown aberrant phosphorylation of tau in response to treatment with the O-GlcNAcase inhibitor PUGNAc.[83] Interestingly, O-GlcNAc is an endogenous inhibitor of the proteasome, and corollary studies suggested that aberrant O-GlcNAc glycosylation in the brain induced proteome malfunction and neuronal apoptosis.[83]

Whether O-GlcNAc plays a role in neurodegeneration remains unclear, but even more fundamentally, its contribution to non-pathological brain function is still unknown. Specifically, although found throughout the neuron, OGT is particularly abundant in

synaptosomes, especially in the cytosol of the axon terminal and around dendritic microtubules, consistent with *O*-GlcNAc modification of several microtubule-associated proteins.[81,84] Both enzymes are found throughout brain substructures, but are most prevalent in the hippocampus and the Purkinje neurons of the cerebellum.[83,84] Indeed, a study from 1999 identified a panel of phosphatase and kinase inhibitors that had significant effects on *O*-GlcNAc levels in the cytoskeletal fractions of cultured cerebellar neurons suggesting that, as in other tissues, *O*-GlcNAc can be dynamic in the brain.[85]

One possible role for *O*-GlcNAc in the brain may involve glucose sensing in active brain areas. The brain consumes a disproportionally large amount of glucose and oxygen compared to the rest of the body.[86] Contrary to the prevailing model of relatively invariant glucose uptake in the brain, recent studies suggest that glucose levels increase in specific regions of the adult brain, in response to high levels of activity.[86] For example, in rodent hippocampi, glucose uptake from the cerebrospinal fluid was shown to increase significantly in response to maze-learning paradigms.[86] Because of its exquisite sensitivity to glucose, one physiological role for *O*-GlcNAc may be as a mechanism to tag or mark neurons or neuronal networks with high metabolic rates, in turn affecting diverse processes such as axonal outgrowth or synaptic vesicle cycling. In addition to glucose, other factors may contribute to *O*-GlcNAc regulation and function in the brain. For example, OGT appears to be tyrosine phosphorylated[87] and may respond to extracellular cues through tyrosine kinase/phosphatase signaling cascades. Finally, *O*-GlcNAc levels are known to respond to stress, and protect cells from stress-induced apoptosis.[88] In neurons, *O*-GlcNAc may serve analogously to protect cells from

apoptosis (in response to excitotoxicity, for example) potentially in both a UDP-GlcNAc dependent and independent manner.

1.4 Methodologies to Study *O*-GlcNAc

Despite emerging evidence for its significance in the brain, identifying *O*-GlcNAc glycosylated proteins in neurons or understanding the regulation of *O*-GlcNAc in the brain has proven challenging. Methodologies to detect the *O*-GlcNAc modification endogenously have traditionally suffered from both lack of selectivity and sensitivity (discussed in more detail in Chapter 2). The first tool to study *O*-GlcNAc harnessed the selectivity of the enzyme β1,4-galactosyltransferase to catalyze the transfer of galactose to terminal GlcNAc sugars.[36] Highly selective for GlcNAc, the enzyme was shown to effectively transfer UDP-[^{3}H]-galactose to GlcNAc-modified proteins for detection by autoradiography.[37] However, compared to other commonly used radioisotopes such as [^{32}P], UDP-[^{3}H]-galactose has low specific activity, such that it can take weeks or months to detect even the most highly abundant glycosylated proteins.[64,89] This limitation led to the development of several antibodies for *O*-GlcNAc. Specifically, the RL-2 antibody was raised against *O*-GlcNAc modified nuclear pore proteins, and was demonstrated to show good selectivity for *O*-GlcNAc.[90] However, because it was raised against particular *O*-GlcNAc glycosylated peptides, the antibody also recognizes peptide determinants and either fails to detect certain *O*-GlcNAc proteins entirely, or requires a higher concentration of protein than found endogenously for detection.[64,90] Hart and colleagues developed an *O*-GlcNAc-specific antibody shown to bind selectively to *O*-GlcNAc peptides with no cross-reactivity for *O*-linked *N*-acetyl-d-galactosamine (*O*-GalNAc)

peptides or peptides lacking the GlcNAc sugar.[89] While this CTD110.6 antibody has been employed as a strategy to assess global O-GlcNAc glycosylation, it still fails to detect many O-GlcNAc-modified proteins, and often exhibits weak binding affinity by Western blotting.[64]

The traditional strategies are further limited by their inability to identify specific amino acid sites of O-GlcNAc glycosylation on proteins of interest, which is a prerequisite for understanding the functional role of the modification on individual targets. Because OGT lacks a consensus sequence for glycosylation[79] and because the O-GlcNAc moiety is highly labile during mass spectrometry,[91] predicting sites of glycosylation or identifying sites analytically has proven challenging.

1.5 Chemical Biology and Bioanalytical Approaches to Study Proteins and PTMs

The study of PTMs, as exemplified by O-GlcNAc, poses a number of challenges. They are not genetically encoded, (often with no consensus amino acid sequence for modification), present at substoichiometric amounts or on low-abundance proteins, are dynamic in the cell, and often labile during protein analysis. As such, the PTM field has spurred the development of many new and exciting chemical biology and bioanalytical techniques, greatly enhancing our ability to understand this rich area of biology.

PTM analysis in recent years has benefited greatly from the application of well-established chemical strategies to study the biophysical and biochemical properties of proteins. In particular, synthetic chemistry and molecular biology have been used to design non-canonical, unnatural amino acid analogues for incorporation into proteins *in vivo*.[92] For example, site-specific incorporation, via nonsense suppression, has been used to perturb ligand binding sites in receptor proteins and ion channels to glean molecular

insights into ion channel properties.[93] Residue-specific incorporation, via global replacement of a given amino acid, has been used in protein engineering contexts to modulate protein-fold and protein-protein interactions.[92] Unnatural amino acids have also been used as bioorthogonal chemical handles, to permit ready detection and isolation of proteins from cells.[94] In a related vein, chemical strategies have been used to expand the scope of traditional peptide synthesis in order to gain access to large polypeptides or even proteins *in vitro* for functional analysis.[95] Importantly, approaches like native chemical ligation, (which uses transthioesterification to couple peptides) have allowed chemists to build proteins with nonnatural amino acids, such as fluorescent tags.[95] Native chemical ligation and related techniques are particularly attractive strategies for the creation of post-translationally modified proteins *in vitro*. They can be used to create homogenously modified proteins for biological studies as exemplified by the generation of the *O*-linked glycosylated protein, and antimicrobial agent, diptericin[96].

Bertozzi and colleagues elegantly applied the principles of unnatural analogue incorporation to the study of carbohydrates. In particular, they have demonstrated that carbohydrates with bioorthogonal functionalities such as ketones or azides can be incorporated in cells, and even in a limited subset of animal tissues *in vivo*.[97] The bioorthogonal functionality on these sugars can be used as a tag to identify cell surface expression of such molecules, which may be aberrant in certain disease states, such as cancer.[98]

Other synthetic approaches have been geared toward the study of phosphorylation. For example, the 'bump-hole'-type strategies have been used to incorporate unnatural adenosine triphosphate (ATP) analogues into mutant kinases in

order to identify new physiological targets of such kinases.[6] Moreover, lipid modifications, specifically farnesylation, have been the targets of azido analogues incorporated in cells for the identification of new, modified proteins from cell culture.[99]

Such synthetic approaches have developed in tandem with analytical techniques for the discovery of PTMs. The last twenty years have seen the robust expansion of mass spectrometry, not just in the identification of isolated proteins, but in the evaluation of whole-cell proteomes, and the discovery of corresponding sets of PTMs.[100] For example, multidimensional protein identification technology (MudPIT), which uses multiple chromatographic steps coupled to mass spectrometry, has enabled global studies of posttranslationally modified proteins in yeast.[101] Chemical derivitization methods for a number of PTMs, such as phosphates and *N*-linked carbohydrates, as well as techniques such as metal ion affinity chromatography have facilitated the enrichment of low-stoichiometry modified peptides and the sequencing of modification sites.[102,103,104] In another example, Dietrich and colleagues used the incorporation of an unnatural methionine analogue, with azide functionality, in order to specifically isolate and identify newly translated proteins from cell culture.[105] Finally, emerging strategies in mass spectrometry have involved not just the identification of proteomes and PTMs, but the quantification of changes in both protein expression and PTMs in response to cellular stimulation. In the best-characterized strategies such as stable isotope labeling with amino acids in cell culture (SILAC)[106] or isobaric tags for relative and absolute quantification (iTRAQ),[107] cells or cell extracts are labeled with stable isotopes and the ratio of signal intensities of differentially labeled, but essentially chemically identical, peptides are used to determine relative abundance (Figure 1.5).
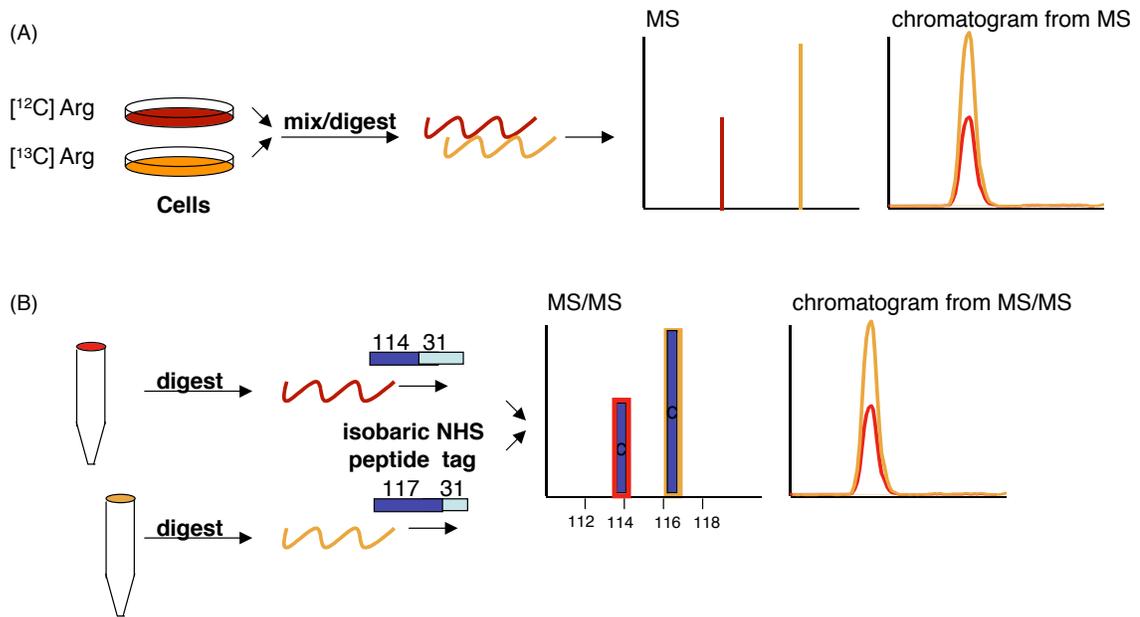
Figure 1.5. Mass spectrometry (MS)-based approaches for quantitative proteomics. (A) In methods such as stable isotopic labeling with amino acids in cell culture (SILAC), differential isotopes of amino acids such as arginine are added to two or more different populations of cells. Samples are mixed and proteolytically digested. Relative signal intensities of mass differentiated but chemically identical peptides are used to determine relative abundance. (B) In another method, isobaric tags for relative and absolute quantification (iTRAQ), cell lysates from two or more different cell populations are proteolytically digested. Peptide n-terminal amines (and ε-amines from lysine residues) are subsequently labeled with isobaric mass tags. The tags are cleaved during tandem MS (MS/MS) allowing mass differentiation, and the differences in signal intensity between portions of the tags are used to determine relative abundance.

In order to monitor quantitative changes in PTMs, techniques such as SILAC have been coupled to enrichment strategies, as in the case of a recent study that used antityrosine antibodies to specifically capture SILAC-labeled tyrosine-modified proteins. In this case, the authors were able to monitor temporal changes in tyrosine phosphorylation in response to cell stimulation.[108]

Overall, the wealth of techniques that have recently emerged to study protein and PTM networks should prove fruitful as the field moves from identifying modified proteins toward addressing the more challenging questions of how PTMs are regulated,

spatially and temporally, and in turn, how they modulate 'proteome' and cellular

function.

1.6 Conclusion

The covalent attachment of post-translational modifications to protein side chains is a critical means by which protein function is elaborated, and by which key cellular and organismal processes are controlled. In the brain, well-characterized phosphorylation signaling cascades mediate molecular changes proposed to form the basis of learning and memory. More recently, other PTMs such as ubiquitination and acetylation have come to be recognized for their contribution to neuronal function. One emerging PTM is *O*-GlcNAc glycosylation, the covalent attachment of the single monosaccharide, *N*-acetyl-d-glucosamine to serine/threonine residues of proteins. Unique among carbohydrate modifications, *O*-GlcNAc is both intracellular and dynamic. The enzymes responsible for the modification are necessary for life at the single cell level and are particularly abundant in the brain and pancreas. Although it may serve global functions in nutrient sensing and transcriptional repression, the role of *O*-GlcNAc in the brain and the brain-specific proteins it modifies are largely unknown. Owing to the difficulty in detecting the modification and identifying *O*-GlcNAc modification sites, this PTM has proven challenging to study. As such, the development of sensitive techniques for both the discovery of *O*-GlcNAc proteins and the identification of stimuli that regulate *O*-GlcNAc in the brain are critical. In this regard, emerging strategies involving synthetic chemistry, chemical biology and analytical chemistry should prove to be powerful tools to understand not just *O*-GlcNAc but the complex regulatory network of PTMs that underlie brain function.

References

1.	International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
2.	The C. elegans sequencing consortium. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science* **282**, 2012-18 (1998).
3.	Becker, M., Schindler, J. & Nothwang, H. G. Neuroproteomics - the tasks lying ahead. *Electrophoresis* **27**, 2819-29 (2006).
4.	Khidekel, N. & Hsieh-Wilson, L. C. A 'molecular switchboard'--covalent modifications to proteins and their impact on transcription. *Org Biomol Chem* **2**, 1-7 (2004).
5.	Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-72 (2005).
6.	Johnson, S. A. & Hunter, T. Kinomics: methods for deciphering the kinome. *Nat Methods* **2**, 17-25 (2005).
7.	Cohen, P. The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-30 (2002).
8.	Turner, K. M., Burgoyne, R. D. & Morgan, A. Protein phosphorylation and the regulation of synaptic membrane traffic. *Trends Neurosci* **22**, 459-64 (1999).
9.	Esteban, J. A., Shi, S. H., Wilson, C., Nuriya, M., Huganir, R.L. & Malinow, R. PKA phosphorylation of AMPA receptor subunits controls synaptic trafficking underlying plasticity. *Nat Neurosci* **6**, 136-43 (2003).
10.	Derkach, V., Barria, A. & Soderling, T. R. Ca2+/calmodulin-kinase II enhances channel conductance of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionate type glutamate receptors. *Proc Natl Acad Sci USA* **96**, 3269-74 (1999).
11.	Pak, D. T. & Sheng, M. Targeted protein degradation and synapse remodeling by an inducible protein kinase. *Science* **302**, 1368-73 (2003).
12.	Lisman, J., Schulman, H. & Cline, H. The molecular basis of CaMKII function in synaptic and behavioural memory. *Nat Rev Neurosci* **3**, 175-90 (2002).
13.	Tomita, S., Stein, V., Stocker, T. J., Nicoll, R. A. & Bredt, D. S. Bidirectional synaptic plasticity regulated by phosphorylation of stargazin-like TARPs. *Neuron* **45**, 269-77 (2005).
14.	Carlezon, W. A., Jr., Duman, R. S. & Nestler, E. J. The many faces of CREB. *Trends Neurosci* **28**, 436-45 (2005).
15.	Bito, H., Deisseroth, K. & Tsien, R. W. CREB phosphorylation and dephosphorylation: a Ca(2+)- and stimulus duration-dependent switch for hippocampal gene expression. *Cell* **87**, 1203-14 (1996).
16.	Barco, A., Alarcon, J. M. & Kandel, E. R. Expression of constitutively active CREB protein facilitates the late phase of long-term potentiation by enhancing synaptic capture. *Cell* **108**, 689-703 (2002).
17.	Nguyen, P. V. & Woo, N. H. Regulation of hippocampal synaptic plasticity by cyclic AMP-dependent protein kinases. *Prog Neurobiol* **71**, 401-37 (2003).
18.	Kassai, H., Aiba, A., Nakao, K., Nakamura, K., Katsuki, M., Xiong, W. H., Yau, K. W., Imai, H., Shichida, Y., Satomi, Y., Takao, T., Okano, T. &

Fukada, Y. Farnesylation of retinal transducin underlies its translocation during light adaptation. *Neuron* **47**, 529-39 (2005).

19. Bingol, B. & Schuman, E. M. Synaptic protein degradation by the ubiquitin proteasome system. *Curr Opin Neurobiol* **15**, 536-41 (2005).

20. Yi, J. J. & Ehlers, M. D. Ubiquitin and protein turnover in synapse function. *Neuron* **47**, 629-32 (2005).

21. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-80 (2001).

22. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).

23. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**, 2364-9 (2001).

24. Hang, H. C. & Bertozzi, C. R. The chemistry and biology of mucin-type *O*-linked glycosylation. *Bioorg Med Chem* **13**, 5021-34 (2005).

25. Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K. & Kitagawa, H. Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr Opin Struct Biol* **13**, 612-20 (2003).

26. Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328**, 307-17 (2003).

27. Pellegrini, L. Role of heparan sulfate in fibroblast growth factor signalling: a structural view. *Curr Opin Struct Biol* **11**, 629-34 (2001).

28. Selleck, S. B. Shedding light on the distinct functions of proteoglycans. *Sci STKE* **2006**, pe17 (2006).

29. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370-6 (2001).

30. Hu, H. Cell-surface heparan sulfate is involved in the repulsive guidance activities of Slit2 protein. *Nat Neurosci* **4**, 695-701 (2001).

31. Silver, J. & Miller, J. H. Regeneration beyond the glial scar. *Nat Rev Neurosci* **5**, 146-56 (2004).

32. Massey, J. M., Hubscher, C. H., Wagoner, M.R., Decker, J.A., Amps, J., Silver, J. & Onifer, S. M. Chondroitinase ABC digestion of the perineuronal net promotes functional collateral sprouting in the cuneate nucleus after cervical spinal cord injury. *J Neurosci* **26**, 4406-14 (2006).

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

34. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: signal transduction and *O*-GlcNAc. *Science* **291**, 2376-8 (2001).

35. Zachara, N. E. & Hart, G. W. Cell signaling, the essential role of *O*-GlcNAc! *Biochim Biophys Acta* (2006).

36. 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).

37.   Holt, G. D. & Hart, G. W. The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, *O*-linked GlcNAc. *J Biol Chem* **261**, 8049-57 (1986).

38.   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).

39.   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).

40.   Kneass, Z. T. & Marchase, R. B. Neutrophils exhibit rapid agonist-induced increases in protein-associated *O*-GlcNAc. *J Biol Chem* **279**, 45759-65 (2004).

41.   Iyer, S. P. N. & Hart, G. W. Dynamic nuclear and cytoplasmic glycosylation: Enzymes of *O*- GlcNAc cycling. *Biochemistry* **42**, 2493-2499 (2003).

42.   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).

43.   Shafi, R., Iyer, S. P., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart, G. W., & Marth J. D. 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).

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

45.   Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J.A. & Conti, E. The superhelical TPR-repeat domain of *O*-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* **11**, 1001-7 (2004).

46.   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).

47.   Andrali, S. S., Marz, P. & Ozcan, S. Ataxin-10 interacts with *O*-GlcNAc transferase OGT in pancreatic beta cells. *Biochem Biophys Res Commun* **337**, 149-53 (2005).

48.   Bauman, A. L., Goehring, A. S. & Scott, J. D. Orchestration of synaptic plasticity through AKAP signaling complexes. *Neuropharmacology* **46**, 299-310 (2004).

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

50.   Wells, L., Gao, Y., Mahoney, J. A., Vosseller, K., Chen, C., Rosen, A. & Hart G. W. 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).

51.   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).

52.   Slawson, C. Zachara, N. E., Vosseller, K., Cheung, W. D., Lane, M. D. & Hart, G. W. Perturbations in *O*-linked beta-N-acetylglucosamine protein modification

cause severe defects in mitotic progression and cytokinesis. *J Biol Chem* **280**, 32944-56 (2005).

53.   Hart, G. W., Greis, K. D., Dong, L. Y., Blomberg, M. A., Chou, T. Y., Jiang, M.S., Roquemore, E.P., Snow, D. M., Kreppel, L. K., Cole, R. N., et al. *O*-linked N-acetylglucosamine: the "yin-yang" of Ser/Thr phosphorylation? Nuclear and cytoplasmic glycosylation. *Adv Exp Med Biol* **376**, 115-23 (1995).

54.   Guinez, C., Losfeld, M. E., Cacan, R., Michalski, J. C. & Lefebvre, T. Modulation of HSP70 GlcNAc-directed lectin activity by glucose availability and utilization. *Glycobiology* **16**, 22-8 (2006).

55.   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**, 10416-24 (1993).

56.   Comer, F. I. & Hart, G. W. Reciprocity between *O*-GlcNAc and *O*-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40**, 7845-52 (2001).

57.   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**, 19229-35 (2002).

58.   Cheng, X. & Hart, G. W. Alternative *O*-glycosylation/*O*-phosphorylation of serine-16 in murine estrogen receptor beta: post-translational regulation of turnover and transactivation activity. *J Biol Chem* **276**, 10570-5 (2001).

59.   Gewinner, C., Hart, G. W., Zachara, N., Cole, R., Beisenherz-Huss, C. & Groner, B. The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *J Biol Chem* **279**, 3563-72 (2004).

60.   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).

61.   Parker, G. J., Lund, K. C., Taylor, R. P. & McClain, D. A. Insulin resistance of glycogen synthase mediated by *O*-linked N-acetylglucosamine. *J Biol Chem* **278**, 10022-7 (2003).

62.   Zachara, N. E. & Hart, G. W. The emerging significance of *O*-GlcNAc in cellular regulation. *Chem. Rev.* **102**, 431-8 (2002).

63.   Zhang, F., Su, K., Yang, X., Bowe, D. B., Paterson, A. J. & Kudlow, J. E. *O*-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **115**, 715-25 (2003).

64.   Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., Qasba, P. K. & Hsieh-Wilson L. C. A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc* **125**, 16162-3 (2003).

65.   Wice, B. M., Trugnan, G., Pinto, M., Rousset, M., Chevalier, G., Dussaulx, E., Lacroix, B. & Zweibaum, A. The intracellular accumulation of UDP-N-acetylhexosamines is concomitant with the inability of human colon cancer cells to differentiate. *J Biol Chem* **260**, 139-46 (1985).

66.	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).

67.	Haltiwanger, R. S., Holt, G. D. & Hart, G. W. Enzymatic addition of *O*-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *J Biol Chem* **265**, 2563-8 (1990).

68.	Walgren, J. L., Vincent, T. S., Schey, K. L. & Buse, M. G. High glucose and insulin promote *O*-GlcNAc modification of proteins, including alpha-tubulin. *Am J Physiol Endocrinol Metab* **284**, E424-34 (2003).

69.	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**, 313-23 (2006).

70.	Clark, R. J., McDonough, P. M., Swanson, E., Trost, S. U., Suzuki, M., Fukuda, M. & Dillmann, W. H. Diabetes and the accompanying hyperglycemia impairs cardiomyocyte calcium cycling through increased nuclear *O*-GlcNAcylation. *J Biol Chem* **278**, 44230-7 (2003).

71.	Vosseller, K., Wells, L., Lane, M. D. & Hart, G. W. Elevated nucleocytoplasmic glycosylation by *O*-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes. *Proc Natl Acad Sci USA* **99**, 5313-8 (2002).

72.	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**, 921-30 (2004).

73.	McClain, D. A., Lubas, W. A., Cooksey, R. C., Hazel, M., Parker, G.J., Love, D. C. & Hanover, J. A. Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *Proc Natl Acad Sci USA* **99**, 10695-9 (2002).

74.	Hanover, J. A., Forsythe, M. E., Hennessey, P. T., Brodigan, T. M., Love, D. C., Ashwell, G. & Krause, M. A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. *Proc Natl Acad Sci USA* **102**, 11266-71 (2005).

75.	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).

76.	Yao, D., Taguchi, T., Matsumura, T., Pestell, R., Edelstein, D., Giardino, I., Suske, G., Ahmed, N., Thornalley, P. J., Sarthy, V. P., Hammes, H. P. & Brownlee, M.l. Methylglyoxal modification of mSin3A links glycolysis to angiopoietin-2 transcription. *Cell* **124**, 275-86 (2006).

77.	Kelly, W. G. & Hart, G. W. Glycosylation of chromosomal proteins: localization of *O*-linked N-acetylglucosamine in Drosophila chromatin. *Cell* **57**, 243-51 (1989).

78.	Yang, X., Su, K., Roos, M. D., Chang, Q., Paterson, A. J. & Kudlow, J. E. *O*-linkage of N-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability. *Proc Natl Acad Sci USA* **98**, 6611-6 (2001).

79.	Iyer, S. P. & Hart, G. W. Dynamic nuclear and cytoplasmic glycosylation: enzymes of *O*-GlcNAc cycling. *Biochemistry* **42**, 2493-9 (2003).

80. 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).

81. Cole, R. N. & Hart, G. W. Cytosolic *O*-glycosylation is abundant in nerve terminals. *J. Neurochem.* **79**, 1080-9 (2001).

82. 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).

83. Liu, K., Paterson, A. J., Zhang, F., McAndrew, J., Fukuchi, K., Wyss, J.M., Peng, L., Hu, Y. & Kudlow JE. 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).

84. Akimoto, Y., Comer, F.I., Cole, R.N., Kudo, A., Kawakami, H., Hirano, H. & Hart GW. Localization of the *O*-GlcNAc transferase and *O*-GlcNAc-modified proteins in rt cerebellar cortex. *Brain Res* **966**, 194-205 (2003).

85. Griffith, L. S. & Schmitz, B. *O*-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to pertubations of phosphorylation. *Eur J Biochem* **262**, 824-31 (1999).

86. Ewan C. Mcnay & Gold, P. E. Food for Thought: Fluctuations in Brain Extracellular Glucose Provide Insight Into the Mechanisms of Memory Modulation. *Behavioral and Cognitive Neuroscience Reviews* **1**, 264-280 (2002).

87. 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).

88. Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D. & Hart G. W. Dynamic *O*-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* **279**, 30133-42 (2004).

89. Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for *O*- linked N-acetylglucosamine. *Analytical Biochemistry* **293**, 169-177 (2001).

90. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Gerace, L. & Hart, G. W. Nuclear pore complex glycoproteins contain cytoplasmically disposed *O*-linked N-acetylglucosamine. *J Cell Biol* **104**, 1157-64 (1987).

91. 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**, 1106-13 (2001).

92. Link, A. J., Mock, M. L. & Tirrell, D. A. Non-canonical amino acids in protein engineering. *Curr Opin Biotechnol* **14**, 603-9 (2003).

93. Beene, D. L., Dougherty, D. A. & Lester, H. A. Unnatural amino acid mutagenesis in mapping ion channel function. *Curr Opin Neurobiol* **13**, 264-70 (2003).

94. Cornish, V. W., Benson, D. R., Altenbach, C. A., Hideg, K., Hubbell, W. L. & Schultz, P. G. Site-specific incorporation of biophysical probes into proteins. *Proc Natl Acad Sci USA* **91**, 2910-4 (1994).

95. Nilsson, B. L., Soellner, M. B. & Raines, R. T. Chemical synthesis of proteins. *Annu Rev Biophys Biomol Struct* **34**, 91-118 (2005).

96. Shin, Y., Winans, K. A., Backes, B. J., Kent, S. B. H., Ellman, J. A. & Bertozzi, C. R. Fmoc-Based Synthesis of Peptide-alphaThioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J Am Chem Soc* **121**, 11684-11689 (1999).

97. Prescher, J. A., Dube, D. H. & Bertozzi, C. R. Chemical remodelling of cell surfaces in living animals. *Nature* **430**, 873-7 (2004).

98. Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat Chem Biol* **1**, 13-21 (2005).

99. Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J. & Zhao, Y. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci USA* **101**, 12479-84 (2004).

100. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212-7 (2006).

101. Washburn, M. P., Wolters, D. & Yates, J. R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7 (2001).

102. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F. & White, F. M. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-5 (2002).

103. Oda, Y., Nagasu, T. & Chait, B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379-82 (2001).

104. Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**, 660-6 (2003).

105. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci USA* **103**, 9482-7 (2006).

106. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-86 (2002).

107. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-69 (2004).

108. Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **22**, 1139-45 (2004).

Chapter 2

Development of a New Chemoenzymatic Strategy to Study *O*-GlcNAc Glycosylation[*]

2.1 Background

Protein glycosylation is one of the most abundant posttranslational modifications and plays a fundamental role in the control of biological systems. For example, carbohydrate modifications are important for host-pathogen interactions, inflammation, development, and cancer.[2-4] Unique among carbohydrates is the covalent attachment of the monosaccharide *N*-acetyl-D-glucosamine to serine and threonine residues of proteins. The *O*-GlcNAc modification is found in all higher eukaryotic organisms from *C. elegans* to man and has been shown to be ubiquitous, inducible, and highly dynamic, suggesting a regulatory role analogous to phosphorylation. However, the regulatory nature of the modification (i.e., dynamic, low cellular abundance) also represents a central challenge in its detection and study.[5]

The first methodology applied to detect *O*-GlcNAc glycosylation relied on the exquisite specificity of the enzyme β1,4-galactosyltransferase-I (GalT).[6] *In vivo*, the GalT enzyme is a Golgi-resident type II membrane glycosyltransferase that specifically transfers galactose from UDP-galactose (UDP-gal) to the C-4 hydroxyl of terminal GlcNAc groups.[7†] In the *O*-GlcNAc field, early work by Hart and colleagues used UDP-

---

[*] Portions of this chapter were taken from N. Khidekel et al. A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc* **125**, 16162-3 (2003).[1]

[†] The GalT-I enzyme was one of the first glycosyltransferases to be biochemically characterized and is of particular interest in evolutionary biology. In the presence of its allosteric effector, α-lactalbumin (LA), expressed in mammary glands, the enzyme switches specificity from GlcNAc to the C-4 hydroxyl of glucose, forming lactose. The evolutionary development of the GalT-LA complex, also known as lactose synthase, is associated with the appearance of mammals.[7]

[3H]-galactose and a commercially available, soluble form of the GalT catalytic domain, as a means to detect O-GlcNAc modified proteins from cell lysate.[8] In a typical experiment, GlcNAc-modified proteins are labeled with the enzyme and UDP-[3H]-galactose substrate, subjected to SDS-PAGE electrophoresis and the gel is exposed to film for autoradiography. Because the GalT transfers galactose to all terminal GlcNAc residues including those projecting from long N or O-linked carbohydrate chains, it can not distinguish between these types of glycosylation and the O-GlcNAc modification.[8] Although non-O-GlcNAc-type glycosylation is very rare intracellularly several methods exist to confirm the exact carbohydrate linkage detected by the GalT assay.[9]   In particular, proteins may be treated with PNGase F (N-linked-glycopeptide-(N-acetyl-beta-D-glucosaminyl)-L-asparagine amidohydrolase), an endoglycosidase that specifically removes N-linked carbohydrates by cleaving the canonical GlcNAc from the asparagine to which it is linked. Additionally, O-linked carbohydrate linkages may be cleaved by β elimination and analyzed by chromatography with comparison to known carbohydrate standards. While the GalT technique is effective, UDP-[3H]-galactose is both very expensive and a weak β-emitter, with a significantly lower specific activity than other typically used radioisotopes such as [32]P. Given the low abundance of many O-GlcNAc proteins, and substoichiometric nature of the modification, it may take weeks or even months to detect an O-GlcNAc modified protein on film.[10]

Another early strategy to identify O-GlcNAc glycosylated proteins relied on the carbohydrate-binding protein (lectin) wheat germ agglutinin (WGA), which is specific for terminal GlcNAc sugars.[8] Although WGA has the capacity to detect O-GlcNAc, it achieves high affinity interactions with complex carbohydrates by several simultaneous

interactions in its binding pocket.[11]   In contrast, the dissociation constant for free GlcNAc and small GlcNAc-containing disaccharides is in the millimolar range,[12,13] which helps explain the low affinity and poor detection capabilities seen for many *O*-GlcNAc proteins.[1]

To circumvent these difficulties, several antibodies have been developed to detect *O*-GlcNAc proteins.[10] The first of these, HGAC85, was raised against streptococcal group A antigens and showed selectivity for *O*-GlcNAc.[14]   But according to some published reports, it recognizes other carbohydrates as well.[10] The second, the RL-2 antibody, was specifically raised against *O*-GlcNAc-modified nuclear pore proteins, and although it has been shown to selectively detect *O*-GlcNAc by Western blotting, it lacks the capability to detect many *O*-GlcNAc proteins, suggesting that peptide determinants may be necessary for detection.[15,16] Most recently, Hart and colleagues have developed a third antibody, CTD110.6, raised against a synthetic, *O*-GlcNAc-modified peptide. Although the CTD antibody is effective by Western blotting and has demonstrated some capacity to immunoprecipitate glycosylated proteins, it also fails to detect numerous *O*-GlcNAc proteins, and thus lacks universal applicability.[1] Moreover, the commercial antibody exhibits batch-to-batch variability and often fails to detect *O*-GlcNAc proteins in whole-cell lysates that migrate below 50 kDa by SDS- PAGE.[17]

In light of these challenges, we sought to design a multifaceted strategy for the rapid and sensitive detection of *O*-GlcNAc proteins that would also be amenable to identification of *O*-GlcNAc glycosylated sites, a prerequisite for understanding *O*-GlcNAc function on protein targets.   Further, we aimed for an approach that could be coupled to proteomic and quantitative proteomic strategies for the wide-scale discovery

of *O*-GlcNAc proteins and analysis of their regulation.

2.1 The Chemoenzymatic Approach Takes Advantage of a Mutant GalT and Unnatural, Synthetic Analogue of UDP-galactose

Because conventional biological techniques for detecting *O*-GlcNAc were limited, we aimed to apply the tools and strategies of chemical biology to advance the study of this PTM. In that regard, we chose to use synthetic chemistry to develop a bioorthogonal UDP-galactose substrate analogue for the GalT enzyme described above. The ideal substrate would capitalize on the selectivity of the GalT enzyme. At the same time, it would surpass the sensitivity of the [^{3}H]-galactose assay while forgoing the cost and safety risks of working with a radioactive substrate.

Our approach allows for chemoselective installation of unnatural ketone functionality to *O*-GlcNAc modified proteins via the generation of a synthetic analogue for UDP-galactose, analogue **1** (Figure 2.1). The ketone moiety has been well-characterized in cellular systems as a neutral, yet versatile, chemical handle.[18-20] Because it is normally absent from protein side chains, the ketone serves as a unique marker to "tag" *O*-GlcNAc glycosylated proteins once it is reacted with aminooxy or hydrazide-containing molecules such as biotin. Once biotinylated, the glycoconjugates can be readily detected by chemiluminescence using streptavidin conjugated to horseradish peroxidase (HRP). The method takes advantage of the affinity of the biotin-streptavidin interaction, one of the strongest non-covalent interactions known in nature (K_d ~4 × 10^{-14} M).[21] The commercial availability of aminooxy biotin and streptavidin-HRP as well as other detection reagents such as aminooxy fluorescent dyes make this a particularly advantageous system.

Analogue **1** was designed on the basis of previous biochemical and structural studies of GalT. We chose to append the ketone functionality at the C-2 position of the galactose ring because GalT has been shown to tolerate unnatural substrates containing minor substitutions at the C-2 position, including 2-deoxy, 2-amino and 2-N-acetyl subsitituents.[22, 23] Moreover, UDP-2deoxy-Gal was transferred at rates comparable to Gal, whereas 3-, 4-, and 6-deoxy-Gal were transferred at reduced rates.[22-24]
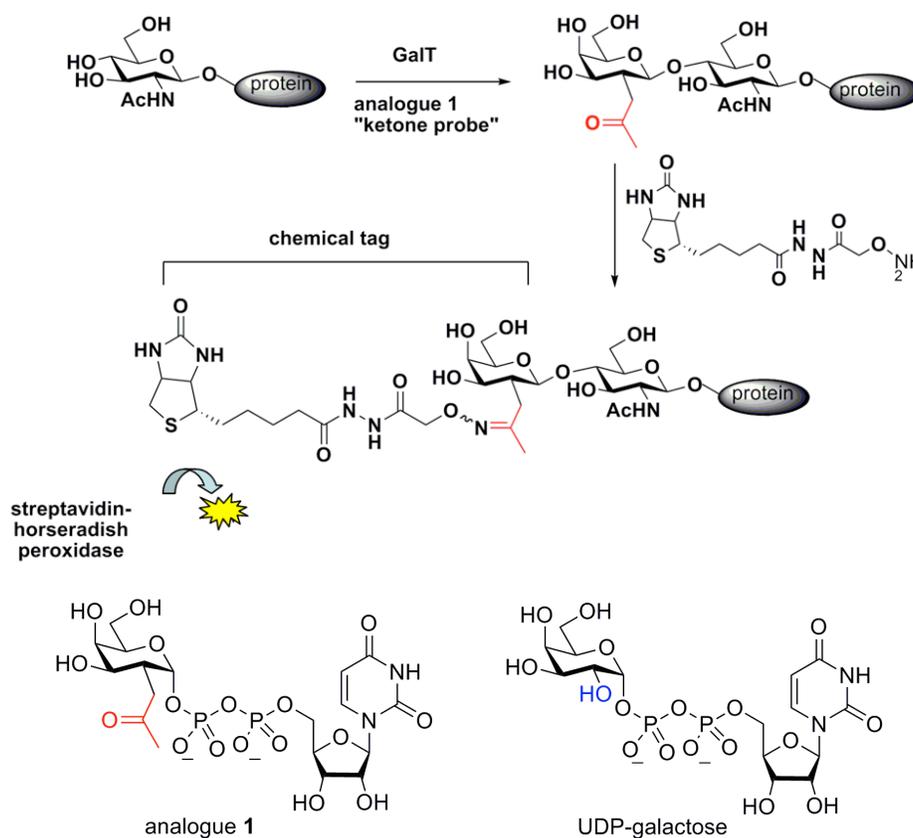
Figure 2.1. Strategy for detection of *O*-GlcNAc glycosylated proteins.   The approach takes advantage of an unnatural, synthetic analogue for UDP-galactose to enzymatically tag *O*-GlcNAc proteins for subsequent detection via biotin and streptavidin.

Analysis of the crystal structures of the GalT complexed with UDP-GalNAc revealed that the C-2 *N*-acetyl moiety is accommodated in a shallow pocket within the active site.[25] Importantly, recent efforts by Ramakrishnan and Qasba had revealed that the single

Y289L mutant enlarged the donor binding pocket of the GalT and removed a key

hydrogen bond that was shown to interact with N-acyl oxygen of the UDP-galactosamine

(UDP-GalNAc). The Y289L substitution afforded an enzyme that could transfer UDP-

GalNAc with equal efficiency to the natural substrate.[25] Because of the structural

similarity between analogue 1 and the UDP-GalNAc (Figure 2.2), we reasoned that the

mutant enzyme, which is readily expressed in milligram quantities from *E .coli* inclusion

bodies, could also be utilized with our unnatural analogue.
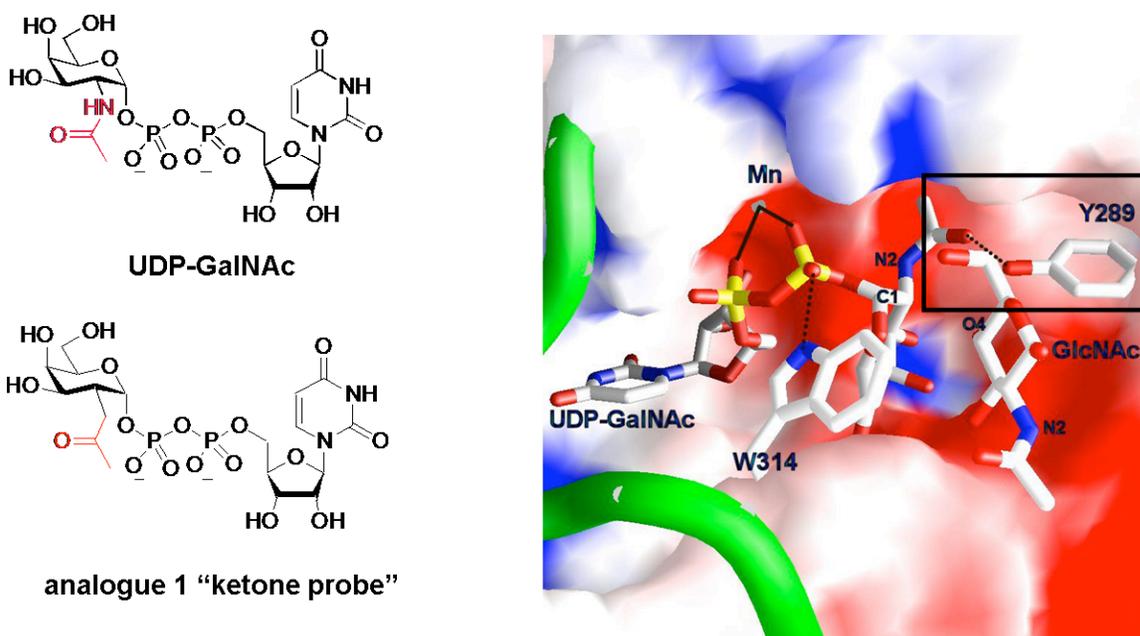
Figure 2.2. Crystal Structure of the β1,4-galactosyltransferase complexed with UDP-GalNAc reveals a hydrogen bond between Y289 and the *N*-acyl oxygen of the UDP-GalNAc. The Y289L mutant transfers UDP-GalNAc with equal efficiency to the natural substrate and may likewise transfer analogue 1 in view of their structural similarity.

Analogue 1 was initially synthesized by postdoctoral scholar Sabine Arndt, as reported

in Khidekel et al.[1] A scheme for the synthesis is reported in the experimental methods

section of this chapter. The Y289L GalT was expressed in *E. coli* as previously

described,[26] with some modification as detailed in the methods section of this chapter.

Typical yields ranged from 2.5 to 3.5 mg of active, soluble protein per liter of bacterial culture.

2.2 Quantitative Labeling of an *O*-GlcNAc-Modified Peptide

We assayed the enzymatic conversion of the UDP-keto-galactose substrate using the peptide TAPTS(*O*-GlcNAc)TIAPG, which encompasses an *O*-GlcNAc modification site within the protein CREB previously identified in our group.[27] Using wildtype GalT, only partial transfer of the keto-sugar was observed by liquid-chromatography-mass spectrometry (< 1.5 %) (Figure 2.3).
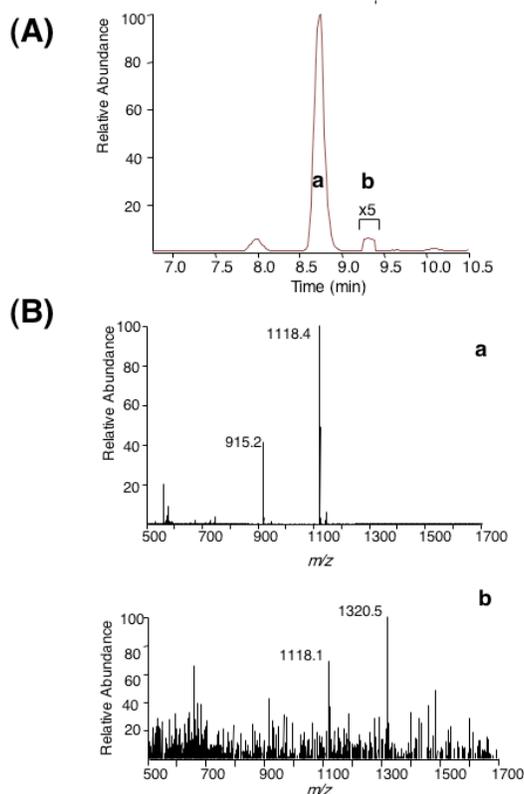
Figure 2.3. Reverse phase LC-MS analysis of *O*-GlcNAc labeling reactions with wildtype GalT. (A) Reverse phase LC-MS analysis and accompanying mass spectra of the labeling reaction 12 h after the addition of **1** and wild-type GalT. Both the starting material (a) and ketone labeled peptide product peak (b) are visible in the base peak chromatogram. The latter peak intensity has been amplified 5-fold for clarity. (B) The ESI mass spectra of peaks a and b confirm the identities of the *O*-GlcNAc glycosylated peptide, $[M_{GlcNAc}+H]^+ = 1118$ *m/z*, and the product, $[M_{ketone\text{-}GlcNAc}+H]^+ = 1320.635$ m/z and $[M_{ketone\text{-}GlcNAc}+2H]^{2+} = 661$ *m/z*, respectively.

As anticipated, however, the Y289L mutant showed greater activity and afforded complete conversion after 6 h at 4 °C (Figures 2.4 and 2.5). Subsequent reaction of the ketone-labeled peptide with the aminooxy derivative, N-(aminooxyacetyl)N 2-(D-biotinoyl) hydrazine at pH ~ 4.5 gave complete formation of the corresponding O-alkyl oxime (Figures 2.4 and 2.5).
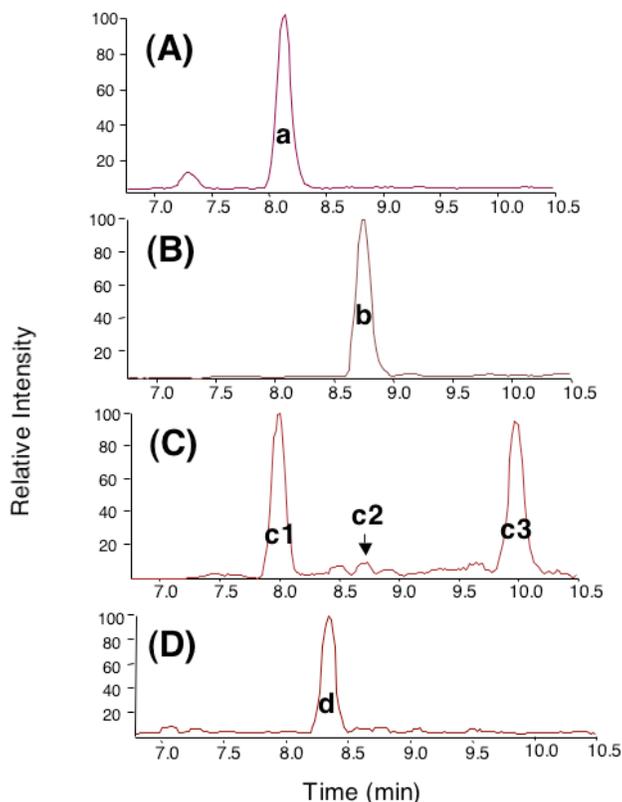
Figure 2.4. Reverse phase LC-MS analysis of O-GlcNAc peptide labeling reactions with Y289L GalT at (A) time 0, (B) 6 h after the addition of **1** and Y289L GalT, (C) 8 h after aminooxy biotin addition. Trace D shows aminooxy biotin in the absence of **1**, Y289L GalT and O-GlcNAc peptide. A and B represent base peaks chromatograms, and C and D represent the extracted ion chromatograms within the mass range 1319.0-1321.0 m/z and 1633.0-1635.5 m/z. As shown in Figure 2.5, peaks c1 and d represent the same biotin impurity. The slight difference in their retention times is due to minor differences in column equilibration time.
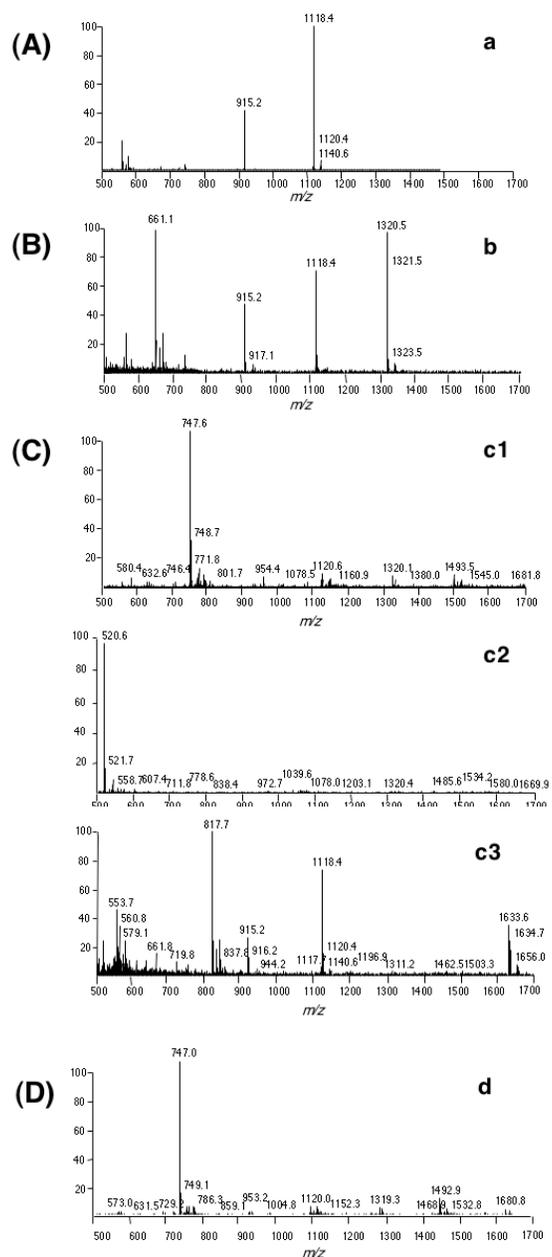
Figure 2.5. Electrospray ionization mass spectra of the LC-MS peaks in Figure 2.4. (A) Spectrum of the peptide starting material (peak a), $[M_{GlcNAc}+H]^+$ =1118.4. The fragment ion at 915.2 m/z represents the deglycosylated peptide $[M+H]^+$, which was induced during ionization in the mass spectrometer. (B) Spectrum of the ketone product (peak b), $[M_{ketone-GlcNAc}+H]^+$ =1320.5. Ions at 1118.4, 915.2, and 661.1 m/z represent the *O*-GlcNAc glycosylated peptide, the deglycosylated peptide and the doubly charged species of the ketone labeled peptide, respectively. (C) Spectra of the biotin impurity (peak c1), peak c2, and the oxime product (peak c3) The identity of the product was confirmed by ions 1633.6 and 817.7 *m/z*, which represent the singly and doubly charged species of the *O*-alkyl oxime product, respectively. The additional fragment ions at 1118.4 and 915.2 *m/z* correspond to the *O*-GlcNAc glycosylated and deglycosylated peptide, respectively. (D) Spectrum of the biotin impurity (peak d), obtained by incubating biotin in the absence of **1**, Y289L GalT and *O*-GlcNAc peptide. Note that the spectrum matches that of c1, indicating that these peaks arise from aminooxy biotin.

Several technical aspects of the methodology merit attention. First, we found the Y289L GalT to be highly tolerant of a variety of buffer and pH conditions, including MOPS and HEPES at pHs ranging from 6.7 to 7.9 (data not shown). Second, we found the oxime-forming reaction to be highly pH dependent as monitored by LC-ESI-MS (Table 2.1).

Table 2.1. Characterizing oxime formation using the O-GlcNAc-keto-galactose peptide TAPTS(O-GlcNAc)TIAPG and 6 mM aminooxy biotin as a function of reaction pH

Temp (°C)	Final pH	Conversion at 11 h (%)	Conversion at 24.5 h (%)
4	4.5	77	100
4	4.75	42	73
4	5	31	66
4	5.25	18	43
25	5	47	79

The optimal pH, near pH 4.5, is consistent with literature reports that the acidic pH maxima of oxime formation (near pH 5 for the reaction of hydroxylamine and acetone) results from the rate-limiting attack of free nitrogen base on the carbonyl compound and its rapid acid-catalyzed dehydration.[28] Finally, we also found that oxime formation was dependent on aminooxy biotin concentration at room temperature, pH 4.5 as monitored by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). (Table 2.2), which suggests an optimum concentration at or above 3 mM.

Table 2.2. Characterizing biotin-oxime formation of the O-GlcNAc-keto-galactose peptide TAPTS(O-GlcNAc)TIAPG and aminooxy biotin as a function of aminooxy biotin concentration

Aminooxy Biotin(mM)	Conversion at 25 h (%)
0.75	50-60
1.875	82-85
3	92-95
3.75	93

2.3 Rapid and Sensitive Detection of *O*-GlcNAc-Modified Proteins

Having demonstrated the quantitative labeling of a peptide, we applied our strategy to the *O*-GlcNAc glycosylated protein CREB. Previous work in the laboratory, using the traditional galactosyltransferase [³H]-labeling methodology had shown that both endogenous CREB from brain and a recombinant form expressed in *Sf9* (*Spodoptera frugiperda)* cells were *O*-GlcNAc-modified.[27] Here, the recombinant form was incubated with **1** and Y289L GalT for 12 h at 4 °C. Following reaction with aminooxy biotin, the mixture was resolved by SDS-PAGE, transferred to nitrocellulose, and probed with streptavidin-HRP. Strong labeling of CREB was observed by chemiluminescence within seconds of exposure to film (Figure 2.3). In contrast, no signal was observed over the same time period for unglycosylated CREB (from *E. coli*) or when reactions were performed in the absence of either **1** or enzyme, demonstrating the selectivity of transfer.
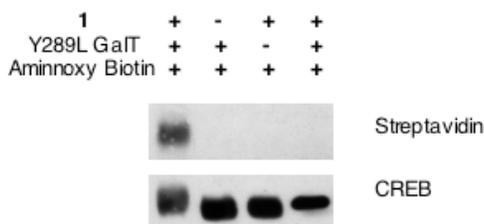
Figure 2.6. The chemoenzymatic strategy affords selective labeling of glycosylated CREB. Streptavidin-HRP signal was only observed in the presence of analogue **1**, Y289L GalT and aminooxy biotin on *O*-GlcNAc-modified CREB from *Sf9* cells (lanes 1-3). Unglycosylated CREB from *E. coli* showed no signal (lane 4).

We next explored the sensitivity of the approach suing the challenging *O*-GlcNAc-modified target αA-cyrstallin.[29] The α and β crystallins are members of a family of small heat shock proteins, which function as molecular chaperones in the eye

lens.[30]   There, the crystallins are responsible for maintaining eye lens integrity, and

defects in these proteins are associated with diseases such as childhood blindness and

cataracts.[30] Detection of the *O*-GlcNAc moiety on αA-crystallin has been reported to be

particularly difficult due to its low stoichiometry of glycosylation (<10%) and the

presence of only one major modification site.[31] Indeed, we found that existing methods

such as WGA lectin and the *O*-GlcNAc-specific antibodies RL-2 and CTD110.65 failed

to detect any *O*-GlcNAc modification on αA-crystallin, even when 5 to 10 μg was used

(Figure 2.7).
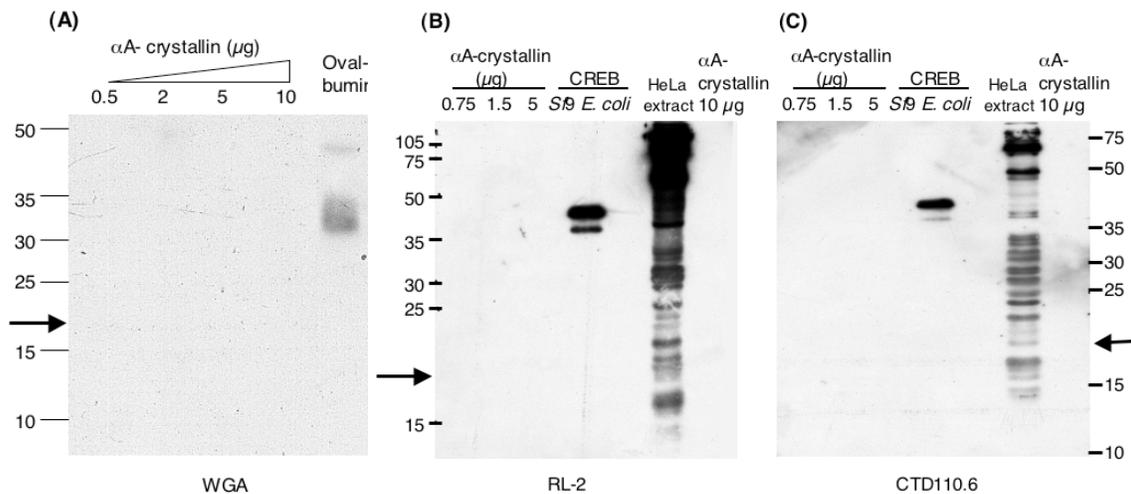
Figure 2.7. Traditional methodology for *O*-GlcNAc detection fails to detect α-crystallin by Western Blotting. (A) Western blot of α-crystallin by WGA lectin. While WGA detected the N-linked terminal GlcNAc groups of the ovalbumin positive control, it could not detect the *O*-GlcNAc moiety on α-crystallin. (B) Western blot of α-crystallin using the RL-2 antibody and (C) CTD110.6 antibody. Both antibodies effectively detected the *O*-GlcNAc present on the CREB positive control and HeLa nuclear lysates. No nonspecific binding to the unglycosylated CREB from *E. coli* was detected. However, the antibodies failed to appreciably detect the *O*-GlcNAc present on α-crystallin, even when 10 *μ*g of protein was used. The arrows mark the expected position of α-crystallin in the gel.

In contrast, our approach enabled detection of the *O*-GlcNAc modification within

minutes using 750 ng of αA-crystallin and subsequent work demonstrated that far less

protein (<50 ng) was necessary for strong detection via streptavidin-HRP.   For

comparison, tritium labeling, with wildtype GalT required 8 days of exposure to film for

comparable signal, demonstrating a significant increase in sensitivity.
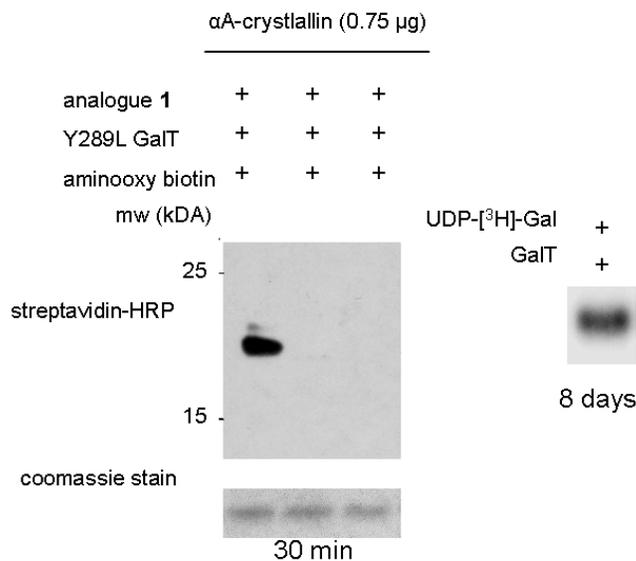
Figure 2.8. The chemoenzymatic strategy permits rapid and sensitive detection of αA-crystallin. In comparison, [³H]-labeling required 8 days of exposure to film via autoradiography for comparable signal on the same amount of protein, 0.75 μg.

2.4 Implications of the Chemoenzymatic Strategy

The greatly enhanced speed and sensitivity of our chemoenzymatic method

suggested that it would be an exciting new platform for studying *O*-GlcNAc in cellular

systems. We envisioned that the biotin tag would be useful not only for detection but

also for isolation of *O*-GlcNAc-modified proteins and peptides. Moreover, because of

the unique tolerance of the Y289L GalT enzyme, we thought that it might accept other

unnatural bioorthogonal UDP-galactose analogues, including those modified with azides,

alkynes, etc., for chemoselective ligation to *O*-GlcNAc.[32] The strategy described in this

chapter has since been patented and the Hsieh-Wilson Laboratory is investigating the

utility of a new synthetic analogue of UDP-Gal with azide functionality at the C2-

hydroxyl position (Figure 2.9), to tag *O*-GlcNAc proteins with biotin and fluorescent

alkynes.
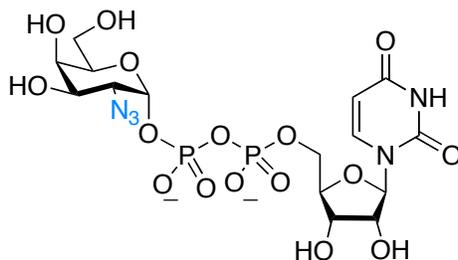
Figure 2.9. The Y289L enzyme may be able to utilize other unnatural analogues for chemoselective ligation, such as azide-functionalized UDP-gal to detect and isolate *O*-GlcNAc glycosylated proteins with alkyne or phosphine-conjugated biotin and fluorescent tags.

One of the advantages of the ketone handle is its chemical reactivity in the absence of any catalyst or cofactor, such as Cu(I) in the case of the copper-catalyzed [3+2] azide-alkyne cycloaddition 'click' chemistry.[33-35] Such cofactors, which can interact nonspecifically with proteins, may present challenges in protein detection by SDS-PAGE or during mass spectrometry analysis. However, the poor reactivity of the ketone at physiological pH makes it less ideal in certain contexts, thus making the UDP-galactose an attractive synthetic target for the installation of other functionalities. Finally, although the Y289L GalT enzyme selectivity attaches the synthetic analogue **1** to GlcNAc-containing proteins, as with the wildtype GalT, it can not distinguish between *O*-GlcNAc and terminal GlcNAc sugars emerging from long carbohydrate chains. This is on one hand advantageous, because it allows this system to be used for the detection of membrane-bound or secreted *N*-linked and *O*-linked-terminating GlcNAc glycosylated proteins as well. On the other, it prevents the selective identification of only *O*-GlcNAc modified proteins when membrane-bound proteins are assayed. This issue can be addressed by elimination of *N*-linked sugars by treatment with the endoglycosidase PNGase F (discussed in detail in Chapter 3). However, it also presents an interesting enzyme

engineering problem – GalT appears to have acceptor site preferences toward specific complex carbohydrates terminating in GlcNAc,[36] but could the GalT enzyme acceptor binding site be likewise manipulated to prefer GlcNAc directly linked to peptides versus that attached to a sugar manifold?

While our work was in preparation, Vocadlo and colleagues reported the extension of their carbohydrate tagging methodology to O-GlcNAc glycosylated proteins. In their approach, an azido-GlcNAc analogue is incubated with cell culture and converted intracellularly to a UDP-azido-GlcNAc, which can then be utilized by OGT in cells.[37] After lysis, proteins are reacted with a phosphine derivative of biotin for tagging via the Staudinger ligation, and, as in our work, are detected by Western blotting with streptavidin. Our chemoenzymatic strategy complements the metabolic labeling approach and is distinct in several key respects. First, the use of an engineered GalT and analogue **1** enables near stoichiometric labeling, resulting in higher sensitivity than possible with the native OGT.[37] Enhanced sensitivity is crucial in studying O-GlcNAc as the regulatory nature of the modification means that it is often present in low cellular abundance. Second, the use of an engineered GalT rather than the native OGT allows one to capture glycosylated species directly and avoid perturbation of metabolic pathways. This is important when studying physiological O-GlcNAc glycosylation levels as OGT's activity and substrate preference is directly effected by UDP-GlcNAc concentration.[38] In contrast, we envisioned that our approach would permit the observation of O-GlcNAc signaling pathways after cellular stimulation, an important frontier in the field.

2.5 Conclusion

We have described the development of a new chemoenzymatic strategy that detects O-GlcNAc modifications with an efficiency and sensitivity that is unrivaled by existing methods. Our approach capitalizes on an engineered galactosyltransferase, and a synthetic unnatural substrate containing a bioorthogonal ketone handle. Using this strategy, we were able to demonstrate quantitative reaction with an O-GlcNAc peptide substrate. Subsequent biotinylation of the ketone to form the biotin-oxime was also quantitative. The approach was then applied to the rapid chemiluminescent detection of the O-GlcNAc protein CREB by Western blotting via streptavidin. The new strategy was then applied to a challenging target, the low stoichiometry O-GlcNAc protein αA-crystallin, where it far surpassed the capabilities of traditional methodologies while at the same time forgoing the need for costly and dangerous radioactivity.

Given the chemical versatility of the ketone handle, we can envision a variety of applications, including affinity enrichment, and isotopic labeling for comparative proteomics. Moreover, the study of other enzymes (e.g., farnesyltransferases and other glycosyltransferases) may also benefit from this approach. The following chapters will highlight our efforts to extend the approach to the discovery of new O-GlcNAc proteins, as well as to the first proteomic and quantitative proteomic studies of O-GlcNAc in neuronal tissue.

2.6 Experimental Methods

General Methods

The peptide TAPTS(*O*-GlcNAc)TIAPG was synthesized at the Beckman Institute Biopolymer Synthesis Center using standard Fmoc chemistry. The Fmoc-protected, peracetylated *O*-GlcNAc serine amino acid was kindly synthesized by S. Tully as reported by Seitz et al., and Comer et al.[39,40,10] Baculovirus preparation and protein expression of CREB in *Spodoptera frugiperda* (Sf9) cells were performed by Dr. P. Snow at the Beckman Institute Protein Expression Facility at the California Institute of Technology.[27] HeLa cell nuclear extracts were kindly prepared by H.-C. Tai according to published procedures.[41] Wildtype GalT and Y289L were initially kindly provided by B. Ramakrishnan with subsequent expression and purification as described below. All protein concentrations were measured using the Bradford assay (Bio-Rad Laboratories, Hercules, CA).

General Reagents

Unless otherwise noted, reagents were purchased from the commercial suppliers Fisher (Fairlawn, NJ) and Sigma-Aldrich (St. Louis, MO) and were used without further purification. Protease inhibitors were purchased from Sigma-Aldrich or Alexis Biochemicals (San Diego, CA). Bovine GalT, ovalbumin, and α-crystallin were obtained from Sigma-Aldrich. Uridine diphospho-D-[6-^{3}H]galactose, Hyperfilm ECL and Amplify reagent were purchased from Amersham Biosciences (Piscataway, NJ). WGA lectin was purchased from E-Y Laboratories (San Mateo, CA). RL-2 antibody was purchased from Affinity Bioreagents (Golden, CO). Alkaline Phosphatase was purchased from New England Biolabs (Beverly, MA), and BSA was obtained from Fisher (Fairlawn, NJ).

SuperSignal West Pico chemiluminescence reagents and secondary antibodies were from Pierce (Rockford, IL), and CTD110.6 antibody was from Covance Research Products (Berkeley, CA). Nitrocellulose was from Schleicher and Schuell (Keene, NH), and PVDF was from Millipore (Bedford, MA). Complete protease inhibitor cocktail tablets were from Roche (Indianapolis, IN). *N*-(aminooxyacetyl)-*N'*-(D-biotinoyl) hydrazine (Aminooxy biotin) was obtained from two sources, Molecular Probes (Eugene, OR) and Dojindo (Gaithersburg, MD). The former material was in the form of a TFA salt and had an effect on final buffer pH in those reactions in which it was used. The latter had no effect on reaction pH.

Uridine 5'-diphosph*O*-2-acetonyl-2-deoxy-α-D-galactopyranose Diammonium Salt (1)

Analogue **1** was synthesized by Dr. Sabine Arndt essentially as described in Khidekel et al (Scheme 2.1). [1]
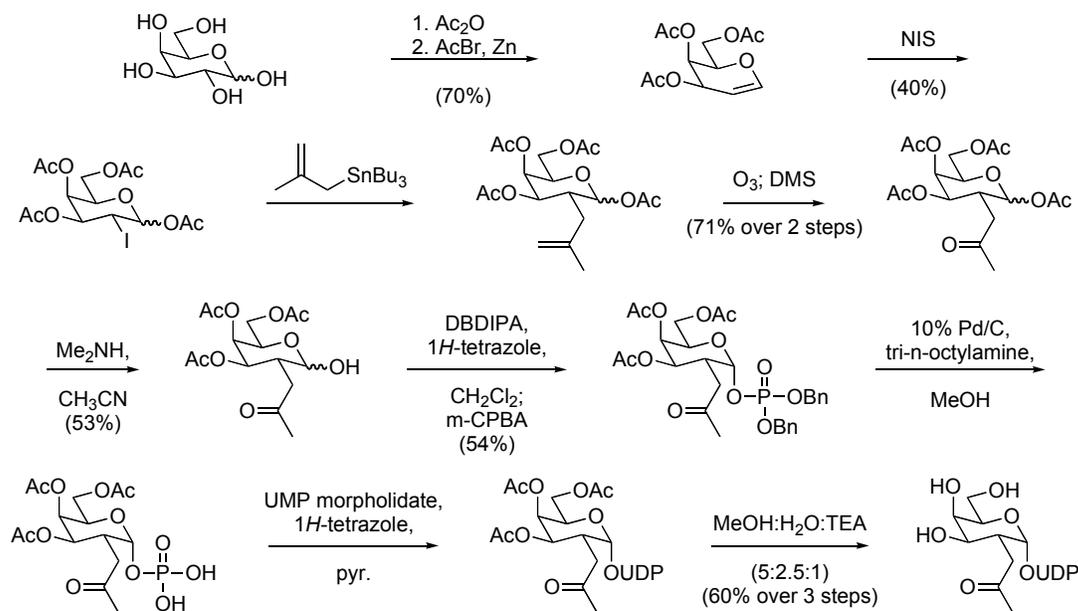
Scheme 2.1. Synthesis of analogue **1** with modifications 5/2006.

Expression, Purification and Activity Assay of Y289L GalT

Y289L cDNA, cloned into the prokaryotic expression vector pET23a, (Novagen, Madison, WI),[25] was kindly provided by B. Ramakrishnan and amplified in electrocompetent DH5α cells using the Qiagen Plasmid Maxi Kit (Qiagen, Valencia, CA) to a final DNA concentration of 5.25 µg/mL. To generate the bacterial protein, electrocompetent BL21(DE3) cells were electroporated and grown in Luria-Bertani (LB) media supplemented with 100 mg/L ampicillin. Typically, a 100 mL culture containing 100 µg/mL ampicillin was grown overnight and diluted to 1 liter using LB media. Growth was allowed to continue until an O.D. of 0.7 at A_{600} was attained. At this point, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein production. The culture was grown for another 4 h, and the cells were harvested by centrifugation at $600 \times g$ for 10 min. The bacterial pellet from the 1 liter culture was resuspended in 10 mL PBS (10.1 mM Na_2HPO_4, 1.76 mM KH_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) with Complete protease inhibitor tablets. The pellet was sonicated on ice, 5×30 sec, to lyse the cells and release inclusion bodies. In some cases, the bacterial pellet was frozen at –20 °C prior to lysis. Subsequently, the suspension was diluted to 80 mL using PBS and centrifuged at $14,000 \times g$ for 30 min. The inclusion body pellet was thoroughly washed four times with PBS containing 25% (*w/v*) sucrose. The protein pellet was washed one final time with 10 mM sodium phosphate buffer, pH 7.4 and resuspended in this buffer. Subsequently, inclusion bodies were precipitated out of suspension (characterized by a white, fluffy precipitant) with 80% EtOH, and lyophilized overnight. Inclusion bodies were dissolved in 10 mL of 5 M guanidine hydrochloride (Gu-HCl) and 0.3 M Na_2SO_3 at room temperature, with

Complete protease inhibitor tablets. In some cases, not all material was soluble in Gu-Hcl, in which cases resuspended inclusion bodies were briefly centrifuged to remove particulate.

To sulfonate all the free thiols in the protein molecule, 1 mL of 50 mM S-sulfonating agent, 2-nitr$O$-5-(sulfothio)-benzoate (NTSB), was added to this solution and stirred vigorously. Completion of sulfonation was judged by the color change of the solution from red to pale yellow (~1-1.5 h). (NTSB was prepared essentially as described in Thannhauser et al.,[42] by bubbling a mixture of 95% O_2/5% CO_2 through a solution of 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) dissolved in 0.3 M Na_2SO_3. The reaction was incubated in a 38 °C water bath and reaction was judged complete by a color change from deep orange to pale yellow, ~1 h. Maintaining this temperature during reaction was particularly critical for NTSB formation).

After sulfonation, the protein solution was diluted 10-fold with water in order to precipitate the sulfonated protein. The protein precipitate was collected by centrifugation at 10,000 × g for 5 min, and washed three times in water, followed by centrifugation to remove any remaining sulfonating agent. The sulfonated protein was re-dissolved in 5 M Gu-HCl to a protein concentration of 1 mg/mL, (which has an absorption of 1.9 to 2.0 OD at 275 nm).

For refolding, the protein solution was diluted 10-fold, in 9 portions over the course of 15 min in a folding solution (5 M D,L–Arginine, 50 mM Tris-HCl, 5 mM EDTA, 4 mM cysteamine and 2 mM cystamine, pH 8.0) at 4 °C, to give a final concentration of 0.5 M Gu-HCl, 0.5 M D,L–Arginine, 50 mM Tris-HCl, 5 mM EDTA, 4 mM cysteamine and 2 mM cystamine, pH 8.0. The solution was gently stirred during the

refolding process. In some cases, the protein solution was slowly refolded, over the course of 48 h at 4 °C, via addition of refolding solution by peristaltic pump. After refolding, the protein solution was dialyzed against 3 × 4 liters of 50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 4 mM cysteamine and 2 mM cystamine at 4 °C to remove Gu-HCl and arginine. The protein that precipitated during dialysis (white fluffy residue) was removed by centrifugation and the supernatant was concentrated using the Amicon concentrator (Millipore, Bedford, MA) to a final concentration of ~ 1 mg/mL and subsequently used for peptide and protein labeling experiments.

After protein concentration, enzyme activity toward analogue **1** was assayed via reaction with the peptide TAPTS(*O*-GlcNAc)TIAPG as follows: the peptide TAPTS(*O*-GlcNAc)TIAPG (10 μM) was dissolved in 10 mM HEPES buffer, pH 7.9 containing 5 mM MnCl$_2$. Ketone analogue **1** and mutant Y289L GalT were added to final concentrations of 500 μM and 20-100 ng/μL, respectively. Reactions were incubated at 4 °C for 6-12 h and stopped by acidification to a final concentration of 0.1% TFA. Extent of TAPTS(*O*-GlcNAc-keto-galactose)TIAPG formation was judged via MALDI-TOF analysis. For the analysis, peptide samples were concentrated on C18 zip tips (Millipore, Bedford, MA) and combined with the MALDI matrix (2,5-dihydroxybenzoic acid in 20% CH$_3$CN, 0.1% TFA in water). Spectra were acquired on a PerSeptive Biosystems Voyager-DE Pro at 20,000 kV in the reflector mode. Typically, 100% conversion was observed at 12 h at 4 ℃ using 20 ng/μL of enzyme.

Labeling of the *O*-GlcNAc Peptide

The peptide TAPTS(*O*-GlcNAc)TIAPG (10 μM) was dissolved in 25 mM MOPS buffer,

pH 6.7 containing 5 mM MnCl$_2$. Ketone analogue **1** and mutant Y289L GalT were added

to final concentrations of 1 mM and 100 ng/μL, respectively. Prior to enzyme addition,

an aliquot of the reaction was removed as an initial time point for LC-MS analysis.

Reactions were incubated at 4 °C for 6 h, after which an aliquot of the reaction mixture

was removed for product analysis by LC-MS. The remainder of the reaction was diluted

5-fold into PBS. *N*-(aminooxyacetyl)-*N'*-(D-biotinoyl) hydrazine (Molecular Probes,

Eugene, OR) was added to a final concentration of 12 mM. (Final pH of this reaction

was ~pH 4.5). After 8 h at 25 °C, the extent of biotin-oxime product was measured by

LC-MS. Labeling reactions with wildtype GalT were performed identically, with the

exception that reactions were incubated at 37 °C for 12 h.

Labeling of the *O*-GlcNAc Peptide to Monitor pH and Concentration Dependence of Oxime Formation

 The peptide TAPTS(*O*-GlcNAc)TIAPG (10 μM) was dissolved in 10 mM HEPES

buffer, pH 7.9 containing 5 mM MnCl$_2$. Ketone analogue **1** and mutant Y289L GalT

were added to final concentrations of 500 μM and 20-100 ng/μL, respectively. Reactions

were incubated at 4 °C for 24 h and then diluted, 3- to 5-fold, into 50 mM NaOAc,

buffered with AcOH, at the pHs described in Table 2.1 and Table 2.2 in the text. *N*-

(aminooxyacetyl)-*N'*-(D-biotinoyl) hydrazine from Dojindo (Gaithersburg, MD) was

added to a final concentrations as described in the tables. The extent of biotin-oxime

product was monitored by LC-MS for reactions measuring pH dependence and MALDI-TOF for those measuring concentration dependence.

LC-MS Monitoring of *O*-GlcNAc Peptide Labeling Reactions

Liquid chromatography and mass spectrometry (LC-MS) were performed on an LCQ Classic ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) interfaced with a Surveyor HPLC system (ThermoFinnigan, San Jose, CA) at the California Institute of Technology Chemistry Mass Spectrometry facility. Approximately 10 pmol of peptide from each labeling reaction was loaded onto a Luna column (2mm i.d. X 50mm) prepacked with 3 μm 100Å C18 RP particles. Flow rate was maintained at 190 μL/min with a gradient optimized for separation of the *O*-GlcNAc peptide from labeled products. LC buffer A consisted of 2% CH_3CN in 0.1M aqueous AcOH and buffer B consisted of 90% CH_3CN in 0.1M aqueous AcOH. The gradient consisted of 0-3 min, 2% B; 3-11 min, 2%-11%B; 11-14.5 min, 11%-27.5% B; 14.5-18min, 27.5%-100% B; 18-22min, 100% B, where the initial 5 minutes of flow were diverted to waste in order to avoid contamination of the mass spectrometer with salts. The LCQ was operated in automated mode using Xcalibur software. The electrospray voltage was 4.5 kV and the heated capillary was 200 °C. Ion injection time was set at 200 ms for full MS scan mode of operation (3 microscans per scan). The ion selection window was set at 500-1700 *m/z* for all experiments.

Labeling of CREB Protein

Recombinant O-GlcNAc glycosylated CREB was generated by coexpression of CREB with OGT in Sf9 cells as described previously.[27] 500 ng of CREB, in 20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 15% glycerol was added to 50 mM MOPS pH 6.45. $MnCl_2$ (50 mM) was added to a final concentration of 5 mM and alkaline phosphatase (2.5 mU/ μL) was added to a final concentration of 0.25 mU/μL to hydrolyze UDP produced by the GalT reaction in order to prevent product inhibition.[43] Analogue **1** and Y289L GalT were then added to final concentrations of 1 mM and 40 ng/μL, respectively. Control reactions without enzyme or analogue **1** were treated identically. Following incubation at 12 h at 4 °C, the reactions were diluted 5-fold into PBS containing protease inhibitors (5 µg/mL pepstatin, 5 µg/mL chymostatin, 20 µg/mL leupeptin, 20 µg/mL aprotinin, 20 µg/mL antipain, 0.2 mM PMSF). Aminooxy biotin (Molecular Probes, Eugene, OR) was added to a final concentration of 2 mM, and the biotinylation reactions were incubated with gentle shaking for 12 h at 37 °C. Reactions were aliquoted for analysis and stopped by boiling in SDS-PAGE loading dye. Proteins were resolved by 10% SDS-PAGE, electrophoretically transferred to nitrocellulose, and probed with streptavidin-HRP.

Nitrocellulose blots were blocked for 1 h at RT using 3% periodated-BSA[44] in PBS, rinsed once with TBS (50 mM Tris·HCl, 150 mM NaCl, pH 7.4) containing 0.05% (v/v) tween-20, and probed with streptavidin-HRP (1:2500 to 1:5000) in TBS-0.05% tween for 1 h at RT. Note that we found some variability among different batches of streptavidin. In some cases, blots were probed for 1 h with streptavidin-HRP, rinsed several times with TBS-0.05% tween, and reprobed with another aliquot of streptavidin-

HRP. After probing with streptavidin, membranes were rinsed and washed 5×10 min with TBS-0.1% tween containing 0.05% BSA. Streptavidin-HRP signal was visualized by chemiluminescence upon exposure to film. After streptavidin visualization, membranes were stripped in 5 mM Na_2HPO_4 pH 7.5, 2% SDS, and 2 mM (β-mercaptoethanol) βME, for 45 min at 60 °C, rinsed several times with dH_2O, and re-probed with anti-CREB antibody as previously described[27] with the modification that the antibody was used at a concentration of 1:400.

Labeling reactions with CREB expressed in *E. coli* were performed identically. To generate the bacterial protein, rat CREB cDNA was cloned into the prokaryotic expression vector pET23b(+) (Novagen, Madison, WI) using *HindIII* and *NdeI* restriction endonucleases. Electrocompetent BL21(DE3) cells were electroporated and grown in LB media supplemented with 100 mg/L ampicillin. Protein expression was induced with 0.3 mM IPTG. Recombinant CREB was purified using Ni-NTA agarose (Qiagen, Valencia, CA) as described previously.[27]

Labeling of α-Crystallin

Bovine lens αA-crystallin (a mixture of A and B chains) was resolved by SDS-PAGE electrophoresis and Coomassie-stained with standards in order to quantify the amount of A chain in the mixture. For reactions, 8.7 μg of α-crystallin (6.5 μg of A chain) in 20 mM HEPES pH 7.9 was added to 50 mM MOPS pH 6.45 containing 5 mM $MnCl_2$ and 0.25 mU/μL alkaline phosphatase. Analogue **1** and Y289L GalT were added to final concentrations of 1 mM and 100 ng/μL, respectively. Reactions were incubated at 4 °C for 18 h and then diluted 5-fold with PBS pH 6.7, protease inhibitors, and aminooxy

biotin (Molecular Probes, Eugene, OR) (6.5 mM final concentration). Biotinylation reactions were incubated with gentle shaking at 25 °C for 12 h. After biotinylation, reactions were aliquoted for analysis and subsequently boiled in SDS-PAGE loading dye. Proteins were resolved on 15% SDS-PAGE gels, transferred to nitrocellulose, and probed with streptavidin-HRP or stained with Coomassie Brilliant Blue. Blotting with streptavidin-HRP was performed as described above. Importantly, with larger quantities of protein, in the absence of Y89L and **1**, a background signal was observed. This background was dependent upon incubation with aminooxy biotin, as it was not observed with streptavidin-HRP blotting of keto-galactose labeled proteins. Stringent dialysis after biotinylation failed to remove the background signal, which was readily distinguished from true labeling due to its weaker intensity compared to labeled proteins.

UDP-[³H]galactose Labeling of α-Crystallin

³H-labeling was performed essentially as described.[8] Briefly, 12.5 μg of α-crystallin (6.5 μg of A chain) in 20 mM HEPES pH 7.9 was added to 10 mM HEPES pH 7.9 containing 5 mM $MnCl_2$ and protease inhibitors. UDP-[³H]-galactose was added to a final concentration of 0.03 μCi/μL, and the reaction was initiated with the addition of 25 mU bovine β1,4-galactosyltransferase, autogalacosylated as described by Roquemore et al.[8] Reactions were incubated at 37 °C for 1 h 15 min. Reactions were subsequently aliquoted for analysis and stopped by boiling with SDS-PAGE loading dye. Proteins were resolved on 15% SDS-PAGE gels, stained with Coomassie Brilliant Blue, incubated with Amplify reagent, and dried for subsequent exposure to Hyperfilm MP at –80 °C.

Western Blotting of α-crystallin using Antibodies RL-2 and CTD110.6

α-Crystallin, and appropriate positive and negative controls were resolved by 15% SDS-PAGE. All Western blotting steps were performed at RT unless otherwise noted. Western blotting with the RL-2 antibody was performed according to reported methods [46] with some changes suggested by the manufacturer to reduce background noise. α-Crystallin and controls were electrophoretically transferred to nitrocellulose blots, and the blots were blocked for 1 h in 5% BSA in high salt (250 mM NaCl) TBS-1% tween-20 (hsTBS-T). RL-2 antibody, at a concentration of 1:2000, was subsequently added in blocking buffer and blots were incubated for 1.5-2 h. Blots were then rinsed with hsTBST and washed 6X5 min. Secondary goat anti-mouse IgG antibody was applied at a concentration of 1:10,000 in hsTBS-T containing 1% BSA. After 1 h, blots were rinsed and washed as described before chemiluminescence detection on film (Figure 2.9). Western blotting with the CTD110.6 antibody was performed according to manufacturer's recommendations. Briefly, α-crystallin and controls were transferred to PVDF and washed 2 × 15 min with TBS-0.1% tween-20 (TBST). Blots were blocked in TBST containing 3% BSA for 1 h, rinsed 2× with TBST, and probed with CTD110.6 (1:2500) in blocking buffer for 1 h. Blots were then rinsed 2× with TBST and washed 2X5 min with the same buffer. Secondary goat anti-mouse IgM antibody was applied at a concentration of 1:10,000 in blocking buffer for 1 h, and blots were subsequently rinsed with TBST and washed 5 × 5 min before chemiluminescence detection on film.

WGA Lectin Blotting of α-crystallin

WGA western blotting was performed essentially as described.[46,8] Briefly, α-crystallin and controls were resolved by 15% SDS-PAGE and electrophoretically transferred to nitrocellulose. Blots were blocked for 1 h in 3% periodate-treated BSA in PBS, rinsed 2 × 15 min with PBS-0.05% tween-20 (PBST), and probed for 2 h with WGA-HRP (1:8000) in PBST. Subsequently, blots were rinsed with PBST, washed 3 × 10 min, then 3 × 20 min before chemiluminescence detection on film.

References:

1. Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., Qasba, P. K. & Hsieh-Wilson L. C. A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc* **125**, 16162-3 (2003).
2. Capila, I. & Linhardt, R. J. Heparin-protein interactions. *Angew Chem Int Ed Engl* **41**, 391-412 (2002).
3. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370-6 (2001).
4. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).
5. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the "*O*-GlcNAc code". *Sci STKE* **2005**, re13 (2005).
6. 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).
7. Ramakrishnan, B., Boeggeman, E., Ramasamy, V. & Qasba, P. K. Structure and catalytic cycle of beta-1,4-galactosyltransferase. *Curr Opin Struct Biol* **14**, 593-600 (2004).
8. 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).
9. Zachara, N. E. & Hart, G. W. Cell signaling, the essential role of *O*-GlcNAc! *Biochim Biophys Acta* **1761**, 599-617 (2006).
10. Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for *O*- linked N-acetylglucosamine. *Analytical Biochemistry* **293**, 169-177 (2001).
11. Wright, C. S. Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. *J Biol Chem* **267**, 14345-52 (1992).
12. Leickt, L., Bergstrom, M., Zopf, D. & Ohlson, S. Bioaffinity chromatography in the 10 mM range of Kd. *Anal Biochem* **253**, 135-6 (1997).
13. Ohlson, S., Bergstrom, M., Leickt, L. & Zopf, D. Weak affinity chromatography of small saccharides with immobilised wheat germ agglutinin and its application to monitoring of carbohydrate transferase activity. *Bioseparation* **7**, 101-5 (1998).
14. Turner, J. R., Tartakoff, A. M. & Greenspan, N. S. Cytologic assessment of nuclear and cytoplasmic *O*-linked N-acetylglucosamine distribution by using anti-streptococcal monoclonal antibodies. *Proc Natl Acad Sci USA* **87**, 5608-12 (1990).
15. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Gerace, L. & Hart, G. W. Nuclear pore complex glycoproteins contain cytoplasmically disposed *O*-linked N-acetylglucosamine. *J Cell Biol* **104**, 1157-64 (1987).).
16. Snow, C. M., Senior, A. & Gerace, L. Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. *J Cell Biol* **104**, 1143-56 (1987).

17. Khidekel, N. Notably, the strategy has been applied to HeLa, HIT-T15 and neuronal cells, *unpublished results* (2003).

18. Cornish, V. W., Hahn, K. M. & Schultz, P. G. Site-Specific Protein Modification Using a Ketone Handle. *J Am Chem Soc* **118**, 8150-8151 (1996).

19. Datta, D., Wang, P., Carrico, I. S., Mayo, S. L. & Tirrell, D. A. A designed phenylalanyl-tRNA synthetase variant allows efficient *in vivo* incorporation of aryl ketone functionality into proteins. *J Am Chem Soc* **124**, 5652-3 (2002).

20. Mahal, L. K., Yarema, K. J. & Bertozzi, C. R. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **276**, 1125-8 (1997).

21. Green, N. M. Avidin and streptavidin. *Methods Enzymol* **184**, 51-67 (1990).

22. Qian, X., Sujino, K., Palcic, M. M. & Ratcliffe, R. M. in *Glycochemistry: Principles, Synthesis, and Application*, New York: Marcel Dekker, P. G. Wang & C. R. Bertozzi, 535-565 (2001).

23. Wong, C.-H., Halcomb, R. L., Ichikawa, Y. & Kajimoto, T. *Angew Chem Int Ed Engl* **34**, 521-546 (1995).

24. Bulter, T., Schumacher, T., Namdjou, D. J., Gutierrez Gallego, R., Clausen, H., & Elling, L. Chemoenzymatic synthesis of biotinylated nucleotide sugars as substrates for glycosyltransferases. *Chembiochem* **2**, 884-94 (2001).

25. 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).

26. Ramakrishnan, B., Shah, P. S. & Qasba, P. K. alpha-Lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase I (beta 4Gal-T1) to transfer glucose from UDP-glucose to N-acetylglucosamine. Crystal structure of beta 4Gal-T1 x LA complex with UDP-Glc. *J Biol Chem* **276**, 37665-71 (2001).

27. 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).

28. Jencks, W. Studies on the Mechanism of Oxime and Semicarbazone Formation. *J Am Chem Soc* **81**, 475-481 (1959).

29. Roquemore, E. P., Dell, A., Morris, H.R., Panico, M., Reason, A. J., Savoy, L. A., Wistow, G. J., Zigler, J. S. Jr., Earles, B. J. & Hart, G. W. Vertebrate lens alpha-crystallins are modified by *O*-linked N-acetylglucosamine. *J Biol Chem* **267**, 555-63 (1992).

30. Sun, Y. & MacRae, T. H. The small heat shock proteins and their role in human disease. *Febs J* **272**, 2613-27 (2005).

31. 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**, 1106-13 (2001).

32. Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat Chem Biol* **1**, 13-21 (2005).

33. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004-2021 (2001).

34.	Link, A. J., Vink, M. K. S. & Tirrell, D. A. Presentation and Detection of the Azide Functionality in Bacterial Cell Surface Proteins. *J Am Chem Soc* **126**, 10598-10602 (2004).

35.	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**, 2596-9 (2002).

36.	Ramasamy, V., Ramakrishnan, B., Boeggeman, E., Ratner, D. M., Seeberger, P. H. & Qasba, P. K. Oligosaccharide preferences of beta1,4-galactosyltransferase-I: crystal structures of Met340His mutant of human beta1,4-galactosyltransferase-I with a pentasaccharide and trisaccharides of the N-glycan moiety. *J Mol Biol* **353**, 53-67 (2005).

37.	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).

38.	Iyer, S. P. N. & Hart, G. W. Dynamic nuclear and cytoplasmic glycosylation: Enzymes of *O*- GlcNAc cycling. *Biochemistry* **42**, 2493-2499 (2003).

39.	Meinjohanns, E., Meldal, M. & Bock, K. Efficient Synthesis of *O*-(2-acetamid*O*-2-deoxy-beta-D-glucopyranosyl)-Ser/Thr building blocks for SPPS of *O*-GlcNAc glycopeptides. *Tetrahedron Lett.* **36**, 9205-9208 (1995).

40.	Seitz, O. & Wong, C.-H. Chemoenzymatic Solution- and Solid-Phase Synthesis of *O*-Glycopeptides of the Mucin Domain of MAdCAM-1. A General Route to *O*-LacNAc, *O*-Silayl-LacNAc, and *O*-Sialyl-Lewis-X Peptides. *J Am Chem Soc* **119**, 8766-8776 (1997).

41.	Arts, J., Herr, I., Lansink, M., Angel, P. & Kooistra, T. Cell-type specific DNA-protein interactions at the tissue-type plasminogen activator promoter in human endothelial and HeLa cells *in vivo* and *in vitro*. *Nucleic Acids Res* **25**, 311-7 (1997).

42.	Thannhauser, T. W., Konishi, Y. & Scheraga, H. A. Sensitive quantitative analysis of disulfide bonds in polypeptides and proteins. *Anal Biochem* **138**, 181-8 (1984).

43.	Unverzagt, C., Kunz, H. & Paulson, J. C. High-efficiency synthesis of siayloligosaccharides and sialoglycopeptides. *J Am Chem Soc* **112**, 9308-9309 (1990).

44.	Glass, W. F., 2nd, Briggs, R. C. & Hnilica, L. S. Use of lectins for detection of electrophoretically separated glycoproteins transferred onto nitrocellulose sheets. *Anal Biochem* **115**, 219-24 (1981).

45.	Konrad, R. J., Janowski, K. M. & Kudlow, J. E. Glucose and streptozotocin stimulate p135 *O*-glycosylation in pancreatic islets. *Biochem Biophys Res Commun* **267**, 26-32 (2000).

46.	Freeze, H. H. in *Current Protocols in Molecular Biology*, F. M. Ausubel et al. 17.7.1-17.7.8 (John Wiley & Sons, Inc., New York, 1999).

Chapter 3

Discovery of *O*-GlcNAc-Modified Proteins from Cell Lysates and Identification of *O*-GlcNAc Glycosylated Sites via Mass Spectrometry[*]

3.1 Background and Introduction

Dynamic glycosylation of proteins by *O*-linked b-*N*-acetylglucosamine (*O*-GlcNAc) has been increasingly implicated in the regulation of cellular physiology and function.[3] Although discovered more than 20 years ago, the elucidation of *O*-GlcNAc as a posttranslational modification has been slow, a feature attributed to the lack of effective tools for its detection and study.[4] Despite recent advances,[5-7] few methods exist to rapidly assess the presence or absence of the *O*-GlcNAc modification on proteins isolated directly from cells. Many approaches involve protein purification or overexpression, procedures that are time-consuming and specific to each protein.[8,9] In addition, no methodology exists for the rapid identification of *O*-GlcNAc glycosylated sites, an important prerequisite for understanding *O*-GlcNAc function. Here, we discuss the development of a parallel strategy for the detection of the *O*-GlcNAc modification in cells.[1] As an illustration of the method, we show that the transcription factors c-Fos, c-Jun and ATF-1 are *O*-GlcNAc modified, and we identify *O*-GlcNAc on an entirely new class of proteins, transcriptional coactivators. We also discover that the methyl-CpG-binding protein (MeCP2), a critical regulator of neuronal development and gene

---

[*]Parts of this chapter taken from H.-C. Tai, N. Khidekel, S. B. Ficarro, E. C. Peters & L. C. Hsieh-Wilson. Parallel identification of *O*-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc* **126**, 10500-1 (2004).[1] and N. Khidekel, S. B. Ficarro, E. C. Peters & L. C. Hsieh-Wilson. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA* **101**, 13132-7 (2004).[2]

expression, is *O*-GlcNAc-modified. Finally, we illustrate the ability of the chemoenzymatic strategy to readily isolate *O*-GlcNAc-modified peptides from CREB and αA-crystallin, as well as identify glycosylation sites on the *O*-GlcNAc transferase (OGT) and the transcription factor ΔFosB, implicated in the physiology of addiction.

In the previous chapter, we described a chemoenzymatic method to tag purified *O*-GlcNAc proteins with a biotin moiety.[4] We envisioned exploiting this tagging chemistry toward the development of a new parallel strategy to identify *O*-GlcNAc glycosylated proteins and the identification of *O*-GlcNAc glycosylated sites. Specifically, *O*-GlcNAc modified proteins from cell lysates would be biotinylated and then selectively captured by affinity chromatography To establish whether a given protein was *O*-GlcNAc glycosylated, one would simply examine whether the protein of interest was captured. Using this approach, multiple proteins could be readily interrogated in parallel by Western blotting using antibodies selective for the proteins of interest (Figure 3.1).
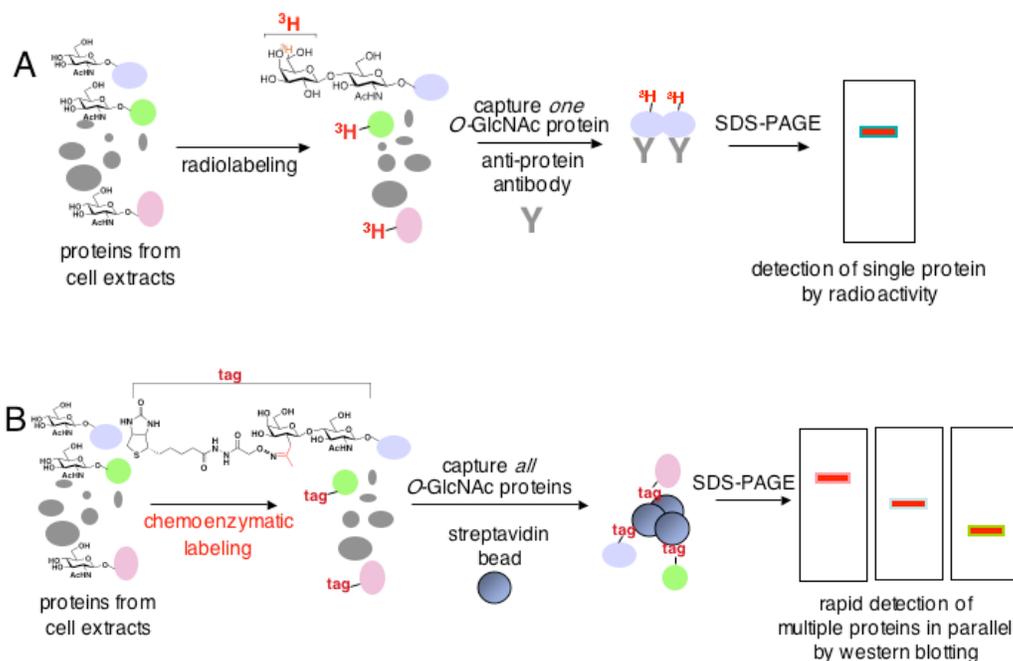
Figure 3.1.   Strategy for identifying *O*-GlcNAc proteins from cells. (A) Traditional methods for isolating and identifying *O*-GlcNAc proteins typically rely on the GalT UDP-[³H]-galactose labeling methodology described in Chapters 1 and 2, in which a single *O*-GlcNAc protein is immunoprecipitated and detected via autoradiography. Detection often requires weeks to months of exposure to film for endogenous proteins. (B) The parallel approach takes advantage of the chemoenzymatic strategy to biotinylate *O*-GlcNAc proteins for capture via streptavidin. Western blotting of captured material can be conducted for multiple proteins simultaneously and the high affinity biotin-streptavidin interaction enhances sensitivity of detection.

This approach would have several notable advantages over traditional strategies for *O*-GlcNAc detection. It would accelerate the discovery of *O*-GlcNAc proteins by circumventing the time-consuming step of overexpressing or immunoprecipitating individual proteins of interest. Virtually any protein of interest could be examined for the modification as a wide variety of antibodies are available for Western blotting. The significantly enhanced sensitivity of our tagging chemistry relative to existing methods would enable identification of even low-abundance regulatory proteins.[4]

Because selective biotinylation and isolation of *O*-GlcNAc species can also be applied to peptides, we sought to extend the chemoenzymatic strategy to the mapping of glycosylation sites. Over 100 *O*-GlcNAc proteins have been identified,[10] but the modification sites for many remain unknown. Notably, cases in which sites have been identified resulted in the discovery of important functional details regarding *O*-GlcNAc, including its role in transcriptional repression.[11,12,13] Thus, identification of glycosylated sites is an important goal in the field. Unfortunately, the *O*-GlcNAc transferase lacks a consensus sequence for glycosylation,[14] precluding facile site prediction, and as with other PTMs, analytical detection of *O*-GlcNAc-modified sites has proven challenging. The traditional strategy to map sites involves [^{3}H] labeling of *O*-GlcNAc proteins followed by protease digestion and HPLC purification/edman sequencing of radiolabeled peptides.[15] This approach lacks sensitivity and often necessitates purification of large quantities of protein and peptide. In some cases, *O*-GlcNAc glycosylated peptide have also been directly observed by mass spectrometry during collision-associated dissociation (CAD).[16,17] During CAD, protonated peptides undergo collisions with an inert gas in which the translational energy of collision is converted into vibrational energy, distributed along all covalent bonds. This mode of excitation results primarily in amide bond cleavage and is routinely used for peptide sequence identification.[18] However, CAD-induced bond cleavage proceeds along lowest energy pathways, and for *O*-GlcNAc peptides, this results primarily in loss of the labile PTM. This both precludes site identification and prevents substantial peptide bond cleavage, often yielding poor-quality peptide sequences.[19] Finally, because *O*-GlcNAc peptides and *O*-phosphate peptides are often substoichiometric, they may be difficult to detect and identify in a mixture of other

peptides, even with reverse-phase chromatographic separation prior to mass spectrometry.

In the phosphorylation field, several techniques have evolved to address these difficulties. These have included selective enrichment of phosphorylated peptides through metal affinity chromatography[20] and antibodies.[21] In addition, chemical strategies have been used that remove the phosphate linkage through β-elimination, replacing it with a stable species that includes an affinity tag such as biotin, for peptide isolation.[22] Recently, Wells and co-workers applied a similar strategy to O-GlcNAc. In their 'mild ß-elimination followed by Michael addition of dithiothreitol' (BEMAD)[7] approach, O-GlcNAc is eliminated and the resulting α,β unsaturated carbonyl is attacked with the nucleophile DTT. The resulting peptide can then be specifically enriched through thiol affinity chromatography. Importantly, BEMAD was shown to successfully enrich O-GlcNAc peptides from several purified proteins. However, the technique is inherently destructive and does not retain information as to whether the isolated peptide was originally modified by O-GlcNAc, O-phosphate or another O-linked carbohydrate. Although O-GlcNAc is more susceptible to ß-elimination than O-phosphate, extensive controls are required to determine whether a modified peptide originally contained a phosphate or a carbohydrate. Moreover, O-GlcNAc-Thr followed by Pro, was found to be difficult to eliminate, requiring conditions harsh enough to cleave phosphates and even peptide bonds.[7]

In light of these issues, we reasoned that the chemoenzymatic strategy would nicely complement existing strategies and provide unique advantages. Specifically, the ketone handle serves as a unique identifier of O-GlcNAc peptides and allows for their selective

isolation and enrichment in the presence of other unmodified, or differentially modified peptides (Figure 3.2). Because the chemoenzymatic labeling is non-destructive it also allows for detection of multiple PTMs on *O*-GlcNAc peptides. Finally, we envisioned that the approach would be amenable to the broader study of proteomics as well as quantitative analysis of site-specific changes in *O*-GlcNAc in response to cell stimuli.
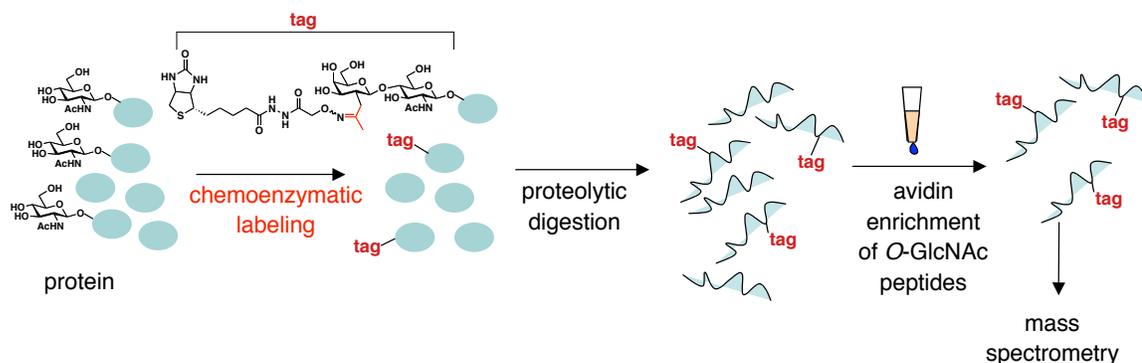
Figure 3.2. Strategy for identifying *O*-GlcNAc glycosylation sites. The approach utilizes the chemoenzymatic strategy to tag *O*-GlcNAc proteins. Subsequent protease digestion and enrichment via avidin chromatography separates *O*-GlcNAc peptides from the total peptide mixture, which improves detection sensitivity during mass spectrometry.

3.2 The Parallel Approach for Discovery of *O*-GlcNAc Proteins from Complex Mixtures

Implementation of the parallel approach required the extension of the chemoenzymatic strategy, described in Chapter 2, from purified proteins to complex mixtures. H.-C. Tai first applied the approach to cells.[1] HeLa cells were rapidly lysed under denaturing conditions to preserve the physiological glycosylation state of the proteins. The cell extract was then labeled with the UDP-ketone analogue **1** and mutant GalT for 12 h at 4 °C. We found that *N*-linked glycans could be removed simultaneously during this incubation period by treatment with PNGase F. Following reaction with

aminooxy biotin, the biotinylated *O*-GlcNAc proteins were captured with streptavidin-agarose beads, resolved by SDS-PAGE, and transferred to nitrocellulose membrane. To determine whether the captured proteins had been biotinylated, the membrane was blotted with streptavidin conjugated to horseradish peroxidase (HRP). A strong chemiluminescence signal was observed, indicating successful labeling of proteins from extracts (Figure 3.3A). Little signal was detected in the absence of either enzyme or **1**, strongly suggesting that *O*-GlcNAc modified proteins had been specifically labeled and captured.

To confirm the results, we examined whether the transcription factor CREB (cAMP-responsive element binding protein) was among the captured proteins. CREB is present in low cellular abundance and contains only two major *O*-GlcNAc glycosylation sites.[12] As such, it represents a challenging cellular target. We readily detected CREB in the captured fraction by Western blotting using an anti-CREB antibody (Figure 3.3B). In contrast, the catalytic domain of PKA (cAMP-dependent protein kinase), which has been shown to lack any detectable *O*-GlcNAc,[23] was not detected. These results demonstrate that low-abundance *O*-GlcNAc glycoproteins from cells can be selectively captured and identified.
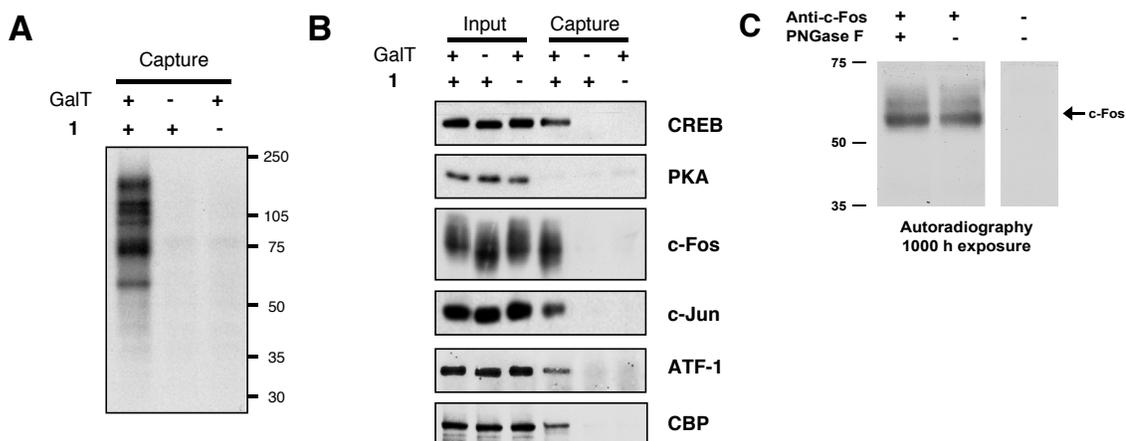
Figure 3.3. Selective isolation of chemoenzymatically tagged *O*-GlcNAc glycosylated proteins via streptavidin. (A) Captured HeLa cell lysate is only detected by streptavidin blotting in the presence of both Y289L GalT and analogue **1**. (B) Western blotting for individual proteins prior to (Input) and following (Capture) affinity capture. (C) Traditional tritium labeling and immunoprecipitation of c-Fos with the anti-c-Fos antibody required nearly 1000 hours of autoradiography exposure for detection.

H. -C. Tai next applied the approach toward the parallel identification of novel protein targets. Although the AP-1 transcription factor complex has been shown to be GlcNAc glycosylated, the specific proteins and nature of the glycosidic linkage (i.e., *N*-linked glycosylation, *O*-linked glycosylation, or direct *O*-GlcNAc) have remained unresolved.[24] Figure 3.3B shows that the two AP-1 family members c-Fos and c-Jun were captured, indicating that both proteins are *O*-GlcNAc glycosylated. As independent confirmation, we used the traditional approach of UDP-[^{3}H]galactose and GalT, followed by immunoprecipitation of c-Fos. Notably, tritium labeling required 1000 h of exposure to film for strong detection (Figure 3.3C). In contrast, our strategy permitted the detection of c-Fos within minutes.

Importantly, the approach enables study of the *O*-GlcNAc modification across structurally or functionally related protein families. ATF-1, a structural homolog and

dimerization partner of CREB,[25] shares only partial sequence identity within the region of CREB glycosylation. Nonetheless, ATF-1 was present in the captured fraction, indicating that both family members are subject to *O*-GlcNAc glycosylation in HeLa cells.

Our strategy also permitted the identification of an entirely new class of *O*-GlcNAc glycosylated proteins, histone acetyltransferases (HAT). CREB-Binding Protein (CBP) is a HAT involved in chromatin remodeling and activation of numerous transcription factors.[26] As shown in Figure 3.3B, we found that CBP is *O*-GlcNAc glycosylated. This finding is interesting in light of recent observations that *O*-GlcNAc transferase (OGT), the enzyme that catalyzes the modification, interacts with a histone deacetylase complex to promote gene silencing. Our results demonstrate that a broader set of transcriptional components are *O*-GlcNAc modified, and they support the notion that *O*-GlcNAc may serve as a general mechanism for transcriptional control.

Having demonstrated that the parallel approach could be used to identify new *O*-GlcNAc glycosylated proteins in HeLa cells, we applied the strategy to primary cultured neurons. In particular, given its role in transcriptional regulation, we investigated whether *O*-GlcNAc glycosylation affected neuron-specific proteins important for gene expression. In collaboration with Yi Sun's laboratory at UCLA, we used the parallel strategy to probe for *O*-GlcNAc glycosylation on the transcriptional repressor MeCP2. MeCP2 specifically binds methylated DNA where it helps to silence transcription by recruiting chromatin remodeling complexes such as histone deacetylases (HDACs).[27] Interestingly, MeCP2 plays a critical role in the developmental X-linked mental retardation disorder, Rett Syndrome, (RTT).[27] Mutations in both the DNA-binding and protein interaction domain of the MeCP2 gene appear to be the primary causes of RTT

and evidence suggests that changes in expression of one of the targets of MeCP2 repression, the brain derived neurotrophic factor, BDNF, may be involved in the RTT phenotype.[28,29]

   As shown in Figure 3.4, we isolated O-GlcNAc glycosylated proteins from cultured neurons using the parallel approach, and showed that MeCP2 appears to be O-GlcNAc-modified in glucosamine-treated neurons. Additional studies in the lab showed that MeCP2 is likewise O-GlcNAc-modified in the basal state (data not shown). Interestingly, depolarization appears to induce MeCP2 phosphorylation and DNA dissociation,[30] so it will be important to understand the effect of neuronal activity on glycosylation as well. Current efforts in the lab, (Appendix II), are aimed at identifying the site(s) of MeCP2 glycosylation in order to understand the significance of O-GlcNAc glycosylation for MeCP2 function.
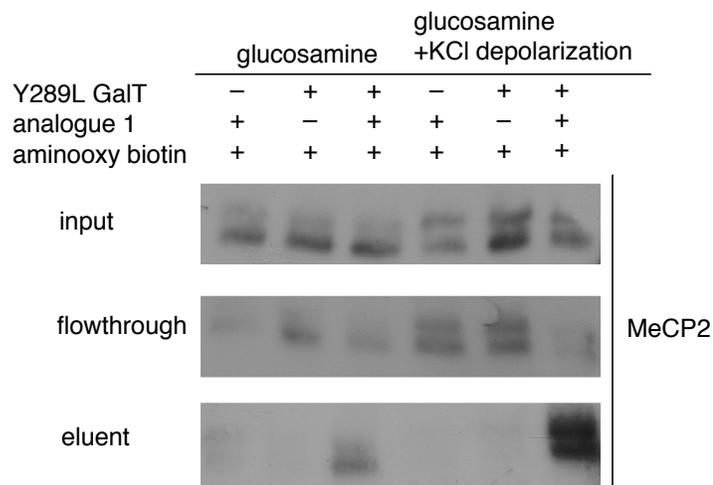
Figure 3.4. The transcriptional repressor MeCP2 is O-GlcNAc glycosylated in neurons. MeCP2 is detected in the input lysate of glucosamine-treated as well as glucosamine-treated, depolarized cortical neurons. It is specifically captured by streptavidin after chemoenzymatic labeling with Y289L GalT, analogue **1,** and the aminooxy biotin, demonstrating that it is O-GlcNAc modified (anti-MeCP2 Western blotting was conducted by Jessica Zellhoefer, in the laboratory of Yi Sun).

3.3 Identification of *O*-GlcNAc Glycosylated Peptides via the Chemoenzymatic Strategy

In order to apply the chemoenzymatic strategy to *O*-GlcNAc-modified peptides, we used an approach similar to the streptavidin protein capture approach described above. In the case of peptide capture, however, we employed monomeric avidin, which has previously been applied to biotinylated cysteine-containing peptides in the isotope-code affinity tag (ICAT) quantitative proteomic strategy.[31]   Specifically, because tetrameric streptavidin binds biotin with such high affinity, eluting biotinylated material requires harsh conditions, usually in the presence of excess biotin. These conditions are generally not compatible with mass spectrometry. In contrast, monomeric avidin, with a significantly-reduced K_d readily releases biotinylated peptides in acidic/organic commixtures, which allows for direct mass spectrometry, and reuse of the avidin resin.[32]

To demonstrate the efficacy of the approach, CREB from Sf9 cells was incubated with **1** and mutant GalT, reacted with aminooxy biotin, and digested with trypsin. Following avidin affinity chromatography, enrichment of the expected CREB glycopeptide, [256]TAPTSTIAPGVVMASSPALPTQPAEEAAR[284] was observed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Figure 3.5).
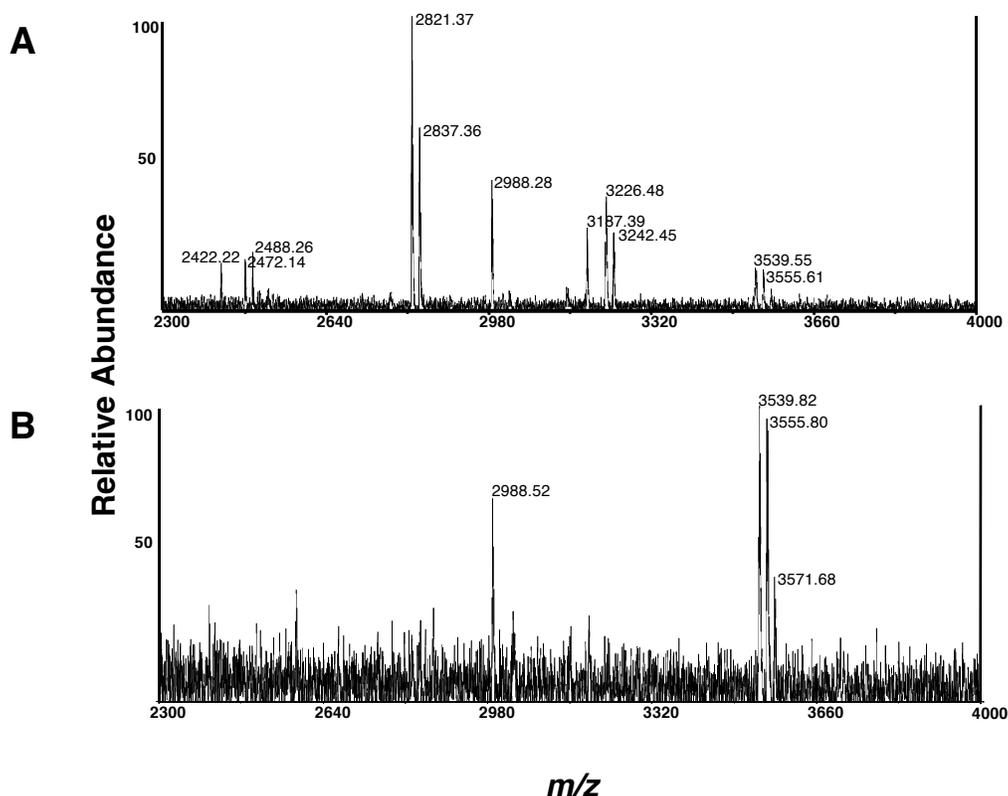
Figure 3.5. Enrichment of CREB *O*-GlcNAc peptides via the chemoenzymatic strategy. (A) MALDI-TOF spectrum of CREB tryptic peptides prior to avidin chromatography. The peak at *m/z* 3539.55 corresponds to the mass of the *O*-GlcNAc glycosylated peptide labeled with the ketone-biotin moiety. (B) MALDI-TOF spectrum of the eluent following avidin affinity capture of CREB peptides. The spectrum reveals enrichment of the labeled CREB peptide at *m/z* 3539.82 as well as two peaks at *m/z* 3555.80 and 3571.68 that correspond to oxidized forms of this peptide. The peptide at *m/z* 2988.52 displays some nonspecific interaction with the avidin column and can be readily discerned as unlabeled by LC-MS/MS.

Importantly, during liquid chromatography-tandem mass spectrometry (LC-MS/MS) of the enriched glycopeptide, the biotin-ketone moiety facilitated the identification of the *O*-GlcNAc peptide by providing a unique mass signature in the various stages of MS (Figure 3.6). Specifically, the primary fragmentation at the MS/MS stage of CAD was loss of the biotin-ketogalactose moiety, at the labile glycosidic linkage. This resulted in detection of the GlcNAc-modified peptide and unmodified peptide. In the MS^3 stage (not shown), the GlcNAc-modified peptide was specifically isolated and fragmented to generate the unmodified peptide, which could then be

sequenced in the MS^4 stage for peptide identification. Notably, the charge loss of the biotin-ketogalactose moiety and the neutral loss of the GlcNAc group are characterized by mass differences that could readily and unambiguously be used to diagnose GlcNAc-containing peptides.

$_{256}$ TAPTSTIAPGVVMASSPALPTQPAEEAAR $_{284}$
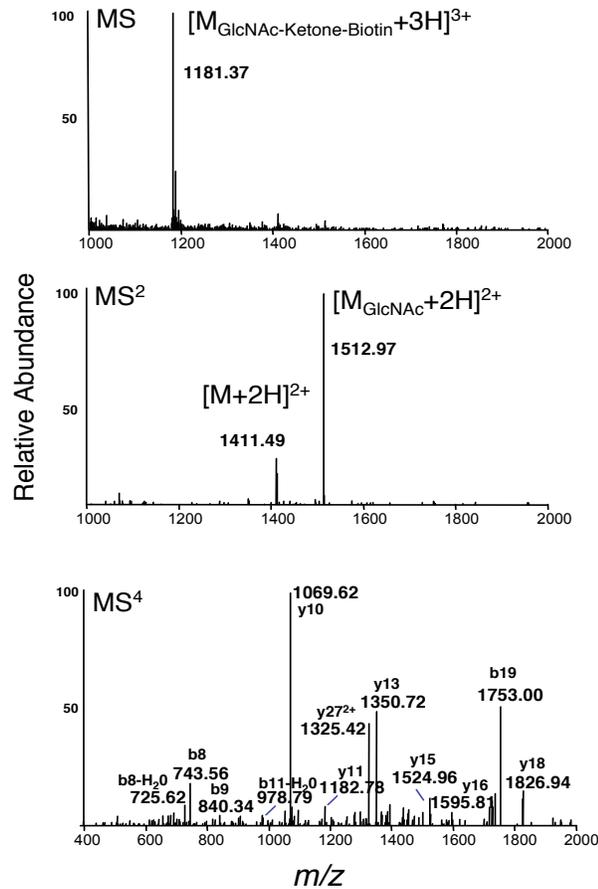
Figure 3.6. Identification of the *O*-GlcNAc modified peptide on CREB by LC-MS/MS. The *O*-GlcNAc-modified CREB peptide, [256]TAPTSTIAPGVVMASSPALPTQPAEEAAR[284] (m/z 1181.37). is observed during the MS stage. CAD revealed signature losses of the biotin-ketone moiety (m/z 1512.97) and the GlcNAc moiety (m/z 1411.49) during MS/MS. Higher-order MS analysis verified the identification of this peptide from the resultant y and b ions.

Having demonstrated the strategy with CREB, we applied it to the identification of the low-stoichiometry site on αA-crystallin.[4,17] Previous studies had demonstrated that

αA-crystallin was no more than 10% glycosylated at its major site. Thus, it had served as

a challenging target to test the sensitivity of several methods applied toward *O*-GlcNAc

peptide detection.[4,17] Here we found that the *O*-GlcNAc-modified peptide of αA-

crystallin was readily observed by LC-MS after chemoenzymatic labeling and

enrichment. In this case, the unique characteristics of the glycosylated peptide afforded

direct detection of an *O*-GlcNAc-modified fragment in the MS[3] stage of MS, allowing the

characterization of the exact modification site (Figure 3.7A, B).
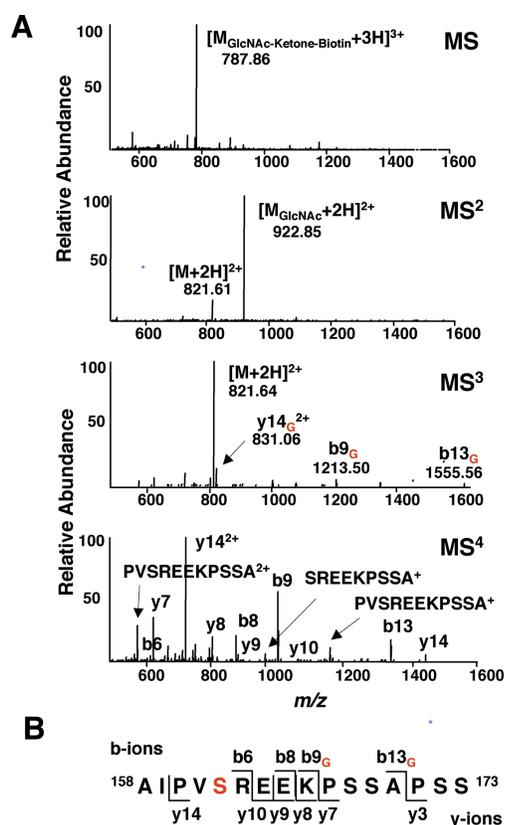
Figure 3.7. Application of the chemoenzymatic strategy toward the identification of the αA-crystallin peptide. (A) MS analysis revealed the tagged *O*-GlcNAc peptide [158]AIPVSREEKPSSAPSS[173] (*m/z* 787.86). The tag provided a diagnostic signature by tandem MS. The MS/MS spectrum of the triply charged parent ion revealed the signature loss of the ketone-biotin moiety to yield the doubly charged GlcNAc-modified peptide (*m/z* 922.85) as the predominant species. The MS[3] revealed the loss of the GlcNAc moiety to yield the unmodified peptide (*m/z* 821.64) as well as several y and b fragment ions containing the GlcNAc moiety that were used to establish the glycosylation site as Ser162. Glycosylated y and b ions are indicated with the subscript G. MS[4] analysis generated additional y and b ions as well as several internal fragment ions that were used to sequence the peptide. (B) Summary of the y and b fragment ions.

Importantly, although αA-crystallin demonstrates that *O*-GlcNAc site mapping is possible, CAD-type-fragmentation generally prevents detection of modified amino acids, because loss of the PTM dominates the fragmentation. Therefore, we envisioned that combining our chemoenzymatic strategy with chemical labeling approaches such as β-elimination (discussed in Chapter 4) or new modes of mass spectrometry fragmentation such as electron transfer dissociation (ETD),[33] (Chapter 5), would allow us to both selectively isolate *O*-GlcNAc peptides and map sites of modification.

Having established that we could identify known *O*-GlcNAc glycosylated peptides and sites via the chemoenzymatic strategy, we aimed to identify new glycosylation sites on interesting targets. We first examined *O*-GlcNAc transferase, of which both the endogenous[34] and recombinant[23,35] forms have been shown to be glycosylated. In order to identify the specific sites, we treated the short splice form of human OGT[36] from *Sf9* insect cells via the chemoenzymatic approach and analyzed the captured peptides by MALDI-TOF MS. Prior to avidin enrichment, *O*-GlcNAc-modified peptides from OGT showed low signal intensity and were difficult to detect among the mixture of other OGT peptides. Enrichment revealed several prominent peptides that corresponded to chemoenzymatically-tagged *O*-GlcNAc peptides from OGT (Figure 3.8A, B).
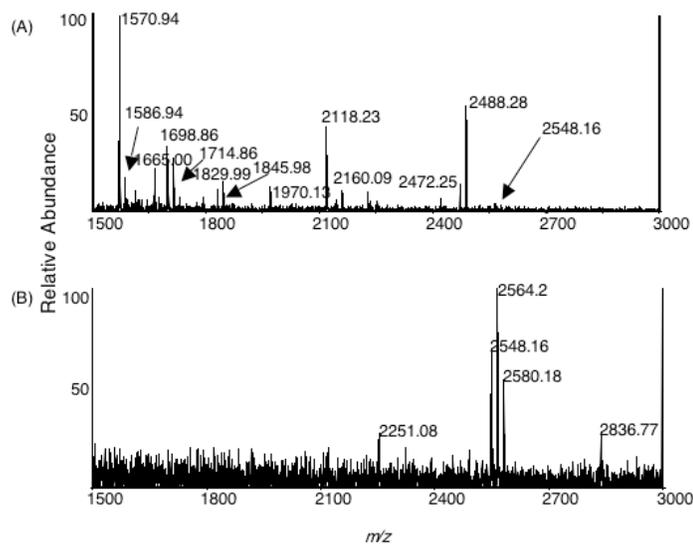
Figure 3.8. Enrichment of OGT *O*-GlcNAc peptides via the chemoenzymatic labeling strategy. (A) MALDI-TOF spectrum of OGT tryptic peptides prior to avidin chromatography reveals a number of OGT peptides while no labeled *O*-GlcNAc modified peptides are visible. The arrow with *m/z* 2548.16 the location within the spectrum that enrichment subsequently shows contains *O*-GlcNAc-modified peptides. (B) Affinity chromatography reveals enrichment of a peak at m/z 2548.16 and two oxidized forms of the same peptide. This mass corresponds to the labeled *O*-GlcNAc peptide [390]ISPTFADAYSNMGNTLK[406], whose sequence was confirmed by LC-MS/MS. The mass at m/z 2836.77 corresponds to the labeled *O*-GlcNAc form of the OGT tryptic peptide [421]AIQINPAFADAHSNLASIHK[440]. The mass at m/z 2251.08 does not correspond to theoretical OGT tryptic modified or unmodified peptides and may be a contaminant.

Subsequent LC-MS analysis identified a number of regions of glycosylation within OGT. Specifically, we found three in the tetratricopeptide repeat N-terminal domain within the peptides: [390]ISPTFADAYSNMGNTLK[406] [407]EMQDVQGALQCYTR[420] and [421]AIQINPAFADHSNLASIHK[440]. Additionally, we identified two in the C-terminal domain, [826]TIIVTTR[832] and [1037]IKPVEVTESA[1046]. Sequencing confirmed the identity of two of these peptides (Figure 3.9A,B). LC-MS/MS experiments described in Chapter 5 and Appendix II confirmed the other sequences as well.
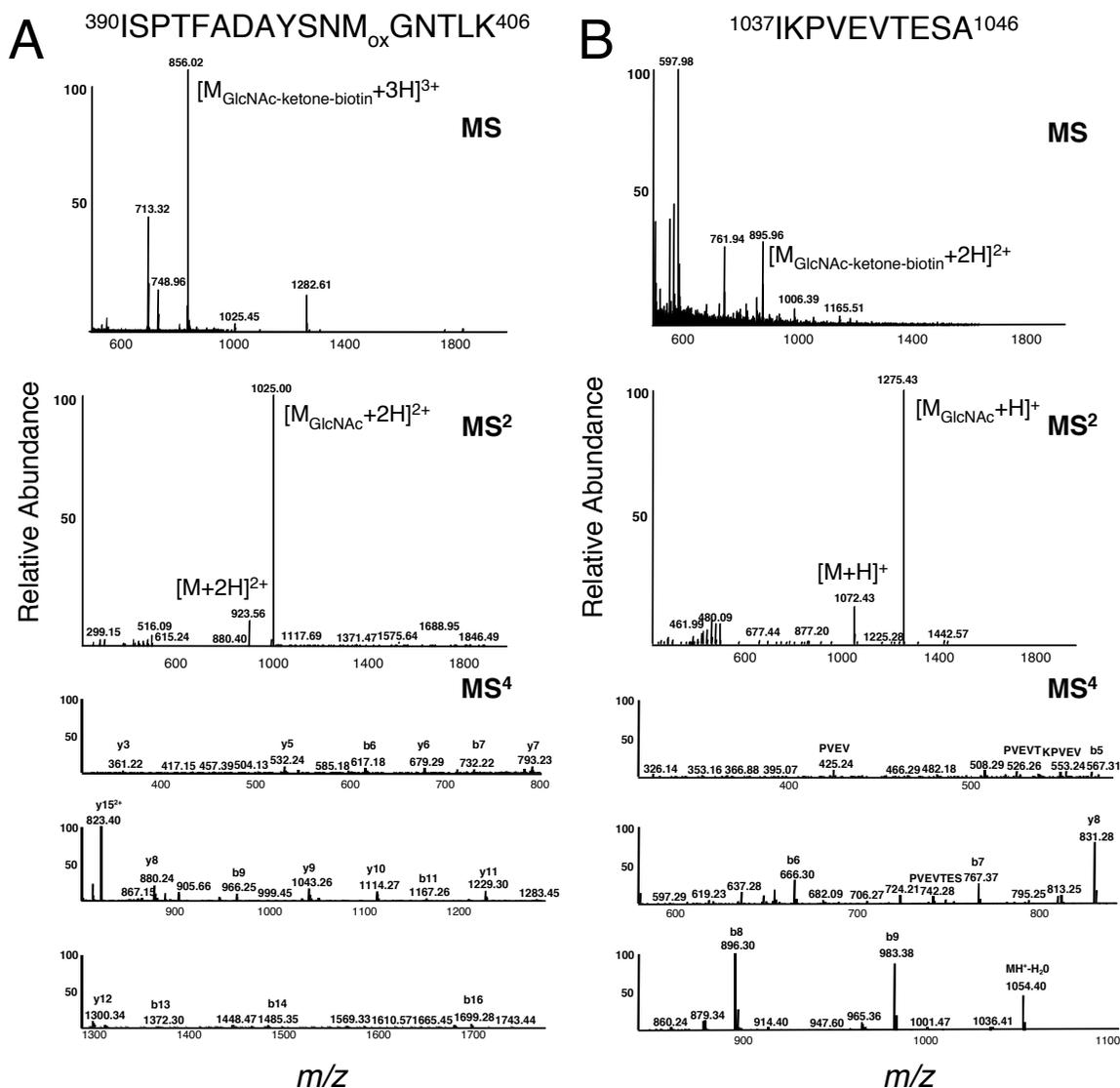
Figure 3.9. Identification of *O*-GlcNAc modified peptides on OGT by LC-MS/MS. (A) Tandem mass spectra of the labeled *O*-GlcNAc peptide [390]ISPTFADAYSNM$_{ox}$GNTLK[406] (m/z 856.02). CID revealed signature losses of the ketone-biotin moiety (m/z 1025.00) and the GlcNAc moiety (m/z 923.56). Higher-order MS analysis provided conclusive identification of this peptide from the resultant y and b ions. (B) Tandem mass spectra of the labeled *O*-GlcNAc peptide [1037]IKPVVEVTESA[1046] (m/z 895.96). CID revealed signature losses of the ketone biotin moiety (m/z 1275.43) and the GlcNAc moiety (m/z 1072.43). Higher-order MS analysis provided conclusive identification of this peptide from the resultant y and b ions as well as internal fragment ions.

As a final application of our chemoenzymatic strategy toward O-GlcNAc glycosylated peptides, we investigated the glycosylation on the transcription factor ΔFosB. Like other Fos family transcription factors, this c-terminal truncation variant of the *FosB* gene functions as a transcriptional regulator. Uniquely, however, ΔFosB protein expression is induced only upon chronic stimulus, and persists in response to chronic drug treatments. In fact, a number of studies suggest that it may mediate the molecular foundations of drug abuse, reward and addiction.[37,38] In collaboration with Eric Nestler's laboratory, Dr. H.-Y. Cheng in the lab discovered that ΔFosB was glycosylated. To gain insight into the funtional role of O-GlcNAc on ΔFosB, we tagged the recombinant protein from *Sf9* cells with the keto-galactose probe and searched for modified peptides in the LC-MS/MS of digested ΔFosB. We identified two distinct sites of glycosylation on ΔFosB, one of which is in the N-terminal transactivation domain.[39] Importantly, as with the biotin-keto-galactose moiety, the loss of the keto-galactose group and keto-galactose-GlcNAc group, provides a unique signature with which to identify tagged peptides by mass spectrometry. Although the glycosylated peptides were not specifically captured in this case, we were able to identify them by searching spectra for the distinct mass signatures (Figure 3.10).

Figure 3.10. Discovery of *O*-GlcNAc glycosylated sites on ΔFosB. (A) LC-MS/MS signature of the ketogalactose labeled *O*-GlcNAc peptide [69]AITTSQDLQW[78] from ΔFosB including summary of the b and y fragment ions identified by MS[3] (B) LC-MS/MS signature of the ketogalactose labeled *O*-GlcNAc peptide [122]STGGASGSGGPSTSTTTSGPVSAR[145] from ΔFosB including b and y fragment ions identified by MS[4].

3.4 Discussion

We have developed a parallel chemoenzymatic strategy that allows isolation and identification of *O*-GlcNAc glycosylated proteins from cell lysate. Our strategy detects low-abundance proteins containing only a few modification sites, and circumvents the need to purify individual proteins. Using this approach, H.-C. Tai in the lab, identified *O*-GlcNAc modified transcription factors such as the members of the AP-1 transcription factor complex, as well as the transcriptional coactivator/histone acetyltransferase CBP. We extended the approach to other cell types including neurons, by demonstrating glycosylation on the transcriptional repressor MeCP2.

In addition, we applied the technique to the analysis of *O*-GlcNAc peptides in order to facilitate mapping of glycosylation sites. Here, we were able to selectively isolate *O*-GlcNAc glycosylated peptides from CREB and αA-crystallin, as well as identify new sites on OGT and the transcription factor ΔFosB.

At least three splice variants of human OGT are known and detected in HeLa cells[40]: the nucleo-cytoplasmic 110kDa form, the 103kDa protein, which appears to be targeted to the mitochondria, and a short 75kDa form. Here, we examined *O*-GlcNAc glycosylation of the short recombinant form, where we identified sites of modification within the N-terminal tetratricopeptide repeat domain (TPR) and within the extreme C-terminus. The TPR motif is known to mediate protein-protein interactions.[41] In the case of OGT, the domain is necessary for OGT dimerization[35] as well as for interaction with the transcriptional repressor mSin3A, where the mSin3A-OGT complex mediates gene silencing in concert with HDACs.[42] Moreover, although the OGT catalytic domain is in its C-terminus, the TPR repeats are required for glycosylation of certain protein

substrates, including the p62 nucleoporin protein.[23,43]   Thus, in lieu of a primary consensus sequence for glycosylation, OGT may select substrates partially through larger protein-recognition motifs. The localization of O-GlcNAc sites within the TPR repeats suggests that glycosylation of OGT may serve to modulate OGT-protein interactions, possibly regulating OGT substrate interaction.   We also found two sites of glycosylation in the C-terminal domain of OGT.   Removal of the extreme C-terminus of OGT, including the peptide IKPVEVTESA on which we find glycosylation, ablates catalytic activity.[23]   Interestingly, Lubas and Hanover found that removal of all but the last three TPR repeat of OGT, as in this short transcript, reduced glycosylation of protein substrates and enhanced autocatalytic activity of the enzyme *in vitro*.[23]   Therefore, it will be important to analyze the nucleo-cytoplasmic, long-form of OGT for O-GlcNAc glycosylation as distinctions in glycosylation sites and levels of glycosylation may help define the functional differences between these two isoforms.

After we identified O-GlcNAc-modified sites on ΔFosB, Dr. H.-Y. Cheng in the lab created Thr->Ala or Ser-> Ala mutations for the dominant site of glycosylation and found that as with the transcription factors CREB and Sp1, O-GlcNAc glycosylation appeared to have a repressive affect on ΔFosB transactivation potential. Future work in the laboratory may explore the mechanisms by which O-GlcNAc affects ΔFosB transcriptional activity, perhaps by modulating protein-protein interactions with as yet undetermined N-terminal domain interacting partners. In addition, it will be important to determine if glycosylation on ΔFosB and FosB is differentially regulated in a manner that influences their strikingly different half-lives in the brain. Although the peptides on which we find glycosylation are conserved between and ΔFosB and FosB, OGT-

substrate interactions may result in differences between ΔFosB and FosB recognition and glycosylation. Alternatively, these glycosylation sites may be conserved, but FosB may also be glycosylated in its extreme c-terminus in a manner that affects its degradation, while the c-terminal truncation variant, ΔFosB, is protected from degradation. Finally, it will be important to discover if endogenous ΔFosB in the brain is likewise glycosylated, and how this glycosylation is regulated in response to activity and drugs of abuse.

One of the key advantages of the peptide tagging approach is the ability to selectively enrich glycosylated peptides from the mixture of significantly more abundant unmodified peptides, which may prevent O-GlcNAc peptide detection. Moreover, in contrast to other strategies, the unique signature produced during LC-MS^n allows us to unambiguously assign O-GlcNAc peptides, sometimes in the absence of enrichment, as seen with ΔFosB. Overall, we anticipate that the strategy will accelerate both the discovery of O-GlcNAc modified-proteins and our understanding of this modification in a physiological context.

3.5 Experimental Methods

General Reagents
Unless otherwise noted, reagents were purchased from the commercial suppliers Fisher (Fairlawn, NJ) and Sigma-Aldrich (St. Louis, MO), and were used without further purification. Protease inhibitors were purchased from Sigma-Aldrich or Alexis Biochemicals (San Diego, CA). Peptide N-glycosidase F (PNGase F) was purchased from New England Biolabs (Beverly, MA). Sequencing grade trypsin was from Promega (Madison, WI). Sequencing-grade chymotrypsin was from Roche (Indianapolis, IN). Agarose-conjugated protein A, agarose-conjugated streptavidin, SuperSignal West Pico chemiluminescence reagents, horseradish peroxidase (HRP)-conjugated streptavidin and

anti-rabbit IgG antibody were from Pierce (Rockford, IL). Nitrocellulose membrane was from Schleicher and Schuell (Keene, NH). N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazine was purchased from Dojindo (Gaithersburg, MD). CTD 110.6 anti-*O*-GlcNAc antibody was from Covance (Princeton, NJ). Mutant GalT (Y289L) was expressed and purified as described in Chapter 1. His-tagged ΔFosB from Sf9 cells was kindly provided by E. Nestler. All protein concentrations were measured using the Bradford assay (Bio-Rad Laboratories, Hercules, CA) or BCA assay, (Pierce, Rockford, IL). Preparation of HeLa Cell Extracts, Labeling and Capture of *O*-GlcNAc Modified Proteins from HeLa Cells, and Radiolabeling and Immunoprecipitation of c-Fos were all performed exactly as described in Tai et al.[1]

Preparation of Neuronal Extracts and Chemoenzymatic Labeling and Capture of *O*-GlcNAc Modified Proteins for MeCP2 analysis

Embryonic (E17) rat cortical neurons, cultured in 37 $^{\circ}$C humidified air with 5% CO_2 in basal eagle medium (Sigma) with 5% fetal bovine serum, 1 X B27 supplement (GIBCO), 2 mM L-glutamine (GIBCO), and penicillin (100 U/mL)/streptomycin (100 U/mL) (GIBCO) were provided by J. Zellhoefer from Yi Sun's laboratory at UCLA. Prior to lysis, neurons were treated with 10mM glucosamine (in 100mM HEPES pH 7.4) for 8.5 h and, in some cases, 50mM KCl for 1 h 50 min for depolarization experiments. After induction, cells from a 100 mm dish were trypsinized and pelleted. After removing media by aspiration, the pellet was resuspended in boiling lysis buffer (1% SDS and protease inhibitors, 15 µg/mL antipain, 15 µg/mL leupeptin, 7.5 µg/mL chymostatin, 7.5 µg/mL pepstatin, 1 mM phenylmethylsulfonyl fluoride, sonicated for 10 sec, and boiled

for 8 min. After centrifugation at 21,500 × g for 15 min, the supernatant was collected as denatured neuronal extract.

A small fraction of neuronal extract was saved and used for CTD110.6 anti-*O*-GlcNAc Western blotting to ensure successful glucosamine uptake (data not shown). The remainder of the neuronal extract (typically 600 µg of total protein in 80 µL) was diluted 5-fold into 10 mM HEPES pH 7.9, 2.5 mM DTT, 1.25 mM adenosine 5'-diphosphate, 1.8% NP-40, 5 mM $MnCl_2$ containing protease inhibitors (15 µg/mL antipain, 15 µg/mL leupeptin, 7.5 µg/mL chymostatin, 7.5 µg/mL pepstatin, 0.75 mM phenylmethylsulfonyl fluoride). Diluted extract was then supplemented with and 0.5 mM analogue **1** and 20 µg/mL Y289L GalT. The reaction mixture was incubated at 4 °C for 12 h, and dialyzed into buffer A (10 mM HEPES pH 7.9, 5 M urea) 3 × 3 h at 4 °C. The sample was then acidified to pH 4.8 by adding 3 M NaOAc pH 3.9 to a final concentration of 50 mM and mixed for 10 min. Aminooxy biotin derivative was added to a final concentration of 3 mM. After incubation at room temperature for 20 h, the sample was neutralized by adding 0.5 M HEPES pH 7.9 to a final concentration of 33 mM, followed by dialysis into buffer B (10 mM HEPES pH 7.9, 6 M urea, 100 mM NaCl) 2 × 15 h and 1 × 3 h at RT, and into buffer C (10 mM HEPES 7.9, 100 mM NaCl, 0.2% triton-X100) 2 × 3 h, 1 × 15 h at 4º C. Dialyzed sample was collected as labeled neuronal extract.

Labeled neuronal extract was supplemented with 1 mM DTT and protease inhibitors (10 µg/mL antipain, 10 µg/mL leupeptin, 5 µg/mL chymostatin, 5 µg/mL pepstatin, 0.5 mM phenylmethylsulfonyl fluoride), and 25 µg of material was saved as input to the streptavidin capture. The remainder of the material was precleared with sepharose 6B

beads (30 µL/100 µg of proteins) for 1 h at 4 °C. After centrifugation at 5,000 × g for 3 min, the supernatant was collected and incubated with agarose-conjugated streptavidin (30 µL/100 µg of proteins) for 2 h at 4 °C. Following centrifugation at 5,000 × g for 3 min, the supernatant was removed, and the beads were washed 3× with 8 volumes of low salt wash buffer (0.1 M Na_2HPO_4 pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) and 3× with high salt wash buffer (0.1 M Na_2HPO_4 pH 7.5, 0.5 M NaCl, 0.2% Triton X-100). After washing, the beads were boiled for 10 min in 2.5 volumes of elution buffer (50 mM Tris-HCl 6.8, 2.5% SDS, 100 mM DTT, 10% glycerol, 2 mM biotin). After centrifugation at 2,000 × g for 1 min, the supernatant was collected as the captured material.

Control reactions without analogue 1 or Y289L GalT were treated identically. After capture, Streptavidin-HRP Western blots were conducted to ensure successful labeling and capture of *O*-GlcNAc-modified proteins (data not shown). Anti-MeCP2 Western blots were conducted by J. Zellhoefer.

Labeling of CREB and *O*-GlcNAc Transferase (OGT) for Mass Spectrometry
Baculovirus preparation and protein expression were performed as described previously.[12] CREB (2 µg) or OGT (10 µg) in 20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, and 15% glycerol were supplemented with 5 mM $MnCl_2$. Analogue **1** and Y289L GalT were added to final concentrations of 750 µM and 40 ng/µL, respectively. Control reactions without enzyme or analogue **1** were treated identically except water was added in place of one of the components. Following incubation at 12 h at 4 °C, the reactions were diluted 2-fold with saturated urea. NaOAc (2.7 M) pH 3.9 was added to a

final concentration of 50 mM and a final pH of 4.8. Aminooxy biotin derivative was added to a final concentration of 5 mM, and the biotinylation reactions were incubated with gentle shaking for 20-24 h at 23 °C. Reactions were aliquoted for analysis by Western blotting or mass spectrometry and stopped by boiling in SDS-PAGE loading dye. Proteins were resolved by 10% SDS-PAGE and either electrophoretically transferred to nitrocellulose or stained with Coomassie Brilliant Blue. Western blotting with streptavidin-HRP was performed as described above to confirm successful labeling (data not shown).

In-Gel Trypsin Digestion, Avidin Enrichment and MALDI-TOF Analysis of Labeled CREB and OGT

CREB and OGT bands were excised from Coomassie-stained gels and treated essentially as described by Shevchenko et al.[44] Briefly, excised bands were destained overnight in 50% MeOH, 5% AcOH. Destained bands were dehydrated in CH_3CN, dried by vacuum, and rehydrated in 10 mM DTT in 50 mM NH_4HCO_3 pH 8.0. After 30 min reduction at room temperature, excess DTT was removed, and proteins were alkylated in 50 mM iodoacetamide in 50 mM NH_4HCO_3 pH 8.0 for 30 min at room temperature in the dark. After alkylation, excess iodoacetamide was removed and protein bands were washed in 100 mM NH_4HCO_3 pH 8.0 for 10 min, followed by two successive dehydrations in CH_3CN. Wash and dehydration steps were repeated once more, and excess CH_3CN was removed under vacuum. Protein bands were rehydrated in 15 ng/μL trypsin dissolved in 50 mM NH_4HCO_3 pH 8.0. Excess trypsin solution was removed after rehydration, and 20-30 μL of 50 mM NH_4HCO_3 pH 8.0 was then added to cover the gel slices. Proteins

were digested overnight at 37 °C. Following digestion, peptides were extracted with two successive washes of water (50 μL each) followed by 50% acetonitrile/5% formic acid in water (100 μL), and dried by vacuum centrifugation.

A small portion of each sample was saved prior to affinity chromatography for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The remainder was subjected to avidin affinity chromatography (Applied Biosystems, Foster City, CA). Chromatography was performed essentially as described by the manufacturer except that the volume of washes was doubled. Eluted peptides were partially dried by vacuum centrifugation, and a small portion of the eluted peptides was analyzed by MALDI-TOF MS. For the analysis, peptide samples were concentrated on C18 zip tips (Millipore, Bedford, MA) and combined with the MALDI matrix (2,5-dihydroxybenzoic acid in 20% CH_3CN, 0.1% TFA in water). Spectra were acquired on a PerSeptive Biosystems Voyager-DE Pro at 20,000 kV in the reflector mode.

LC-MS/MS Analysis of Avidin-Enriched CREB and OGT Peptides

Automated nanoscale liquid chromatography and tandem mass spectrometry (LC-MS/MS) were conducted using a ThermoFinnigan Surveyor HPLC and LTQ ion trap mass spectrometer along with a variation of the "vented column" approach described by Licklider et al.[45] Avidin-enriched peptides were loaded onto a 5 cm-long X 75 μm i.d. precolumn packed with 5 μm C-18 silica (Monitor 100 Å) retained by a Kaisel frit. After thorough washing, the vent was closed and the sample was transferred to a 12 cm long X 75 μm i.d. column with a pulled 5 μm tip packed with the same material. The chromatographic profile was from 100% solvent A (0.1% aqueous AcOH) to 50%

solvent B (0.1% AcOH in CH_3CN) in 30 min at approximately 200 nL/min (manual split from 300 μL/min). Additional time was allotted for column washing and reequilibration. The LTQ was operated in automated mode using Xcalibur software. The acquisition method during MS/MS analysis involved one MS precursor ion scan followed by five data-dependent MS/MS scans. Higher-order MS analyses involved an MS precursor scan followed by targeted MS^4 scans of those masses that specifically demonstrated loss of the ketone-biotin moiety and ketone-biotin-GlcNAc moiety in the MS/MS analysis. In the case of the OGT sample peptides, MS^4 data was used to search against an OGT sequence database using SEQUEST.[46] All potential peptide identifications were manually verified. In the case of the CREB sample, the acquisition method involved targeted MS/MS analysis of the presumptive ketone-biotin-GlcNAc modified peptide at *m/z* 1181.2, with simultaneous targeted MS^3 analysis of the GlcNAc modified peptide at *m/z* 1513.6 and MS^4 analysis of the unmodified peptide at *m/z* 1412.1.

The electrospray voltage was set at 1.6 kV and the heated capillary was set at 250 °C. The ion selection window was set at 500-2000 *m/z* for all experiments. For MS/MS and higher-order MS analyses, the relative collision energy for collision-induced dissociation (CID) was preset to 35% and a default charge state of +2 was selected to calculate the scan range for acquiring tandem MS spectra. The precursor ion isolation window was set at 3.5 for maximum sensitivity.

Chemoenzymatic Labeling, Biotinylation and Avidin Enrichment of α-Crystallin
Bovine lens α-crystallin (8.7 μg, Sigma-Aldrich) was incubated with analogue 1 (750 μM), and Y289L GalT in 20 mM HEPES pH 7.9 containing 5 mM $MnCl_2$ and 100 mM

NaCl for 12 h at 4 °C. The reactions were then diluted 2-fold with saturated urea, 2.7 M

NaOAc pH 3.9 (50 mM final concentration, pH 4.8) and N-(aminoxyacetyl)-N'-(D-

biotinoyl) hydrazine (5 mM final concentration, Dojindo), and incubated with gentle

shaking for 20-24 h at 23 °C. The tagged α-A-crystallin was excised from a Coomassie-

stained gel and digested with trypsin (Promega) essentially as described by Shevchenko

et al.[44]   Avidin affinity chromatography and LC-MS/MS were performed as described

above.

Chemoenzymatic Labeling and LC-MS/MS Analysis of ΔFosB

Recombinant ΔFosB from *Sf9* cells (12.5 μg) was incubated with the unnatural analogue

1 (500 μM), and Y289L GalT in 20 mM HEPES pH 7.9 containing 5 mM $MnCl_2$ and 100

mM NaCl for 12 h at 4 °C. The reactions were then diluted 3-fold with 9M urea, 2.7 M

NaOAc pH 3.9, 100% THF and 2-Aminooxy-N-(3-perfluorolhexyl)propyl-acetamide

(Fluorous Technologies, Pittsburgh PA) in 100% THF (final concentration 3.8 M urea, 50

mM NaOAc, 20% THF, 3mM 2-Aminooxy-N-(3-perfluorolhexyl)propyl-acetamide, final

pH 4.8) and incubated with gentle shaking for 20-24 h at 23 °C. Excess reagents were

removed via microcon (Millipore, Billerica, MA) by repeated dialfiltration into 5 M

urea, 50 mM NH_4HCO_3, pH 8.1. The protein solution was then reduced in 10 mM DTT

for 1 h at RT, alkylated in 40 mM iodoacetamide for 1 h at RT, and reacted with 40 mM

DTT for 1 h at RT to remove excess iodoacetamide. The protein solution was then

diluted to a final urea concentration of 2M, and trypsin (Promega) was added to a final

concentration of 0.015 μg/μL. Trypsin digestion was carried out at 37º C for 6.5 h, and

chymotrypsin (Roche, Indianapolis IN) was subsequently added to a final concentration

of 0.015 μg/μL, with 10 mM $CaCl_2$. Digestion was allowed to proceed for 6 h at 37º C, and then stopped by acidification. Fluorous solid phase extraction chromatography (as described in appendix II) failed to detect modified ΔFosB peptides, so we inspected sample input via LC-MS/MS for presence of O-GlcNAc keto-galactose modified peptides. (Later experiments suggested that the presence of ketone and aldehyde by-products in THF may have interfered with aminooxy labeling).

Automated nanoscale liquid chromatography and tandem mass spectrometry (LC-MS/MS) were conducted using a ThermoElectron Surveyor HPLC and LCQ Deca ion trap mass spectrometer along with a variation of the "vented column" approach as described for CREB and OGT above. The LCQ Deca XP was operated in automated mode using Xcalibur software. The acquisition method during MS/MS analysis involved one MS precursor ion scan followed by three data-dependent MS/MS scans. Higher-order MS analyses involved an MS precursor scan followed by targeted MS^3 scans of those masses that specifically demonstrated loss of the keto-galactose moiety and ketone-galactose-GlcNAc moiety in the MS/MS analysis. MS^3 data were used to search against an FosB sequence database using SEQUEST. All potential peptide identifications were manually verified.

References:

1.   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).
2.   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).
3.   Zachara, N. E. & Hart, G. W. Cell signaling, the essential role of *O*-GlcNAc! *Biochim Biophys Acta* (2006).
4.   Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., Qasba, P. K. & Hsieh-Wilson, L. C. A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc* **125**, 16162-3 (2003).
5.   Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for *O*- linked N-acetylglucosamine. *Anal Biochem* **293**, 169-177 (2001).
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 USA* **100**, 9116-21 (2003).
7.   Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. & Hart, G. W. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **1**, 791-804 (2002).
8.   Chou, T. Y., Dang, C. V. & Hart, G. W. Glycosylation of the C-Myc Transactivation Domain. *Proc Natl Acad Sci USA* **92**, 4417-4421 (1995).
9.   Zhang, F., Su, K., Yang, X., Bowe, D. B., Paterson, A. J. & Kudlow, J. E. *O*-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **115**, 715-25 (2003).
10.   Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D. & Hart, G. W. Dynamic *O*-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* **279**, 30133-42 (2004).
11.   Cheng, X. & Hart, G. W. Alternative *O*-glycosylation/*O*-phosphorylation of serine-16 in murine estrogen receptor beta: post-translational regulation of turnover and transactivation activity. *J Biol Chem* **276**, 10570-5 (2001).
12.   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).
13.   Yang, X., Su, K., Roos, M. D., Chang, Q., Paterson, A. J. & Kudlow, J. E. *O*-linkage of N-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability. *Proc Natl Acad Sci USA* **98**, 6611-6 (2001).

14. 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).

15. 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).

16. Haynes, P. A. & Aebersold, R. Simultaneous detection and identification of *O*-GlcNAc-modified glycoproteins using liquid chromatography-tandem mass spectrometry. *Anal Chem* **72**, 5402-5410 (2000).

17. 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**, 1106-13 (2001).

18. Steen, H. & Mann, M. The ABCs (and XYZs) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699-711 (2004).

19. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-61 (2003).

20. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F. & White, F. M. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* **20**, 301-5 (2002).

21. Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **22**, 1139-45 (2004).

22. Oda, Y., Nagasu, T. & Chait, B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* **19**, 379-82 (2001).

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

24. Jackson, S. P. & Tjian, R. *O*-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**, 125-33 (1988).

25. Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat Rev Mol Cell Biol* **2**, 599-609 (2001).

26. Vo, N. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. *J Biol Chem* **276**, 13505-8 (2001).

27. Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat Rev Genet* **7**, 415-26 (2006).

28. Chang, Q., Khare, G., Dani, V., Nelson, S. & Jaenisch, R. The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. *Neuron* **49**, 341-8 (2006).

29. Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G. & Sun, Y. E. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* **302**, 890-3 (2003).

30. Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R. & Greenberg, M. E. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* **302**, 885-9 (2003).

31.   Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. & Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-9 (1999).

32.   Green, N. M. & Toms, E. J. The properties of subunits of avidin coupled to sepharose. *Biochem J* **133**, 687-700 (1973).

33.   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).

34.   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).

35.   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).

36.   Hanover, J. A., Yu, S., Lubas, W. B., Shin, S. H., Ragano-Caracciola, M., Kochran, J. & Love, D. C. Mitochondrial and nucleocytoplasmic isoforms of *O*-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* **409**, 287-97 (2003).

37.   McClung, C. A., Ulery, P. G., Perrotti, L. I., Zachariou, V., Berton, O. & Nestler, E. J. DeltaFosB: a molecular switch for long-term adaptation in the brain. *Brain Res Mol Brain Res* **132**, 146-54 (2004).

38.   Nestler, E. J. Molecular neurobiology of addiction. *Am J Addict* **10**, 201-17 (2001).

39.   Jooss, K. U., Funk, M. & Muller, R. An autonomous N-terminal transactivation domain in Fos protein plays a crucial role in transformation. *Embo J* **13**, 1467-75 (1994).

40.   Love, D. C., Kochan, J., Cathey, R. L., Shin, S. H. & Hanover, J. A. Mitochondrial and nucleocytoplasmic targeting of *O*-linked GlcNAc transferase. *J Cell Sci* **116**, 647-54 (2003).

41.   Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A. & Conti, E. The superhelical TPR-repeat domain of *O*-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* **11**, 1001-7 (2004).

42.   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).

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

44.   Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-8 (1996).

45.   Licklider, L. J., Thoreen, C. C., Peng, J. & Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal Chem* **74**, 3076-83 (2002).

46.   Eng, J. K. M., A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).

Chapter 4

Exploring the *O*-GlcNAc Proteome:

Direct Identification of *O*-GlcNAc-Modified Proteins from the Brain[*]

4.1 Background and Introduction

Protein PTMs represent an important mechanism for the regulation of cellular physiology and function. The covalent addition of chemical groups (e.g., phosphate, acetate, carbohydrate) extends the capabilities of proteins and provides a selective and temporal means of controlling protein function.[2,3] Despite the importance of PTMs, their extent and significance are only beginning to be understood. We have been investigating *O*-GlcNAc glycosylation, the covalent attachment of β-*N*-acetylglucosamine to serine or threonine residues of protein.[4] Unlike most carbohydrate modifications, *O*-GlcNAc is dynamic and intracellular and, as such, shares common features with protein phosphorylation.[5] Over 100 proteins bearing the *O*-GlcNAc group have been identified to date, including transcription factors, cytoskeletal proteins, protein kinases, and nuclear pore proteins.[6] Recent studies have elucidated diverse roles for the *O*-GlcNAc modification, ranging from nutrient sensing to the regulation of proteasomal degradation and gene silencing.[4] Moreover, perturbations in *O*-GlcNAc levels have been associated with disease states such as cancer, Alzheimer's, and diabetes.[7,6]

Several lines of evidence suggest an important role for *O*-GlcNAc in the brain. First, activation of protein kinase A or C pathways leads to reduced levels of *O*-GlcNAc in certain protein fractions from cerebellar neurons,[8] suggesting an intriguing, dynamic

---

[*] Parts of this chapter taken from N. Khidekel, S. B. Ficarro, E. C. Peters & L. C. Hsieh-Wilson. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA* **101**, 13132-7 (2004).

interplay between the two modifications. Second, *O*-GlcNAc transferase (OGT), the enzyme that catalyzes the modification, is most abundant in the brain and pancreas.[9] Interestingly, the activity of OGT appears to be modulated by complex mechanisms, including differential splicing, interaction with regulatory partners, and regulation via PTMs.[9] In the brain OGT forms a stable complex with protein phosphatase 1 (β and γ)[10] highlighting the apparent interrelationship between *O*-GlcNAc and phosphorylation, one of the best characterized PTM's regulating brain function. Third, OGT and the corresponding beta-*N*-acetylglucosaminidase (*O*-GlcNAcase) are particularly abundant in the nerve terminal,[11] or synaptosome, where they are enriched in the presynaptic cytosol[11] and around synaptic vesicles,[12] suggesting neuron-specific functions for the modification. Finally, a critical role for *O*-GlcNAc in the brain is suggested by its presence on proteins important for neuronal function and pathogenesis such as cAMP-responsive binding protein (CREB)[13] and β-amyloid precursor protein (APP).[14]

Despite tantalizing evidence of its significance, the *O*-GlcNAc modification has been definitively linked to only a handful of proteins from the brain.[11] Efforts to identify proteins have been challenged by the difficulty of detecting the modification *in vivo*. Like many PTMs, *O*-GlcNAc is often dynamic, substoichiometric, and prevalent on low-abundance regulatory proteins. The sugar is both enzymatically and chemically labile, being subject to reversal by cellular glycosidases and cleavage on the mass spectrometer. 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 alone.

Several powerful methods have been reported for the identification of *O*-GlcNAc modified proteins. Proteins have been tritium labeled,[15] enriched using lectins or antibodies,[16,17] or chemically tagged by metabolic labeling or BEMAD (β-Elimination followed by Michael Addition with Dithiothreitol).[17,18] However, none of the existing methods is ideally suited to the direct, high-throughput identification of *O*-GlcNAc proteins from tissues or cell lysates. For instance, the tritium methodology is labor intensive and lacks sensitivity, necessitating purification of relatively large amounts of protein. Enrichment of *O*-GlcNAc proteins using antibodies has not afforded direct observation of *O*-GlcNAc glycosylated peptides and thus cannot rule out false positives.[17] Initially, metabolic labeling was shown to identify only the highly glycosylated protein p62.[18] Moreover, the method has not yet been applied to map glycosylation sites, and it may not be broadly applicable to tissues due to cellular uptake requirements. Although the BEMAD approach can be used to map sites on purified proteins and protein complexes, it is an inherently destructive technique that requires extensive controls to establish whether a peptide contains a phosphate, *O*-GlcNAc or complex *O*-linked carbohydrate group.[17]

A robust strategy for exploring the *O*-GlcNAc proteome would permit investigations into the breadth of the modification and its potential functions across various tissues and species. Direct detection of the *O*-GlcNAc moiety would enable conclusive identification of the glycoproteins and localize the modification to specific functional domains, a prerequisite for understanding the physiological role of the modification. Moreover, such an approach might also allow for quantitative comparisons of glycosylation levels in cellular or disease states.

In Chapter 2 we described a chemoenzymatic strategy for the rapid and sensitive detection of purified *O*-GlcNAc proteins.[19] In Chapter 3, the approach was applied to the discovery of individual *O*-GlcNAc proteins from cells, as well as to the isolation and identification of *O*-GlcNAc peptides and sites of modification.[20] Here, we describe extension of the approach to the first, direct, high-throughput identification of *O*-GlcNAc proteins from the mammalian brain. Using this strategy, 23 new *O*-GlcNAc modified proteins have been identified, including regulatory proteins associated with gene expression, neuronal signaling, and synaptic plasticity.[1] In addition, we have utilized the chemoenzymatic strategy to discover *O*-GlcNAc proteins specific to the synaptosome, an area enriched in the enzymes that regulate *O*-GlcNAc cycling. The synaptosomal *O*-GlcNAc proteins identified here strongly implicate *O*-GlcNAc in a role in synaptic vesicle cycling and neurotransmitter release. Overall, the diversity represented by this set of proteins provides new insight into the role of *O*-GlcNAc in neuronal function and should yield exciting targets for future study.

4.2 Proteome-Wide Identification of *O*-GlcNAc Proteins in the Brain

Previously, we described a chemoenzymatic strategy for the detection of purified *O*-GlcNAc glycosylated proteins. Our approach took advantage of an engineered β-1,4-galactosyltransferase (GalT) enzyme to transfer a ketone-containing galactose analogue selectively to the 4-position of GlcNAc on glycosylated proteins. Once transferred, the ketone functionality was reacted with an aminooxy biotin nucleophile, permitting rapid, sensitive detection of *O*-GlcNAc modified proteins by chemiluminescence. We applied this methodology to the parallel discovery of *O*-GlcNAc proteins from cell lysate via streptavidin capture and western blotting for proteins of interest. We likewise

demonstrated that this approach could be used to isolate *O*-GlcNAc modified peptides and identify sites of glycosylation from single proteins such as CREB and OGT (Chapter 3). We reasoned that this biotin tagging approach could be extended to complex mixtures, in order to enrich *O*-GlcNAc glycosylated peptides and discover *O*-GlcNAc proteins (Figure 4.1). Numerous studies have demonstrated the importance of enrichment strategies for the detection of PTMs.[21] In our case, proteins from cellular lysates would be selectively labeled with the ketone-biotin handle, digested, and glycopeptides captured using avidin affinity chromatography. Mass spectrometric analysis of the enriched glycopeptides would afford the proteome-wide identification of novel glycosylated proteins. Importantly, the approach would also permit the direct detection of modified peptides, enabling simultaneous mapping of *O*-GlcNAc to specific functional domains within a protein.
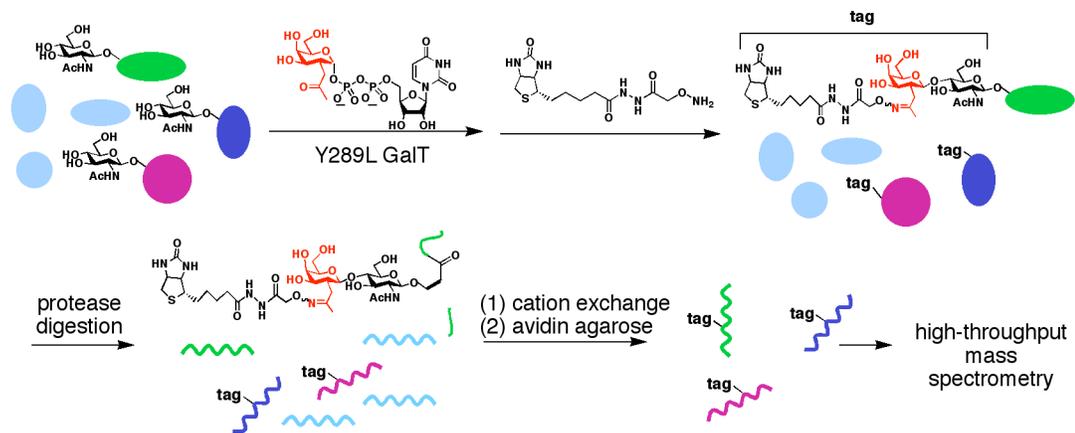
Figure 4.1. Chemoenzymatic strategy for identifying *O*-GlcNAc glycosylated peptides from complex mixtures.

Having demonstrated the selective tagging and capture of *O*-GlcNAc glycosylated peptides from purified proteins (Chapter 3), we extended the approach to the *O*-GlcNAc proteome of the mammalian brain. Rat brain lysates were separated into nuclear and S100 cytoplasmic fractions, labeled with the tag, and digested with trypsin. We also subjected a portion of the samples to proteolytic digestion with GluC to broaden the scope of analysis and generate confirmatory peptide sequences. Due to the overall complexity of the sample, the digested peptides were fractionated via strong cation exchange chromatography prior to avidin affinity chromatography.

Nearly 100 peptides containing the characteristic signature loss of the ketone-biotin tag were observed by LC-MS/MS. Figure 4.2A shows an averaged ESI spectrum of ions eluting from the LC column with retention time 17.0 to 18.1 minutes. Peaks corresponding to peptides with the diagnostic signature were subsequently selected for targeted MS^4 analysis. Notably, the vast majority of peaks in this region contained the GlcNAc-ketone-biotin moiety, demonstrating significant enrichment of this low-abundance modification. Figure 4.2B shows the MS/MS spectrum of a representative peptide (*m/z* = 789.2), indicating the characteristic loss of a ketone-biotin moiety (*m/z* = 925.5) and GlcNAc-ketone-biotin moiety (*m/z* 823.9). Higher-order MS analysis generated a definitive series of b and y ions (Fig. 4.2C), and database searching identified the peptide as belonging to the protein synaptopodin. Notably, alternative MS instrumentation and techniques such as Q-TOF[22] can be utilized to obtain sequencing information of species exhibiting the characteristic loss signature.
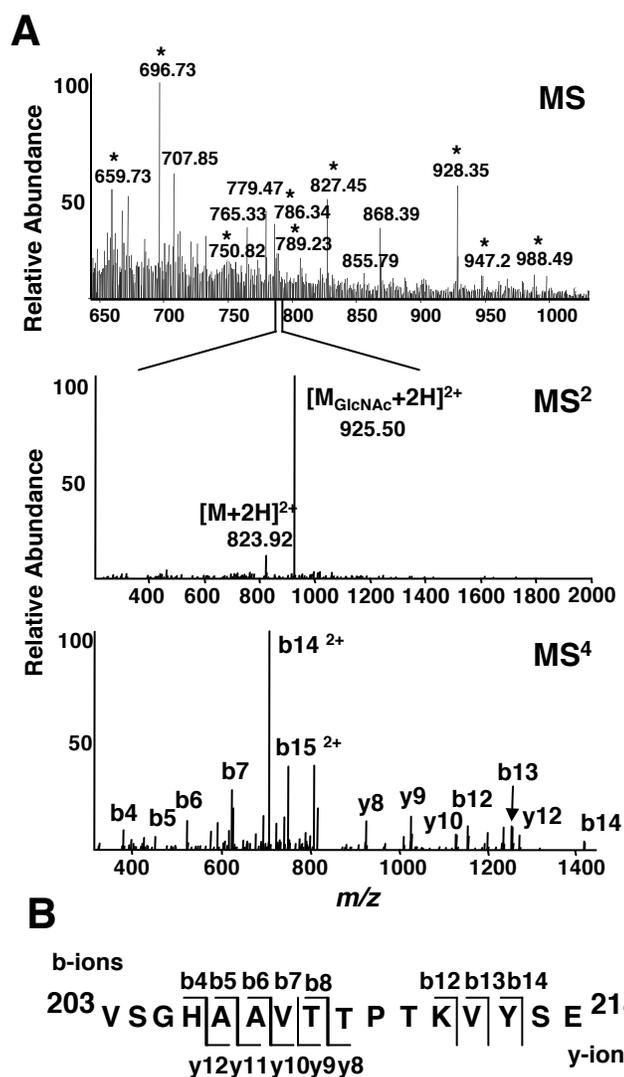
Figure 4.2. Analysis of tagged *O*-GlcNAc peptides from brain lysates. (A) Summed *m/z* spectrum of ions eluting from the LC column with retention time 17.0 to 18.1 minutes. Peaks indicated with an asterisk represent peptides that yielded the diagnostic ketone-biotin and GlcNAc-ketone-biotin loss signature upon MS/MS. (B) MS/MS spectrum of a representative peak (*m/z* = 789.23), showing loss of a ketone-biotin moiety (*m/z* = 925.50) and GlcNAc-ketone-biotin moiety (*m/z* 823.92). Fragmentation during MS[4] analysis yielded numerous b and y ions, which permitted sequencing of the peptide. (C) Prominent fragment ions used to identify the peptide as [203]VSGHAAVTTPTKVYSE[218] from synaptopodin.

Using this approach, we successfully sequenced 34 unique peptides corresponding to 25 proteins from rat brain (Table 4.1). Importantly, two of the proteins, microtubule-associated protein 2B (MAP2B) and host cell factor (HCF) have previously been reported to be O-GlcNAc glycosylated,[23,24] providing strong validation of our methodology. In addition, our results extend earlier reports by establishing distinct amino acid stretches within each protein that bear the modification. Two sites of glycosylation were identified in the N-terminal region of MAP2B. In accordance with a demonstrated interaction between the N-terminal region of HCF and both wheat germ agglutinin lectin and an anti-O-GlcNAc antibody,[24] we observed four distinct sites within three peptides in the N-terminal region of HCF. We also identified O-GlcNAc on erythrocyte protein band 4.1-like 3 within a region that shares significant sequence identity to a reported glycopeptide from human erythrocyte membrane protein band 4.1 ([1026]TITSETTSTTTTTHITK[1042] and [772](TAQ)TITSETPSSTTTTQITK[791], respectively).[25]

In addition to known proteins, our approach enabled the identification of 23 novel O-GlcNAc glycosylated proteins from the mammalian brain (Table 4.1). The proteins fall into a broad range of functional classes, including those involved in transcriptional regulation, neuronal signaling, and synaptic plasticity. Consistent with studies demonstrating that O-GlcNAc modifies transcription factors and RNA polymerase II, we identified a large number of proteins involved in transcription. In addition to low-abundance transcription factors, we found O-GlcNAc on novel transcriptional proteins such as coactivators, corepressors, and chromatin remodeling enzymes, which suggests expanded roles for O-GlcNAc in transcriptional control.

Table 4.1. *O*-GlcNAc glycosylated proteins identified from the mammalian brain

Protein	NCBI Entry	Function	Peptide Sequence	Residues
Transcriptional Regulation				
SOX2 (sex-determining region Y-box 2)	31543759‡	transcription factor	SEASSSPPVVTSSSHSR	248-264
ATF-2 (activating transcription factor 2)	13591926	transcription factor, histone acetyltransferase	AALTQQHPPVTDGDTVK	262-278
HCF (host cell factor)	34881756	transcriptional regulator, chromatin associated factor	TAAAQVGTSVSSAANTSTRPIITVHK**	620-645
HCF (host cell factor)	34881756	transcriptional regulator, chromatin associated factor	SPITIITTK	802-810
HCF (host cell factor)	34881756	transcriptional regulator, chromatin associated factor	VMSVVQTK	691-698
SRC-1 (steroid receptor coactivator-1)	34863079	transcriptional coactivator for nuclear receptors	INPSVNPGISPAHGVTR	188-204
CCR4-NOT4	34855140	global transcriptional regulator, mRNA metabolism	SNPVIPISSSNHSAR	329-343
CCR4-NOT subunit 2	34864872	global transcriptional regulator, mRNA metabolism	SLSQGTQLPSHVTPTTGVPTMSLHTPPSPSR	79-109
TLE-4 (transducin-like enhancer protein 4)	9507191	transcriptional corepressor	TDAPTPGSNSTPGLRPVPGKPPGVDPLASSLR	298-329
RNA binding motif protein 14	16307494‡	transcriptional coregulator for steroid receptors	AQPSVSLGAAYR	239-250
Nucleic Acid Binding Proteins				
NFRKB (nuclear factor related kappaB)	34862978	DNA binding protein	VPVTATQTK	896-904
zinc finger RNA binding protein	34854400	RNA binding protein	AGYSQGATQYTQAQQAR	58-74
Intracellular Transport				
Hrb (HIV-1 Rev binding protein)	34859394	RNA trafficking	APVGSVVSVPSHSSASSDK*	360-378
GRASP55 (Golgi reassembly stacking protein 2)	20301956	membrane protein transport, Golgi cisternae stacking	VPTTVEDR	423-430
Cellular Organization / Dynamics				
erythrocyte protein band 4.1-like 3	16758808	cytoskeletal protein	TITSETTSTTTTTHITK	1026-1042
erythrocyte protein band 4.1-like 4	16758808	cytoskeletal protein	TTSTTTTTHITKTVKGGISE	1031-1050
erythrocyte protein band 4.1-like 1, isoform L	11067407	cytoskeletal protein	DVLTSTYGATAETLSTSTTTHVTK	1460-1483
erythrocyte protein band 4.1-like 1, isoform L	11067407	cytoskeletal protein	TLSTSTTTHVTKTVKGGFSE	1472-1491
spectrin beta chain, (fodrin beta chain)	34879632	axonal/pre-synaptic cytoskeletal protein	HDTSASTQSTPASSR	2354-2368
MAP1B (microtubule-associated protein 1B)	19856246	axonogenesis	TTTKTTRSPDTSAYCYE	2018-2034
MAP2B (microtubule-associated protein 2B)	111965	dynamic assembly of microtubules at dendrites	SSKDEEPQKDKADKVADVPVSE	366-387
MAP2B (microtubule-associated protein 2B)	111965	dynamic assembly of microtubules at dendrites	KADKVADVPVSE	376-387
MAP2B (microtubule-associated protein 2B)	111965	dynamic assembly of microtubules at dendrites	TSSESPFPAKE	788-798
Cellular Communication / Signal Transduction				
WNK-1 (lysine deficient protein kinase)	16758634	signal transduction, ion homeostasis	DGTEVHVTASSSGAGVVK	1584-1601
WNK-1 (lysine deficient protein kinase)	16758634	signal transduction, ion homeostasis	MGGSTPISAASATSLGHFTK	2043-2062
PDZ-GEF	34857578	guaninine nucleotide exchange factor for RAP1/2	ISSRSSIVSNSSFDSVPVSLHDE	1211-1233
PDZ-GEF	34857578	guaninine nucleotide exchange factor for RAP1/2	SSFDSVPVSLHDER	1221-1234
PDZ-GEF	34857578	guaninine nucleotide exchange factor for RAP1/3	SVPVSLHDE	1225-1233
synaptopodin	11067429	dendritic spine formation	VSGHAAVTTPTKVYSE	203-218
bassoon	9506427	synaptic vesicle cycling	VTQHFAK***	1338-1444
Uncharacterized Proteins				
hypothetical protein FLJ31657	34855501	unknown	IGGDLTAAVTK	196-206
1300019H17Rik protein	34880180	unknown	EAALPSTK	286-293
KIAA1007 protein	34851212	unknown	TVTVTKPTGVSFK	1051-1063
DACA-1 homolog	34861007	unknown	IGDVTTSAVK	271-280

*The site of modification was localized to Ser372 or Ser373 using a combination of chemoenzymatic tagging and β-elimination. ** We identified two distinct sites of *O*-GlcNAc glycosylation on this peptide. ***Confirmed by peptide synthesis and MS sequencing analysis (see Supporting Information). ‡Mouse proteins identified in the NCBI database. Corresponding rat orthologs were identified in the Celera database.

Notably, our methodology also afforded the simultaneous detection of multiple PTMs. For instance, we observed an *O*-GlcNAc modified peptide with a characteristic loss of 98 Da upon CID, consistent with phosphorylation within the same peptide (data not shown). Moreover, two *O*-GlcNAc modifications were identified within the same peptide of HCF (data not shown).

Although we readily isolated *O*-GlcNAc-modified peptides, the mapping of specific *O*-GlcNAc glycosylation sites is inherently difficult due to the lability of the glycosidic linkage upon CAD and the preference of OGT for sequences rich in serine, threonine, and proline residues. To address this issue, we combined precedented β-elimination strategies with our methodology to localize specific modification sites. Previous studies have shown that glycosylated and phosphorylated serine/threonine

residues as well as carboxyamido-modified cysteine residues undergo β-elimination to form dehydroalanine/β-methyldehydroalanine under strong alkaline conditions.[17,26] Subsequent Michael addition of a thiol nucleophile generates a stable sulfide adduct. We first labeled S100 cytoplasmic lysates with our ketone-biotin tag and enriched the *O*-GlcNAc glycopeptides using avidin chromatography as described. One of the enriched fractions was then selected for β-elimination, followed by butanethiol addition (Figure 4.3). Tandem MS analysis of the resultant peptides permitted localization of the glycosylation site on HIV-1 Rev binding protein from seven possible residues on the peptide [360]APVGSVVSVPSHSSASSDK[378] to Ser372 or Ser373. Notably, tandem MS analysis prior to β-elimination conclusively demonstrated that the original peptide was *O*-GlcNAc glycosylated, rather than phosphorylated or modified with a complex carbohydrate. With further refinement of the β-elimination methodology toward complex mixtures, we anticipate that the combined ketone labeling and β-elimination approaches will be a powerful tool for identifying specific *O*-GlcNAc modification sites.

Figure 4.3. Combining the chemoenzymatic and β-elimination approaches to map specific glycosylation sites. (A) Strategy for the formation of a stable sulfide adduct from tagged *O*-GlcNAc peptides. (B) MS/MS analysis of the sulfide adduct of peptide [360]APVGSVVSVPSHSSASSDK[378] from HIV-1 Rev binding protein. Several y and b fragment ions containing the sulfide moiety (indicated with the subscript S) were observed, permitting peptide sequencing and identification of Ser372 or Ser373 as the glycosylation site. (C) MS/MS spectrum of the corresponding peptide prior to β-elimination shows the characteristic ketone-biotin signature, indicating that the original peptide was *O*-GlcNAc glycosylated. MS[4] analysis of the peptide confirmed its amino acid sequence. (D) Summary of the prominent b and y-ions from MS/MS analysis of the β-eliminated peptide. Fragment ions y7, y10, and b15 localized the site of modification to one of two serines within the peptide. An additional ion, consistent with b13, strongly suggested Ser372 as the specific site of glycosylation.

Having demonstrated the effectiveness of the chemoenzymatic strategy for protein

mixtures from the brain, we sought to identify *O*-GlcNAc proteins at the synaptosome, a

region critical for neuronal communication, which is enriched in the enzymes of *O*-GlcNAc cycling.[11]   Rat brain synaptosomal lysate was prepared, labeled and trypsin digested as described in the experimental methods. Overall, we identified 34 peptides from 15 unique proteins, including several that we had identified via the initial screen above (Table 4.2).   As expected, proteins involved in vesicle cycling and neurotransmitter release, such as the scaffolding protein bassoon were enriched in the synaptosome, allowing us to identify a host of *O*-GlcNAc glycosylated peptides within it and the related scaffolding protein piccolo. Previous studies *O*-GlcNAc antibody precipitation had suggested that Piccolo was a putative *O*-GlcNAc-modified protein;[17] here we found at least seven unique sites of modification. Consistent with the presence of OGT at synaptic vesicles,[12] many of the identified proteins are involved in synaptic or intracellular vesicle recycling. Moreover, we identified *O*-GlcNAc on the protein ankyrin-B, a closely-related protein to ankyrin-G, which localizes to nodes of ranvier of myelinated axons and was previously shown to be *O*-GlcNAc modified.[27]

Table 4.2. *O*-GlcNAc proteins identified from the synaptosome

Protein	NCBI Entry	Function	Peptide Sequence	Residues
NAD+ ADP-ribosyltransferase 3*	4808552	DNA binding enzyme/cell cycle control in mitotic cells	APVRTVTK	171-179
Hrb (HIV-1 Rev binding protein)	34859394	RNA trafficking	APVGSVVSVPSHSSASSDK	360-378
GRASP55 (Golgi reassembly stacking protein 2)	20301956	membrane protein transport, Golgi cisternae stacking	VPTTVEDR	423-430
ankyrin b, brain isoform 2	77681962	intracellular vesicle cycling/protein sorting	VGLQAQPMHSELVK	323-336
cappuccino*	45767878	vesicle tethering in organelle biogenesis	RAAAGYAACLLPGAGAR	63-79
erythrocyte protein band 4.1-like 1, isoform L	11067407	cytoskeletal protein	DVLTSTYGATAETLSTSTTTHVTK	1460-1483
AAK1 (AP2 associated kinase 1 isoform 1)	91992157	regulates clathrin-mediated endocytosis	LTDPIPTTETSIAPR	353-367
ProSAP1/Shank 2 (proline-rich synaptic-associated protein 1)	32491882	pdz-domain scaffolding/adaptor protein	SPEVMSTVSGTR	1651-1662
Neuron Navigator 1	109498680	microtubule-associated/neurite outgrowth	TPPVAVTSPITHTAQSALK	537-555
SCAMP1 (secretory carrier-associated membrane protein)	3914958	cell surface vesicle recycling	MPNVPNTQPAIMKPTEEHPAYTQITK	53-78
ankyrin repeat domain 40	109488798	contains ankyrin protein-protein interaction domain	TPESTKPGPVCQPPVSQNR	295-313
AP180 (assembly protein 180)	55725	clathrin-mediated synaptic vesicle endocytosis	SSPATTVTSPNSTPAK	305-320
NF-kB essential modulator	37576796	regulatory subunit of IkappaB complex	MQNQSVEAALR	204-214
bassoon	9506427	synaptic vesicle cycling	STAPAASHPAGKPQPGPTTAPGPQPAGLPR	3823-3852
bassoon	9506427	synaptic vesicle cycling	HEASASSSAAAAAAR	2639-2653
bassoon	9506427	synaptic vesicle cycling	LYSSMSDTNLAEAGLNYHAQR	1973-1993
bassoon	9506427	synaptic vesicle cycling	HSYSLGFADGR	2014-2024
bassoon	9506427	synaptic vesicle cycling	ASGAGGPPRPELPAGGAR	2283-2300
bassoon	9506427	synaptic vesicle cycling	SSVSQSPAPTYPSDSHYTSLEQNVPR	3198-3223
bassoon	9506427	synaptic vesicle cycling	HPTDLLSHPLPMR	2040-2052
bassoon	9506427	synaptic vesicle cycling	GLAGPTTVPATK	2920-2931
bassoon	9506427	synaptic vesicle cycling	QLLPSTATVR	2172-2181
bassoon	9506427	synaptic vesicle cycling	VTQHFAK	1338-1344
bassoon	9506427	synaptic vesicle cycling	FPFGSSCTGTFHPAPSAPDK	1900-1919
bassoon	9506427	synaptic vesicle cycling	VSPAIHITAATDPK	2678-2691
bassoon	9506427	synaptic vesicle cycling	LLDTSFASSER	2800-2810
bassoon	9506427	synaptic vesicle cycling	ISSVPGTSR	1636-1644
piccolo	7493836	neurotransmitter release	NQPLMIAPVSTDNTYAVSHLGSK	3929-3951
piccolo	7493836	neurotransmitter release	PAVPQIPVTTQKPTDTCPK	2463-2481
piccolo	7493836	neurotransmitter release	TVVTMDESTSNVVTK	2852-2866
piccolo	7493836	neurotransmitter release	VSTGEVMDYSSK	2975-2986
piccolo	7493836	neurotransmitter release	QVISGVGISTPQYSTAR	2996-3012
piccolo	7493836	neurotransmitter release	SCTAQQPATTLPEDR	2897-2911
piccolo	7493836	neurotransmitter release	ITSNYEVIR	3920-3928

*Human protein; residues marked in red are potential glycosylation sites as narrowed down by beta-elimination

4.3. Discussion

Here we have described the first direct, high-throughput analysis of *O*-GlcNAc glycosylated proteins from the mammalian brain. The proteins were identified using a chemoenzymatic approach that exploits an engineered galactosyltransferase enzyme to

selectively label *O*-GlcNAc proteins with a ketone-biotin tag. The tag provides both a straightforward means to enrich low-abundance *O*-GlcNAc peptides from complex mixtures and a unique signature upon tandem MS for unambiguous identification of the *O*-GlcNAc glycosylated species. In contrast to reported antibody and metabolic labeling methods,[17,18] the strategy provides direct evidence of *O*-GlcNAc glycosylation and permits mapping of modification sites to short amino acid sequences. The ability to localize *O*-GlcNAc is essential for surveying its distribution across the proteome as well as understanding its functional significance on a given protein or family of proteins.

An exciting feature of our approach is its potential to explore the interplay among PTMs.[3,28,29] In this study, we identified two peptides that contained more than one PTM. For instance, the N-terminal domain of HCF showed two *O*-GlcNAc moieties within the same peptide, and a second peptide exhibited evidence of both phosphorylation and glycosylation. Notably, all *O*-GlcNAc proteins known to date are phosphoproteins, and evidence suggests that glycosylation functionally antagonizes phosphorylation in many cases.[30,31] The approach reported herein involves a non-destructive technique that does not require the removal of other PTMs in order to study *O*-GlcNAc. As such, this strategy should permit a direct examination of whether specific glycosylation and phosphorylation events are mutually exclusive *in vivo*, as suggested for the C-terminal domain of RNA polymerase II,[30,32] or whether the two modifications coexist, as recently reported for the transcription factor signal transducer and activator of transcription 5 (Stat5).[33] Thus, this strategy is complementary to top-down MS approaches that can be used to simultaneously interrogate multiple PTMs from intact proteins.[34]

The chemoenzymatic approach can also be combined with existing β-elimination strategies, providing a powerful tool to identify precise sites of glycosylation. Notably, emerging MS techniques such as electron transfer dissociation (ETD), which has been successfully used to map phosphopeptide sites, could also be combined with our methodology to directly map glycosylation sites and abrogate the need for β-elimination.[35] Notably, following publication of our work, Vosseller and colleagues reported the discovery of *O*-GlcNAc post-synaptic density (PSD) proteins via WGA lectin chromatography.[36] The authors identified 18 *O*-GlcNAc-modified proteins in their extract preparation. Interestingly, many of the modified peptides detected in the study overlap directly with those we identified in the synaptosomal study described in this chapter, as well those we describe in a quantitative study in Chapter 5. While the lectin study was effective, the weak affinity of WGA for *O*-GlcNAc necessitated the generation of a 40 ft affinity column to separate *O*-GlcNAc and unmodified peptides. Such a method is probably not feasible for most laboratories and is not obviously amenable to the detection of intact *O*-GlcNAc proteins.

In this work, we demonstrated the power of the chemoenzymatic approach by identifying 25 *O*-GlcNAc glycosylated proteins from the mammalian brain, as well as another 11 proteins from synaptosomal preparations. Over the last 20 years, the *O*-GlcNAc modification has been established on approximately 100 proteins.[37] Thus, our results represent a significant expansion in the number of known *O*-GlcNAc proteins, and they provide new insights into the breadth of the modification and its potential functions in the brain.

Consistent with previous studies demonstrating an important role for *O*-GlcNAc in transcriptional regulation, we identified two novel transcription factors, sex determining factor Y box (SOX2) and activating transcription factor-2 (ATF-2). SOX2 is a member of the high mobility group (HMG) box superfamily of minor groove DNA-binding proteins,[38] which are implicated in guiding cell fate during diverse developmental processes. Although primarily known for its role in embryogenesis, SOX2 has also been detected in the adult central nervous system.[39] ATF-2 is a DNA-binding transcription factor that is ubiquitous but enriched in the brain.[40] It also appears to possess an intrinsic histone acetyltransferase (HAT) activity that is required for activating transcription.[41] As *O*-GlcNAc has been implicated in nutrient sensing and the development of insulin-resistant diabetes,[4,5] it is interesting that ATF-2 appears to play multiple roles in glucose homeostasis. For instance, ATF-2 has been shown to up-regulate transcription from the insulin promoter in human pancreatic β-cells in a Ca^{2+}/calmodulin-dependent protein kinase IV (CaMKIV)-dependent manner.[42] In addition, recent studies indicate that ATF-2 activates the gluconeogenic gene phosphoenolpyruvate carboxykinase (PEPCK) in HepG2 hepatic cells upon retinoic acid induction.[43]

While transcription factors and RNA polymerase II have been shown to be glycosylated, other important elements of the transcriptional machinery have not been well documented. In this study, we demonstrated *O*-GlcNAc on novel transcriptional proteins, including coactivators and corepressors. This finding suggests broader roles for *O*-GlcNAc in regulating transcription than previously recognized. For instance, we found the modification on two proteins (including a ubiquitin ligase) in the carbon

catabolite repression 4-negative on TATA-less (CCR4-NOT), a large protein complex involved in mRNA metabolism and the global control of gene expression.[44] In addition, *O*-GlcNAc was identified on the steroid receptor coactivator-1 (SRC-1), a chromatin remodeling protein that functions as a transcriptional coactivator for estrogen, thyroid, and other nuclear receptors.[45] Finally, *O*-GlcNAc was found on HCF, a chromatin-associated factor that interacts with both OGT and the Sin3A histone deacetylase (HDAC) complex *in vivo*.[24] Mammalian Sin3A interacts with OGT and thereby synergistically represses transcription from both basal and Sp-1 driven promoters.[46] We identified four distinct sites of glycosylation within the N-terminal domain of HCF, a region required for interaction with both OGT and Sin3A.[24] It will be interesting to examine the functional impact of HCF glycosylation on its binding to Sin3A and OGT, and on gene silencing.

Importantly, our results demonstrate that a number of proteins involved in neuronal signaling and synaptic function are the targets of *O*-GlcNAc glycosylation. For instance, we identified the modification on PDZ-GEF, a guanine nucleotide exchange factor that activates the Ras-related GTPases Rap1 and Rap2.[47] PDZ-GEF contains a PDZ domain, a protein-interacting module often involved in the assembly of signal transduction complexes at the synapse.[48] Another *O*-GlcNAc protein is WNK-1 (With No Lysine K), a serine/threonine protein kinase whose activation has been linked to ion transport and hypertension.[49] Moreover, we identified two brain-enriched proteins important for synaptic function, synaptopodin and bassoon. The actin-associated protein synaptopodin is essential for dendritic spine formation, with synaptopodin-deficient mice exhibiting a lack of spine apparatuses as well as impaired long-term potentiation and

spatial learning.[50] Bassoon, a scaffolding protein of the cytomatrix assembled at the active zone (CAZ) plays a critical role in synaptic vesicle cycling.[51]

Given the enrichment of the *O*-GlcNAc cycling enzymes to neuronal synaptosomes, and to synaptic vesicles, we sought to identify synaptosomal-specific proteins that might be *O*-GlcNAc modified. Indeed, we found 15 such proteins and 34 *O*-GlcNAc peptides. Here, we recapitulated a result from whole brain lysate, by identifying the ^{1338}VTQHFAK^{1344} peptide from bassoon as *O*-GlcNAc modified. In addition, we identified another 13 regions of *O*-GlcNAc modification within bassoon. Along with bassoon, we also found *O*-GlcNAc modification on the protein piccolo, which likewise localizes to the CAZ and appears to serve both a scaffolding and Ca(2+)-sensing role important in synaptic plasticity.[52]   Given the extensive glycosylation of these two proteins, what might be the significance of glycosylation for proteins involved in vesicle cycling and release?

Interestingly, piccolo is also found in pancreatic β-cells, where it appears to mediate cAMP-dependent exocytosis and insulin secretion.[53] The *O*-GlcNAc transferase is most abundant in pancreas and brain, and as discussed in Chapter 1, evidence suggests that it may function as a glucose sensor.[54] OGT is uniquely poised to respond to shifts in extracellular glucose as its substrate, UDP-GlcNAc, is the major product of the hexosamine biosynthesis pathway, and varying the UDP-GlcNAc concentration changes OGT's K_m and peptide substrate preference.[5] As with synaptic vesicles in the brain, OGT is found to localize with secretory granules of endocrine islet cells in the pancreas.[55] Thus, one possibility is that *O*-GlcNAc may function to regulate vesicle trafficking in both tissues in a glucose-dependent manner. In the brain, this may serve to modulate

neurotransmitter release or protein trafficking in a manner that corresponds to changes in cellular glucose uptake. Studies suggest that vesicle cycling and neurotransmitter release are altered in response to extracellular glucose concentration, well before detectable changes in ATP levels.[56] *O*-GlcNAc cycling is an attractive potential mediator of this process. In addition, it may function in a protective manner to prevent excitotoxic injury by inhibiting excess neurotransmitter release. Indeed, *O*-GlcNAc has been shown to mediate survival in response to various cell stresses in cultured cells.[57]

Our studies support a robust role for *O*-GlcNAc in vesicle recycling, as many of the other identified proteins are likewise involved in this process, such as SCAMP 37, which has been implicated in vesicle budding during clathrin-mediated endocytosis,[58] and AAK1, a serine/threonine kinase whose phosphorylation of the adaptor protein AP2 negatively regulates endocytosis *in vitro*.[59] In addition, we also find *O*-GlcNAc on proteins associated with the post-synaptic density, such as Shank2, a scaffolding protein important in mediating glutamate receptor function.[60,61] Taken together, these findings strongly suggest that *O*-GlcNAc glycosylation likely plays critical roles in neuronal communication and synaptic function.

4.3 Conclusion

In summary, we demonstrate a chemoenzymatic strategy for the high-throughput identification of *O*-GlcNAc glycosylated proteins from the mammalian brain. The approach permits the enrichment and direct identification of *O*-GlcNAc glycosylated peptides from complex mixtures and can be combined with existing technologies to map specific glycosylation sites. The generality of the method should enable explorations of the *O*-GlcNAc proteome in any cell type or tissue. Moreover, studies of the dynamic

interplay among PTMs and future extension of the methodology to quantitative proteomics should be possible. Using the approach, we discovered 23 new *O*-GlcNAc glycosylated proteins from the brain, as well as ten synaptosomal-specific proteins, including regulatory proteins associated with gene expression, neuronal signaling, and synaptic plasticity. The functional diversity represented by this set of proteins suggests an expanded role for *O*-GlcNAc in regulating neuronal function. Moreover, the identification of *O*-GlcNAc on proteins specifically involved in vesicle cycling suggests a unique role for *O*-GlcNAc in regulating cellular trafficking and neurotransmitter release. We anticipate that further investigations of the proteins identified in this study, coupled with the continued development of chemical tools, will provide new insights into the physiological importance of this posttranslational modification.

4.4 Experimental Methods and Supporting Figures

Preparation of Rat Forebrain Extracts

The forebrains of Sprague Dawley rats (Charles River Laboratories) were dissected on ice, lysed into 10 volumes of homogenization buffer, and fractionated into nuclear and S100 cytoplasmic components as described by Dignam et al.*,[62]* except that protease inhibitors, phosphatase inhibitors, and a hexosaminidase inhibitor (50 mM GlcNAc)[63] were added to the buffers. Prior to labeling, the extracts were dialyzed into 20 mM HEPES pH 7.3, 0.1 M KCl, 0.2 mM EDTA, 0.2% Triton X-100, 10% glycerol.

Preparation of Synaptosomal Rat Forebrain Extracts

The forebrains of 8-10 week-old Sprague Dawley rats (Charles River Laboratories) were dissected on ice and sliced by razor into 3 pieces per hemisphere. Forebrains were fractionated as described by Kiebler et al.,[64] with some modification. Importantly, all steps were conducted at 4 ºC unless otherwise stated. Briefly, two forebrains at a time were lysed into 7 ml of ice-cold homogenization buffer (0.32 M sucrose, 10 mM TRIS-HCl pH 7.4, containing protease inhibitors) via manual homogenization (5 strokes) with a glass dounce homogenizer. Homogenized extract was combined with extract from another two forebrains, diluted to 30 ml with homogenization buffer, and mechanically homogenized (700 rpm) for eight strokes. Homogenate was spun at $1000 \times g$ for 10 min (SS34 rotor) to pellet nuclei. Supernatant was recovered and spun at $14,000 \times g$ for 20 min (SS34 rotor) to pellet the P2 fraction. The P2 pellet was washed by resuspension in fresh homogenization buffer and spun again at $14,000 \times g$ for 20 min (SS34). Supernatant was decanted and the P2 pellet was resuspended in 4 ml of homogenization buffer and diluted two-fold with 50% optiprep solution. (50% optiprep solution was prepared by mixing 5 volumes of Optiprep (Accurate Chemicals, Westbury NY) with 1 volume of diluent, 0.32 M sucrose, 60 mM Tris-HCl pH 7.4). An optiprep gradient was layered into tubes for use in the HB-6 rotor as follows: 9mL of 15% optiprep (a mixture of 3 volumes of 50% optiprep and 7 volumes of homogenization buffer) was loaded on the bottom via serological pipet. Then, 9 ml of 12.5% optiprep (a mixture of 2.5 volumes of 50% optiprep and 7.5 volumes of homogenization buffer) was gently layered above the 15% layer, via 18-gauge needle and syringe, while slightly tipping the rotor tube.

Finally, 9 ml of 9% optiprep (a mixture of 1.8 volumes of 50% optiprep and 8.2 volumes of homogenization buffer) was layered above the 12.5% layer as described above. The resuspended P2 pellet was gently loaded, (via syringe and 18 gauge 6-inch blunt end stainless steel needle (Popper, New Hyde Park NY)), from the bottom of the prepared gradient. The sample was centrifuged at 18,000 × g for 20 min (HB-6 rotor). Synaptoneurosomes were collected as the first white voluminous band (below the top yellow band of myelin), resuspended in 5/2 volume of preincubation buffer (10 mM HEPES pH 7.4, 10 mM Glucose, 4.8 mM KCl, 1.2 mM Na_2HPO_4, 2.4 mM $MgSO_4$, 132 mM NaCl, 1.2 mM EGTA) and spun at 2000 × g, 15 min (SS34 rotor). Supernatant was discarded and synaptosomes were lysed directly in boiling 1% SDS with Complete Protease Inhibitors (Roche, Indianapolis, IN).

Chemoenzymatic Labeling of Rat Forebrain Extracts

Extract (1-10 mg; 1-3 mg/mL) was incubated with 5 mM $MnCl_2$, 1.25 mM ADP, 0.5 mM unnatural UDP substrate, and Y289L GalT (25 ng/µL) for 12-14 h at 4 °C. Following enzymatic labeling, extracts were dialyzed into denaturing buffer (5 M urea, 50 mM NH_4HCO_3 pH 7.8, 100 mM NaCl; 3 × 2 h). The pH was adjusted with 2.7 M NaOAc pH 3.9 (final concentration 50 mM, pH 4.8). Aminooxy biotin (2.75 mM) was added, and the reactions were incubated for 20-24 h at RT. Extracts were diluted with 3 M NH_4HCO_3 pH 9.6 (50 mM final concentration, pH 8) and dialyzed (1× 2 h, 1 × 10 h) into 6 M urea, 50 mM NH_4HCO_3 pH 7.8, 100 mM NaCl, followed by either denaturing (4 M urea, 50 mM NH_4HCO_3 pH 7.8, 10 mM NaCl) or non-denaturing buffer (50 mM NH_4HCO_3 pH 7.8, 10 mM NaCl).

Chemoenzymatic Labeling of Synaptosomal Extracts

Extracts (10-15 mg/mL) were diluted 5-fold with 100mM HEPES pH 7.9, 25mM

adenosine 5'-diphosphate, 20% Triton X-100, 100mM $MnCl_2$ into labeling buffer (10mM

HEPES pH 7.9, 1.25mM adenosine 5'-diphosphate, 1.8% Triton X-100, 5mM $MnCl_2$

containing Complete Protease Inhibitors, and 0.75 mM phenylmethylsulfonyl fluoride).

Diluted extract was then supplemented with 0.5 mM analogue **1**, 40 μg/mL Y289L GalT.

The reaction mixture was incubated at 4 °C for 12 h, and dialyzed into buffer A (10 mM

HEPES pH 7.9, 5 M urea) 3 × 3 h at 4 °C. The sample was then acidified to pH 4.8 by

adding 2.7 M NaOAc pH 3.9 to a final concentration of 50mM. Aminooxy biotin

derivative (30 mM) was added to a final concentration of 2.75 mM. After incubation at

room temperature for 20-24 h, the sample was neutralized by adding 0.5 M HEPES pH

7.9 to a final concentration of 33 mM, followed by dialysis into buffer B (10 mM HEPES

pH 7.9, 7 M urea, 100mM NaCl) 1× 12 h and 2 × 3 h at RT, and into buffer C (50mM

NH_4HCO_3 pH 8, 2M urea) 3 × 4 h at 4 ℃.

Proteolytic Digestion and Cation Exchange/Avidin Affinity Chromatography

Non-denatured extracts from the previous step were concentrated and denatured/reduced

as described in the ICAT protocol from Applied Biosystems. Proteins were then

alkylated with 15 mM iodoacetamide for 45 min in the dark, diluted to 0.04% SDS with

50 mM NH_4HCO_3 pH 7.8, and digested with trypsin or GluC (20-30 ng/μL) for 12-14 h

at 37 °C. Urea-denatured extracts were diluted with 50mM NH_4HCO_3 pH 8 to a final

concentration of less than 1 M urea following the reduction (10 min) and alkylation steps,

and subjected to protease digestion as described above. Synaptosomal extracts were reduced in 10mM DTT (from a stock of 500 mM in 50mM NH_4HCO_3 pH 8) for 1h at RT, alkylated in 20mM iodoacetamide (from a stock of 500 mM in 50mM NH_4HCO_3 pH 8) for 1 hr at RT, and incubated with 20mM DTT for 1 h at RT to react with excess iodoacetamide. The extract solution was then centrifuged at 15,000 rpm for 5 min to remove any insoluble material. Protein concentration was measured via Biorad assay, and then extracts were diluted to a final urea concentration of 1M. Sequencing-Grade Trypsin (Promega) was added to a final concentration of 0.015μg/μL for proteolytic digestion at 37 ºC. In all cases, proteolytic enzymes were added at a ratio of 1:20-1:30 to protein extracts.

Proteolytic digests conducted in the presence of urea were desalted with peptide macrotrap cartridges (Michrom Bioresources). Digests conducted without urea were acidified with 1% aqueous TFA and diluted into cation exchange load buffer (Applied Biosystems). In some cases, synaptosomal samples were dimethyl labeled for quantitative proteomics as described in Chapter 5. Cation exchange chromatography was performed on 0.5-3 mg of lysate as described by the manufacturer Applied Biosystems, except that peptides were eluted with a step gradient of 100 mM, 250 mM, and 350 mM KCl in 5 mM KH_2PO_4 containing 25% CH_3CN. Fractionated peptides were enriched via avidin chromatography (Applied Biosystems) as described by the manufacturer except that the washes were tripled in volume.

β-Elimination of Avidin-Purified Peptides

Following avidin chromatography, a portion of the S100 lysate fraction (40 mM KCl elution) was subjected to β-elimination exactly as described by wells et al.,[17] using 25 mM butanethiol, and reactions were stopped with AcOH.

LC-MS Analysis of Avidin-Enriched Biotinylated Peptides

Automated nanoscale reversed-phase HPLC/ESI/MS was performed using an HPLC pump, autosampler (Agilent Technologies), and linear ion trap mass spectrometer (ThermoElectron) with a variation of the "vented column" approach described by Licklider et al.[65] For data-dependent experiments, the mass spectrometer was programmed to record a full-scan ESI mass spectrum (*m/z* 500-2000) 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 ketone-biotin and ketone-biotin-GlcNAc moieties. Up to eight candidate peptides at a time were analyzed in subsequent targeted MS^4 experiments to derive sequence information.

Approximately 20% of the avidin-enriched peptides from each cation exchange fraction were loaded onto a 360 μm O.D. X 75 μm i.d. precolumn packed with 4 cm of 5 μm Monitor C18 particles (Column Engineering) at a flow rate of 4 μL/min. After desalting, the vent was closed and peptides eluted to a 360 μm O.D. X 75 μm i.d. analytical column with integrated emitter tip (10 cm of 5 μm C18, ca. 5 μm tip, see Martin et al. for more details.[66] The chromatographic profile was from 100% solvent A

(0.1% aqueous AcOH) to 50% solvent B (0.1% AcOH in CH$_3$CN) in 30 min. The flow rate through the analytical column was approximately 100 nL/min. The mass spectrometer was programmed to record a full-scan ESI mass spectrum (*m/z* 500-2000), followed by MS^4 scans of each peptide (relative collision energy = 35%; 3.5 Da isolation window for each fragmentation). In the first stage of tandem MS, precursor ions were isolated and fragmented to yield *O*-GlcNAc modified peptide ions. In the second stage, these ions were isolated and fragmented to yield ions corresponding to the unmodified peptide. In the third stage, unmodified peptide ions were isolated and fragmented, and the fragment ions were detected in the final stage. For all MS experiments, the electrospray voltage was set at 1.6 kV and the heated capillary was maintained at 250 °C. For database analysis to identify *O*-GlcNAc proteins, Bioworks Browser 3.1SR1 (Thermoelectron)software was used to create files from MS^4 data. These files were then directly queried, using the SEQUEST algorithm,[67] against amino acid sequences in the NCBI rat/mouse protein database.

In the course of manual analysis, it was discovered that SEQUEST assignment of an MS^4 targeted peptide at precursor ion mass *m/z* 899.6 incorrectly identified the peptide as AALTQQHPPVT*N*GDTVK from the protein ATF-2. Manual analysis demonstrated that peptide was AALTQQHPPVT*D*GDTVK, which failed to generate sequence identity to any peptide in either the NCBI or Celera non-redundant databases. Therefore, we concluded that the peptide represents the chemically deamidated product of ATF-2. Notably, chemical deamidation is substantiated by recent data for deamidation rates of specific amino acid sequences.[68] We confirmed the identity of the singly charged peptide VTQHFAK of bassoon by chemically synthesizing the corresponding peptide and

comparing its MS/MS spectrum to the MS[4] spectrum of the tagged peptide isolated from

the brain. The spectra were nearly identical, verifying the assignment.

Annotated MS[4] spectra for nuclear and cytoplasmic S100 fractions are presented

in the supporting information of Khidekel et al., 2004.[1]

References:

1.   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).
2.   Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* 425, 475-9 (2003).
3.   Greengard, P. The neurobiology of slow synaptic transmission. *Science* 294, 1024-30 (2001).
4.   Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the "*O*-GlcNAc code". *Sci STKE* 2005, re13 (2005).
5.   Zachara, N. E. & Hart, G. W. Cell signaling, the essential role of *O*-GlcNAc! *Biochim Biophys Acta* (2006).
6.   Whelan, S. A. & Hart, G. W. Proteomic approaches to analyze the dynamic relationships between nucleocytoplasmic protein glycosylation and phosphorylation. *Circ Res* 93, 1047-58 (2003).
7.   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).
8.   Griffith, L. S. & Schmitz, B. *O*-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to pertubations of phosphorylation. *Eur J Biochem* 262, 824-31 (1999).
9.   Iyer, S. P. N. & Hart, G. W. Dynamic nuclear and cytoplasmic glycosylation: Enzymes of *O*- GlcNAc cycling. *Biochemistry* 42, 2493-2499 (2003).
10.   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).
11.   Cole, R. N. & Hart, G. W. Cytosolic *O*-glycosylation is abundant in nerve terminals. *J Neurochem* 79, 1080-9 (2001).
12.   Akimoto, Y., Comer, F. I., Cole, R. N., Kudo, A., Kawakami, H., Hirano, H. & Hart, G. W. Localization of the *O*-GlcNAc transferase and *O*-GlcNAc-modified proteins in rat cerebellar cortex. *Brain Res* 966, 194-205 (2003).
13.   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).
14.   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).
15.   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).
16.   Cieniewski-Bernard, C., Bastide, B., Lefebvre, T., Lemoine, J., Mounier, Y. & Michalski, J. C. Identification of *O*-linked N-acetylglucosamine proteins in rat skeletal muscle using two-dimensional Gel electrophoresis and mass spectrometry. *Mol Cell Proteomics* (2004).

17. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. & Hart, G. W. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Molec Cell Proteomics* 1, 791-804 (2002).

18. 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).

19. Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., Qasba, P. K. & Hsieh-Wilson, L. C. A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc* 125, 16162-3 (2003).

20. 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).

21. Kalume, D. E., Molina, H. & Pandey, A. Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol* 7, 64-9 (2003).

22. 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, 1106-13 (2001).

23. Ding, M. & Vandre, D. D. High molecular weight microtubule-associated proteins contain *O*-linked-N-acetylglucosamine. *J Biol Chem* 271, 12555-61 (1996).

24. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* 17, 896-911 (2003).

25. Inaba, M. & Maede, Y. *O*-N-acetyl-D-glucosamine moiety on discrete peptide of multiple protein 4.1 isoforms regulated by alternative pathways. *J Biol Chem* 264, 18149-55 (1989).

26. Greis, K. D., Hayes, B. K., Comer, F. I., Kirk, M., Barnes, S., Lowary, T. L. & Hart, G. W. Selective detection and site analysis of *O*-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry. *Anal Biochem* 234, 38-49 (1996).

27. Zhang, X. & Bennett, V. Identification of *O*-linked N-acetylglucosamine modification of ankyrinG isoforms targeted to nodes of Ranvier. *J Biol Chem* 271, 31391-8 (1996).

28. Khidekel, N. & Hsieh-Wilson, L. C. A 'molecular switchboard'--covalent modifications to proteins and their impact on transcription. *Org Biomol Chem* 2, 1-7 (2004).

29. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 44, 7342-72 (2005).

30. Comer, F. I. & Hart, G. W. Reciprocity between *O*-GlcNAc and *O*-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* 40, 7845-52 (2001).

31. 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, 19229-35 (2002).

32. 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, 10416-24 (1993).

33. Gewinner, C., Hart, G., Zachara, N., Cole, R., Beisenherz-Huss, C. & Groner, B. The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *J Biol Chem* 279, 3563-72 (2004).

34. Kelleher, N. L. Top-down proteomics. *Anal Chem* 76, 197A-203A (2004).

35. 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).

36. Vosseller, K., Trinidad, J. C., Chalkley, R. J., Specht, C. G., Thalhammer, A., Lynn, A. J., Snedecor, J. O., Guan, S., Medzihradszky, K. F., Maltby, D. A., Schoepfer, R. & Burlingame, A. L. *O*-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 5, 923-34 (2006).

37. Whelan, S. A. & Hart, G. W. Proteomic approaches to analyze the dynamic relationships between nucleocytoplasmic protein glycosylation and phosphorylation. *Circ Res* 93, 1047-58 (2003).

38. Pevny, L. H. & Lovell-Badge, R. Sox genes find their feet. *Curr Opin Genet Dev* 7, 338-44 (1997).

39. Gure, A. O., Stockert, E., Scanlan, M. J., Keresztes, R. S., Jager, D., Altorki, N. K., Old, L. J. & Chen, Y. T. Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer. *Proc Natl Acad Sci USA* 97, 4198-203 (2000).

40. Herdegen, T. & Leah, J. D. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. *Brain Res Brain Res Rev* 28, 370-490 (1998).

41. Kawasaki, H., Schiltz, L., Chiu, R., Itakura, K., Taira, K., Nakatani, Y. & Yokoyama, K. K. ATF-2 has intrinsic histone acetyltransferase activity which is modulated by phosphorylation. *Nature* 405, 195-200 (2000).

42. Ban, N., Yamada, Y., Someya, Y., Ihara, Y., Adachi, T., Kubota, A., Watanabe, R., Kuroe, A., Inada, A., Miyawaki, K., Sunaga, Y., Shen, Z. P., Iwakura, T., Tsukiyama, K., Toyokuni, S., Tsuda, K. & Seino, Y. Activating transcription factor-2 is a positive regulator in CaM kinase IV-induced human insulin gene expression. *Diabetes* 49, 1142-8 (2000).

43. Lee, M. Y., Jung, C. H., Lee, K., Choi, Y. H., Hong, S. & Cheong, J. Activating transcription factor-2 mediates transcriptional regulation of gluconeogenic gene PEPCK by retinoic acid. *Diabetes* 51, 3400-7 (2002).

44. Collart, M. A. Global control of gene expression in yeast by the Ccr4-Not complex. *Gene* 313, 1-16 (2003).

45. Xu, J. & Li, Q. Review of the *in vivo* functions of the p160 steroid receptor coactivator family. *Mol Endocrinol* 17, 1681-92 (2003).

46. 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).

47. Rebhun, J. F., Castro, A. F. & Quilliam, L. A. Identification of guanine nucleotide exchange factors (GEFs) for the Rap1 GTPase. Regulation of MR-GEF by M-Ras-GTP interaction. *J Biol Chem* 275, 34901-8 (2000).

48. Zhang, M. & Wang, W. Organization of signaling complexes by PDZ-domain scaffold proteins. *Acc Chem Res* 36, 530-8 (2003).

49. Wilson, F. H., Disse-Nicodeme, S., Choate, K. A., Ishikawa, K., Nelson-Williams, C., Desitter, I., Gunel, M., Milford, D. V., Lipkin, G. W., Achard, J. M., Feely, M. P., Dussol, B., Berland, Y., Unwin, R. J., Mayan, H., Simon, D. B., Farfel, Z., Jeunemaitre, X. & Lifton, R. P. Human hypertension caused by mutations in WNK kinases. *Science* 293, 1107-12 (2001).

50. Deller, T., Korte, M., Chabanis, S., Drakew, A., Schwegler, H., Stefani, G. G., Zuniga, A., Schwarz, K., Bonhoeffer, T., Zeller, R., Frotscher, M. & Mundel, P. Synaptopodin-deficient mice lack a spine apparatus and show deficits in synaptic plasticity. *Proc Natl Acad Sci USA* 100, 10494-9 (2003).

51. Altrock, W. D., tom Dieck, S., Sokolov, M., Meyer, A. C., Sigler, A., Brakebusch, C., Fassler, R., Richter, K., Boeckers, T. M., Potschka, H., Brandt, C., Loscher, W., Grimberg, D., Dresbach, T., Hempelmann, A., Hassan, H., Balschun, D., Frey, J. U., Brandstatter, J. H., Garner, C. C., Rosenmund, C. & Gundelfinger, E. D. Functional inactivation of a fraction of excitatory synapses in mice deficient for the active zone protein bassoon. *Neuron* 37, 787-800 (2003).

52. Gerber, S. H., Garcia, J., Rizo, J. & Sudhof, T. C. An unusual C(2)-domain in the active-zone protein piccolo: implications for Ca(2+) regulation of neurotransmitter release. *Embo J* 20, 1605-19 (2001).

53. Fujimoto, K., Shibasaki, T., Yokoi, N., Kashima, Y., Matsumoto, M., Sasaki, T., Tajima, N., Iwanaga, T. & Seino, S. Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of cAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis. *J Biol Chem* 277, 50497-502 (2002).

54. Iyer, S. P. & Hart, G. W. Dynamic nuclear and cytoplasmic glycosylation: enzymes of *O*-GlcNAc cycling. *Biochemistry* 42, 2493-9 (2003).

55. Akimoto, Y., Kreppel, L. K., Hirano, H. & Hart, G. W. Localization of the *O*-linked N-acetylglucosamine transferase in rat pancreas. *Diabetes* 48, 2407-13 (1999).

56. Fleck, M. W., Henze, D. A., Barrionuevo, G. & Palmer, A. M. Aspartate and glutamate mediate excitatory synaptic transmission in area CA1 of the hippocampus. *J Neurosci* 13, 3944-55 (1993).

57. Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D. & Hart, G. W. Dynamic *O*-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* 279, 30133-42 (2004).

58. Fernandez-Chacon, R., Achiriloaie, M., Janz, R., Albanesi, J. P. & Sudhof, T. C. SCAMP1 function in endocytosis. *J Biol Chem* 275, 12752-6 (2000).

59.   Conner, S. D. & Schmid, S. L. Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J Cell Biol* 156, 921-9 (2002).

60.   Boeckers, T. M., Bockmann, J., Kreutz, M. R. & Gundelfinger, E. D. ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *J Neurochem* 81, 903-10 (2002).

61.   Hwang, J. I., Kim, H. S., Lee, J. R., Kim, E., Ryu, S. H. & Suh, P. G. The interaction of phospholipase C-beta3 with Shank2 regulates mGluR-mediated calcium signal. *J Biol Chem* 280, 12467-73 (2005).

62.   Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11, 1475-89 (1983).

63.   Chou, T. Y., Hart, G. W. & Dang, C. V. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J Biol Chem* 270, 18961-5 (1995).

64.   Kiebler, M. A., Lopez-Garcia, J. C. & Leopold, P. L. Purification and characterization of rat hippocampal CA3-dendritic spines associated with mossy fiber terminals. *FEBS Lett* 445, 80-6 (1999).

65.   Licklider, L. J., Thoreen, C. C., Peng, J. & Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal Chem* 74, 3076-83 (2002).

66.   Martin, S. E., Shabanowitz, J., Hunt, D. F. & Marto, J. A. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 72, 4266-74 (2000).

67.   Eng, J. K. M., A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5, 976-989 (1994).

68.   Robinson, N. E. & Robinson, A. B. Molecular clocks. *Proc Natl Acad Sci USA* 98, 944-9 (2001).

Chapter 5

Probing the Dynamics of *O*-GlcNAc Glycosylation Using the Chemoenzymatic Strategy and Quantitative Proteomics

5.1 Background and Introduction

The addition of post-translational modifications (PTMs) to proteins represents a fundamental mechanism by which protein function is elaborated.[1,2] In the brain, finely-tuned regulation of PTMs such as phosphorylation govern mechanisms of neuronal plasticity, the proposed basis for learning and memory.[3,4] Roles are also emerging for other PTMs, such as polyubiquitination, which modulates protein degradation in response to neuronal activity.[5] Even more remarkably, recent studies highlight the dynamic equilibrium of PTMs whose switch-like behavior alters protein function and has dramatic consequences for cells.[6] For example, the dynamic shuttling of sumoylation and acetylation on the transcription factor MEF2A appears to regulate its transactivation potential and, ultimately, neuronal morphology.[7]

Our work has focused on *O*-GlcNAc glycosylation, the covalent attachment of β-*N*-acetylglucosamine to serine or threonine residues of proteins.[8] *O*-GlcNAc is intracellular and dynamic, sharing more in common with phosphorylation than with ER or Golgi-mediated glycosylation. The modification is mediated by a unique nucleo-cytoplasmic *O*-GlcNAc transferase (OGT) necessary for life at the single cell level.[9] A diverse group of over 100 proteins has been has been identified as *O*-GlcNAc-modified, including transcription factors, proto-oncogenes, cytoskeletal-associated proteins, and protein kinases.[10,11] Recent studies highlight a role for *O*-GlcNAc in many cellular processes, such as nutrient sensing, cell division, the stress response and gene

silencing.[8,12] Moreover, perturbations in *O*-GlcNAc levels have been associated with disease states such as cancer, Alzheimer's, and diabetes.[11,13]

Evidence suggests that *O*-GlcNAc may play a particularly important role in the brain. OGT is most abundant in the brain and pancreas,[14] and neuron-specific knockout of the OGT gene induces early postnatal lethality.[15] The activity of OGT appears to be modulated by complex mechanisms, including differential splicing, interaction with regulatory partners, and changes in UDP-GlcNAc concentration.[14] In the brain, OGT forms a stable complex with protein phosphatase-1 (β and γ),[16] highlighting the potential interrelationship between *O*-GlcNAc and phosphorylation, one of the best characterized PTMs regulating neuronal function. Finally, OGT and the corresponding β-*N*-acetylglucosaminidase (*O*-GlcNAcase) are particularly abundant in the nerve terminal,[17] or synaptosome, where they are enriched in the presynaptic cytosol and around synaptic vesicles,[18] suggesting neuron-specific functions for the modification.

Despite tantalizing evidence for its significance, the regulation of *O*-GlcNAc glycosylation, particularly in the brain, is poorly understood. Early studies suggested that *O*-GlcNAc is a highly dynamic modification, with a turnover rate that exceeds that of the protein backbone.[19] Moreover, *O*-GlcNAc levels in lymphocyte cells were shown to respond dynamically to mitogens.[20] Recent evidence suggests that *O*-GlcNAc levels can rapidly respond (within minutes) to ligand-binding on neutrophils.[21] Finally, *O*-GlcNAc levels showed marked upregulation in response to numerous cell stresses, and perturbation of the *O*-GlcNAc enzymes decreased cell survival in response to stress.[22] In the brain, *O*-GlcNAc regulation has received only limited attention thus far. In cultured cerebellar neurons, *O*-GlcNAc levels of cytoskeletal-associated proteins were shown to

respond reciprocally to PKA and PKC activators and inhibitors, highlighting the apparent antagonism between the two modifications, in certain contexts.[23] Most of the studies assaying *O*-GlcNAc regulation took advantage of the traditional [³H]-labeling methodology[24] or antibodies with varying degrees of *O*-GlcNAc specificity[25] to detect *O*-GlcNAc. These approaches are generally insensitive,[24] and not applicable to the direct identification of specific proteins, and sites of modification on which *O*-GlcNAc glycosylation might be dynamically regulated.[26]

Recently, several new, significantly advanced approaches for the identification of *O*-GlcNAc glycosylated proteins have been described. *O*-GlcNAc proteins have been chemically tagged by metabolic labeling[27,28] or BEMAD (β-Elimination followed by Michael Addition with Dithiothreitol)[26] and *O*-GlcNAc peptides have been directly identified via weak lectin affinity chromatography.[29] However, no method currently exists for unbiased, quantitative proteomics analyses of *O*-GlcNAc glycosylation. While metabolic labeling has been used to identify new *O*-GlcNAc proteins, it is not ideally suited for the study of *O*-GlcNAc dynamics as it requires the use of high concentrations of an unnatural GlcNAc analogue, which may perturb both physiological *O*-GlcNAc levels and substrate targets of the OGT.[30] Moreover, it may not useful for the study of O-GlcNAc dynamics in tissue due to cellular and organismal uptake requirements. Although the BEMAD approach can be used to map sites on purified proteins and protein complexes, it is a destructive technique that requires extensive controls to establish whether a peptide contains a phosphate, *O*-GlcNAc or complex *O*-linked carbohydrate group.[26] Finally, the latest lectin-based strategy effectively identified over 60 *O*-GlcNAc glycosylated peptides from the brain, but required the use of a 40 ft affinity column to

separate *O*-GlcNAc and unmodified peptides.[29] Such a strategy may not be technically feasible for many laboratories, and is not directly applicable to the detection and identification of intact *O*-GlcNAc proteins.

In the previous chapters, we described the development of a new chemoenzymatic strategy to rapidly and sensitively detect *O*-GlcNAc modified proteins. We applied this strategy to the discovery of single proteins from cell lysate as well as the identification of *O*-GlcNAc-modification sites. Finally, we combined the approach with high-throughput mass spectrometry for the first explorations of the *O*-GlcNAc proteome of the brain. We simultaneously discovered 23 *O*-GlcNAc proteins, as well as another 12 specific to the synaptosomal compartment, which is enriched in the enzymes of *O*-GlcNAc cycling, and is critical for neuronal communication.

As our approach enables sensitive detection and specific isolation of *O*-GlcNAc proteins and peptides from both cells and tissue, we felt that it would be an ideal platform on which to study *O*-GlcNAc dynamics and regulation. In this chapter, we first demonstrate the effectiveness of the approach to identify changes in *O*-GlcNAc glycosylation on single protein targets. We find that *O*-GlcNAc levels on cyclic AMP response element (CRE)-binding protein (CREB), a transcription factor implicated in pancreatic cell survival and diabetes,[31] can be modulated in cultured pancreatic cells. Second, we combine the chemoenzymatic strategy with quantitative mass spectrometry approaches for the first proteomic studies of *O*-GlcNAc dynamics, both from cultured primary neurons and brain tissue.

Several elegant strategies have been described for quantitative mass-spectrometry-based proteomics such as stable isotope labeling with amino acids in cell

culture (SILAC),[32] and isobaric tags for relative and absolute quantification (iTRAQ).[33] Moreover, these approaches have been successfully coupled with phosphoprotein[34] and phosphopeptide[35] enrichment to monitor changes in phosphorylation in response to extracellular stimuli. Nonetheless, these quantification strategies have some limitations. SILAC requires multiple cell divisions for highest incorporation of amino acid isotopes and therefore is not suited to post-mitotic cells, such as neurons, or tissue. ITRAQ requires the use of mass spectrometry instruments that can detect very low mass ions (e.g., Q-TOF) that do not allow multiple stages of mass spectrometry and thus may not be conducive to the sequencing of chemoenzymatically-tagged O-GlcNAc-modified peptides. Therefore, in order to study O-GlcNAc in the brain, we chose to combine the chemoenzymatic approach with a modified dimethyl labeling strategy,[36] which incorporates stable isotopes onto peptide N-terminal amines and ε-amino groups of lysine residues by reductive amination. This approach is extremely rapid, proceeds quantitatively, and in contrast to other strategies, is very inexpensive.

In order to facilitate the sequencing of quantified peptides, we also investigated the utility of electron-transfer dissociation (ETD) for the mapping of O-GlcNAc glycosylation sites. ETD relies on a mode of fragmentation (e.g., the transfer of electrons from an anion donor source and subsequent intermolecular rearrangement and cleavage) that is independent of peptide sequence or the presence of PTMs.[37] In contrast to collision-associated dissociation (CAD), ETD has the capacity to preserve PTMs on the peptide backbone for exact site identification. In addition, unlike the similar strategy of electron capture dissociation (ECD), ETD is conducted on a modified ion trap mass

spectrometer, a comparatively inexpensive instrument that will be accessible to many laboratories.

In this chapter, we demonstrate the quantitative proteomics strategy on neurons treated with the O-GlcNAcase inhibitor O-(2-acetamid$O$-2-deoxy-D-glucopyranosylidene)amino-N-phenylcarbamate (PUGNAc), where we find site-specific changes in O-GlcNAc glycosylation on several proteins. Moreover, we show that O-GlcNAc levels respond rapidly, in the cortex, to kainic-acid induced seizures. We also highlight the power of ETD to identify exact sites of O-GlcNAc modification, which forgoes the need for destructive chemical approaches for site mapping. Overall, the strategy permits concomitant discovery of new O-GlcNAc proteins and detection of O-GlcNAc glycosylation changes in response to stimuli. These discoveries show that O-GlcNAc is indeed reversible and dynamic in the brain and provide new molecular targets for understanding O-GlcNAc regulatory pathways in neuronal tissue.

5.2 O-GlcNAc Dynamics on Individual Proteins

The transcription factor CREB is implicated in a host of cellular processes, including glucose homeostasis, cell survival, and synaptic plasticity.[38] Multiple signal transduction cascades converge on CREB, and N. Lamarre-Vincent, in our group, discovered that CREB was O-GlcNAc-modified in a region necessary for interaction with the TFIID complex of the basal transcriptional machinery.[39] The interaction with TFIID is effectively blocked by O-GlcNAc glycosylation *in vitro*. Moreover, cell stimulation with agents that increase levels of O-GlcNAc (such as the UDP-GlcNAc precursor glucosamine and PUGNAc) repress endogenous CREB-mediated transcription in cells.

In pancreatic cells, this repression is associated with decreases in transcription of pro-survival factors and increased levels of cellular apoptosis.[40]

Global upregulation of O-GlcNAc levels can be monitored using the anti-O-GlcNAc antibody CTD110.6. However, traditional methods to identify O-GlcNAc changes directly on CREB were ineffective. [^{3}H]-labeling with β-1,4-galactosyltransferase required hundreds or thousands of hours of exposure for detection, while the O-GlcNAc antibodies, CTD110.6 and RL-2 could not detect endogenous CREB.[41]

Previously, we had developed a chemoenzymatic strategy to efficiently biotinylate and thereby detect endogenous O-GlcNAc-modified CREB from cell lysates.[42] Here, we sought to apply this approach to measure changes in CREB glycosylation in response to cellular simulation. In this approach (Figure 5.1), O-GlcNAc-modified proteins from two (or more) different cell populations are labeled, and the protein of interest, in this case CREB, is immunoprecipitated from each cell state. Streptavidin labeling of the immunoprecipitated protein detects the O-GlcNAc modification, which can be compared across cell states, while Western blotting for the protein controls for variation in immunoprecipitation between samples.
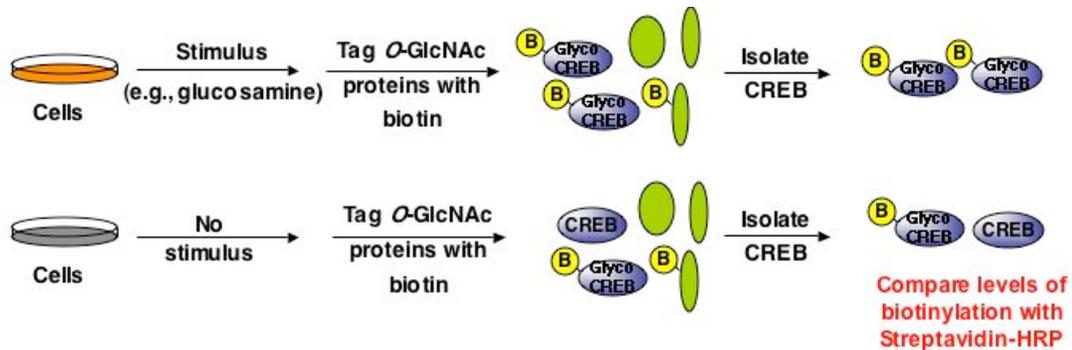
Figure 5.1. Strategy for detection of dynamic changes in O-GlcNAc glycosylation on individual proteins where "B" represents the chemoenzymatically transferred biotin tag).

We applied this strategy to pancreatic HIT-T15 cell lysate treated with glucosamine, a precursor to UDP-GlcNAc, the substrate for OGT. Glucosamine, which bypasses the rate-limiting enzyme of hexosamine synthesis, generally increases global O-GlcNAc levels. HIT-T15 cells were treated with forskolin, an adenylate cyclase activator known to activate cAMP pathways and increase CREB transactivation potential, and glucosamine or control vehicle. After treatment, control and experimental cells were lysed and chemoenzymatically labeled. Western blotting with streptavidin-HRP of immunoprecipitated CREB showed a statistically significant increase in O-GlcNAc glycosylation on CREB in response to glucosamine treatment (Figure 5.2A). This was consistent with a glucosamine-induced decrease in CREB-mediated transcription as measured by a CRE-promoter luciferase reporter assay in HIT-T15 cells (Figure 5.2B, courtesy of N. Lamarre-Vincent).
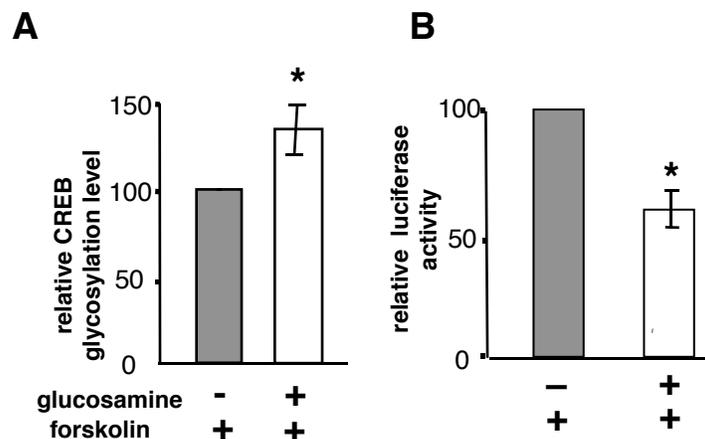
Figure 5.2. The chemoenzymatic strategy allows direct detection of changes in O-GlcNAc glycosylation on the transcription factor CREB. (A) CREB glycosylation increases 33% ± 12% with glucosamine treatment, *$P<0.006$; n=4 (B) CREB-mediated transcription of a CRE-luciferase reporter is repressed in response to glucosamine treatment, *$P<0.0001$; n=4.

PUGNAc treatment of cells likewise resulted in an increase in CREB *O*-GlcNAc

modification, although a more modest one, and a concomitant decrease in CREB-

mediated transcription (Figure 5.3A, B courtesy of N. Lamarre-Vincent).
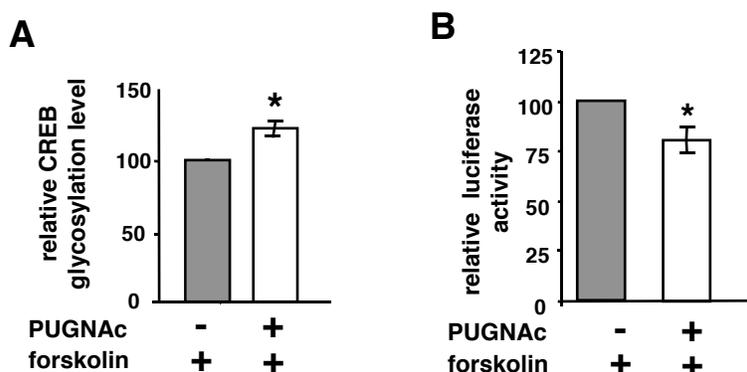
Figure 5.3. The chemoenzymatic strategy allows direct detection of changes in *O*-GlcNAc glycosylation on the transcription factor CREB in response to PUGNAc. (A) CREB glycosylation increases 21%± 5% with PUGNAc treatment,*P<0.0005; n=5 (B) CREB-mediated transcription of a CRE-luciferase reporter is repressed in response to glucosamine treatment, * P< 0.03; n = 3.

Because the chemoenzymatic approach is also amenable to tissues, current efforts

in the laboratory are focused at understanding *O*-GlcNAc dynamics on CREB *in vivo*, in

animal models of diabetes. Overall, the approach is readily amenable to other proteins

and should provide a sensitive way to monitor and understand *O*-GlcNAc regulation on

individual protein targets.

5.3 Quantitative Proteomics of *O*-GlcNAc Glycosylation in Neurons

Having demonstrated that the chemoenzymatic strategy could be used to quantify

changes in glycosylation on individual endogenous proteins, we applied the approach

toward the *O*-GlcNAc proteome of cultured primary neurons and the brain. In the

approach, proteins are chemoenzymatically labeled and proteolytically digested. Peptides from two different cell states are differentiated by the incorporation of stable isotopes via reductive amination of lysine ε-amino groups and peptide N-terminal amines. The use of either formaldehyde and NaCNBH$_3$ or deuterated formaldehyde and NaCNBD$_3$ creates mass differences of 6 or 12 Da, depending on the presence of a lysine residue, which readily permits differentiation of peptides at even higher charge state. Subsequent mass spectrometry analysis of avidin enriched, dimethylated *O*-GlcNAc peptides allows for quantification of changes in glycosylation state (Figure 5.4).
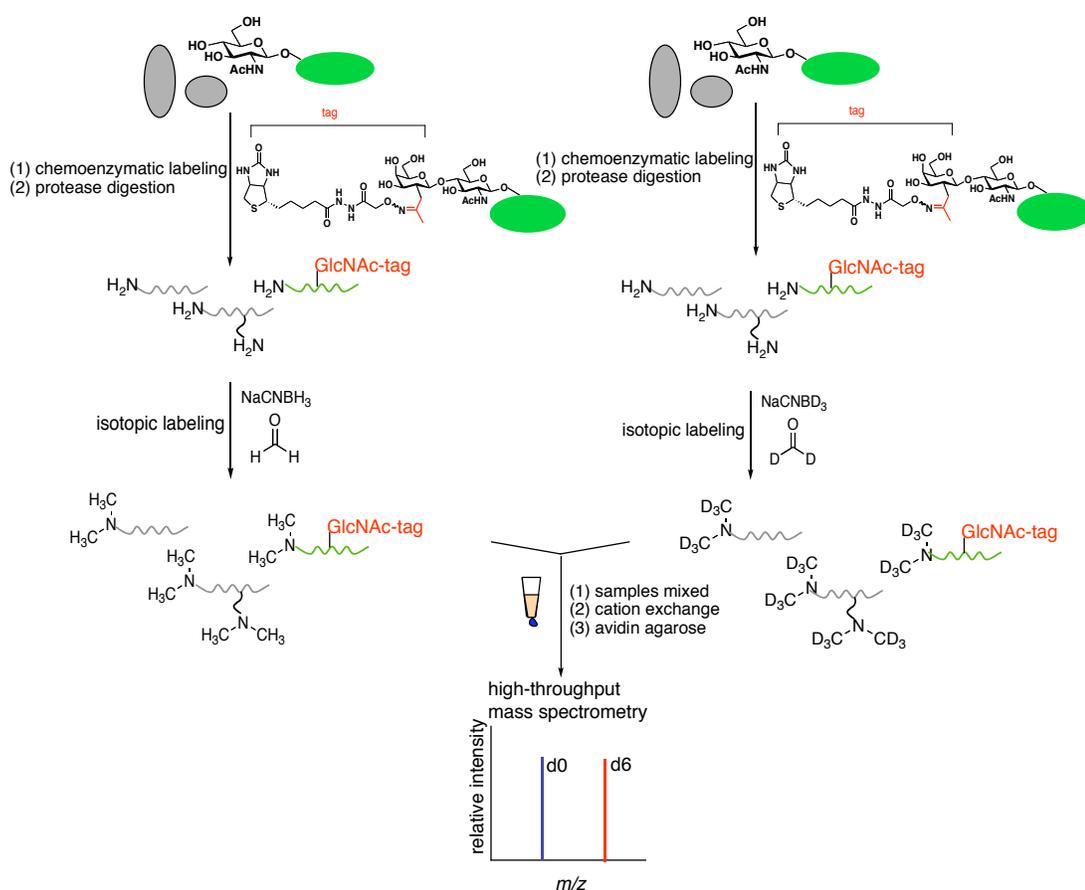
Figure 5.4. Chemoenzymatic strategy for proteome-wide quantification of *O*-GlcNAc dynamics.

In order to evaluate the effectiveness of the dimethyl labeling strategy, we first labeled a single protein digest of α-casein and evaluated the extent of labeling by liquid-chromatography mass spectrometry (LC-MS). We found that reductive amination proceeded quantitatively for both lysine and N-terminal primary amines in less than 10 min at pH 7.5 (data not shown). Unlike other strategies, which incorporate stable isotopes by acylation of N-terminal primary amines or ε-amino groups of lysines with groups such as succinic anhydride,[43] this approach does not change the overall charge of the peptide, thus preserving high ionization efficiency.

Having established the best conditions for dimethyl labeling, we investigated our ability to capture and quantify known *O*-GlcNAc peptides from complex mixtures.[10,42] α-crystallin (ca. 300 pmol) and OGT (ca. 10 pmol) proteins were mixed with two separate samples of rat brain or primary neuronal culture lysate in ratios of 1:1 for each protein. The samples were chemoenzymatically labeled and proteolytically digested with trypsin. After digestion, peptides were dimethyl labeled, mixed and fractionated by strong-cation exchange to reduce sample complexity. *O*-GlcNAc peptides were captured by avidin affinity chromatography, and the sample was analyzed by (LC-MS) on the hybrid LTQ/orbitrap mass spectrometer. Accurate quantification of peptide ratios was achieved in MS mode via the orbitrap mass analyzer, which takes advantage of new ion trapping principles[44] to generate mass accuracy upwards of 20 ppm.[45] In order to identify quantified ions, peptides that specifically showed the biotin-ketone and biotin-ketone-GlcNAc loss during MS/MS analysis (described in Chapters 3 and 4) were targeted for sequencing by MS^4 analysis.

A

19.03

do

19.03

d12

Relative Intensity

17.5 18.0 18.5 19.0 19.5 20.0
Time (min)

100

806.37402

806.70630

806.04169

807.03937

807.37347

807.70844

809.72687

810.39355 d12

810.06134

810.72595

811.05981

811.39398

811.72888

do

50

Relative Intensity

0

806 807 808 809 810 811 812
m/z

B

24.94

do

24.94

d12

Relative Intensity

24.5 25.0 25.5 26.0 26.5
Time (min)

100

1303.07019

1303.57056

1302.56921

1304.07104

1304.57288

1305.07520

1308.60461

1308.10291

1309.10535 d12

1309.60608

1310.10693

1310.60791

1311.11121

do

50

Relative Intensity

0

1302 1304 1306 1308 1310 1312
m/z

Figure 5.5. Application of the chemoenzymatic quantitative proteomics strategy toward α-crystallin and *O*-GlcNAc transferase peptides isolated from a complex mixture. (A) Chromatogram and corresponding orbitrap MS spectrum of avidin affinity captured αA-crystallin *O*-GlcNAc glycosylated peptide [158]AIPVSREEKPSSAPSS[173] in light form (*m/z* 806.4159) and heavy form (*m/z* 810.06134). αA-crystallin protein was mixed with two samples of lysate in a heavy:light ratio of 1:1 and the mixture was chemoenzymatically labeled. The *O*-GlcNAc glycosylated peptide was captured and observed at a ratio of 0.97 - 0.09, 0.97 + 0.10 (g.s.d. of 1.10), n = 7. (B) Chromatogram and corresponding orbitrap MS spectrum of the avidin affinity captured OGT *O*-GlcNAc glycosylated peptide [390]ISPTFADAYSNMGNTLK[406] in light form (*m/z* 1302.5691) and heavy form (*m/z* 1308.60461). OGT protein was mixed and labeled as above and this *O*-GlcNAc glycosylated peptide was observed at a ratio of 0.93 - 0.12, 0.93 + 0.14 (g.s.d. of 1.15), n = 7.

As shown in Figure 5.5, we were successfully able to isolate and quantify *O*-GlcNAc-modified peptides [158]AIPVSREEKPSSAPSS[173] from α-crystallin and [390]ISPTFADAYSNMGNTLK[406] from OGT, from the complex lysate mixture (peptide identity was confirmed by MS[4] sequencing analysis). The ratio of signal intensities from the heavy to light peptide forms, across the entire chromatographic profile of each form,

allowed direct, peptide-specific quantification. This α-crystallin peptide was observed at a mean ratio of 0.97 - 0.09, 0.97 + 0.10 (geometric standard deviation, g.s.d. 1.10), while the OGT peptide was observed at a mean ratio of 0.93 - 0.12, 0.93 + 0.14 (g.s.d. of 1.15).

Importantly, hydrogen-based isotopic tags are known to be subject to isotope resolution effects during reverse-phase chromatography such that deuterium-labeled peptides can elute in advance of non-deuterated counterparts, which can interfere with quantification.[46] In general, we found that deuterated and non-deuterated dimethylated peptides (Figure 5.5) coeluted in the LC. This is consistent with reports that deuterium incorporated into a dimethyl label does not produce an isotope effect.[36,46]

In addition to the two peptides shown in Figure 5.5, we reproducibly captured and quantified two more peptides from α-crystallin, encompassing the major known glycosylation sites of both the A[47] and B[19] forms of α-crystallin. Additionally, we also captured another seven glycosylated peptides from OGT that encompassed all the known glycosylated sites within that protein.[42] Table 5.1 shows all the α-crystallin and OGT peptides and their mean and standard deviations across all experiments in which they were accurately quantified. The mean ratio across all peptides was 0.91 - 0.17, 0.91 + 0.21 (g.s.d. of 1.23), which compares favorably with the reported mean ratios of peptides from standard protein mixtures used in similar iTRAQ and SILAC quantification (mean observed ratios of 1.03 ± 0.16 and 1.03 ± 0.17 for an expected 1:1 ratio, respectively).[33]

Table 5.1. Summary of *O*-GlcNAc glycosylated peptides from α-crystallin and *O*-GlcNAc transferase (OGT) isolated and quantified via the chemoenzymatic quantitative proteomics strategy

Protein	Peptide Sequence	n	Ratio[a]	g.s.d.[b]
crystallin 1	AIPVSREEKPSSAPSS	7	0.97	1.10
crystallin 2	AIPVSREEKPSSAPS	7	0.90	1.15
crystallin 3	EEKPVVTAAPK	4	0.81	1.13
OGT1	IKPVEVTESA	7	0.91	1.36
OGT 2	AIQINPAFADAHSNLASIHK	7	0.77	1.20
OGT 3	ISPTFADAYSNMGNTLK	7	0.93	1.15
OGT 4	EMQDVQGALQCYTR	5	0.98	1.11
OGT 5	AIQINPAFADAHSNLASIHKDSGNIPEAIASYR	4	1.01	1.29
OGT 6	AIQINPAFADAHSNLASIHKDSGNIPEAIAS	3	0.72	1.21
OGT 7	AATGEEVPRTIIVTTR	7	0.96	1.19
OGT 8	EAIRISPTFADAYSNMGNTLK	2	1.12	1.16

[a] Geometric mean
[b] Geometric standard deviation

Having demonstrated the selective capture and accurate quantification of *O*-GlcNAc glycosylated peptides, we applied our strategy to the quantification of *O*-GlcNAc peptides from neuronal lysate. Consistent with its ability to inhibit the *O*-GlcNAcase enzyme, we found that PUGNAc strongly upregulated *O*-GlcNAc levels on numerous proteins in primary cortical neurons, as determined by Western blotting with an anti-*O*-GlcNAc antibody (Figure 5.6).
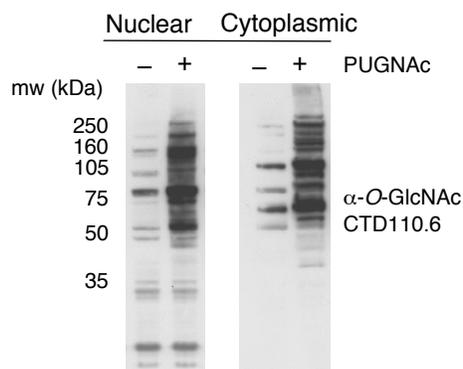
Figure 5.6. PUGNAc treatment of cultured primary cortical neurons significantly elevates *O*-GlcNAc glycosylation levels, as monitored by the anti-*O*-GlcNAc antibody CTD110.6.

In order to identify the proteins undergoing dynamic regulation, neurons treated in the presence or absence of PUGNAc were separated into nuclear and S100 cytoplasmic fractions, chemoenzymatically labeled and digested with trypsin. Digested proteins were dimethyl labeled, fractionated by strong cation-exchange chromatography, and tagged *O*-GlcNAc peptides were enriched via avidin chromatography. In order to normalize for procedural errors, we mixed α-crystallin and OGT into lysates at a ratio of 1:1, prior to chemoenzymatic labeling.

Quantification by orbitrap MS was conducted in tandem with MS/MS analysis, which was used to identify peaks bearing the distinctive biotin-ketone-GlcNAc loss signature characteristic of *O*-GlcNAc peptides. Twenty-two peptides from the nuclear sample and 11 peptides from the corresponding cytoplasmic sample showed an increase in *O*-GlcNAc glycosylation upon PUGNAc stimulation. Interestingly, we found that the presence of PUGNAc did not result in increased *O*-GlcNAc glycosylation on all proteins universally, suggesting that not all *O*-GlcNAc sites are subject to direct reversibility. 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. We also observed decreases in glycosylation on 5 nuclear and 4 cytoplasmic *O*-GlcNAc peptides.

Portions of dynamically regulated peaks from cytoplasmic and nuclear lysates were targeted for sequencing by MS^4 CAD analysis. Figure 5.7A shows a representative MS spectrum of an upregulated *O*-GlcNAc peptide. The CAD MS/MS spectrum of the deuterated, triply charged peptide (*m/z* = 862.3889), indicates the characteristic loss of a biotin-ketone moiety (*m/z* = 1208.43) and biotin-ketone-GlcNAc moiety (*m/z* = 1005.34) (Figure 5.7B). Higher-order MS analysis generated a series of internal cleavages, as well as b and y ions along the amide backbone, that enabled definitive sequencing of the peptide (Figure 5.7C). Database searching identified the peptide as belonging to the protein elongation initiation factor 4G (EIF4G) (Figure 5.7C).
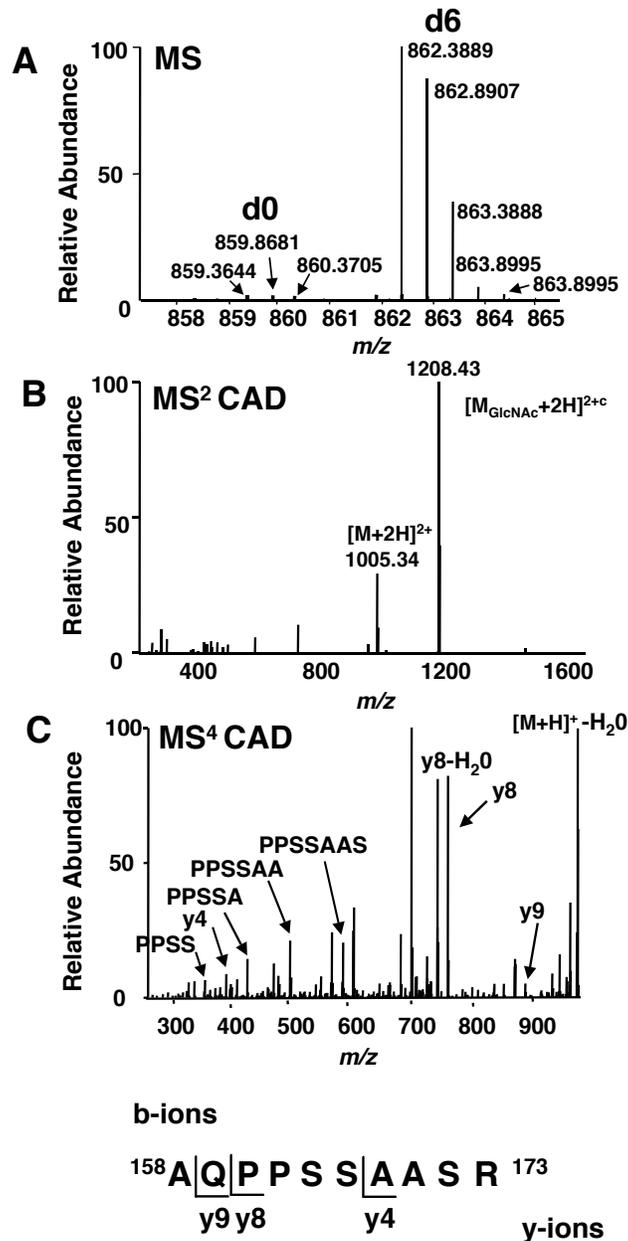
Figure 5.7. Sequencing of quantified *O*-GlcNAc peptides from neuronal lysates by CAD. (A) Orbitrap MS spectrum of a significantly upregulated peptide in response to PUGNAc treatment. (B) CAD MS/MS spectrum of the deuterated peak (*m/z* = 862.3889), showing loss of a ketone-biotin moiety (*m/z*=1208.43) and GlcNAc-ketone-biotin moiety (*m/z* = 1005.34). (C) Fragmentation during MS[4] analysis yielded numerous internal cleavages and several prominent b and y ions, which permitted identification of the peptide as [158]AQPPSSASSR[173] from elongation initiation factor 4G.

As an alternative to MS[4] CAD sequencing, we explored the utility of ETD to

sequence chemoenzymatically tagged *O*-GlcNAc peptides and directly identify sites of

modification. Figure 5.8A shows the MS spectrum of different upregulated O-GlcNAc

modified peptide. In this case, ETD sequencing provided nearly complete sequence

coverage in MS/MS mode, through the generation of c and z ions produced by cleavage

between the amide nitrogen and the peptide backbone α-carbon (Figure 5.8B). Database

searching identified this peptide as belonging to the transcriptional repressor p66.

Importantly, ETD fragmentation preserved the linkage to the tagged O-GlcNAc, which

narrowed down the possible O-GlcNAc site to serines 584 and 586.
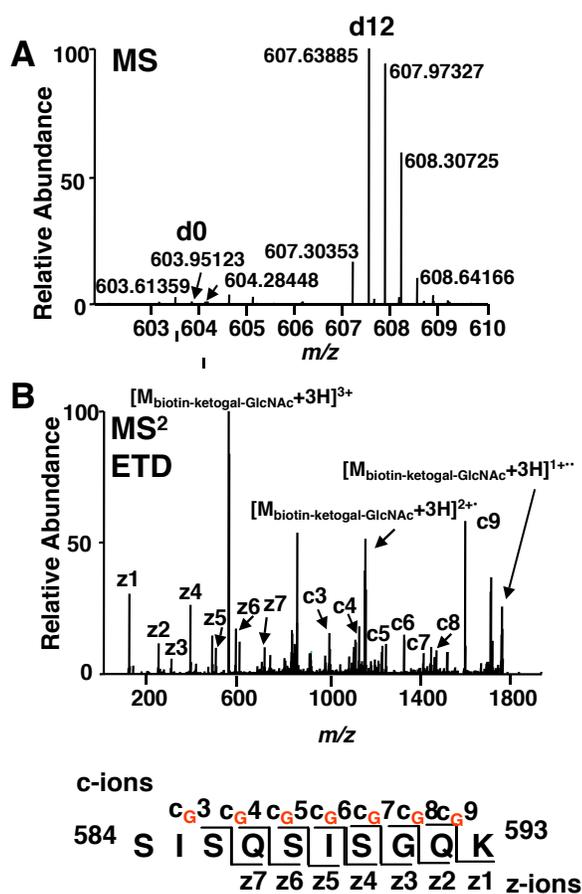
Figure 5.8. Sequencing of quantified O-GlcNAc peptides from neuronal lysate via ETD. (A) MS spectrum of a second upregulated O-GlcNAc peptide from neuronal lysate. (B) ETD MS/MS analysis of the deuterated peak at 607.3885 yielded c and z ions that permitted identification of the peptide as [584]SISQSISGQK[593] from the transcriptional repressor p66. The presence of the tagged GlcNAc moiety on the c series of ions narrowed site identification to S584 and S586.

Using a combination of CAD and ETD, we successfully sequenced seven of the dynamically upregulated peptides (Table 5.2). In addition, we identified one more peptide by ETD MS/MS, which was not observed by in the orbitrap MS analysis.

Table 5.2. Quantification of *O*-GlcNAc glycosylated proteins from PUGNAc-treated neuronal culture

Table 1 Identification and quantification of changes in *O* -GlcNAc glycosylation induced by PUGNAc

Protein	NCBI Entry	Fold Change[a]	Function	Peptide Sequence[b]	Residues	MS Method
BHC80	62645406	1.4	neuronal gene repression, scaffolding	FTPTTLPTSQNSIHPVR	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	*N* -acetyl-D-glucosaminidase	QVAHSGAK	401-408	MS4
p66β	67846054	40.3	transcriptional repression	SISQSISGQK	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

[a] Fold change represents the observed heavy:light ratio averaged over all experiments. See Supplementary Methods for details on statistical analysis.
[b] Potential glycosylation sites determined by ETD are shown in red.

In our previous work, we had identified the steroid receptor coactivator 1 (SRC-1) as *O*-GlcNAc modified in the brain.[1] Here we showed that the same peptide can be regulated by PUGNAc treatment in primary cortical cultures. Importantly, sequencing by ETD permitted identification of the exact site of *O*-GlcNAc-modification as Thr-203. Likewise, we had previously demonstrated that the zinc finger RNA binding protein was *O*-GlcNAc modified.[10] Here, we isolated the site of modification as Ser-61, and found that in cortical neurons, this site of glycosylation could be dramatically upregulated by inhibiting the *O*-GlcNAcase enzyme.

We also find that *O*-GlcNAc glycosylation is reversible on the nucleic acid binding nucleoporin 153 (NUP153). Nucleoporins have been well characterized as *O*-GlcNAc-modified proteins.[48,49] Here, we specifically identified a site of glycosylation in the C-terminal domain of NUP153, which is necessary for docking and trafficking mRNA.[50] Consistent with a role for *O*-GlcNAc in transcriptional repression, we also

found a site of glycosylation on p66β, which interacts with histone tails and mediates

transcriptional repression by the methyl-CpG-binding domain protein MBD2.[51]

In order to confirm that the observed changes were specific to O-GlcNAc rather

than due to PUGNAc-induced changes in protein expression, we treated neurons with

PUGNAc and chemoenzymatically labeled the lysate of these and untreated control

neurons. We then specifically captured O-GlcNAc proteins by streptavidin and probed

for changes in streptavidin-binding after PUGNAc treatment by Western blotting.

Western blot analysis for the proteins O-GlcNAcase and p66β demonstrated that

PUGNAc treatment of neurons had little impact on protein expression. However, it

dramatically influenced streptavidin binding, in a manner consistent with changes in O-

GlcNAc observed by mass spectrometry (Figure 5.9).
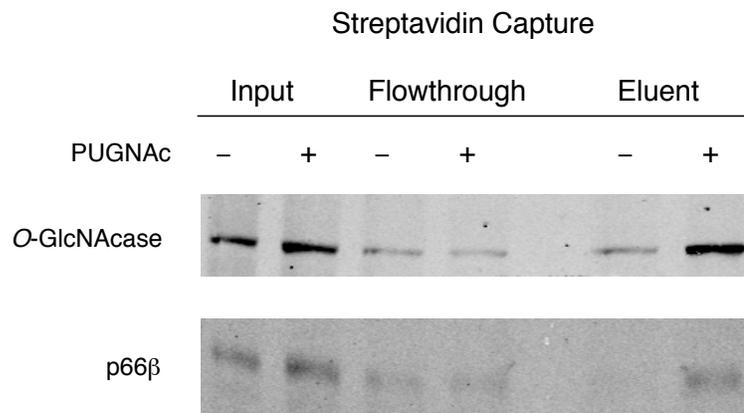
Figure 5.9. Western blot analysis of chemoenzymatically labeled and streptavidin-captured O-GlcNAc neuronal proteins following PUGNAc treatment. PUGNAc induced minimal changes in protein expression in O-GlcNAcase and the transcriptional repressor p66β (input), but had a significant impact on streptavidin-binding (eluent), which signifies upregulation of O-GlcNAc glycosylation on these two proteins. (Western Blot courtesy of P. Clark).

Having demonstrated that we could use the chemoenzymatic strategy to examine

O-GlcNAc regulation from cultured neurons, we applied the strategy to the intact brain.

In particular, given the emerging role of *O*-GlcNAc in the stress response,[22] and its prevalence on many proteins involved in synaptic vesicle cycling (Chapter 4), we examined whether *O*-GlcNAc was responsive to a strong neuronal stimulation. Kainic acid seizure induction is a well-characterized model for the study of epilepsy[52] and has also been used to activate excitatory pathways in the brain that lead to gene expression.[53] We administered kainic acid to rats via intraperitoneal injection and monitored their behavioral responses. We isolated cerebral cortices at three time points: the peak of seizure (~2.5 h post-injection), when animals had resumed some normal resting behavior (~6 h post-injection) and when animals showed nearly identical behavior to saline-injected controls (~10 h post-injection). Western blotting of fractionated cortical lysate with the CTD110.6 antibody revealed that *O*-GlcNAc levels were elevated at ~2.5 h post-injection and returned essentially to basal levels by ~10 h post-injection (Figure 5.10).
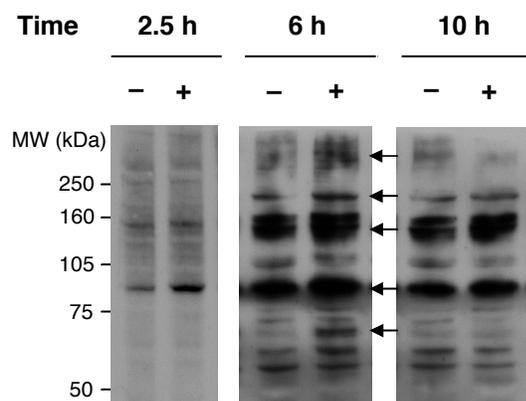
Figure 5.10. Representative *O*-GlcNAc antibody Western blots of nuclear cortical lysate from control and kainic acid treated animals. At ~2.5 h and ~6h post-injection *O*-GlcNAc levels appear to be elevated on several protein bands (marked with arrows), but after prolonged recovery, (10 h post-injection, ~8.5 h after onset of seizing behavior) *O*-GlcNAc levels return to a basal state.

In order to characterize the proteins undergoing dynamic *O*-GlcNAc glycosylation in response to neuronal hyperactivity, we chemoenzymatically-labeled

cortical lysate from the ~6 h post-injection animals, then proteolytically digested and dimethyl labeled the peptides. *O*-GlcNAc peptides were then selectively enriched by avidin affinity chromatography. Thirteen of 83 *O*-GlcNAc peptides detected by MS underwent a robust, reproducible increase in response to excitatory kainic acid stimulation of rats. Using CAD tandem mass spectrometry, we successfully identified 4 of these proteins as well as two proteins that did not appear to undergo changes in glycosylation under these conditions (Table 5.3).

Table 5.3. *O*-GlcNAc glycosylated peptides upregulated in the mammalian cerebral cortex in response to kainic acid administration

Protein	NCBI Entry	Fold Change	*P* Value	n	Function	Peptide Sequence	Residues
EGR-1	6978799	10.1	0.02	2	gene transcription, stress response	ALVETSYPSQTTR	87-99
eIF4G	62658155	5.3	0.02	2	translation elongation	AQPPSSAASR	63-72
GRASP55	51259254	1.8	0.06	2	membrane protein transport, golgi stacking	VPTTVEDR	423-430
Hrb	90101424	1.6	0.03	4	RNA trafficking	SSSADFGSFSTSQSHQTASTVSK	291-313
bassoon	9506427	1.3	0.11	4	synaptic vesicle cycling	SPSTSSTIHISYGQPPTTANYGSQ-TEELPHAPSGPAGSGR	1402-1440
bassoon	9506427	1.5	0.12	2	synaptic vesicle cycling	ASGAGGPPRPELPAGGAR	2283-2300
inositol polyphosphate-4-phosphatase	13591898	1.2	0.56	4	lipid phosphatase	SDQQPPVTR	177-186

[a] Fold change represents the observed heavy:light ratio averaged over all experiments.

We identified a peptide on elongation initiation factor 4G (EIF4G), which was dynamically regulated in response to kainic acid administration. Glycosylation at EIF4G, a molecular scaffold which mediates mRNA interaction with the ribosome,[54] was also strongly upregulated by PUGNAc treatment. Notably, previous work on *O*-GlcNAc had found that other components of the translational machinery are modified by glycosylation, such as the EIF4A binding protein p67, which controls the phosphorylation state of eukaryotic initiation factor 2α (EIF2α) and its activity.[55] Our work suggests that as with transcription, the *O*-GlcNAc modification may play several roles in the translation process and that the modification can be dynamically regulated in response to robust neuronal activity. We also found that the transcription factor early growth response 1 (EGR-1) was *O*-GlcNAc modified and that *O*-GlcNAc levels change

dramatically upon kainate acid treatment. EGR-1 is a known immediate early gene whose expression is activated by a host of stimuli such as fear conditioning and stress. Consistent with earlier reports that EGR-1 expression is upregulated ~2-fold in the cerebral cortex following kainic acid administration,[56] we found that its expression was elevated 1.8 ± 0.2-fold at ~ 6 h post-injection (data not shown). Given that O-GlcNAc glycosylation is upregulated ~ 10-fold, our mass spectrometry data suggest that gene expression changes alone can not account for the dramatic effect of kainate treatment on EGR-1 glycosylation.

Importantly, MS[4] sequencing also identified 20 O-GlcNAc peptides from 6 new and 12 previously characterized O-GlcNAc proteins. Changes in glycosylation for these proteins were either within experimental error, or within a low signal to noise in the orbitrap MS spectra, preventing accurate quantification (Table 5.4).

Table 5.4. O-GlcNAc glycosylated proteins identified from kainic-acid treated animals

Protein	NCBI Entry	Function	Peptide Sequence	Residues
bassoon	9506427	synaptic vesicle cycling	VTQHFAK	1338-1444
CCR4-NOT4	34855140	global transcriptional regulation, mRNA metabolism	SNPVIPISSSNHSAR	329-343
CRMP-2	599966	axonal guidance, neuronal polarity	TVTPASSAK	512-520
erythrocyte protein band 4.1-like 1, isoform L	11067407	cytoskeletal protein	DVLTSTYGATAETLSTSTTTHVTK	1460-1483
HCF	109511332	chromatin-associated factor	QPETYHTYTTNTPTTAR	1232-1248
LMP-1	62988302	contains PDZ and LIM domain	AQPAQSKPQK	28-37
MAP2b	547890	dynamic assembly of microtubles at dendrites	VADVPVSEATTVLGDVHSPAVEGFVGENISGEEK	380-413
O -GlcNAcase	18777747	N -acetyl-D-glucosaminidase	QVAHSGAK	401-408
PDZ-GEF	34857578	GTP/GDP exchange factor for RAP1/2	SSIVSNSSFDSVPVSLHDER	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	AAAATTTATTTTTSGGHPLEFLR	176-198
SH3p8	2293466	SH3 domain binding protein, synaptic vesicle cycling	ITASSSFR	283-290
SRC-1	34863079	coactivation of nuclear receptor transcription	INPSVNPGISPAHGVTR	188-204
SynGAP	34098355	inhibitory regulation of Ras pathway, synaptic strength regulation	QHSQTPSTLNPTMPASER	1121-1138
Ythdf3	109466336	contains YtH domain	IGGDLTAAVTK	145-155

Not all O-GlcNAc glycosylated sites will be amenable to direct peptide-specific quantification, due to low abundance or accessibility to protease digestion. These difficulties underscore the power of the chemoenzymatic strategy to specifically isolate both O-GlcNAc modified peptides and proteins. In a complementary approach to

peptide-specific quantification, current efforts in the laboratory are focused on using the chemoenzymatic strategy to capture intact *O*-GlcNAc glycosylated protein*s* after kainic acid induction. Changes in protein association with streptavidin can then be detected either directly by Western blotting with known antibodies, or my mass spectrometry analysis of captured and proteolytically digested proteins.

5.4. Discussion

In this chapter, we have described an efficient, sensitive strategy for the quantification of *O*-GlcNAc on single proteins, as demonstrated on the transcription factor CREB. The approach capitalizes on the biotin affinity tag for sensitive detection and thus should be applicable for the quantitative study of other low-abundance *O*-GlcNAc proteins that are not amenable to traditional techniques.[42]

Using the approach, we have provided direct evidence that *O*-GlcNAc levels on CREB can be upregulated in response to glucosamine and PUGNAc treatment of pancreatic cells, in agreement with the effect of these compounds on CREB-mediated transcription. The effect of PUGNAc is modest and may represent the fact that glycosylated CREB is generally protected from dynamic removal of *O*-GlcNAc under basal conditions. Different populations of CREB (e.g., within different parts of the cell, or on different promoters) may be differentially regulated by OGT/*O*-GlcNAcase, an attractive possibility given the multiple interacting partners of both enzymes, which may affect their localization and substrate specificity.[57,58]

In addition to studying the quantification of individual *O*-GlcNAc proteins, we have also developed the first quantitative proteomic assays of *O*-GlcNAc glycosylation.

O-GlcNAc dynamics were assayed by a chemoenzymatic strategy that facilitated both the discovery and quantification of *O*-GlcNAc peptides from cultured neurons and the mammalian brain.   In contrast to other approaches, such as lectin affinity chromatography,[29] antibody affinity[26] or BEMAD,[26] the approach is applicable to the study of *O*-GlcNAc dynamics both on intact proteins and at specific sites of modification. Moreover, the approach does not rely on metabolic labeling[28] and thus is amenable to both tissues and whole organisms.

Our proteomic approach featured the use of a simple, efficient and commercially available isotopic label, which had previously been shown to quantify changes in protein expression.[36] Reductive amination of lysines with formaldehyde and $NaCNBH_3$, or their deuterated counterparts, proceeds quantitatively within minutes and shows no side reaction or significant 'isotopic effect' during LC-MS. Because the strategy does not require cell division for incorporation[32] it is particularly useful for the study of postmitotic cells, such as neurons and tissue. Here, it was readily coupled with our chemoenzymatic strategy to specifically quantify enriched *O*-GlcNAc glycosylated peptides cultured neurons and brain tissue.

In our initial experiments we examined the strategy on PUGNAc-treated neuronal lysate. PUGNAc was one of the first *O*-GlcNAcase inhibitors described, and has been used extensively for the study of *O*-GlcNAc.[59] We showed that it could significantly elevate endogenous *O*-GlcNAc levels in neurons. In order to identify affected proteins, we isolated *O*-GlcNAc peptides from both basal and treated cells and examined peptides for changes in glycosylation by orbitrap LC-MS analysis. Interestingly, many peptides did not show an elevation in *O*-GlcNAc levels in response to PUGNAc treatment. This

may be due to the localization and targeting of the *O*-GlcNAcase[58] or the constitutive nature of glycosylation for some protein targets.

In order to identify dynamically regulated peptides, we employed two modes of sequencing, CAD MS/MS analysis followed by multistage MS or ETD MS/MS. We had previously described that the biotin-ketone tag provides a unique signature by CAD MS/MS, which allows unambiguous characterization of *O*-GlcNAc peptides.[10] Peptides displaying the unique signature can then be targeted for sequencing by MS^4 analysis. In contrast to CAD, ETD fragmentation does not release the tagged *O*-GlcNAc group and can be used to narrow down or identify exact sites of glycosylation within peptides.[37] The peptide backbone is directly fragmented even in the presence of the PTM so identification can be made directly from MS/MS analysis on an ion trap mass spectrometer. We found ETD to be highly effective for the fragmentation of higher charged species ($[M+4H]^{4+}$ and above) but less effective for $[M+3H]^{3+}$ and $[M+2H]^{2+}$ species. This was particularly pronounced for tagged *O*-GlcNAc peptides bearing the bulky biotin group and two carbohydrates, and may be due to inadequate charge to mass distribution, which is necessary for effective ETD fragmentation.[*] Our efforts to resolve these issues with the use of endoproteinase Lys-C to generate larger, and more highly charged peptide fragments were hampered by the high rate of false positives when searching highly charged ETD fragment ions by the database querying program SEQUEST. Advances in programs to query ETD data, such as the publicly available OMSSA program (http://pubchem.ncbi.nlm.nih.gov/omssa/) should help address this issue. In addition, the generation of a cleavable biotinylated linker, which would restore

---

[*] Effective ETD fragmention relies on high charge to mass distribution. For example, doubly charged peptides typically yield poor fragmentation compared with more highly charged species. (J. Coon. Personal Communication, 2006).

the lone *O*-GlcNAc group after affinity capture, may help ETD sequencing of less-highly charged glycosylated peptides. With further development, we envision that the ETD strategy will become a powerful tool for the identification of *O*-GlcNAc glycosylated peptides and the mapping of *O*-GlcNAc glycosylation sites.

Using a combination of CAD and ETD, we sequenced 7 of the *O*-GlcNAc peptides that undergo significant increases in glycosylation upon PUGNAc treatment. In addition, we identified another peptide by ETD that was not observed in the orbitrap MS analysis and thus could not be quantified. Among the identified proteins undergoing reversible glycosylation was the enzyme *O*-GlcNAcase. Inhibition of *O*-GlcNAcase using PUGNAc led to a robust increase in OGA glycosylation at Ser-405, suggesting that OGT may be capable of regulating the activity of its antagonizing enzyme. This finding is consistent with a recent report that the two enzymes form a stable association as a gene transcription regulatory complex, which requires a region of OGA that encompasses the identified glycosylation site.[60]

In addition, we identified several proteins involved in mRNA binding and transport. Such processes are of particular interest in neurons, where regulated transport of mRNA from the cell body to dendrites and local translation of mRNA are involved in the strengthening of individual synapses to give rise to synaptic plasticity.[61] In particular, reversible *O*-GlcNAc glycosylation occurred on the zinc finger RNA-binding protein, which is associated with staufen2 granules in neurons and may be important in the early stages of RNA translocation from the nucleus to the dendrites.[62] We also identified a dynamically glycosylated peptide from the C-terminal domain of nucleoporin 153, which is necessary for docking and trafficking of mRNA.[50] The presence of reversible and

potentially regulatory sites of glycosylation on proteins involved in RNA transport supports an expanded role for *O*-GlcNAc glycosylation in translational control and may have important consequences for neuronal signaling and synaptic plasticity.

In order to investigate the potential contribution of *O*-GlcNAc glycosylation to neuronal communication, we examined the effects of excitatory neuronal stimulation on *O*-GlcNAc levels in the brain. We demonstrated for the first time that *O*-GlcNAc glycosylation is regulated *in vivo* by robust excitatory stimulation. The *O*-GlcNAc levels on several proteins increased upon treatment of rats with the glutamate analogue kainic acid and returned to basal levels as the behavioral effects of the treatment dissipated. Using our quantitative proteomics strategy, we found that EGR-1, an immediate early gene and transcription factor important for long-term memory formation[63] and cell survival, undergoes a 10-fold increase in glycosylation.[64] 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, which play critical roles in synaptic plasticity.[65]

We also observed an increase in *O*-GlcNAc glycosylation on the translation factor eIF4G upon kainic acid stimulation. The observation that eIF4G is reversibly glycosylated in neurons is consistent with an emerging role for *O*-GlcNAc in regulation of the stress response.[22] As kainic acid treatment induces excitotoxicity in addition to synaptic potentiation and suppressed translation is a known marker for neuronal excitoxicity,[66] 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 regulated by O-GlcNAc glycosylation, including p67, which binds to the eukaryotic initiation factor 2α (eIF2α) in its glycosylated form and promotes protein synthesis by preventing inhibitory phosphorylation of eIF2α.[55]

Notably, our studies indicate that only a fraction of the O-GlcNAc-modified proteins undergo dynamic glycosylation in response to specific stimuli. Similar heterogeneous responses have been observed in the case of phosphorylation. For example, kainic acid stimulation exerts bidirectional effects on the phosphorylation state of proteins, both enhancing and inhibiting phosphorylation in a substrate-dependent manner.[67,68] These results suggest that OGT and OGA are 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 (including our observation of reversible glycosylation on OGA itself). 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.

The ability of O-GlcNAc to respond to specific extracellular stimuli also 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.[10] 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[29] and colleagues, we also find that 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, is glycosylated.

Finally, our work highlights the emergent interplay between O-GlcNAc glycosylation and phosphorylation. For example, we identified a glycosylated peptide on the protein Bassoon, ^{1402}SPSTSSTIHSYGQPPTTANYGSQTEELPHAPSGPAGSGR^{1440} that was likewise shown to be phosphorylated by phospho-proteomic methods.[69] In addition, we also identified a glycosylated peptide in the axonal guidance protein CRMP-2, which is known to be phosphorylated at two residues within the identified glycopeptide. Interestingly, when hyperphosphorylated within the residues of this peptide, CRMP-2 appears as a component of the neurofibrillary tangles (NFTs) associated with Alzheimer's disease.[70] This is reminiscent of the microtubule-associated protein tau, which is also O-GlcNAc glycosylated, but which exists in hyperphosphorylated form in the AD brain.[71] Deciphering the mechanisms that regulate the interplay of glycosylation and phosphorylation for these two and other proteins may have important ramifications for the study of Alzheimer's disease and other neurological disorders.

Overall, we have shown that O-GlcNAc is both reversible in neurons and dynamic in the brain in response to robust neuronal stimulation. Understanding the mechanisms by which kainic acid administration modulates O-GlcNAc levels will be an

important next step in understanding *O*-GlcNAc regulation in the brain. In addition, the
chemoenzymatic approach should prove useful for studying *O*-GlcNAc dynamics in
response to stimuli such as behavioral and memory training protocols, or other modes of
excitotoxicity that might identify *O*-GlcNAc as a mediator of the stress response in
neuronal tissue. Although we have only begun to understand the processes that regulate
O-GlcNAc, this strategy should serve as a valuable tool in the study of this PTM, as well
as in identifying its role within the rapidly expanding network of PTMs that govern cell
function.

5.5 Experimental Methods

Preparation of HIT-T15 Cell Extracts

HIT-T15 cells were cultured in 37 ^{o}C humidified air with 5% CO_2 in DMEM
supplemented with fetal bovine serum (10%), penicillin (100 U/mL) and streptomycin
(100 μg/mL). Nine hours prior to lysis, the culture medium was supplemented with 10
mM glucosamine (500 mM stock in 100 mM HEPES pH 7.4) or 100 μM PUGNAc (10
mM stock in ddH$_2$0) (Toronto research chemicals). Six hours prior to lysis 10 μM
forskolin (Alexis biochemicals) (50 mM stock in DMSO) was added to the culture
medium. After treatment, cells from a 100 mm dish were trypsinized and pelleted. The
pellet was resuspended in 0.1 mL of boiling lysis buffer (1% SDS with protease inhibitor
cocktail Complete, Roche), sonicated for 3 × 3 sec, and boiled for 8 min. After
centrifugation at 21,500 × g for 5 min, the supernatant was collected as denatured HIT-
T15 extract.

Chemoenzymatic Labeling of Proteins from HIT-T15 cells

One volume of denatured HIT-T15 extract was diluted five-fold with 100 mM HEPES pH 7.9, 20% Triton-X100, 5 M NaCl, 100 mM $MnCl_2$, 25 mM adenosine 5'-diphosphate, 25X Complete protease inhibitor cocktail and 250 mM phenylmethylsulfonyl fluoride (PMSF) to a final concentration of 0.2% SDS, 10 mM HEPES pH 7.9, 1.8% Triton-X100, 100 mM NaCl, 8 mM $MnCl_2$, 1.25 mM adenosine 5'-diphosphate, 1X Complete protease inhibitor cocktail and 1 mM PMSF. Diluted extract was then supplemented with 10 mM analogue **1** and Y289L GalT to final concentrations of 0.5mM analogue **1** and 20-40 ng/µL Y289L GalT. The reaction mixture was incubated at 4 °C for 12 h, and dialyzed into buffer A (10 mM HEPES pH 7.9, 5 M urea) 3×4 h at 4 ºC. Following dialysis, the sample was acidified to pH 4.8 by adding of 2.78 M NaOAc pH 3.9 to a final concentration of 50 mM and mixed for 10 min. Pepstatin was added to a final concentration of 5 µg/mL and the aminooxy biotin derivative (30 mM stock in ddH_20, Dojindo) was then added to a final concentration of 2.75 mM. After incubation at room temperature for 20 h, the sample dialyzed into CREB IP buffer (10 mM HEPES pH 7.5, 100 mM KCl, 1 mM EDTA and 0.2% Triton-X100, 1 mM PMSF) 2×2 h and 1×12 h at 4 ºC.

CREB Immunoprecipitation

Dialyzed HIT-T15 cell extract was supplemented with protease inhibitors (1X Complete protease inhibitor cocktail, 1 mM PMSF), and protein concentration was assayed by BCA assay (Pierce). Protein concentration was adjusted for control and experimental samples with excess CREB IP buffer such that the concentrations and total amounts of

lysate for immunoprecipitation were identical. Lysates were pre-cleared with protein A sepharose beads (30 µL/100 µg of protein) for 2 h at 4 °C. After centrifugation by table-top centrifuge, the supernatant was collected and incubated with CREB antibody (rabbit polyclonal, Upstate Biotechnology) (1.5 µg/100 µg of proteins) for 4 h at 4 °C. Extracts and antibody were then incubated with 30 µL protein A sepharose for 1-1.5 h at 4 °C. Flowthrough was removed and beads were washed 3×5min with 1 mL CREB IP buffer at 4 °C, 3×5 min with 1 mL PBS at 4 °C, and 1×5 min with 1 mL 50mM Na_2HPO_4 pH 7.5 at 4 °C. After washing, the beads were boiled for 2×3 min in 3.5 volumes of 1× SDS-PAGE loading buffer (beads were briefly vortexed after the first three minutes and reboiled). After centrifugation at $2,000 \times g$ for 1 min, the supernatant was collected as the captured material.

Western Blotting

Captured material from CREB immunoprecipitation was resolved by SDS-PAGE and transferred to nitrocellulose membranes. Membranes were blocked for 1 h at RT in 5% Milk, PBS-T(0.05%), followed by incubation with anti-CREB antibody (mouse monoclonal, Chemicon International) at 1:1000 dilution in blocking buffer for 1 h at RT. Blots were then rinsed in PBS-T(0.05%) and washed 3×5 min in PBS-T(0.05%), then incubated with G anti-M-HRP IgG secondary antibody at a 1:10,000 dilution in blocking buffer for 1 h at RT. Blots were then rinsed in PBS-T(0.05%), washed 2×15 min in PBS-T(0.05%) and 3×5min in PBS-T(0.05%) and developed by chemiluminescence and exposure to film. Following exposure to film, blots were stripped for 45 min at 60 °C in 250 mL of stripping buffer (2% SDS, 5 mM Na_2HPO_4 pH 7.5, 2 mM β-mercaptoethanol)

and reprobed with Streptavidin-HRP as follows: blots were blocked for 1 h at RT in 5% BSA TBS-T(0.1%) then incubated with Streptavidin-HRP (Pierce) at a 1:25,000 dilution in TBS-T(0.1%) for 1 h at RT (unstripped blots were incubated in Streptavidin-HRP at a concentration of 1:50,000). Blots were subsequently rinsed with TBS-T(0.1%) and washed 2 × 15 min, 3 × 5 min in TBS-T(0.1%) before exposure to film by chemiluminescence.

Animal Care and Kainic Acid Administration

Male Long Evans rats from Charles River laboratories (7 week-old, 190-200 g) were injected with 10-11 mg/kg kainic acid (axxora, San Diego CA) (neutralized with NaOH and sterile-filtered). Kainic acid (5 mg/mL concentration) was administrated to the peritoneum and animals were housed separately, on paper towel bedding and closely monitored for behavioral changes. Peak of seizure activity (usually ~1.5 h post-injection) was marked by shaking, rearing posture and significant frothing by mouth. Animals were sacrificed at three time points post-injection, with paired animals demonstrating similar kainic-acid induced behavior. First, 1 h after onset of seizure, while animals were still displaying seizure behavior. Second, ~4.5 after onset of seizure activity, when animals had begun to display characteristics of controls (responsiveness to cage tapping). Third, ~8.5 h after onset of seizure activity, at which point animals were largely indistinguishable from controls and were responsive to cage tapping, and were eating, and resting. At these time points, animals were decapitated and brains immediately dissected for cortices, hippocampi and cerebella, on dry ice. Brains were flash frozen in liquid N_2 and stored at -80 °C until further use. Control animals were treated identically, except that they were peritoneally injected with phosphate-buffered saline.

The animal protocol was approved by the Institutional Animal Care and Use Committee at Caltech, and all procedures were performed in accordance with the Public Health Service Policy on Humane Care and Use of Laboratory Animals.

Preparation of Rat Cortical Extract from Seized Animals

The cortices of control and seized Long Evans rats (Charles River Laboratories) were weighed frozen, allowed to briefly thaw on ice, and sliced by razor blade into three portions. They were subsequently lysed into 5 volumes of homogenization buffer by 4-5 strokes of manual dounce homogenization on ice, and 8 strokes of mechanical homogenization at 700 rpm at 4 ºC. Homogenate was fractionated into nuclear and S100 cytoplasmic components as described by Dignam et al,[72] except that Complete protease inhibitor cocktail tablets (Roche, Indianapolis IN), phosphatase inhibitors, 10 µM PUGNAc (Toronto Research Chemicals) and the hexosaminidase inhibitor 50 mM GlcNAc[73] were added to the buffers. In some cases, crude nuclear pellets from seized and control animals were washed (with homogenization buffer not containing PUGNAc or GlcNAc) and lysed directly into boiling 1% SDS, by sonication. Prior to labeling, cytoplasmic extracts and nuclei not lysed directly into SDS were dialyzed into 20 mM HEPES pH 7.9, 0.1 M KCl, 0.2 mM EDTA, 0.2% Triton X-100, 0.5mM (PMSF) phenymethylsulfonylfluoride. Before labeling, portions of the nuclear and cytoplasmic samples were removed for Western blotting with the anti-*O*-GlcNAc CTD110.6 antibody (Covance) or the anti-EGR-1 antibody (Cell Signaling).

Preparation of Cortical Culture. Cortical neuronal cultures were prepared using a

modified version of the Goslin and Banker [74] protocol. Embryos at the E18/E19 stage were obtained from timed-pregnant Sprague-Dawley rats, and the cortex from each embryo was dissected. In some cases, two to three animals were combined in one mixed preparation, in which cases volumes below were appropriately scaled. All the cortices from one preparation were transferred to a 15 mL conical tube containing 4.5 mL of ice-cold Calcium and Magnesium Free-Hank's Balanced Salt Solution (CMF-HBSS) (Gibco). Trypsin (2.5%, no EDTA; Gibco) was added to 5 mL, and the tissue was digested for 15 min at 37 °C. The trypsin solution was removed and the tissue rinsed with 5 mL of warm CMF-HBSS 3×. The tissue was then dissociated in 5 mL of CMF-HBSS by passing through a 2 mL serological pipet 5× and a P1000 pipet tip 20×. Tissue was allowed to settle, and supernatant was filtered. Settled material was dissociated in 5 mL of fresh CMF-HBSS by passing through a P1000 pipet tip 10× and undissociated tissue was removed by filtration. The cells were counted with a hemacytometer and plated on 100 mM culture dishes, which had been previously coated with 0.1 mg/mL sterile-filtered poly-D,L-Lysine (Sigma) in 50 mM $Na_2B_4O_7$ pH 10. Cells were plated at a concentration of 8-12 $X10^6$ cells/100 mM dish in 500 μL Eagle's Medium (MEM) (Gibco) supplemented with 10% Fetal bovine serum (Gibco) and penicillin (100 U/mL)/streptomycin (100 μg/mL). (Gibco). After 30 min, cells were supplemented with 10 mL of Neurobasal Medium (Gibco) containing 1 × b27 serum-free supplement (Gibco), 2 mM L-glutamine (Gibco), 1X antibiotic/antimycotic (Gibco), penicillin (100 U/mL)/streptomycin (100 U/mL) (Gibco) and 500 μM kyneuric acid (in 1 N NaOH). Cells were maintained for 4 days at 5% CO_2/ 95% O_2 at 37 ºC and media was replaced on the second day and just before PUGNAc treatment.

PUGNAc (Toronto research chemicals) (10 mM solution) was added to a final concentration of 100 μM. After 12 h of incubation, cells were trypsinized and pelleted. After removing media by aspiration the pellet was washed in 1 mL of HEPES-buffered Saline and lysed as described by Dignam et al,[72] with the following modifications. Complete protease inhibitor cocktail tablets, phosphatase inhibitors, 10 μM PUGNAc and the hexosaminidase inhibitor 50 mM GlcNAc[73] were added to the buffers. Crude nuclear pellets were washed (with homogenization buffer not containing PUGNAc or GlcNAc) and lysed directly into boiling 1% SDS by sonication for 3 × 3 sec. After centrifugation at 21,500 × g for 5 min, the supernatant was collected as denatured nuclear extract. Prior to labeling, cytoplasmic extracts were dialyzed into 20 mM HEPES pH 7.9, 0.1 M KCl, 0.2 mM EDTA, 0.2% Triton X-100, 0.5mM PMSF. Basal neurons were treated identically except that water rather than PUGNAc was added to neuronal culture.

Chemoenzymatic Labeling of Neuronal and Brain Extracts

α-Crystallin, 5 mg/mL (a ~ 1:1 mixture of A and B crystallin), (Sigma) was denatured in 1% SDS with 25 mM DTT, boiled for 5 min and added to extracts prior to labeling (6.5-11μg). OGT (0.8 mg/mL in 50 mM NaH$_2$PO$_4$ pH 7.5, 1 mM DTT, 20% glycerol) from *sf9* cells[39] was denatured in 1% SDS with 25mM DTT, boiled for 5 min and added to extracts prior to labeling (0.8-2 μg). Extracts (500 μg-3 mg; 1-3 mg/mL) containing α-Crystallin and OGT standard proteins were supplemented with 5 mM MnCl$_2$, 0.5 mM ketone probe, and Y289L GalT (60 ng/μL), with 1× Complete protease inhibitors and 1 mM PMSF and incubated for 12-14 h at 4 °C. Denatured extracts, in 1% SDS, were first diluted 5-fold with 100 mM HEPES pH 7.9, 5 M NaCl and 20% Triton-X100 to a

final concentration of 10 mM HEPES pH 7.9, 100 mM NaCl and 1.8% Triton-X100. Following enzymatic labeling, extracts were dialyzed into denaturing buffer (5 M urea, 10 mM HEPES pH 7.5; 3 × 3 h). The pH was adjusted with 2.7 M NaOAc pH 3.9 (final concentration 50 mM, pH 4.8). Aminooxy biotin (30 mM) was added to a final concentration of 2.75 mM, and the reactions were incubated for 20-24 h at RT. Extracts were dialyzed (2 × 2 h, 1× 10 h) into 7 M urea, 10mM HEPES pH 7.5 at RT followed by 2 M urea, 50 mM NH_4HCO_3 pH 8 (3 × 3 h) at 4 °C.

Proteolytic Digestion

Dialyzed extracts were reduced in 10 mM DTT (500 mM stock in 50 mM NH_4HCO_3 pH 8) for 1 h at RT, alkylated in 20mM iodoacetamide (500 mM stock in 50 mM NH_4HCO_3 pH 8) for 1 h at RT, and incubated with 20 mM DTT for 1 h at RT to react with excess iodoacetamide. The extract solution was then centrifuged at 15,000 rpm for 5 min to remove any insoluble material. Protein concentration was measured via Biorad assay. Control and experimental extracts were adjusted to identical volumes/protein concentrations with excess 2 M urea, 50 mM NH_4HCO_3 pH 8 and then diluted with 50 mM NH_4HCO_3 pH 8 to a final urea concentration of 1M. Sequencing-Grade Trypsin (Promega) was added to a final extract: trypsin ratio of 20:1. Trypsin concentration was always maintained above 0.01 mg/mL and digestion was carried out in a water bath at 37 °C for 12-14 h.

Dimethyl Labeling

Digested extracts were desalted using a Sep-Pak C18 cartridge (1 cc bed volume; Waters) as follows: cartridges were equilibrated with 1 mL of 0.1% AcOH in water, and acidified extracts were loaded onto cartridges. Cartridges were washed with 1.5 mL water and peptides were eluted in 500 µL of 60% aqueous CH_3CN, concentrated by speedvac to a volume of 50 µL, and diluted with 450 µL of 1 M HEPES pH 7.5. To begin the reactions, the samples were mixed with 40 µL of a 600 mM stock of $NaCNBH_3$ or $NaCNBD_3$ (Sigma) in water, followed by 40 µL of 4% aqueous formaldehyde (Mallinckrodt Chemicals) or 40 µL of 4% aqueous formaldehyde-d2 (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) as described above and the eluents (500 µL in 60% aqueous CH_3CN, 0.1% AcOH) were concentrated by speedvac to a volume of 100 µL.

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 KH_2PO_4 containing 25% CH_3CN. Fractionated peptides were enriched via 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 Na_2HPO_4, 1.76 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4), 2 mL of 1X PBS,

1.5 mL of manufacturer wash buffer 2 and 1 mL of ddH$_2$0. Avidin-enriched peptides were eluted as described by the manufacturer.

Orbitrap LC-MS Analysis of Avidin-Enriched Biotinylated Peptides

Automated nanoscale reversed-phase HPLC/ESI/MS was performed as described in previous chapters. Approximately 1/5 of the avidin-enriched peptides from each cation exchange fraction was loaded onto a 360 μm O.D. X 75 μm I.D. precolumn packed with 4 cm of 5 μm Monitor C18 particles (Column Engineering) at a flow rate of 4 μL/min. After desalting, the vent was closed and peptides eluted to a 360 μm O.D. X 75 μm I.D. analytical column with integrated emitter tip (10 cm of 5 μm C18, ca. 5 μm tip). The chromatographic profile was from 100% solvent A (0.1% aqueous AcOH) to 50% solvent B (0.1% AcOH in CH$_3$CN) in 30 min. The flow rate through the analytical column was approximately 100 nL/min. 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[4] 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[4] 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. 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. [75] Standard peptide ratios were tested for goodness of fit to the log-normal distribution via the Kolmogorov-Smirnov 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 versus light forms of standard peptides within experiments. Where applicable, paired t-test was performed on normalized, log-transformed peptide ratios to determine significance.

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, and annotated CAD and ETD spectra will be presented in a published manuscript.

References:

1.   Khidekel, N. & Hsieh-Wilson, L. C. A 'molecular switchboard'--covalent
        modifications to proteins and their impact on transcription. *Org Biomol Chem* 2,
        1-7 (2004).
2.   Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational
        modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed
        Engl* 44, 7342-72 (2005).
3.   Griffith, L. C. Calcium/calmodulin-dependent protein kinase II: an unforgettable
        kinase. *J Neurosci* 24, 8391-3 (2004).
4.   Purcell, A. L. & Carew, T. J. Tyrosine kinases, synaptic plasticity and memory:
        insights from vertebrates and invertebrates. *Trends Neurosci* 26, 625-30 (2003).
5.   Ehlers, M. D. Activity level controls postsynaptic composition and signaling via
        the ubiquitin-proteasome system. *Nat Neurosci* 6, 231-42 (2003).
6.   Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes
        in histone biology and beyond. *Nature* 425, 475-9 (2003).
7.   Shalizi, A., Gaudilliere, B., Yuan, Z., Stegmuller, J., Shirogane, T., Ge, Q., Tan,
        Y., Schulman, B., Harper, J. W. & Bonni, A. A calcium-regulated MEF2
        sumoylation switch controls postsynaptic differentiation. *Science* 311, 1012-7
        (2006).
8.   Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the
        "*O*-GlcNAc code". *Sci STKE* 2005, re13 (2005).
9.   Shafi, R., Iyer, S. P., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart,
        G. W. & Marth, J. D. The *O*-GlcNAc transferase gene resides on the X
        chromosome and is essential for embryonic stem cell viability and mouse
        ontogeny. *Proc Natl Acad Sci U S A* 97, 5735-9 (2000).
10.   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, 13132-7 (2004).
11.   Whelan, S. A. & Hart, G. W. Proteomic approaches to analyze the dynamic
        relationships between nucleocytoplasmic protein glycosylation and
        phosphorylation. *Circ Res* 93, 1047-58 (2003).
12.   Zachara, N. E. & Hart, G. W. Cell signaling, the essential role of *O*-GlcNAc!
        Biochim Biophys Acta (2006).
13.   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).
14.   Iyer, S. P. N. & Hart, G. W. Dynamic nuclear and cytoplasmic glycosylation:
        Enzymes of *O*- GlcNAc cycling. *Biochemistry* 42, 2493-2499 (2003).
15.   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).
16.   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).

17. Cole, R. N. & Hart, G. W. Cytosolic *O*-glycosylation is abundant in nerve terminals. *J Neurochem* 79, 1080-9 (2001).

18. Akimoto, Y., Comer, F. I., Cole, R. N., Kudo, A., Kawakami, H., Hirano, H. & Hart, G. W. Localization of the *O*-GlcNAc transferase and *O*-GlcNAc-modified proteins in rat cerebellar cortex. *Brain Res* 966, 194-205 (2003).

19. 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).

20. Kearse, K. P. & Hart, G. W. Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins. *Proc Natl Acad Sci U S A* 88, 1701-5 (1991).

21. Kneass, Z. T. & Marchase, R. B. Neutrophils exhibit rapid agonist-induced increases in protein-associated *O*-GlcNAc. *J Biol Chem* 279, 45759-65 (2004).

22. Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D. & Hart, G. W. Dynamic *O*-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* 279, 30133-42 (2004).

23. Griffith, L. S. & Schmitz, B. *O*-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to pertubations of phosphorylation. *Eur J Biochem* 262, 824-31 (1999).

24. 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).

25. Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for *O*- linked N-acetylglucosamine. *Analytical Biochemistry* 293, 169-177 (2001).

26. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. & Hart, G. W. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Molec Cell Proteomics* 1, 791-804 (2002).

27. 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).

28. Nandi, A., Sprung, R., Barma, D. K., Zhao, Y., Kim, S. C. & Falck, J. R. Global identification of *O*-GlcNAc-modified proteins. *Anal Chem* 78, 452-8 (2006).

29. Vosseller, K., Trinidad, J. C., Chalkley, R. J., Specht, C. G., Thalhammer, A., Lynn, A. J., Snedecor, J. O., Guan, S., Medzihradszky, K. F., Maltby, D. A., Schoepfer, R. & Burlingame, A. L. *O*-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 5, 923-34 (2006).

30. 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).

31. Jhala, U. S., Canettieri, G., Screaton, R. A., Kulkarni, R. N., Krajewski, S., Reed, J., Walker, J., Lin, X., White, M. & Montminy, M. cAMP promotes pancreatic

beta-cell survival via CREB-mediated induction of IRS2. *Genes Dev* 17, 1575-80 (2003).

32. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1, 376-86 (2002).

33. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154-69 (2004).

34. Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22, 1139-45 (2004).

35. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M. & Jensen, O. N. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 4, 310-27 (2005).

36. Hsu, J. L., Huang, S. Y., Chow, N. H. & Chen, S. H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 75, 6843-52 (2003).

37. 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 U S A* 101, 9528-33 (2004).

38. Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat Rev Mol Cell Biol* 2, 599-609 (2001).

39. 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).

40. Lamarre-Vincent, N. (2006).

41. Lamarre-Vincent, N. (2005).

42. 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).

43. Che, F. Y. & Fricker, L. D. Quantitation of neuropeptides in Cpe(fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry. *Anal Chem* 74, 3190-8 (2002).

44. Hardman, M. & Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem* 75, 1699-705 (2003).

45. 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, 977-82 (2006).

46. Zhang, R., Sioma, C. S., Wang, S. & Regnier, F. E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal Chem* 73, 5142-9 (2001).

47. Roquemore, E. P., Dell, A., Morris, H. R., Panico, M., Reason, A. J., Savoy, L. A., Wistow, G. J., Zigler, J. S., Jr., Earles, B. J. & Hart, G. W. Vertebrate lens alpha-crystallins are modified by *O*-linked N-acetylglucosamine. *J Biol Chem* 267, 555-63 (1992).

48. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Gerace, L. & Hart, G. W. Nuclear pore complex glycoproteins contain cytoplasmically disposed *O*-linked N-acetylglucosamine. *J Cell Biol* 104, 1157-64 (1987).

49. Miller, M. W., Caracciolo, M. R., Berlin, W. K. & Hanover, J. A. Phosphorylation and glycosylation of nucleoporins. *Arch Biochem Biophys* 367, 51-60 (1999).

50. Bastos, R., Lin, A., Enarson, M. & Burke, B. Targeting and function in mRNA export of nuclear pore complex protein Nup153. *J Cell Biol* 134, 1141-56 (1996).

51. 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).

52. Ben-Ari, Y. & Cossart, R. Kainate, a double agent that generates seizures: two decades of progress. *Trends Neurosci* 23, 580-7 (2000).

53. Nedivi, E., Hevroni, D., Naot, D., Israeli, D. & Citri, Y. Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature* 363, 718-22 (1993).

54. Hernandez, G. & Vazquez-Pianzola, P. Functional diversity of the eukaryotic translation initiation factors belonging to eIF4 families. *Mech Dev* 122, 865-76 (2005).

55. 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, 5453-60 (2003).

56. Beckmann, A. M. & Wilce, P. A. Egr transcription factors in the nervous system. *Neurochem Int* 31, 477-510; discussion 517-6 (1997).

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

58. Wells, L., Gao, Y., Mahoney, J. A., Vosseller, K., Chen, C., Rosen, A. & Hart, G. W. 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).

59. 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, 3611-7 (1998).

60. Whisenhunt, T. R., Yang, X., Bowe, D. B., Paterson, A. J., Van Tine, B. A. & Kudlow, J. E. Disrupting the enzyme complex regulating *O*-GlcNAcylation blocks signaling and development. *Glycobiology* 16, 551-63 (2006).

61. Ule, J. & Darnell, R. B. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr Opin Neurobiol* 16, 102-10 (2006).

62. Elvira, G., Massie, B. & DesGroseillers, L. The zinc-finger protein ZFR is critical for Staufen 2 isoform specific nucleocytoplasmic shuttling in neurons. *J Neurochem* 96, 105-17 (2006).

63. Jones, M. W., Errington, M. L., French, P. J., Fine, A., Bliss, T. V., Garel, S., Charnay, P., Bozon, B., Laroche, S. & Davis, S. A requirement for the immediate

early gene Zif268 in the expression of late LTP and long-term memories. *Nat Neurosci* 4, 289-96 (2001).

64. Thiel, G. & Cibelli, G. Regulation of life and death by the zinc finger transcription factor Egr-1. *J Cell Physiol* 193, 287-92 (2002).

65. 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).

66. Marin, P., Nastiuk, K. L., Daniel, N., Girault, J. A., Czernik, A. J., Glowinski, J., Nairn, A. C. & Premont, J. Glutamate-dependent phosphorylation of elongation factor-2 and inhibition of protein synthesis in neurons. *J Neurosci* 17, 3445-54 (1997).

67. Lubin, F. D., Johnston, L. D., Sweatt, J. D. & Anderson, A. E. Kainate mediates nuclear factor-kappa B activation in hippocampus via phosphatidylinositol-3 kinase and extracellular signal-regulated protein kinase. *Neuroscience* 133, 969-81 (2005).

68. Misonou, H., Mohapatra, D. P., Park, E. W., Leung, V., Zhen, D., Misonou, K., Anderson, A. E. & Trimmer, J. S. Regulation of ion channel localization and phosphorylation by neuronal activity. *Nat Neurosci* 7, 711-8 (2004).

69. Collins, M. O., Yu, L., Coba, M. P., Husi, H., Campuzano, I., Blackstock, W. P., Choudhary, J. S. & Grant, S. G. Proteomic analysis of in vivo phosphorylated synaptic proteins. *J Biol Chem* 280, 5972-82 (2005).

70. 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, 4267-75 (2000).

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

72. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11, 1475-89 (1983).

73. Chou, T. Y., Hart, G. W. & Dang, C. V. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J Biol Chem* 270, 18961-5 (1995).

74. Goslin, K. & Banker, G. *Culturing Nerve Cells* (G. Banker & K. Goslin, eds.) (MIT Press, Cambridge, MA, 1991).

75. Molina, H., Parmigiani, G. & Pandey, A. Assessing reproducibility of a protein dynamics study using in vivo labeling and liquid chromatography tandem mass spectrometry. *Anal Chem* 77, 2739-44 (2005).

Appendix I

Exploring *O*-GlcNAc Dynamics from Neuron Culture and the Brain

AI.1 Background and Introduction

Several lines of evidence suggest that *O*-GlcNAc glycosylation, like phosphorylation, is a dynamic modification. As described in Chapter 5, early studies suggested that *O*-GlcNAc is a highly dynamic modification, with a turnover rate that exceeds that of the protein backbone.[1] Moreover, *O*-GlcNAc levels in lymphocyte cells were shown be responsive to mitogens.[2] Evidence also suggests that *O*-GlcNAc levels can rapidly respond (within minutes) to ligand binding on neutrophils.[3] Finally, *O*-GlcNAc levels in mammalian cells showed marked upregulation in response to numerous cell stresses, and perturbation of the *O*-GlcNAc machinery decreased cell survival in response to stress.[4] In cultured cerebellar neurons, *O*-GlcNAc levels of cytoskeletal-associated proteins were shown to respond reciprocally to PKA and PKC activators and inhibitors, highlighting the apparent antagonism between the two modifications, in some contexts.[5] In addition, okadaic acid, an inhibitor of protein phosphatase-2A (PP2A) (and less robustly protein phosphatase-1) (PP1)[6] was shown to downregulate *O*-GlcNAc levels in a neuroblastoma cell line.[7] Finally, *O*-GlcNAc levels are decreased in the Alzheimer's brain, which may be a reflection of the altered glucose uptake/metabolism associated with Alzheimer's disease.[8] Interestingly, perturbing *O*-GlcNAc levels with the inhibitor PUGNAc induces phosphorylation of the microtubule associated protein tau, at many sites. As hyperphosphorylated tau has been implicated in the pathology of Alzheimer's disease, researches have speculated that perturbations in *O*-GlcNAc in the Alzheimer's brain is one of the mechanisms by which tau may become dysfunctional. The interplay of

O-GlcNAc and phosphorylation on tau underscores the relationship between the modifications (Chapter 1 and Chapter 5), which may have important consequences for cell function

Our studies have been focused on understanding *O*-GlcNAc in neuronal tissue. As such , we investigated stimuli that might activate or repress *O*-GlcNAc glycosylation, as well the relationship between *O*-GlcNAc and phosphorylation in neurons. We found that global *O*-GlcNAc levels, as monitored by the anti-*O*-GlcNAc antibody CTD110.6, could readily be manipulated by PUGNAc and glucosamine, a precursor of UDP-GlcNAc in neuronal tissue, suggesting that *O*-GlcNAc is reversible in these cells. In addition, we found that *O*-GlcNAc levels were responsive to the PP1/PP2A inhibitor calyculin A within a matter of minutes, suggesting that phosphorylation pathways can impact *O*-GlcNAc and may be involved in its regulation.

AI.2 Characterizing *O*-GlcNAc Dynamics in Adult Hippocampal Slices

For our initial studies, we assayed *O*-GlcNAc dynamics in hippocampal slices, choosing this region of the brain for its particularly significant role in processes such as learning and memory and its reported high abundance of the *O*-GlcNAc enzymes.[9] As Table AI.1 shows, the majority of stimuli had no reproducible effect on *O*-GlcNAc levels as monitored by the anti-*O*-GlcNAc antibody CTD110.6. PUGNAc and glucosamine both elevated *O*-GlcNAc levels suggesting that *O*-GlcNAc is dynamic and reversible in acute slice preparations. Moreover, we did find a pronounced effect on *O*-GlcNAc levels by the PKA inhibitor H89.

Table AI.1. Pharmacological treatment of adult hippocampal slices can modulate *O*-GlcNAc levels as measured by *O*-GlcNAc antibody Western blotting

Stimulus	Known Target of Stimulus	Cell Fractionation	Time	Observed Effect	Control
15/20 mM Gluocsamine	UDP-GlcNAc levels	Whole-Cell SDS	3 h	upregulation	–
200*µ*M PUGNAc	O-GlcNAcase inhibition	Whole-Cell SDS	3 h	upregulation	–
Glucose (5mM versus 10mM)	UDP-GlcNAc	Whole-Cell SDS	3 h	no change	–
50mM Forskolin	cAMP pathway PKA activation	Whole-Cell SDS	5 min&10min	no change	phospho-GlUR1 (s845)
Okadaic acid 0.1*µ*M	PP2A inhibition	Whole-Cell SDS	3 h	no change	phospho-tau (s212)
Genistein 100*µ*M	tyrosine kinase inhibition	Whole-Cell SDS Nuclear/S2	1.5 h	no change	–
H89 20*µ*M	PKA inhibition	Whole-Cell SDS Nuclear/S2	1 h	no change upregulation/upregulation	–
KCl	depolarization mutliple pathways	Whole-Cell SDS	30 min	no change	–
Lavendustin A 10*µ*M	tyrosine kinase inhibition	Whole-Cell SDS	1 h	no change	–
PD98059 50*µ*M	MEK inhibition	Whole-Cell SDS	50 min	no change	–

H89 appeared to activate *O*-GlcNAc glycosylation on several proteins, as seen after fractionation into a cytosolic (S2) and nuclear fraction. SDS lysis of coordinately treated slices shows only a subtle change in *O*-GlcNAc on high molecular weight proteins (Figure AI.1) suggesting that whole-cell lysis may mask compartmentalized changes in glycosylation in the cell.
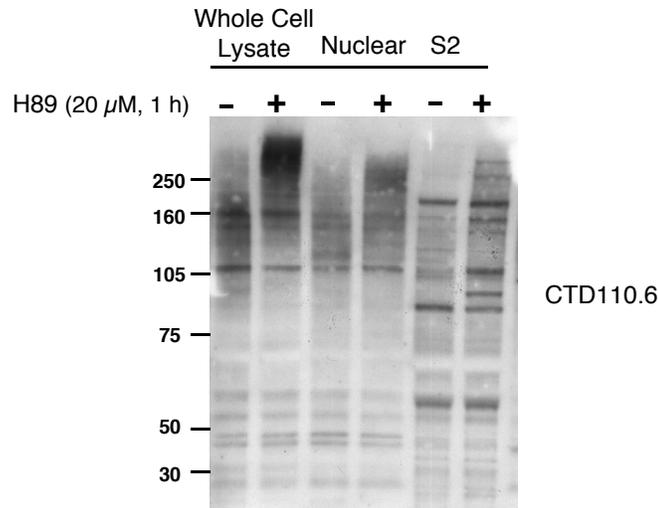

Figure AI.1. The PKA inhibitor H89 upregulates *O*-GlcNAc levels in hippocampal slices as observed by CTD110.6 Western blot of adult acute hippocampal slice lysate.

The mechanism by which H89 might elevate *O*-GlcNAc levels is unclear. Notably, PKA is known to inhibit the activity of Glutamine:fructose-6-phosphate amidotransferase (GFAT), the rate-limiting enzyme in glucosamine synthesis.[10] By blocking PKA activity, H89 may influence UDP-GlcNAc levels and the activity of OGT toward certain substrates. Comparison of fractionated tissue after H89 treatment and concurrent repression of GFAT with the inhibitor azaserine[11] should help address whether the effect is specifically due to elevation of UDP-GlcNAc levels.

AI.3 Characterizing *O*-GlcNAc Dynamics in Cortical Culture

Previous studies on *O*-GlcNAc had shown that phosphorylation pathways could regulate glycosylation levels in cytoskeletal fractions of cultured cerebellar neurons [5]. Here, we asked if similar effects could be observed for mixed cortical/hippocampal neurons in culture.

Table AI.2. Pharmacological treatment of neuronal cortical culture can modulate *O*-GlcNAc levels as measured by *O*-GlcNAc antibody CTD110.6 Western blotting

Glucose Concentration (mM)	Stimulus	Known Target of Stimulus	Cell Fractionation	Time	Observed Effect	Control
25mM	10 mM Gluocsamine	UDP-GlcNAc levels	Whole Cell SDS	3 h/6h/9h	upregulation	–
25mM	100 μM PUGNAc	O-GlcNAcase inhibition	Whole Cell SDS	9h/12h/24h	upregulation	–
25mM	50μM Forskolin	cAMP pathway PKA activation	Whole Cell SDS	1.5 h	no change	phospho-GlUR1 (s845) phospho-CREB(s133)
25mM	Calyculin 0.03 μM	PP2A inhibition	S2/P2 Whole Cell SDS	15 min/1 h	downregulation	phospho-CREB(s133)
25mM	Orthovanadate 10mM	tyrosine phosphatase inhibition	Whole Cell SDS	2 h	no change	–
25mM	H89 10μM	PKA inhibition	Whole Cell SDS	1.5 h	no change	phospho-CREB(s133) phospho-GLUR1(s845)
25mM	50 mM KCl	depolarization	Whole Cell SDS	2 min	no change	phospho-CREB(s133)
5mM/7.5mM/20mM	50 mM KCl	depolarization	Whole Cell SDS	30 s/1min/5min/30min/1h	no change	phospho-CREB(s133)
25mM	Calcium Ionophore A23187	multiple pathways	Nuc/S2/P2	1.5 min	no change	phospho-CREB(s133)
5mM	Calcium Ionophore A23187	multiple pathways	Whole Cell SDS	1/2/5/10 min	no change	phospho-CREB(s133)
25mM	NMDA 100μM	multiple pathways	Whole Cell SDS	2 min/1h	no change	phospho-CREB(s133)
25mM	NMDA 20μM/200μM		Nuclear/S2/P2	3 min	no change	phospho-CREB(s133)
25mM	PDBU 5μM	PKC activation	Whole Cell SDS	5 min	no change	phospho-GlUR1 (s831)
25mM	Insulin 0.02μM	multiple pathways	Whole Cell SDS	15 min/3h/12h	no change	phospho-Akt(s473)
25mM	Insulin Growth Factor-1 0.02μM	multiple pathways	Whole Cell SDS	15 min/3h/12h	no change	phospho-Akt(s473)
25mM	PD98059 10μM	MEK inhibition	Whole Cell SDS	1 h	no change	–

Our results with cortical culture suggest that as with hippocampal slices, *O*-GlcNAc

levels can respond dynamically to glucosamine and PUGNAc treatment. Varying

glucose concentration produced ambiguous effects, in some cases producing subtle

differences in *O*-GlcNAc levels by CTD110.6 Western blotting. Finally, as shown in

Figure AI.2, the PP1/PP2A inhibitor calyculin A was able to significantly downregulate

O-GlcNAc levels on several prominent protein bands as quickly as 15 min after

treatment.
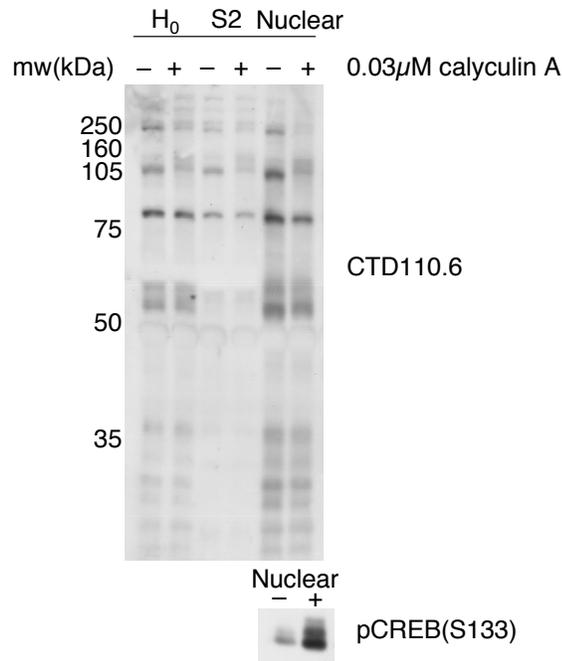
Figure AI.2. Pharmacological treatment of neuronal cortical culture with the phosphatase inhibitor calyculin A decreases *O*-GlcNAc glycosylation as observed by CTD110.6 Western blotting. Neuronal Fractions probed include total cellular homogenate (H$_0$), cytosol (S2) and Nuclear. Western blotting of phosphorylated CREB from the sample samples shows the expected increase in phosphorylation with calyculin A treatment.

AI.4 Characterization of *O*-GlcNAc Cycling at Synaptosomes

The *O*-GlcNAc enzymes are very abundant at nerve terminal synaptosomes, particularly

in the presynaptic cytosol.[12]   Moreover, our own work (Chapter 4) has identified a

number of modified proteins in the synaptosome, many of which are involved in synaptic

vesicle cycling and presynaptic scaffolding. To investigate the dynamics of *O*-GlcNAc in synaptosomes, we fractionated adult rat cortices and isolated synaptosomes. Intact synaptosomes were then treated with a variety of pharmacological stimuli. *O*-GlcNAc levels on SDS-lysed synaptosomes were monitored with the CTD110.6 antibody. As described in Table AI.3, we tested a number of stimuli, including PUGNAc, none of which reproducibly produced changes in *O*-GlcNAc glycosylation. *O*-GlcNAc may be constitutive on the bands detected in SDS lysis of the synaptosomal fraction. Therefore, it will be important to test the effect of these stimuli on fractionated synaptosomes, as SDS lysis may be masking effects on cytosolic *O*-GlcNAc proteins.

Table AI.3. The effect of pharmacological stimuli on *O*-GlcNAc levels of cortical synaptoneurosomes as monitored by SDS-lysis of treated synaptosomes and subsequent CTD110.6 Western blotting

Stimulus	Known Target of Stimulus	Observed Effect	Time	Control
50 *µ*M PUGNAc 200*µ*M	O-GlcNAcase inhibition	no change	1 h/2 h 1 h	–
5 *µ*M/10 *µ*M/50 *µ*M Forskolin	cAMP pathway PKA activation	no change	15 min 15 min	phospho-synapsin(s9)
0.25 *µ*M Calyculin A	PP2A/PP1 inhibition	no change	10 min	phospho-synapsin(s9)
2.5 *µ*M/5 *µ*M KN-62	CaMKII inhibition	no change	5 min	phospho-synapsin(s9)
H89 100*µ*M	PKA inhibition	no change	15 min/1 h	phospho-synapsin(s9)
50 mM KCl	depolarization	no change	30 sec/1 min/5 min 10 min/30 min/1 h	phospho-synapsin(s9)

AI.4 Implications and Future Directions

Overall, we found that *O*-GlcNAc levels can respond directly to the UDP-GlcNAc precursor glucosamine and the *O*-GlcNAcase inhibitor PUGNAc in both neuronal culture and hippocampal slices. This supports a dynamic cycling of *O*-GlcNAc in neuronal tissue. We also found that *O*-GlcNAc levels could respond to perturbations in phosphorylation pathways, specifically to the PKA inhibitor H89 in hippocampal slices

and the protein phosphatase (PP1 and PP2A) inhibitor calyculin A in cultured cortical neurons.

Although previous studies on cerebellar neurons had indicated that *O*-GlcNAc levels are significantly downregulated with short treatments of the calcium ionophore A23187,[5] we found no such effect on cortical neurons. This may be due to technical variation. The cerebellar neuron study used an *O*-GlcNAc antibody coupled to an ElISA assay that has been reported to detect other carbohydrate epitopes.[13] Alternatively, there may be differences in *O*-GlcNAc regulation in the cortex versus the cerebellum

Several additional studies may address the findings reported in this appendix. *In vitro* glycosylation assays with immunoprecipitated OGT have successfully been used to detect changes in enzyme activity in response to cellular stimuli.[4] Those studies would be important to do on slices or neuron culture. Although we could not detect changes in glycosylation within synaptosomes, this may reflect difficulties in detecting such changes from unfractionated material. It will be useful to repeat some of these stimulations and fractionate synaptosomes into the components that showed greatest enrichment of the *O*-GlcNAc enzymes (i.e., the synaptosomal cytosol).[12] Moreover, *in vitro* activity assays for both enzymes after stimulation will likewise offer insight into *O*-GlcNAc regulation in this compartment.

Several studies suggest that *O*-GlcNAc may play a fundamental role in the stress response.[14] With the exception of our study with kainic acid discussed in Chapter 5, this has not been addressed in neuronal tissue. It will be important to examine changes in *O*-GlcNAc in response to neuronal stresses such as hypoxia and excitotoxicity both on the level of enzyme activity and *O*-GlcNAc glycosylation. Finally, ongoing developments in

the chemoenzymatic strategy (as described in Chapter 2) should allow for direct chemiluminescent or fluorescent detection of tagged O-GlcNAc proteins. We envision that this methodology will have significant benefits in sensitivity and scope over the CTD110.6 antibody and will be useful for studying O-GlcNAc dynamics both by Western blot and through in-gel fluorescent detection.

AI.5: Experimental Methods

Preparation of Hippocampal Slices

8-10 week old Sprague-Dawley rats (Charles River Laboratories) were sacrificed (with CO_2 for 1-1.5 min) and decapitated. Brains were removed and rinsed with ice-cold Krebs bicarbonate buffer (124 mM NaCl, 4 mM KCl, 26 mM NaHCO3, 1.5 mM CaCl2, 1.25 mM KH_2PO_4, 1.5 mM $MgSO_4$, and 10 mM D-glucose, pH 7.4), which was buffered with 5% CO_2/95% O_2 that had been aerated for at least 30 min before use. Cerebella were removed by razor blade and 400 µM coronal slices of cerebral cortices were prepared by vibratome (Leica). During slicing, brains were immersed in Krebs bicarbonate buffer, chilled by ice-water bath. Hippocampi were dissected out of slices and placed into 2 mL polypropylene tubes containing chilled Krebs bicarbonate buffer. Buffer was immediately removed and replaced with fresh solution, and slices were incubated for 30 min at 30 ºC with constant bubbling of 5% CO_2/95% O_2. Incubation buffer was removed and new buffer, containing the pharmacological agents described in Table AI.1 were added to slices. All pharmacological agents were purchased from Alexis biochemicals except PUGNAc (Toronto research chemicals) and glucosamine-HCL (Sigma). KCl depolarization experiments were conducted in Krebs bicarbonate buffer except that NaCl

concentration was 55 mM and KCl concentration was 60 mM. Treatment was stopped by removal of the slice from treatment buffer via glass pipet. Slices were immediately flash frozen in liquid N_2 and stored at −80 ºC until further use. Each experimental slice was paired with a control slice from the opposing hemisphere. Control slices were treated identically except that pharmacological agents were replaced with the same volume of vehicle (DMSO, ddH$_2$0, or 100mM HEPES pH 7.4 for glucosamine).

For whole cell SDS lysis, slices were lysed in boiling 1% SDS, with protease inhibitors, sonicated for 3 sec and boiled for 8 min. Slices were kept on dry ice until the addition of boiling SDS. For fractionation, two slices were homogenized by mechanical homogenization for 5 strokes at 700 rpm in 300 µL of homogenization buffer (0.32 M sucrose, phosphatase inhibitors, protease inhibitors, 10 µM PUGNAc, 50 mM glucosamine). Centrifugation at 800 × g for 10 min produced the nuclear pellet, which was washed one time in homogenization buffer and lysed directly into boiling 1% SDS by sonication. Supernatant (S1) was spun at 16,000 × g for 15 min to produce the S2 and P2 fractions.

Preparation of Cortical Culture

Cortical neuronal cultures were prepared as described in experimental methods of Chapter 5. Neurons were grown in 100 mM dishes (8-15X 10^6 cells/dish) or in 6-well 35 mM plates (2-4 X 10^6 cells/dish) pre-coated with poly-D,L-lysine (Sigma). Media was replaced 12 h before addition of pharmacological agents, which were added directly to culture dishes/wells. Neurons were generally grown for 7-10 days before pharmacological treatment. To specifically assess the effect of glucose concentration on

O-GlcNAc levels, neurons were grown in minimum essential media (MEM) (Gibco) containing 10% fetal calf serum (FCS) (Gibco), 1X N2-supplement (Gibco) 1 mM pyruvate, and 5 mM glucose. For different concentrations of glucose, media was supplemented with sterile-filtered glucose (1 M) to final concentrations as described in the text.

For SDS lysis of short pharmacological treatments (≤3 h), treatments were stopped by aspiration of the media and 1% SDS (with protease inhibitors, phosphatase inhibitors and 10 µM PUGNAc) was added directly to dishes/wells (250 µL/well of a 6-well dish). Neurons were scraped into lysis buffer, drawn up with a plastic P1000 pipet (cut at the tip) and sonicated for 2 × 3 sec. Lysate was boiled for 8 min and centrifuged for 5 min at 15,000 × g to remove insoluble material. For SDS lysis of prolonged treatments (≥3 h) in 100 mM dishes, media was removed and neurons were trypsinized, pelleted and frozen at −80 ºC until future use. Pellets were kept on dry ice until lysis in boiling 1% SDS with protease inhibitors, phosphatase inhibitors and 10 µM PUGNAc.

S2/P2 fractionation was conducted on neurons grown in 100 mM dishes. Treatment was stopped by media aspiration. Plates were quickly washes with 1 mL of ice-cold HEPES-buffered Saline and lysed by scraping into 500-800 *µ*L of ice-cold lysis buffer (0.32 M sucrose, 10 mM HEPES pH 7.4 containing protease inhibitors, phosphatase inhibitors, 10 *µ*M PUGNAc and 50 mM GlcNAc). Scraped cells were homogenized by manual dounce homogenization for 5 strokes on ice followed by mechanical homogenization for 5 strokes at 700 rpm. A portion of the lysate was saved as homogenate (H$_0$) and the rest was centrifuged for 10 min at 800 × g. The resultant nuclear pellet was washed one time with lysis buffer (with no PUGNAc or GlcNAc) and

then lysed directly in boiling 1% SDS. The supernatant (S1) was centrifuged for 15 min at 16,000 × g yielding the S2 supernatant and the P2 pellet. P2 pellets were washed one time with 10% sucrose and lysed directly into boiling 1% SDS.

Synaptosome Treatment and Lysis

Intact Synaptosomes were prepared as described in the experimental methods of Chapter 4. After rinsing in pre-incubation buffer as described, synaptosomes were diluted in incubation buffer (10 mM HEPES pH 7.4, 10 mM Glucose, 4.8 mM KCl, 1.2 mM Na_2HPO_4, 2.4 mM $MgSO_4$, 132 mM NaCl, 1.1 mM $CaCl_2$, 0.1 mM EGTA), at a volume of 750-1000 μL per brain originally used for extract preparation. Synaptosomes were split into portions, in eppendorf tubes, for experiments and incubated by rocking at 37 ºC for 15 min. (Typically, 6 experiments, each with a control, were conducted from 8 rats.) After 15 min, synaptosomes were diluted 2-fold with incubation buffer containing pharmacological agents or an equal volume of vehicle and incubated for the prescribed time. In the case of depolarization experiments synaptosomes were diluted two-fold with 10 mM HEPES pH 7.4, 10 mM Glucose, 105 mM KCl, 1.2 mM Na_2HPO_4, 2.4 mM $MgSO_4$, 32 mM NaCl, 1.1 mM $CaCl_2$, 0.1 mM EGTA, such that the final concentration of NaCl was 82 mM and final KCl concentration was 55 mM[15]. Incubations were stopped by placing tubes on ice. Eppendorf tubes were immediately centrifuged by pulse centrifugation for 15 sec (in a refrigerated centrifuge) to collect synaptosomes. Incubation buffer was removed by pipet and synaptosomes immediately lysed by addition of boiling 1% SDS with protease inhibitors. Synaptosomal extract was sonicated for 3 × 3 sec and boiled for 8 min. Synaptosomal extracts were spun at 15,000 × g for 5min to

remove insoluble material.

CTD110.6 Western Blotting

Blots were blocked for 1 h in 5% BSA, TBS-T(0.1%) and incubated with the CTD110.6 antibody (Covance) for 1 h at RT or o/n for 10-12 h at 4 ℃ at a concentration of 1:2500-1:5000 in blocking buffer. Blots were rinsed with TBS-T(0.1%), washed 3 × 5 min and incubated with secondary antibody, (GαM IgG-HRP, Pierce) at a concentration of 1:10,000 at RT for 1 h. Blots were subsequently rinsed with TBS-T(0.1%), washed 2 × 15 min and 3 × 5 min and developed by chemiluminescence.

References:

1.   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).
2.   Kearse, K. P. & Hart, G. W. Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins. *Proc Natl Acad Sci USA* **88**, 1701-5 (1991).
3.   Kneass, Z. T. & Marchase, R. B. Neutrophils exhibit rapid agonist-induced increases in protein-associated *O*-GlcNAc. *J Biol Chem* **279**, 45759-65 (2004).
4.   Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D. & Hart, G. W. Dynamic *O*-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* **279**, 30133-42 (2004).
5.   Griffith, L. S. & Schmitz, B. *O*-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to pertubations of phosphorylation. *Eur J Biochem* **262**, 824-31 (1999).
6.   Nishi, A., Snyder, G. L., Nairn, A. C. & Greengard, P. Role of calcineurin and protein phosphatase-2A in the regulation of DARPP-32 dephosphorylation in neostriatal neurons. *J Neurochem* **72**, 2015-21 (1999).
7.   Lefebvre, T., Alonso, C., Mahboub, S., Dupire, M. J., Zanetta, J. P., Caillet-Boudin, M. L. & Michalski, J. C. Effect of okadaic acid on *O*-linked N-acetylglucosamine levels in a neuroblastoma cell line. *Biochim Biophys Acta* **1472**, 71-81 (1999).
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 U S A* **101**, 10804-9 (2004).
9.   Liu, K., Paterson, A. J., Zhang, F., McAndrew, J., Fukuchi, K., Wyss, J. M., Peng, L., Hu, Y. & Kudlow, J. E. 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).
10.   Chang, Q., Su, K., Baker, J. R., Yang, X., Paterson, A. J. & Kudlow, J. E. Phosphorylation of human glutamine:fructose-6-phosphate amidotransferase by cAMP-dependent protein kinase at serine 205 blocks the enzyme activity. *J Biol Chem* **275**, 21981-7 (2000).
11.   Gao, Y., Miyazaki, J. & Hart, G. W. The transcription factor PDX-1 is post-translationally modified by *O*-linked N-acetylglucosamine and this modification is correlated with its DNA binding activity and insulin secretion in min6 beta-cells. *Arch Biochem Biophys* **415**, 155-63 (2003).
12.   Cole, R. N. & Hart, G. W. Cytosolic *O*-glycosylation is abundant in nerve terminals. *J Neurochem* **79**, 1080-9 (2001).
13.   Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for *O*- linked N-acetylglucosamine. *Analytical Biochemistry* **293**, 169-177 (2001).
14.   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).
15.   Bauerfeind, R., Takei, K. & De Camilli, P. Amphiphysin I is associated with coated endocytic intermediates and undergoes stimulation-dependent dephosphorylation in nerve terminals. *J Biol Chem* **272**, 30984-92 (1997).

Appendix II

Identification of *O*-GlcNAc Glycosylation Sites: Application to Endogenous Proteins and Exploration of Fluorous Enrichment Strategies

AII.1 Background and Introduction

In Chapter 3 we highlighted methodology for identification of glycosylation sites from individual proteins. Glycosylation site mapping is a particularly important endeavor as it helps to establish the functional significance of the *O*-GlcNAc modification. Mapping sites to particular domains of proteins advances research into the mechanisms by which *O*-GlcNAc may alter protein-protein interaction,[1] protein stability,[2] or protein activity.[3]

In Chapter 3 we demonstrated that the chemoenzymatic strategy allows the facile isolation and identification of *O*-GlcNAc peptides from purified proteins such as CREB, OGT and ΔfosB and the low-stoichiometry protein α-crystallin. Here, we discuss our efforts to extend this approach to individual endogenous proteins or proteins transfected at low levels into neuronal culture. This presents a formidable task given the lack of consensus sequence for OGT, the low-abundance of many targets, and the substoichiometric nature of the modification.[4] We also demonstrate the versatility of the chemoenzymatic approach by demonstrating non-biotin streptavidin techniques for isolation of *O*-GlcNAc peptides.

AII.1 *O*-GlcNAc Site Identification on Low Abundance Proteins from Cells

To investigate the applicability of site identification for endogenous proteins, we chemoenzymatically labeled HIT-T15 pancreatic cell lysate and immunoprecipitated

CREB. Control samples were treated identically except that the enzyme β1,4-galactosyltransferase Y289L was not added to reactions. As Figure AII.1 shows, only CREB that was immunoprecipitated from lysates labeled with both the enzyme and analogue **1** displayed a signal by streptavidin Western blotting.
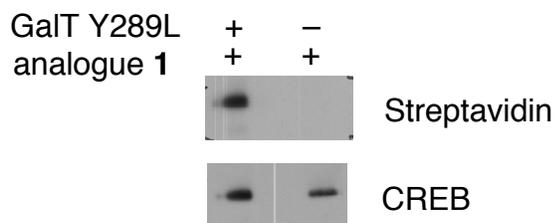

Figure AII.1. Endogenous *O*-GlcNAc glycosylated CREB is chemoenzymatically tagged and immunoprecipitated from HIT-T15 pancreatic cell lysate.

A fraction of the chemoenzymatically tagged lysate was run on an SDS-PAGE gel and silver stained, revealing two discrete bands in the molecular weight vicinity of CREB (Figure AII.2A). In-gel trypsin digestion followed by liquid-chromatography mass spectrometry (LC-MS) revealed that the upper band corresponded to CREB (sequence coverage in Figure AII.2B).

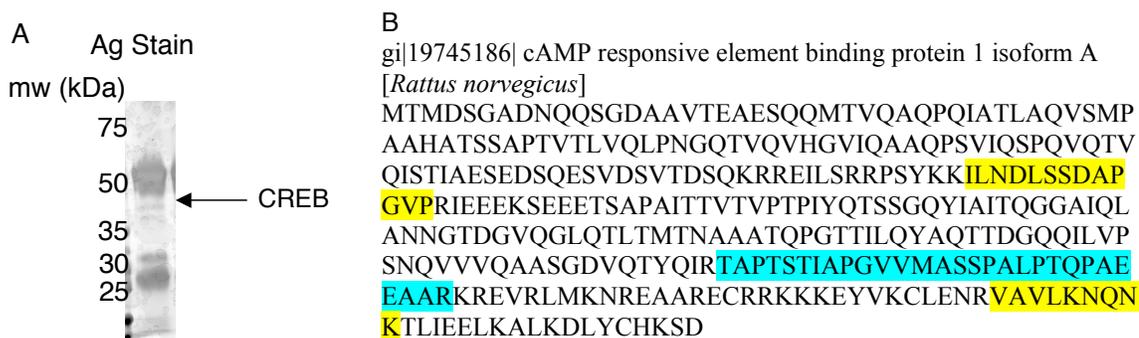**Figure AII.2.** (A) Silver stain of endogenous CREB immunoprecipitated from chemoenzymatically tagged HIT-T15 pancreatic cell lysate. (B) LC-MS/MS sequence coverage of in-gel tryptic digest of the marked band. Sequencing revealed that this band contained peptides corresponding to CREB. Peptides found by LC-MS are highlighted in yellow, including the peptide known to contain the *O*-GlcNAc site (highlighted in blue).

Although we identified the peptide known to contain the *O*-GlcNAc site, avidin enrichment of the digested CREB peptides did not reveal any *O*-GlcNAc-modified species. Because we did not see the modified peptide in the avidin chromatography input we must conclude that it is of considerably lower abundance (and thus stoichiometry) than the unmodified peptide.

Concurrently with CREB, we applied a similar strategy to neuronal lysates that had been transected with a Flag-tagged MeCP2. We had previously shown that the transcriptional repressor MECP2 is an *O*-GlcNAc protein (Chapter 3) but the site of modification is unknown. We chemoenzymatically labeled neuronal lysate and immunoprecipitated using an anti-Flag antibody (Figure AII.3A). We found two bands that had a molecular weight close to that reported for MECP2 and analyzed both by LC-MS. As shown in figure AII.3B, we were able to identify the lower band as MeCP2. However, as with CREB, avidin enrichment of the digested peptides did not reveal an *O*-GlcNAc-modified species.
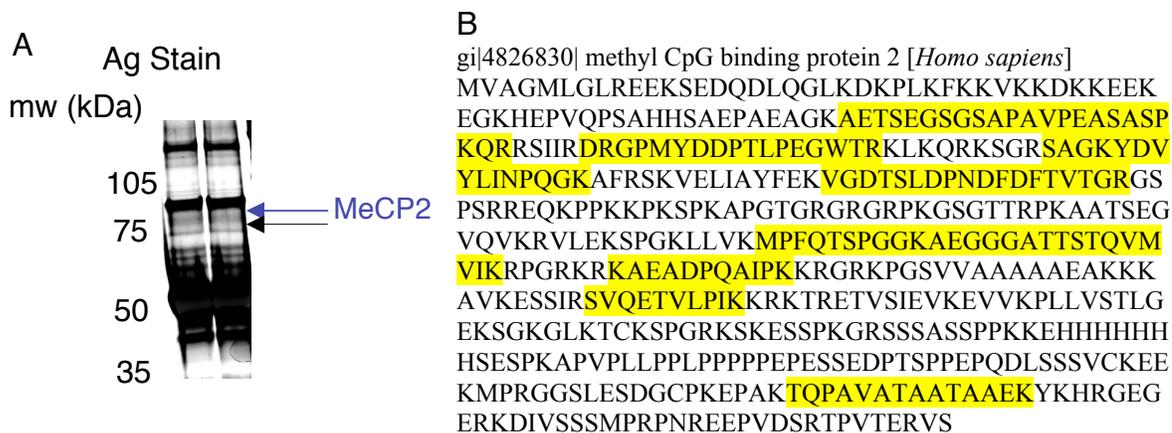
A

Ag Stain

mw (kDa)

105

75

50

35

MeCP2

B

gi|4826830| methyl CpG binding protein 2 [*Homo sapiens*]
MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEK
EGKHEPVQPSAHHSAEPAEAGK<mark>AETSEGSGSAPAVPEASASP</mark>
<mark>KQR</mark>RSIIR<mark>DRGPMYDDPTLPEGWTR</mark>KLKQRKSGR<mark>SAGKYDV</mark>
<mark>YLINPQGK</mark>AFRSKVELIAYFEK<mark>VGDTSLDPNDFDFTVTGR</mark>GS
PSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTTRPKAATSEG
VQVKRVLEKSPGKLLVK<mark>MPFQTSPGGKAEGGGATTSTQVM</mark>
<mark>VIK</mark>RPGRKR<mark>KAEADPQAIPK</mark>KRGRKPGSVVAAAAAEAKKK
AVKESSIR<mark>SVQETVLPIK</mark>KRKTRETVSIEVKEVVKPLLVSTLG
EKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHH
HSESPKAPVPLLPPLPPPPPEPESSEDPTSPPEPQDLSSSVCKEE
KMPRGGSLESDGCPKEPAK<mark>TQPAVATAATAAEK</mark>YKHRGEG
ERKDIVSSSMPRPNREEPVDSRTPVTERVS

Figure AII.3. Immunoprecipitation and LC-MS/MS sequence coverage of Flag-tagged MeCP2 isolated from neuron culture. (A) Silver Stain of chemoenzymatically-tagged and Flag-immunoprecipitated lysate reveals two bands in the molecular weight vicinity of the transcriptional repressor MeCP2. (B) LC-MS/MS of the in-gel tryptic digest of both bands reveals that the bottom band corresponds to MeCP2 as demonstrated by the identified peptides.

Several methodological adjustments may make identification of the *O*-GlcNAc site from very low-abundance proteins such as CREB and neuron-transfected or endogenous MeCP2 technically feasible. First, in-gel digestion is known to incur loss and it may be replaced with on-bead proteolytic digestion. Second, our avidin enrichment strategy currently utilizes a large (0.2 mL, Applied Biosystems) affinity cartridge that includes an 800 µL elution step. The eluent is then dried by vacuum centrifugation before LC-MS. This approach is effective for *O*-GlcNAc glycosylated peptides isolated from >1 µg of protein, or from complex mixtures, but may induce significant loss for low-abundance material. Therefore, it may be useful to add carrier species in picomolar abundance (such as biotinylated peptides, or digested purified *O*-GlcNAc proteins such as OGT) when analyzing endogenous proteins. In addition, it may be effective to pack monomeric avidin (commercially available from several sources) into microcapillary columns for improved sensitivity.

Finally, one advantage of the chemoenzymatic approach is the versatility of the ketone 'handle,' which may be exploited to tag *O*-GlcNAc with other molecules besides biotin for affinity purification. To address some of the problems described with biotin-streptavidin, we explored the use of a fluorous-aminooxy group for fluorous-affinity chromatography and enrichment of *O*-GlcNAc peptides, with some success.

AII.3 Coupling the Chemoenzymatic Strategy with Fluorous Affinity Chromatography

Highly fluorinated or 'fluorous' compounds have been exploited for a number of separation and enrichment strategies in both organic synthesis[5] and more recently in mass spectrometry of peptides.[6] Because of the selectivity of the fluorine-fluorine interaction, fluorous compounds can be readily separated from a mixture by solid-phase extraction over fluorous-functionalized silica gel (fluorous solid-phase extraction (FSPE)) . In order to overcome some of the limitations of streptavidin enrichment described above, we explored the utility of FSPE for the enrichment of *O*-GlcNAc glycosylated peptides that had been tagged with a fluorous compound. We used a commercially available fluorous-aminooxy nucleophile, 2-Aminooxy-N-(3-perfluorolhexyl)propyl-acetamide (Fluorous Technologies, Inc, Pittsburgh PA), to tag keto-galactose labeled *O*-GlcNAc transferase (OGT). After in-gel trypsin digestion, fluorous-*O*-GlcNAc peptides were specifically enriched via FSPE and analyzed in-line with mass spectrometry. Because fluorous peptides could be specifically enriched in a microcapillary column in-line with the mass spectrometer, we did not need to manually collect eluent or conduct additional reverse-

phase chromatography to detect *O*-GlcNAc peptides. Enrichment of a representative
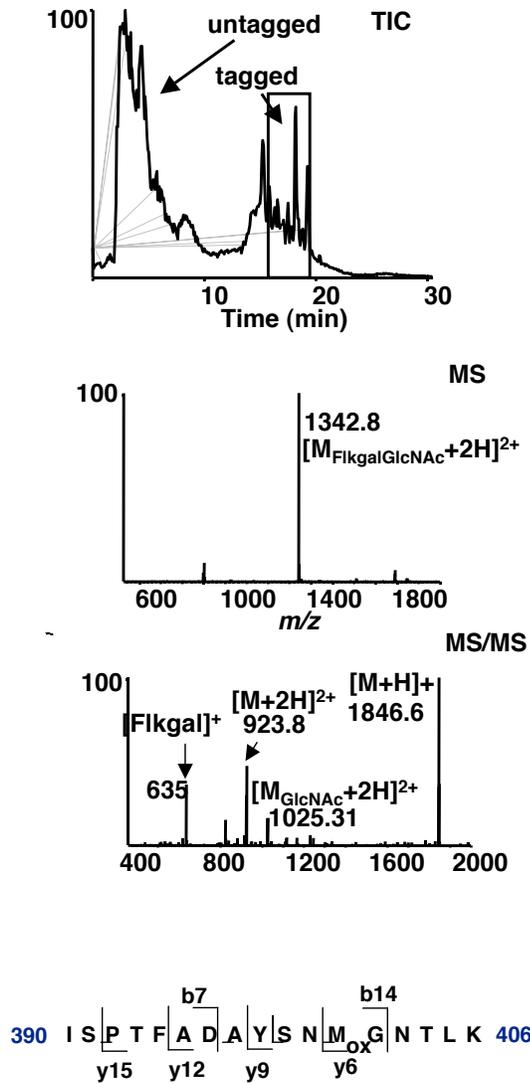
OGT peptide is shown in Figure AII.4

Figure AII.4. FSPE enrichment of tagged *O*-GlcNAc peptides from OGT. The total ion chromatogram (TIC) of trypsin-digested OGT as it eluted from the FSPE column onto the mass spectrometer shows a region of enriched, fluorous-tagged *O*-GlcNAc peptides. MS/MS analysis of one of these peptides shows the characteristic signature loss of the fluorous-ketogalactose group and fluorous-ketogalactose-GlcNAc. Representative b and y ions are shown from the MS^4 sequencing of this peptide, which was identified as ^{390}ISPTFADAYSNMoxGNTLK^{406} of OGT.

Notably, we detected all the *O*-GlcNAc peptides of OGT that had been identified by

biotinylation and streptavidin enrichment (Chapter 3) and additionally found that the

peptides [421]AIQINPAFADAHSNLASIHK[440] and [390]ISPTFADAYSNMoxGNTLK[406] were doubly *O*-GlcNAc modified.

We investigated the fluorous strategy on several other proteins such as the transcription factor ΔfosB (Chapter 3) and CREB as well as on complex mixtures. However, we were confronted with several problems. First, the fluorous compound necessitates organic solvent for solvation. We found that protein lysates were most stable in commixtures of THF/water. However, care had to be taken as ketone byproducts form in THF that compete for oxime formation. We addressed some of these problems by digesting proteins directly after enzymatic addition of the ketogalactose group. Digested peptides were subsequently labeled with the fluorous-aminooxy, (peptides are stable in organic/aqueous co-mixtures such as EtOH/water, forgoing the need for THF). To remove excess fluorous-aminooxy, we used Zwitterionic-HILIC chromatography (SeQuant). However, we had difficulty implementing the methodology, reproducibly, for complex mixtures.

However, as demonstrated with OGT, the strategy may be useful for single proteins as the enrichment step is conducted within microcapillary columns, in-line with mass spectrometry. This forgoes the large volumes and unnecessary handling and drying steps of avidin enrichment and may increase the recovery of *O*-GlcNAc peptides from low-abundance, low-stoichiometry proteins.

AII.3: Experimental Methods

Chemoenzymatic Labeling of HIT-TI5 cells and Immunoprecipitation of CREB

HIT-T15 cells were cultured, chemoenzymatically labeled and CREB

immunoprecipitated as described in Chapter 5. 600μg of lysate was chemoenzymatically labeled for Western blot analysis and 3.8 mg was labeled for mass spectrometry analysis. To immunoprecipitate CREB from lysate for mass spectrometry, 50 μL of protein A sepharose (Pierce) and 7.5 μg of anti-CREB antibody (rabbit polyclonal, Upstate Biotechnology) were used. SDS-PAGE of immunoprecipitated proteins was conducted with 4-12% Bis-Tris gels (Invitrogen).

Chemoenzymatic Labeling of Neuronal Cell Lysate and Immunoprecipitation of Flag-Tagged MeCP2

Harvested and frozen embryonic cortical neurons, which had been transfected with an N-terminal Flag construct of MeCP2 and treated with 10 mM glucosamine for 6 h were provided by J. Zellhoefer from Yi Sun's Laboratory at UCLA. Pellets were lysed in 1% boiling SDS and chemoenzymatically labeled as described in Chapter 3. One milligram of lysate was used for mass spectrometry analysis. After biotinylation, labeled lysates were dialyzed into Flag-IP buffer (0.05 M Tris pH 7.4, 100 mM NaCl, 5 mM EDTA, 0.2% NP-40) 1× 12 h, and 2 × 3 h at 4 ℃. One hundred microliters of Flag-agarose beads (Sigma) were washed with 1 mL Flag-IP buffer 3× and incubated with labeled lysate for 2 h at 4 ℃ by inversion. Flowthrough was removed by centrifugation on a table-top centrifuge. Beads were washed 3× with Flag-IP buffer, 3× with TBS (50 mM Tris-HCl, 150 mM NaCl, pH 7.4) and 1× with 50 mM Na_2HPO_4 pH 7.4 at 4 ℃. Captured proteins were eluted by boiling beads in 150 μL of 2X SDS-PAGE loading dye for 2 × 3min (beads were briefly vortexed after the first 3 min incubation). SDS-PAGE was conducted with 7% Tris-Acetate gels (Invitrogen).

Western Blotting

Streptavidin western blotting and anti-CREB western blotting were performed as described in the experimental methods of Chapter 5. Flag immunoprecipitates were blocked for 1 h at RT in 5% milk-TBS-T(0.1%), and incubated with the anti-Flag antibody (mouse monoclonal, Sigma) at a concentration of 1:2000, in blocking buffer, for 1 h at RT. Blots were subsequently rinsed in TBS-T(0.1%) and washed 3 × 5 min in TBS-T(0.1%). Blots were incubated with a GαM IgG secondary antibody at a concentration of 1:10,000 in blocking buffer for 1 h at RT. They were subsequently rinsed in TBS-T(0.1%), washed 2 × 15 min and 3 × 5 min and developed via chemiluminescence.

Silver Stain, In-Gel Digestion and LC-MS of CREB and MeCP2

SDS-PAGE gels were silver stained with a variantion of the methods described by Shevchenko et al.,[7] and Blum et al.[8] Gels were fixed for 30 min in 50% MeOH/10% AcOH. Solution was removed and gels were fixed in 5% MeOH/1% AcOH for 15 min, then rinsed with 50% MeOH for 1 min. Gels were then washed for 3 × 10 min with ddH_2O, then sensitized for 90 sec with 20 mg/100 mL of $Na_2S_2O_3 \cdot 5H_2O$. Gels were then rinsed for 3 × 30 sec in ddH_2O, and silver stained with $AgNO_3$ (200 mg/100 mL) for 30 min. After rinsing 3 × 30 sec in ddH_2O, gels were developed in Na_2CO_3 (6 g/100 mL), 37% formaldehyde (50 µL/100 mL), $Na_2S_2O_3 \cdot 5H_2O$ (from the sensitization step, 2 mL/100 mL). Development proceeded until bands at the appropriate molecular weight of CREB and MeCP2 were visualized and stopped by removal of the development solution and addition of 6% AcOH, in water. After rinsing in ddH_2O, putative CREB and MeCP2

bands were excised from silver-stained gels. Gel pieces were cut into 5-6 pieces with a razor blade and incubated with destain solution ($K_3Fe(CN)_6$ 0.4 g/200mL in 0.2 g/L $Na_2S_2O_3 \cdot 5H_2O$) for 15 min with shaking (or until gel pieces turned yellow). Gel pieces were washed 4-5 times for 15 min each with ddH_2O until pieces were transparent. Gel pieces were reduced, alkylated and trypsin digested as described in Chapter 3.

LC-MS and avidin enrichment on digested peptides was performed as described for CREB/OGT in the experimental methods of Chapter 3.

Chemoenzymatic (Fluorous) Labeling and LC/MS of OGT

Baculovirus preparation and protein expression were performed as described previously.[1] OGT (10 μg) (in a mixture with CREB (2 μg)) in 20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 15% glycerol was supplemented with 5 mM $MnCl_2$. Analogue **1** and Y289L GalT were added to final concentrations of 500 μM and 40 ng/μL, respectively. Following incubation at 12 h at 4 °C, the reactions were diluted 3-fold with 9M urea, 2.7 M NaOAc pH 3.9, neat THF and 2-Aminooxy-N-(3-perfluorolhexyl)propyl-acetamide (Fluorous Technologies, Inc, Pittsburgh PA) dissolved in neat THF (final concentration 3.8 M urea, 50 mM NaOAc, 20% THF, 3 mM 2-Aminooxy-N-(3-perfluorolhexyl)propyl-acetamide, final pH 4.8) and incubated with gentle shaking for 20-24 h at 23 °C. Tagged OGT was resolved by SDS-PAGE and the protein was excised and digested from a Coomassie-stained gel as described in Chapter 3.

Automated nanoscale liquid chromatography and tandem mass spectrometry (LC-MS/MS) were conducted using a ThermoElectron Surveyor HPLC and LCQDecaXP+ mass spectrometer along with a variation of the "vented column" approach described by

Licklider et al.[9] Labeled digests were combined with an equal volume of 50% methanol

with 10 mM ammonium formate and loaded onto fluorous analytical columns with

integrated ESI emitter tips (360 um OD x 75 um ID fused silica, 5 um tip) packed with 5

cm Fluoroflash media (FTI, Pittsburgh, PA).   Peptides were eluted into the

LCQDecaXP+ ion trap mass spectrometer (ThermoElectron Corp, San Jose, CA) using

an HPLC gradient (50% B for 2 min, then 50%-95% B in 10 minutes, A = 0.1M acetic

acid with 10 mM ammonium acetate, B = 0.1M acetic acid in methanol with 10 mM

ammonium formate). The mass spectrometer was operated in data dependent mode such

that the top 3 ions in each MS scan was subjected to MS/MS (relative collision energy =

35%). Higher-order MS analyses involved an MS precursor scan followed by targeted

MS[4] scans of those masses that specifically demonstrated loss of the ketone-fluorous

moiety and ketone-fluorous-GlcNAc moiety in the MS/MS analysis. MS[4] data were used

to search against an OGT sequence database using SEQUEST.[10] For all experiments, the

spray voltage was 1.8 kV and the capillary temperature was 150 ºC.

References:

1.   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).
2.   Cheng, X. & Hart, G. W. Alternative O-glycosylation/O-phosphorylation of serine-16 in murine estrogen receptor beta: post-translational regulation of turnover and transactivation activity. *J Biol Chem* 276, 10570-5 (2001).
3.   Lamarre-Vincent, N. (2006).
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, 10500-1 (2004).
5.   Luo, Z., Zhang, Q., Oderaotoshi, Y. & Curran, D. P. Fluorous mixture synthesis: a fluorous-tagging strategy for the synthesis and separation of mixtures of organic compounds. *Science* 291, 1766-9 (2001).
6.   Brittain, S. M., Ficarro, S. B., Brock, A. & Peters, E. C. Enrichment and analysis of peptide subsets using fluorous affinity tags and mass spectrometry. *Nat Biotechnol* 23, 463-8 (2005).
7.   Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68, 850-8 (1996).
8.   Blum, H., Hildburg, B. & Gross, H. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93-99 (1987).
9.   Licklider, L. J., Thoreen, C. C., Peng, J. & Gygi, S. P. Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal Chem* 74, 3076-83 (2002).
10.   Eng, J. K. M., A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5, 976-989 (1994).