Structural and Functional Studies of GlcNAc-modified Tau

by

Adrienne Hoyann Cheung

B.Sc., Queen's University, 2010

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2013

Abstract

O-GlcNAcylation is an abundant post-translational modification found on serine and threonine hydroxyl groups of nucleocytoplasmic proteins. *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA) are the sole enzymes responsible for addition and removal of all GlcNAc moieties, respectively. Although *O*-GlcNAcylation is involved in many diverse cellular processes and has been linked to many diseases, the molecular mechanisms by which it can modulate the activities of proteins remain largely unknown. As such, the primary goal of my thesis is to gain a better structural and functional understanding of the consequences of *O*-GlcNAcylation of tau, an intrinsically disordered microtubule-binding protein. Upon hyperphosphorylation, tau aggregates, forming neurofibrillary tangles that are a hallmark of Alzheimer's disease. In contrast, *O*-GlcNAc has a reciprocal relationship with phosphorylation and reduces tau aggregation. To address my goals, I used NMR spectroscopy to probe local structural and dynamic effects of *O*-GlcNAc on a fragment of tau spanning residues 353 to 408 which encompasses a microtubule-binding repeat and Ser400, a key *O*-GlcNAc acceptor. Although chemical shift perturbations were observed near the site of *O*-GlcNAc-modification, based on main-chain chemical shifts, ^{3}J-coupling and NOE interactions, there were no significant local structural changes compared to the wild-type tau polypeptide. However, there was a small decrease in the nsec-psec time scale mobility on the main-chain of tau around residue 400. In order to investigate the functional impact of *O*-GlcNAc on tau, I compared the heparin-induced aggregation of the wild-type tau peptide, the *O*-GlcNAc form, and three serine to aspartate "phosphomimic" mutants. Importantly, the *O*-GlcNAc modification significantly decreased the amount of aggregation, whereas two of the phosphomimic mutants increased aggregation relative to wild-type tau. I postulate that the *O*-

GlcNAc modification increases the solubility of tau, thereby stabilizing the monomer in solution, and reducing the stability of aggregates.

.

Table of contents

List of figures

List of tables

Abbreviations

ε_{280}	molar absorptivity at 280 nm
2D	two-dimensional
3D	three-dimensional
A_{280}	absorption at 280 nm
AD	Alzheimer's disease
BSA	bovine serum albumin
CDK5	cyclin-dependent kinase 5
EPR	electron paramagnetic resonance
FPLC	fast protein liquid chromatography
FRET	fluorescence resonance energy transfer
GlcNAc	β-*N*-Acetylglucosamine
GSK3β	glycogen synthase kinase 3
GH	glycoside hydrolase
GT	glycosyl transferase
HAT	histone acetyltransferase
hOGA	human OGA
hOGT	human OGT
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered peptide
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria broth
MALDI-TOF	matrix-assisted laser desorption/ionization time of flight
MS	mass spectrometry

MT	microtubules
M9	M9 minimal media
NFT	neurofibrillary tangles
Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
OGA	*O*-GlcNAcase
OGT	uridine diphosphate-*N*-acetylglucosamine:polypeptide β-*N*-acetylglucosaminyltransferase
$O.D._{600}$	optical density at 600 nm
PCR	polymerase chain reaction
PHF	paired helical filaments
Pin1	peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
PP2A	protein phosphatase 2A
PTM	post-translational modification
RP-HPLC	reverse phase HPLC
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sWGA	succinylated wheat germ agglutinin
TFA	trifluoroacetic acid
UDP-GlcNAc	uridine diphosphate *N*-acetylglucosamine
WT	wild-type

Acknowledgements

The completion of this thesis would not have been possible without the efforts of many people, and to those people, I express endless gratitude. In particular, Lawrence, thank you for this opportunity, but more importantly, thank you for your guidance, hard work, and unlimited patience, I could not have hoped for a better supervisor. Dr. Mark Okon, NMR genius, and tennis superstar, I would still be sitting and staring at the magnets without your assistance and expertise. To our collaborators, Dr. David Vocadlo and Dr. Scott Yuzwa, your knowledge, enthusiasm for science, and willingness to help have been invaluable. Members of the McIntosh lab, current and past, thanks for the friendships, the science lessons, the wonderful discussions and troubleshooting sessions. Julien, I am so grateful to have you as a mentor and a friend, thank you for the laughter, the Kleenex for the tears, and the immeasurable support you have provided. To my mom, dad, and brother, although you were 3,000 km away, I would not have made it here without your continuous support and words of encouragement.

Chapter 1: Introduction

1.1 Post-translational modifications

Our genome has approximately 25,000 genes encoding for an even larger number of proteins that must be regulated for proper cellular function[1]. One of the most important routes to such regulation involves post-translational modifications (PTMs) of amino acid side chains. It is estimated that there are more than 300 types of PTMs, such as methylation, acetylation, ubiquitinylation, SUMOylation, phosphorylation, and glycosylation, to name only a few[2]. The primary focus of this thesis lies in characterizing the effects of two PTMs, glycosylation and phosphorylation, on the neurofibrillary protein, tau.

1.2 *O*-GlcNAcylation defines a new class of glycoproteins

There are three major classes of glycoconjugates: glycolipids, *O*-linked glycoproteins, and *N*-linked glycoproteins (Figure 1.1). In the latter two cases, saccharides are linked post-translationally to proteins via serine/threonine hydroxyls or asparagine amide groups, respectively. When considering glycoconjugates, it is most common to think of complex branched structures that are used for mediating and modulating cell adhesion and trafficking. However, in the 1980's, a new class of monosaccharide modification was recognized. While studying *N*-acetylglucosamine (GlcNAc) found on the surface of intact lymphocytes, Hart and coworkers discovered intracellular proteins modified with a simple *O*-linked monosaccharide β-*N*-acetylglucosamine (*O*-GlcNAc)[3]. Since then, mono-*O*-GlcNAc modified proteins have been found to be very abundant and widespread.

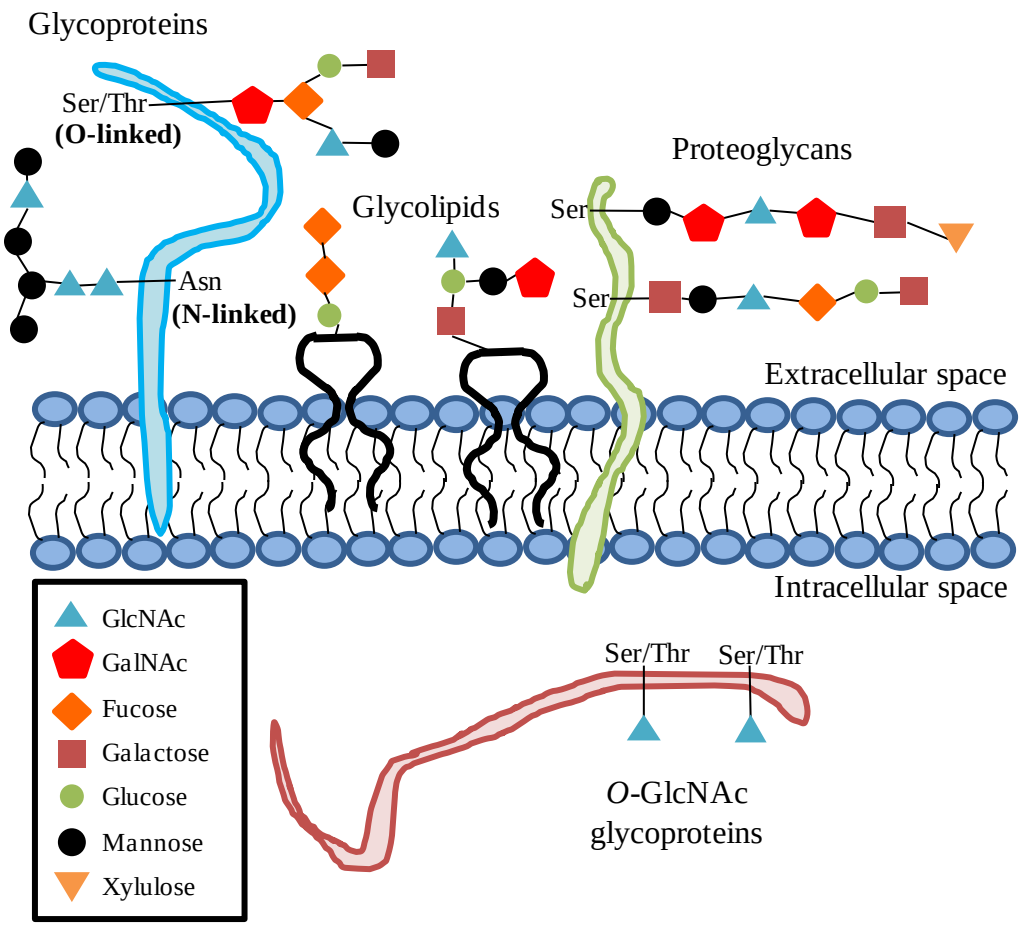

Figure 1.1: Different types of glycoconjugates in which straight-chain or branched saccharides are covalently linked to lipids or proteins.

O-GlcNAc modifications remained elusive for so many years due to the lack of tools and methods available to study this PTM. The addition of a simple *O*-GlcNAc to a serine or threonine side chain does not result in a significant change in size or charge, as required for detection by conventional analytical methods. In addition, the *O*-GlcNAc glycosidic bond is labile, and is preferentially fragmented during ionization, rendering *O*-GlcNAc-modified residues undetectable by regular protein mass spectrometry (MS) approaches at low concentrations[4]. Furthermore, there is preferential ionization for the unmodified peptide, suppressing the signal observed for corresponding *O*-GlcNAc-modified peptides[5]. However,

with the advent of electron-transfer dissociation MS in 2004, *O*-GlcNAc detection is now possible because the labile glycosidic bond is not cleaved[6]. Purifying *O*-GlcNacylated proteins is also challenging. A potential method for purification includes using succinylated wheat germ agglutinin (sWGA) affinity chromatography, since sWGA specifically binds to terminal *N*-acetylglucosamine. However, this method requires closely clustered GlcNAc residues for sufficient binding affinity[7]. Recently selective enrichment of *O*-GlcNAcylated peptides was achieved by using a mutant galactosyl transferase to enzymatically add azido-sugars to GlcNAc. Using the chemically reactive azide group, many different tags could then be introduced[5]. In particular, the addition of a biotin moiety allowed for subsequent purification with streptavidin. Although this increases the detection limits for *O*-GlcNAc modified peptides, it chemically alters them, making subsequent structural and functional studies difficult to interpret.

O-GlcNAc modifications occur on serine and threonine side chain hydroxyl groups. The addition of *O*-GlcNAc from the activated donor uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) is catalyzed by *O*-linked *N*-acetylglucosamine transferase (OGT). The hydrolysis of *O*-GlcNAc is catalyzed by *O*-GlcNAcase (OGA). These processes occur on a time scale similar to that of protein phosphorylation (Figure 1.2a)[7]. Thus far, *O*-GlcNAcylation has been documented in all metazoans, but remarkably there is only one gene that encodes for OGT and one gene that encodes for OGA in any given species[7]. This stands in marked contrast to the plethora of kinases and phosphatases found in any cell.

O-GlcNAcylation has many diverse cellular functions. During metabolism, 2-5% of glucose enters the hexosamine biosynthetic pathway, from which UDP-GlcNAc is synthesized[8]. As such, the level of *O*-GlcNAc modifications is postulated to serve as a

nutrient/stress sensor that modulates signaling, transcription, and cytoskeletal functions[9,10]. *O*-GlcNAc modifications have been linked to cancer, since this PTM is found on many oncogenic proteins and tumor suppressor proteins[9]. Finally, GlcNAcylation levels have also been linked to neurodegenerative diseases[9]. Through knockout studies, OGT has been shown to be necessary for development[11].

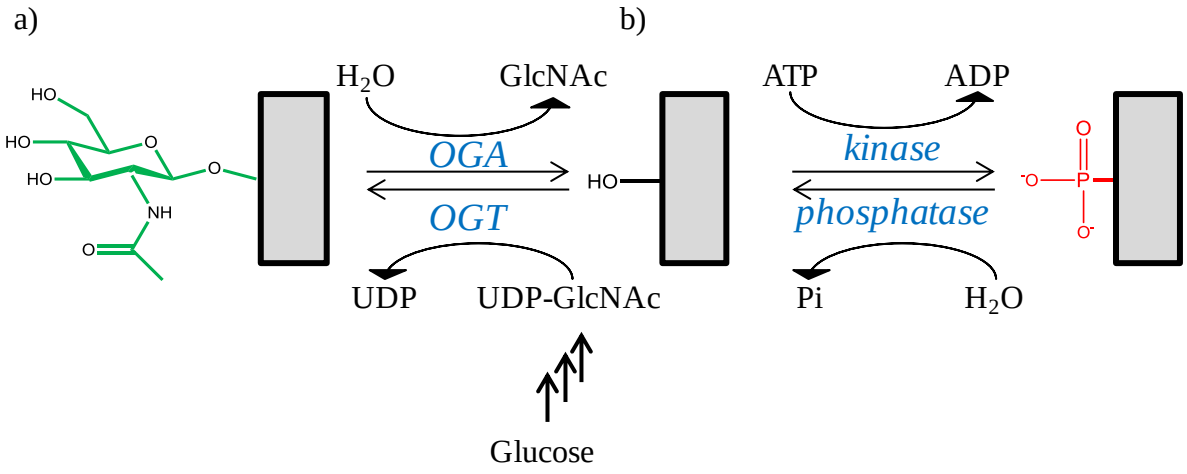

Figure 1.2: Dynamic *O*-GlcNAcylation and phosphorylation. *O*-GlcNAcylation and phosphorylation are both reversible reactions that occur on a similar time scale. a) OGT catalyzes the reversible addition of GlcNAc to a protein from the donor UDP-GlcNAc, whereas OGA catalyzes the hydrolysis of the bound GlcNAc. UDP-GlcNAc is synthesized through the hexosamine biosynthetic pathway, which is linked to the amount of glucose available in a system[7]. b) Hundreds of kinases and phosphatases catalyze the reversible phosphorylation of proteins, often in competition with the same *O*-GlcNAcylation acceptor serine/threonine.

1.3 GlcNAcylation enzymes: OGT and OGA

OGT is an essential glycosyl transferase, in the GT-41 family, composed of a catalytic C-terminal domain and an N-terminal protein-protein interaction domain that is made up of 34-residue helix-loop-helix tetratricopeptide repeats (TPRs)[7,12]. From the single gene that encodes OGT, there exist three splice variants of human OGT (hOGT), differing by the length of the TPRs. ncOGT is mostly found in the nucleus and cytosol, and has 13.5

TPRs; mOGT possesses an N-terminal mitochondrial localization sequence and has nine TPRs; and the shortest form of OGT, sOGT, has three TPRs[13]. Studies in rats have demonstrated stable expression of ncOGT throughout development, although the levels of expression decrease gradually throughout maturation; in contrast, sOGT levels were almost undetectable during early development and increased significantly after 15 days, suggesting that each isoform of OGT may play a different role throughout development[14].

Significant progress has been made towards understanding the enzymatic mechanism of OGT. In 2004, the Conti group reported the X-ray crystal structure of the first 11.5 TPRs which form an elongated superhelix resembling importin[15]. Seven years later, the Walker group reported the binary structures of the catalytic domain and its adjacent 4.5 TPRs in complex with UDP, as well as with a peptide known to be modified by OGT[16]. Together these studies yielded a model of full-length hOGT (Figure 1.3a). Although these crystal structures aided in answering some functional questions with regards to OGT, the exact molecular mechanisms by which OGT recognizes and glycosylates its substrates remains largely unknown[15].

Over 3000 proteins have been shown to have *O*-GlcNAc modifications. Of these, almost half have been reported to be modified at proline-valine-serine sites, whereas the remaining half has no identifiable acceptor motif[7]. Since TPRs resemble importin, a protein that binds to other proteins and transfers them into the nucleus, they have been hypothesized to play a role in establishing the specificity of OGT by serving as a docking site for target proteins (Figure 1.3b)[17].

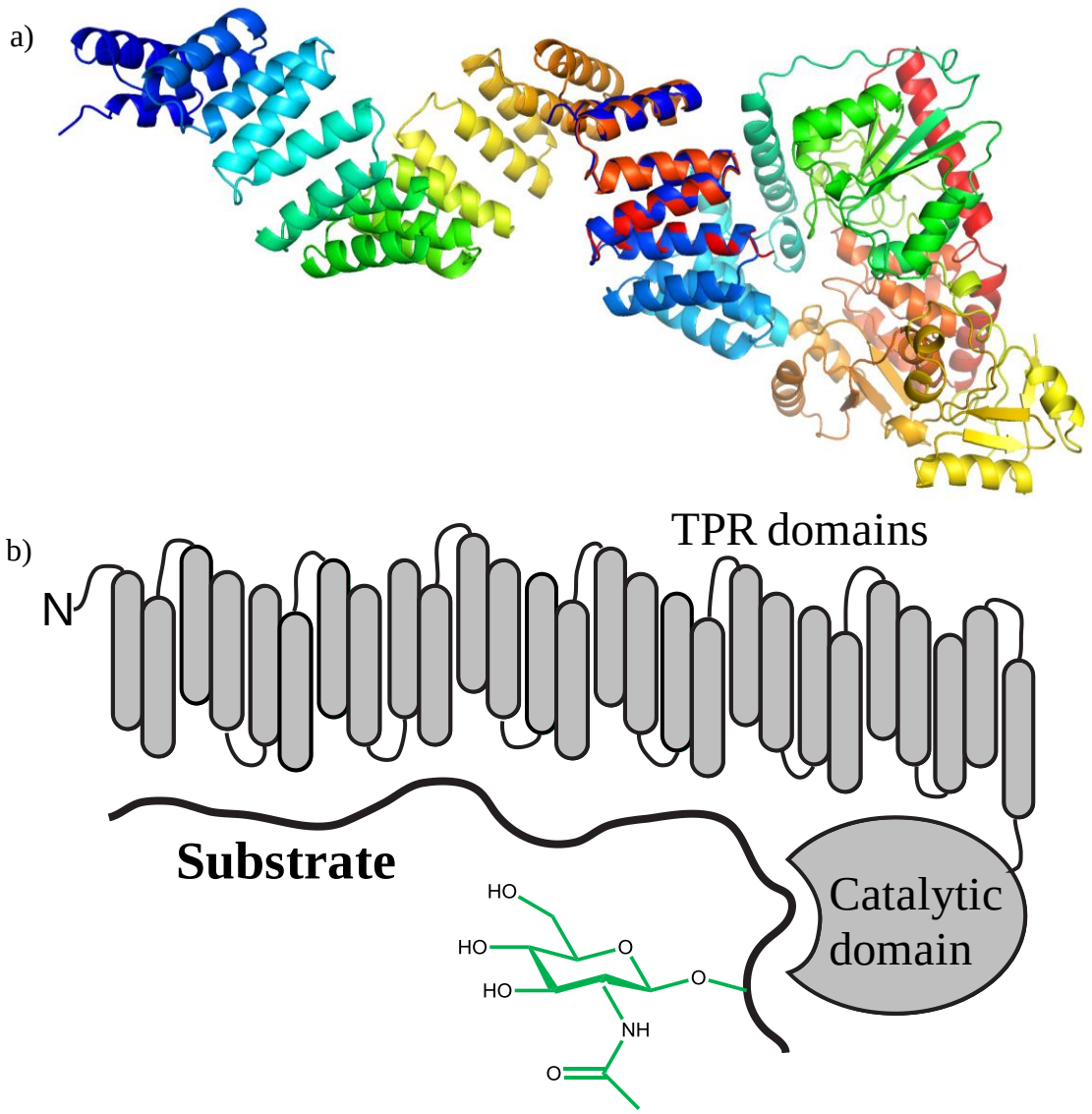

Figure 1.3: a) OGT structure modeled from Conti's 11.5 TPR crystal structure (left side, blue to red) with Walker's catalytic domain and 4.5 TPR (right side, blue to yellow)[15,17]. b) A cartoon model representing the postulated role of the TPR domains in substrate recognition.

OGA also has two domains: an N-terminal GH-84 family glycoside hydrolase domain, and a C-terminal domain that possesses homology to histone acetyltransferases (HATs)[12]. There are two human splice variants, the full length version localizes to the cytosol (hOGA-L), and a shorter variant, (hOGA-S) that localizes to the nucleus. hOGA-S lacks the

HAT-like domain[14]. In rats, it was found that hOGA-L is expressed at low levels during early development and increases during development. hOGA-S is expressed more during early development than later on[14]. Interestingly, hOGA-S seems to exhibit much lower enzyme activity *in vitro*, suggesting that the HAT-like domain is required for OGA to be fully active, although, the role of this domain remains unknown[13,18,19].

Although OGA regulation is not completely understood, the solved structures of two close bacterial homologues of human OGA provided insights on the catalytic mechanism of this enzyme[20-22]. In addition, the OGA active site was found to contain a pocket below the acetamido group of the substrate, typical for GH84s, which was exploited to generate inhibitors with high specificity[22]. These inhibitors now provide an avenue to study the biological relevance of OGA to *O*-GlcNAcylation.

1.4 Crosstalk between *O*-GlcNAcylation and phosphorylation

Crosstalk between *O*-GlcNAcylation and phosphorylation is extensive and has often been described as a "yin and yang" relationship[23]. These two reversible modifications of serine and threonine hydroxyls occur on a similar time scale (Figure 1.2). Furthermore, all the proteins that have been determined to be *O*-GlcNAcylated have also been found to be phosphorylated[7]. There are clear cases where the two modifications compete for the same acceptor site. For example, c-Myc, estrogen receptor β, and RNA polymerase II are reciprocally modified on the same serine/threonine residues[24-26]. Similarly, many proteins, including p53, CAMKIV, and FOXO1, are competitively modified by *O*-GlcNAc or phosphate at proximal residues[27-29].

The addition of *O*-GlcNAc and phosphate is not always antagonistic. Inhibiting a kinase, such as GSK3β, increases the level of *O*-GlcNAcylation of many cytoskeletal and heat shock proteins, but decreases *O*-GlcNAcylation of many transcription factors and RNA-binding proteins[30]. In another study where OGA was inhibited, *O*-GlcNAcylation increased three-fold, and of the 700 monitored sites, levels of phosphorylation either significantly increased or decreased[31].

To further illustrate the extent of crosstalk, *O*-linked GlcNAc cycling enzymes have been found to be phosphorylated. OGT is both tyrosine and serine phosphorylated, and the phosphorylation is thought to activate OGT and assist in substrate recognition[32]. OGA has also been noted to be serine phosphorylated as well as *O*-GlcNAcylated[8,33]. And an ever increasing number of phosphate cycling enzymes are concluded to be *O*-GlcNAc-modified. A functional complex has been found that contains protein phosphatase I and OGT, allowing the enzyme complex to remove a phosphate and add an *O*-GlcNAc to the same substrate[23]. As well, there are examples where OGT and OGA are in complexes with both kinases and phosphatases[7]. Considering all of this data, it is clear that there is no simple explanation describing the relationship between *O*-GlcNAcylation and phosphorylation. Additionally, there have been suggestions of crosstalk between *O*-GlcNAcylation and other PTMs[7].

1.5 *O*-GlcNAc and Alzheimer's disease related protein tau

Second to the pancreas, *O*-GlcNAcylation is most abundant in the brain[7]. Interestingly, many studies have indicated a connection between glucose metabolism in the brain and neurodegenerative diseases. In particular, glucose metabolism is impaired in Alzheimer's disease (AD) neurons, thereby reducing *O*-GlcNAc levels of proteins, such as tau[34]. Tau is the major protein comprising neurofibrillary tangles associated with AD, and is

extensively and reciprocally phosphorylated and *O*-GlcNAcylated[35]. Studies using mouse models have demonstrated that overexpression of OGT in neurons increases *O*-GlcNAcylation and decreases phosphorylation of tau. Conversely, deletion of OGT leads to tau hyperphosphorylation, and ultimately neuronal death[36].

Thus far, the only known function of tau is to bind to and stabilize microtubules (MTs) and is preferentially localized in neuronal axons[37]. The largest isoform of tau is comprised of 441 amino acid residues, and is an intrinsically disordered protein in isolation[38]. Tau possesses an N-terminal domain, referred to as the projection domain, where there is zero, one, or two 29-amino acid long acidic inserts (Figure 1.4a). In the carboxyl half, referred to as the microtubule-binding domain, there are either three or four imperfect 18-residue MT-binding repeats. These differences represent the six isoforms of human tau, all derived from the same gene through alternative splicing[39]. Immediately upstream of the MT-binding repeats is a positively charged proline-rich region. Although its function is not fully characterized, it may contribute to binding negatively-charged MTs as well[40].

Tau has approximately 80 phosphorylation sites, and when hyperphosphorylated is known to be pathogenic[41]. Electrostatic repulsion of the negatively-charged phosphate groups with MTs results in an inability to bind, thereby increasing the concentration of free tau[42]. As the concentration of free tau increases, the probability of aggregation also increases. When tau is misfolded, the likelihood of first forming paired helical filaments (PHFs) and subsequently forming neurofibrillary tangles (NFTs) is increased. Paired helical filaments are formed by the MT-binding repeats stacking to form β-sheets[43].

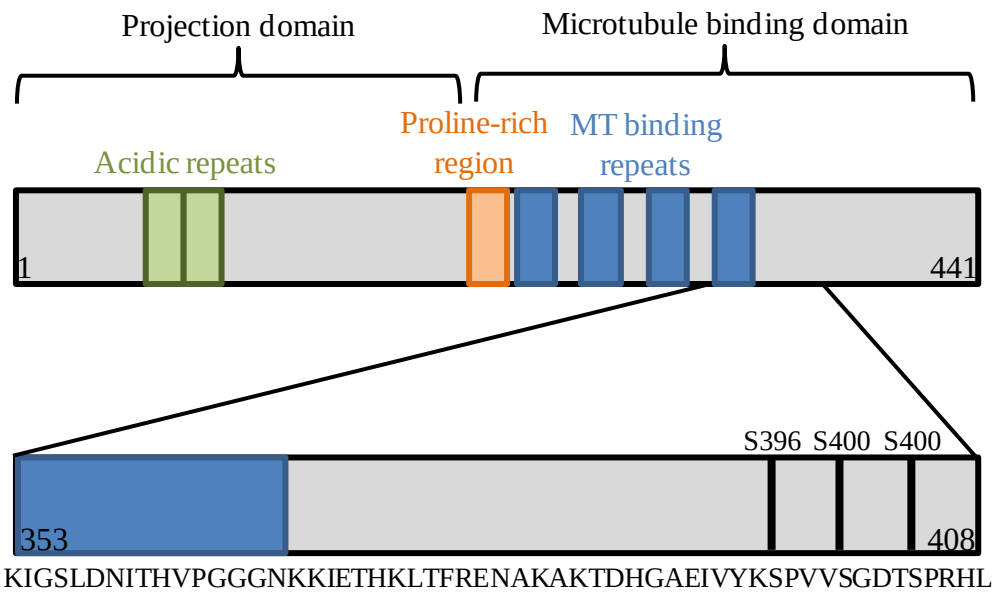

Figure 1.4: Domain structure of tau. a) Relative to the full-length isoform of human tau, six different isoforms resulting from alternative splicing have either zero, one, or two, acidic repeats, and either three or four MT-binding repeats. b) Tau$^{353\text{-}408}$, the fragment studied herein.

PHFs from AD brains contain a protease-resistant core that coincides with the MT-binding repeats[44]. This is corroborated by the increased efficiency of the isolated MT-binding repeats aggregating relative to the full-length isoform of tau[45]. *In vivo* data has suggested that apoptotic caspases cleave tau at the N-terminal and C-terminal ends, yielding fragments that are more prone to aggregation. C-terminal truncations have been found to occur at Glu391 and Asp421 (Figure 1.5b) [46]. As a result, Mandelkow and coworkers suggest a global "paperclip fold" for tau based on fluorescence resonance energy transfer (FRET) and electron paramagnetic resonance (EPR) analyses (Figure 1.5a)[47]. This agrees with the idea that when tau is truncated, the MT-binding repeats become more accessible, thereby increasing the rate of aggregation[47]. One of their more recent FRET and EPR analyses on phosphomimic

mutants to glutamic acid of tau have suggested a further opening and compaction of the paperclip (Figure 1.5c)[48].

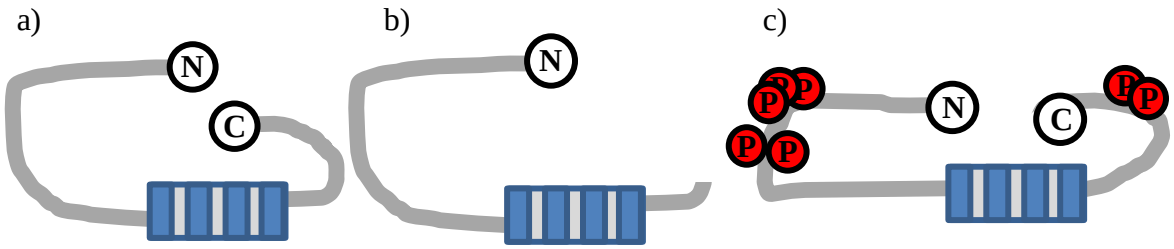

Figure 1.5: A model representing global folding of a) full length wild-type tau, b) C-terminally truncated tau, and c) with phosphomimic glutamic acid mutations. C-terminally truncated tau results in more accessible MT-binding repeats, whereas the phosphomimics result in a more open and compact conformation of tau[47,48].

In addition to extensive phosphorylation, tau is *O*-GlcNAc modified to a level of approximately 4 mol *O*-GlcNAc/mol tau[35]. There has also been a reciprocal relationship observed between *O*-GlcNAc and phosphorylation[49]. For example, brains associated with AD posses lower amounts of *O*-GlcNAc, and intriguingly, *O*-GlcNAc has not been detected on PHFs or NFTs[49,50]. As such, *O*-GlcNAc could be seen as a protective mechanism by limiting the amount of tau phosphorylation, and thus, aggregation[51]. When OGA is inhibited, tau from healthy rats had reduced overall phosphorylation, as well as 1.5-fold increased Ser400 *O*-GlcNAcylation. This reflects a competition between GlcNAcylation and phosphorylation[51,52]. Conversely, Ser396, Ser400, and Ser404 can be phosphorylated[53]. In fact, a sequential phosphorylation pathway has been proposed for this region, where Ser404 phosphorylation by CDK5 primes the cascade, which then facilitates Ser400 and Ser396 phosphorylation by GSK3β[53]. GlcNAcylation is in direct competition with phosphorylation to modify Ser400, and can disrupt the phosphorylation cascade, thus affecting proximal residues (Figure 1.6). Vocadlo and coworkers also showed that *O*-GlcNAcylated Ser400 inhibited tau

oligomerization relative to wild-type species[51]. Due to the reciprocal nature of phosphorylation and *O*-GlcNAcylation on tau, it is an attractive model for studying these PTMs.

Figure 1.6: Post-translational modifications of S396, S400, and S404. a) Sequential phosphorylation pathway, where S404 phosphorylation by CDK5 promotes phosphorylation of S400, then S396 by GSK3β. b) *O*-GlcNAcylation at S400 inhibits this sequential phosphorylation pathway.

1.7 Thesis overview

In this thesis, tau will be used as a model to study the structural and functional effects of *O*-GlcNAcylation. We hypothesize that *O*-GlcNAcylation of Ser400 results in local structural effects that decrease/prevent the polymerization of tau. Therefore, the goals of this thesis were primarily to compare the local structure and dynamic effects of *O*-GlcNAcylation on tau, and consequentially, its functional effects on aggregation. To gain a full understanding of the system, the effects of phosphorylation mimics on tau were also considered.

In order to study *O*-GlcNAc-modified tau, an experimental method to post-translationally modify tau in milligram quantities was required. Our collaborator, Dr. David Vocadlo (SFU) developed a system to *O*-GlcNAcylate tau in bacterial cells by co-transduction with an inducible plasmid encoding hOGT. Upon induction, both tau and hOGT

are produced, and over time, hOGT modifies tau *in vivo*. Because complete and specific phosphorylation is difficult to achieve *in vitro* or *in vivo*, phosphorylation mimics were used to study its effects on tau instead. In mutating the key serine residues (396, 400, and 404) to aspartic acid residues, a single negative charge was introduced, yielding a "phosphomimic" (S396D, S400D, and S404D).

The next step was to design a construct that allowed for structural and functional studies of these PTMs on tau. Since local structure and dynamic studies were to be performed using nuclear magnetic resonance (NMR) spectroscopy, studying full-length tau, with 441 residues, was overly challenging due to line broadening and signal overlap. Furthermore, tau is an intrinsically disordered protein, and these tend to have poor chemical shift dispersion, increasing the difficulty of spectral interpretation. Therefore, a smaller model fragment was needed. However, to perform functional aggregation studies, it is necessary to include at least one MT-binding repeat. Finally, there has been an extra C-terminal *O*-GlcNAcylation site proposed to be on either Ser409, Ser412, or Ser413[54]. To avoid multiple *O*-GlcNAcylation sites, the final construct that I chose to study spanned residues 353-408 and is denoted as tau$^{353\text{-}408}$ (Figure 1.4b).

Complete NMR assignments were achieved for wild-type, *O*-GlcNAc-modified, and S400D tau$^{353\text{-}408}$. Based on these chemical shifts, secondary structure propensity calculations revealed that S400D and *O*-GlcNAc-modified tau$^{353\text{-}408}$ maintains its random coiled conformation. Although its structure was unchanged, ^{15}N relaxation experiments displayed a small decrease in nsec-psec time scale conformational flexibility for amides near position 400. Finally, functional aggregation studies revealed that *O*-GlcNAc-modified tau$^{353\text{-}408}$ had a

reduced amount of aggregation *in vitro* compared to phosphomimic and wild-type tau$^{353\text{-}408}$ peptides.

Chapter 2: Materials and methods

2.1 Cloning

2.1.1 His_6-SUMO-$tau^{353-408}$

Restriction-free cloning was used to create a plasmid encoding His_6-SUMO-$tau^{353-408}$ following published protocols[55]. SUMO is known to enhance the expression and solubility of proteins in *Escherichia coli*, this construct was designed with the goal of increasing the yield of $tau^{353-408}$ [56]. Primers, listed in Table 2.1, were used to PCR amplify the DNA for human $tau^{353-408}$ (HGNC ID: 6893)[57]. PCR protocols and reagents used are listed in Tables 2.2 and 2.3, respectively. The resulting DNA was separated on a 1% agarose gel, purified with a GeneJET Gel Extraction Kit (Fermentas), and inserted into a His_6SUMO-containing vector via linear amplification (Tables 2.2 and 2.3). After purification with a GeneJet PCR purification Kit (Fermentas), the resulting plasmid was transformed into *E. coli* DH5α cells via heat shock. Heat-shocked cells were incubated overnight on Luria broth (LB) agar plates at 37 ºC with selection for kanamycin resistance. The final plasmids were extracted using a GeneJET Plasmid Miniprep Kit (Fermentas) and submitted for DNA sequencing via GENEWIZ, Inc. The confirmed plasmid was transformed into *E. coli* BL21(λDE3) cells by heat shock.

2.1.2 Site-directed mutagenesis for aspartic acid mutants

QuikChange site-directed mutagenesis was performed to generate phosphorylation mimic constructs. Residues 396, 400, and 404 were mutated into aspartic acid residues, independently, using primers listed in Table 2.1. Amplification of desired plasmids was

achieved using protocols and reagents listed in Tables 2.2 and 2.3. Purification, plasmid extraction, sequencing, and transformations were performed as previously mentioned.

Table 2.1: Primers used for cloning

Construct	Primer	Sequence
SUMO-tau$^{353-408}$	Forward	5’- GTTTATCAGGAACAAACGGGGGGTAAGATTGG GTCCCTGGACAATATC-3’
	Reverse	5’-GTTAGCAGCCGGATCTCAGAGATGCCGTGGAGA CGTGTC-3’
S396D	Forward	5'-GCGGAGATCGTGTACAAGGACCCAGTGGTG-3'
	Reverse	5'-CACCACTGGGTCCTTGTACACGATCTCCGC-3'
S400D	Forward	5'-CCAGTGGTGGATGGGGACACG-3'
	Reverse	5'-CGTGTCCCCATCCACCACTGG-3'
S404D	Forward	5'-GGGGACACGGATCCACGGCATCTCAGC-3'
	Reverse	5'-GCTGAGATGCCGTGGATCCGTGTCCCC-3'

Table 2.2: Thermocycling protocols used for cloning

Steps	**PCR amplification**		**Linear amplification**		**Site-directed mutagenesis**	
	Temp.	**Time**	**Temp.**	**Time**	**Temp.**	**Time**
1. Initial	95 ºC	5 min.	95 ºC	5 min.	95 ºC	5 min.
2. Denaturating	95 ºC	30 sec.	95 ºC	30 sec.	95 ºC	30 sec.
3. Annealing	65 ºC	1 min.	65 ºC	1 min.	65 ºC	1 min.
4. Extending	68 ºC	1 min.	72 ºC	13 min.	72 ºC	14 min.
5. Ending	68 ºC	5 min.	72 ºC	15 min.	72 ºC	15 min.
Repeat 2-5	35 times		35 times		35 times	

Table 2.3: Reagents used for cloning

Reagents	PCR amplification		Linear amplification		Site-directed mutagenesis	
	stock	volume	stock	volume	stock	volume
Plasmid/Template	50 ng/µL	1 µL	50 ng/µL	1 µL	50 ng/µL	1 µL
PCR product	-	-	300 ng/µL	2 µL	-	-
Forward primer	100 pM	2 µL	-	-	10 µM	1 µL
Reverse primer	100 pM	2 µL	-	-	10 µM	1 µL
dNTPs	10 mM	4 µL	10 mM	1 µL	10 mM	2 µL
10x buffer	-	10 µL	-	5 µL	-	5 µL
DMSO	-	-	-	-	-	2.5 µL
ddH_2O	-	80 µL	-	40 µL	-	36.5 µL
pfu polymerase	-	-	2.5 u/µL	1 µL	2.5 u/µL	1 µL
Taq polymerase	5 u/µl	1 µL	-	-	-	-
Total		100 µL		50 µL		50 µL

2.2 Protein expression and purification

To express *O*-GlcNAc-modified tau$^{353\text{-}408}$, an ampicillin-selectable plasmid containing the OGT gene was co-transformed with the kanamycin-resistant His_6-SUMO-tau$^{353\text{-}408}$ plasmid into *E. coli* BL21(λDE3)[51]. In instances where there were no *O*-GlcNAc modifications, a catalytically inactive OGT gene was co-transformed as a control instead.

Tau constructs were expressed in *E. coli* BL21(λDE3) cells and grown at 37 ºC until $O.D._{600}$ = 0.6. LB media was used to produce unlabeled protein, whereas M9 minimal media supplemented with 1 g/L $^{15}NH_4Cl$ or 1 g/L $^{15}NH_4Cl$ and 3 g/L $^{13}C_6$-glucose (Sigma-Aldrich) was used to produce ^{15}N- or $^{15}N/^{13}C$-labeled protein, respectively. Protein expression was induced with 0.5 mM IPTG for 16 hours at 16 ºC, and harvested by centrifugation at 4000×g in a GSA rotor (Sorvall) for 15 minutes. After one freeze-thaw cycle, the cell pellet was resuspended in Ni-NTA binding buffer (20 mM NaH_2PO_4 (pH 7.4), 500 mM NaCl, 5 mM imidazole). A minimum of 2 mg/mL of lysozyme was added and the sample was left rocking

on ice for 30 minutes. Cells were then disrupted by sonication (Branson Sonifier 250, VWR Scientific) at 60% duty cycle until clarified. The lysate was spun at 26,000×g for 60 minutes in a SS34 rotor (Sorvall). The supernatant was then passed through a 0.8 µm filter before being applied to a 5 mL Ni-NTA column (Qiagen). The column was washed with 20 mM NaH_2PO_4 (pH 7.4), 500 mM NaCl, 30 mM imidazole, and the protein was resolved using an FPLC (ÄKTA prime plus, General Electric) with 100 mL of buffer increasing the amount of imidazole linearly to 250 mM. Fractions containing the desired His_6-SUMO-$tau^{353-408}$ were identified by 15% SDS-PAGE gels and pooled.

The His_6-SUMO tag was removed, without leaving any residual amino acids, using the catalytic domain of the *Sacchroymces cerevisiae* SUMO hydrolase Ulp1. SUMO hydrolase was expressed and purified as published, using a clone provided by Dr. Keith Vosseller (Drexel University)[58]. Cleavage was performed overnight at room temperature with 5 µg/mL Ulp1 while dialyzing in 20 mM NaH_2PO_4 (pH 7.0) and 150 mM NaCl. Cleavage was verified on a 15% SDS-PAGE gel. The contents of the dialysis bag were centrifuged at 5000×g for 10 minutes, and the supernatant was applied to a Ni-NTA column to separate uncleaved protein and His_6-SUMO from the desired, cleaved $tau^{353-408}$, found in the flow-through.

$Tau^{353-408}$ was then concentrated and purified by reverse phase-HPLC (Dionex). $Tau^{353-408}$ was loaded onto a C_{18} 250×10 mm semi-preparative column (Higgins Analytical, Inc.) and subsequently eluted with a 0%-60% acetonitrile gradient (Fisher Chemical) with 0.1% trifluoroacetic acid (TFA) (Sigma) at 1 mL/min over 80 minutes. Fractions (1 mL) were collected with a Gilson FC205 fraction collector (Mandel Scientific Company Ltd.), and those containing pure $tau^{353-408}$, were pooled and lyophilized.

Since there was only ~10% *O*-GlcNAcylation *in vivo*, *O*-GlcNAc-modified fragments were subjected to a more stringent purification process to separate *O*-GlcNAc-tau$^{353-408}$ from unmodified tau$^{353-408}$. A semi-preparative 250×9.4 mm C_8 column (Agilent) was used in RP-HPLC. The fragments were moderately separated using a 22%-27% acetonitrile with 0.1% TFA over 100 minutes at 1 mL/min. Again, 1 mL fractions were collected. Using MALDI-TOF MS, resulting fractions were analyzed for *O*-GlcNAc-tau$^{353-408}$. Samples containing mostly *O*-GlcNAc-tau$^{353-408}$ were pooled and lyophilized, yielding a ~60% *O*-GlcNAc-modified tau$^{353-408}$ sample.

2.3 NMR spectroscopy

To characterize tau$^{353-408}$, NMR spectroscopy was utilized. Optimal sample conditions were determined initially by recording ^{1}H-^{15}N HSQC spectra as a function of pH from 5.5-8.5 at increments of 0.5 units, followed by optimizing temperature conditions between 5 ºC-35 ºC. Final conditions were chosen to be pH 6.0 and 15 ºC in order to minimize signal loss due to amide hydrogen exchange while yielding good quality spectra under near neutral pH conditions.

All isotopically-labeled samples contained 10 mM NaH_2PO_4 (pH 6.0), 1.6% protease stock inhibitor tablet (Roche), and 10% D_2O for the signal lock. One protease inhibitor tablet was dissolved in 5 mL of 10 mM NaH_2PO_4 (pH 6.0), yielding a 1000× stock. Spectra were recorded at 15 ºC using 600 MHz and 850 MHz Bruker Avance III NMR spectrometers equipped with triple resonance cryo-probes. Spectra were processed and analyzed using NMRpipe[59], nmrDraw[59], and Sparky[60]. The CBCA(CO)NH, HNCACB, HNCO, (H)CC(CO)TOCSY-NH, HNH- (τ_{mix}= 200 ms) and NNH- (τ_{mix}= 200 ms) NOESY-HSQC

are described as per Sattler *et al.*[61]. HNHA and ^{15}N relaxation (T_1, T_2, NOE) experiments were performed as well[62,63].

2.4 Aggregation assays

Aggregation assays were performed at 37 ºC with the same buffer conditions as NMR experiments. Samples (50 µL) contained 30 µM tau$^{353-408}$, 7.5 µM low molecular weight heparin (4.0-6.5 kDa) (International Laboratory, USA), and 0.01 mg/mL thioflavin S (Sigma). These samples were placed in a black flat-bottom 386-well plate (Grenier) and sealed with optically clear crystallography tape (Hampton Research). Fluorescence ($\lambda_{exciation}$ = 440 nm, $\lambda_{emission}$ = 480 nm) was measured every 15 minutes for 40 hours using a Varioskan Flash fluorimeter (Thermo Scientific).

Aggregation assays are extremely sensitive to protein concentrations, thus they were determined by two different spectrophotometric methods in triplicate. The first method involved measuring the absorbance of tyrosine residues at A_{280} with a NanoDrop 200c absorbance spectrophotometer (Thermo Scientific). The concentration was calculated according to Beer's Law, using ε_{280} = 1490 L/mol•cm as predicted by the ExPASy ProtParam tool[64]. The second method utilized a bicinchoninic acid assay (Thermo Scientific) which measures protein concentration based on reduction of Cu^{+2} by peptide bonds[65]. BSA was used to make a standard curve. The final concentration for each protein sample was determined based on the average of the two assays.

Aggregation assays were performed in triplicates with a "no protein" control, as well as a "no heparin" control, since heparin is the aggregation-inducing factor. Three different cultures of ^{15}N-labeled wild-type, S396D, S400D, and S404D were purified individually, and

aggregation assays were performed for each batch in triplicates. To ensure consistency, each batch was checked by recording a ^{1}H-^{15}N HSQC spectrum prior to performing aggregation assays. After the assay was optimized, it was carried out with *O*-GlcNAc-tau$^{353\text{-}408}$ as well. Because *O*-GlcNAc-tau$^{353\text{-}408}$ was extremely difficult to produce, the sample used in the aggregation assays was preserved from prior NMR experiments.

Chapter 3: NMR spectral assignments

3.1 Assignments from main-chain nuclei

A ^{1}H-^{15}N HSQC spectrum is a two-dimensional NMR spectrum that yields a crosspeak at the chemical shifts of each pair of directly bonded ^{1}H and ^{15}N nuclei. Under the conditions used herein, every non-proline amide should yield one signal, and thus the ^{1}H-^{15}N HSQC spectrum provides an "NMR fingerprint" of a polypeptide or protein. Although unstructured polypeptides typically have limited amide $^{1}H^{N}$ chemical shift dispersion, the ^{15}N chemical shift of each residue is affected by its neighboring residues. As a result, there is often sufficient dispersion in the ^{15}N dimension of a ^{1}H-^{15}N HSQC to yield resolved signals from each amide in even a relatively large intrinsically disordered protein. Furthermore, due to the dynamic behavior of unstructured polypeptides, these signals are usually strong and sharp.

Assigning the ^{1}H-^{15}N HSQC spectrum is often the first step in studying protein by NMR spectroscopy. This required the preparation of uniformly ^{13}C/^{15}N-labeled tau and the recording of three-dimensional NMR experiments. Of these, the CBCA(CO)NH and HNCACB experiments are most useful for sequential backbone assignments (Figure 3.1)[66]. The HNCACB correlates the amide $^{1}H^{N}$ and ^{15}N signals of residue *i* to its own $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ as well as those of the preceding residue, *i-1* (Figure 3.2). The complementary CBCA(CO)NH correlates the amide $^{1}H^{N}$ and ^{15}N signals of residue *i* to only the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ signals of the previous residue, *i-1*. Together, these experiments provide sequential connections and allow one to distinguish intra- versus inter- residue correlations. Furthermore, the $^{13}C^{\alpha}$ and 13C^{β} signals may be distinguished as they possess opposite phases in HNCACB experiments. Several additional experiments were also recorded to extend and

verify the spectral assignments of tau. These included an (H)CC(CO)-TOCSY-NH which correlates all of the aliphatic ^{13}C signals of residue *i-1* with the $^{1}H^{N}$ and ^{15}N signals of residue *i*, an HNHA connects the $^{1}H^{\alpha}$, $^{1}H^{N}$, and ^{15}N signals of residue *i*, and the HNH- and NNH-NOESY-HSQC experiments which provide intra-residue NOE correlations between ^{1}H's within 5 Å (Figure 3.2). Finally, the HNCO experiment yields the carbonyl ^{13}C' signal of *i-1* via correlations with the $^{1}H^{N}$ and ^{15}N signals of residue *i*. Collectively, these spectra were used to obtain the chemical shift assignments for wild-type, GlcNAc-modified, and S400D tau$^{353\text{-}408}$, tabulated in Appendix A, and shown in Figure 3.3. All three tau$^{353\text{-}408}$ species gave ^{1}H-^{15}N HSQC spectra with very narrow $^{1}H^{N}$ chemical shift dispersion, indicating that they are predominantly disordered.

In the *O*-GlcNAc-modified tau$^{353\text{-}408}$ sample, there were chemical shift changes in the residues proximal to Ser400, further discussed in section 4.1. Since only 60% of tau$^{353\text{-}408}$ was modified, based on peak intensities, two sets of peaks were observed for residues adjacent to the site of modification. Of particular interest, the signal from the GlcNAc amide group also appeared in the ^{1}H-^{15}N HSQC spectrum (Figure 3.3b). On the other hand, the phosphomimic S400D tau$^{353\text{-}408}$ mutant was 100% "modified," and as such, there was only one set of peaks. The signal from Ser400 no longer existed, and a new peak appeared, corresponding to Asp400. Several residues near Asp400 also exhibited minor chemical shift perturbations.

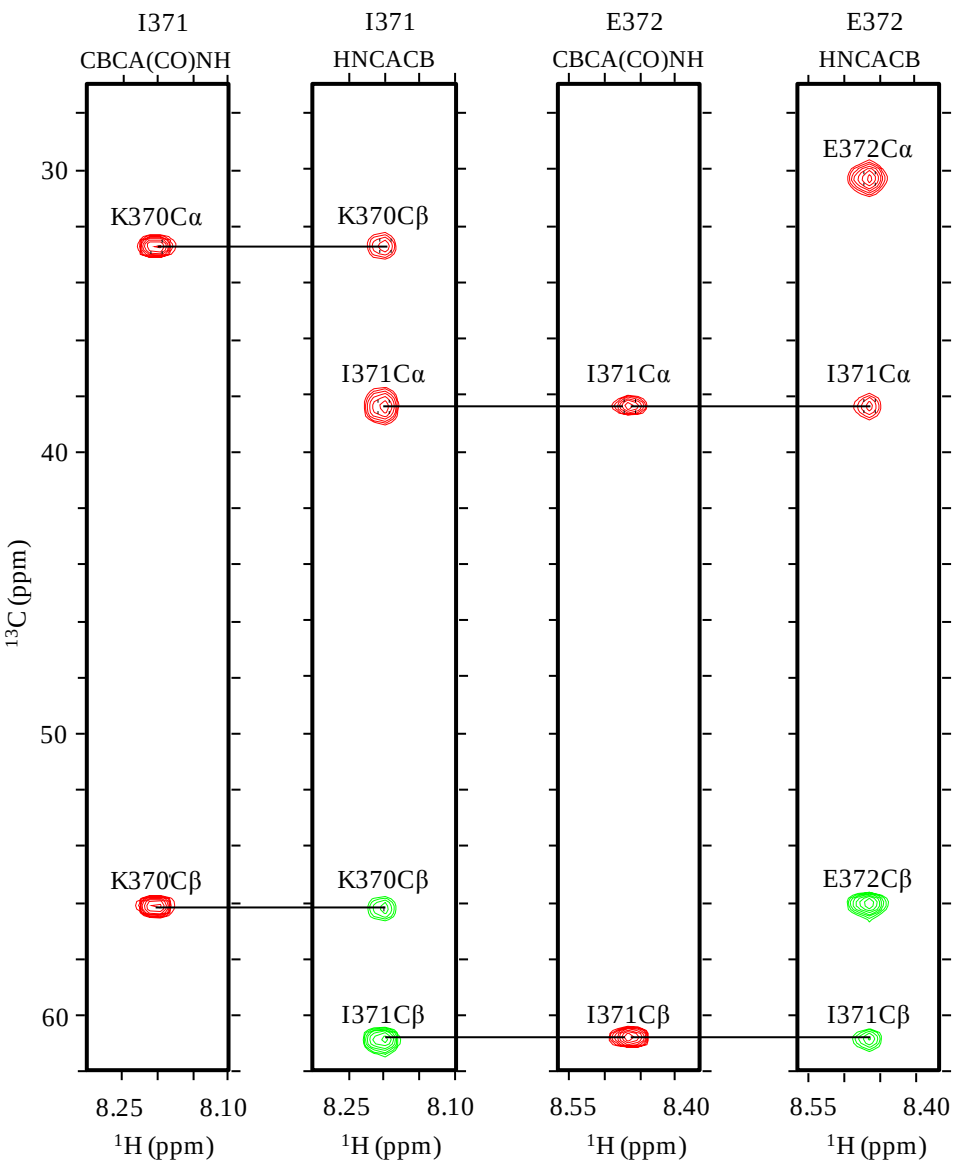

Figure 3.1: Strategies for assigning main-chain resonances. The spectra are sequentially assigned based on the connectivities of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ resonances. The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances of the *i*, and the *i* and *i-1* residues are detected in the CBCA(CO)NH and HNCACB spectra, respectively. Furthermore, in the HNCACB spectrum, the signs of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ peaks are positive (red) and negative (green), respectively.

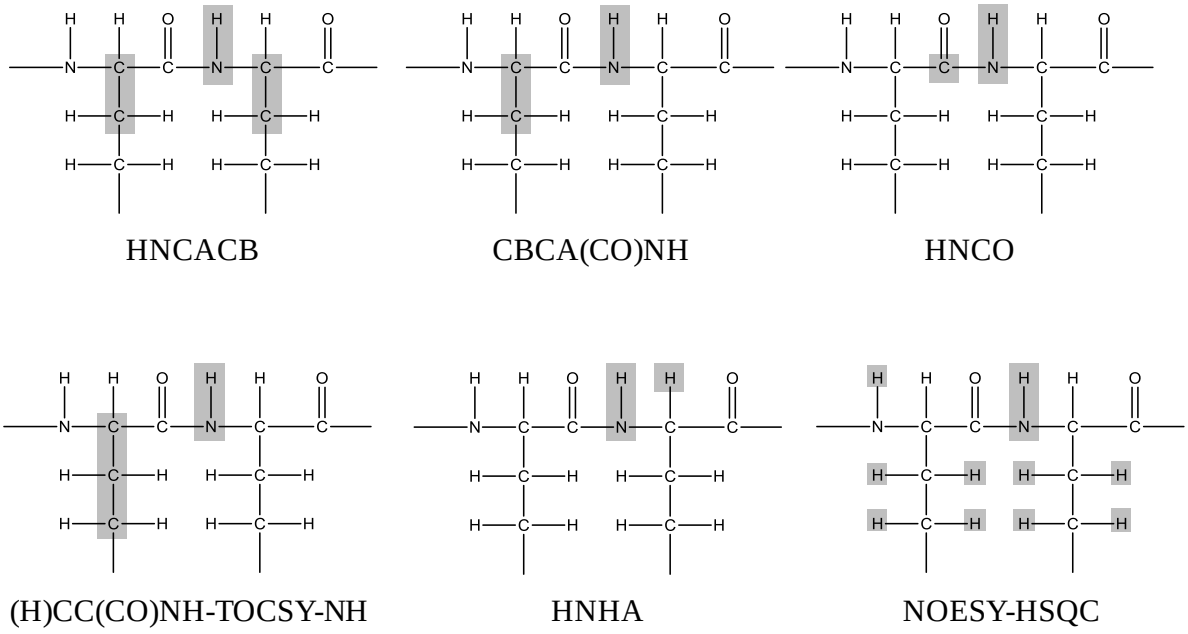

Figure 3.2: Nuclei involved in NMR spectral assignments. The HNCACB experiments correlate amide $^{1}H^{N}$ and ^{15}N signals with the *i* and *i-1* $^{13}C^{\alpha/\beta}$, whereas CBCA(CO)NH and HNCO experiments correlate amide signals with *i-1* $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$, respectively. (H)CC(CO)NH-TOCSY-NH experiments correlate amide signals with all *i-1* aliphatic carbons. The HNHA experiments correlate the amide signals with the $^{1}H\alpha$ of *i*, and also provides a measure of the $^{3}J_{HN\text{-}H\alpha}$ scalar coupling. Finally, NOESY-HSQC experiments resolve through-space NOE interactions between adjacent protons (<5 Å) through the ^{15}N shift of an amide.

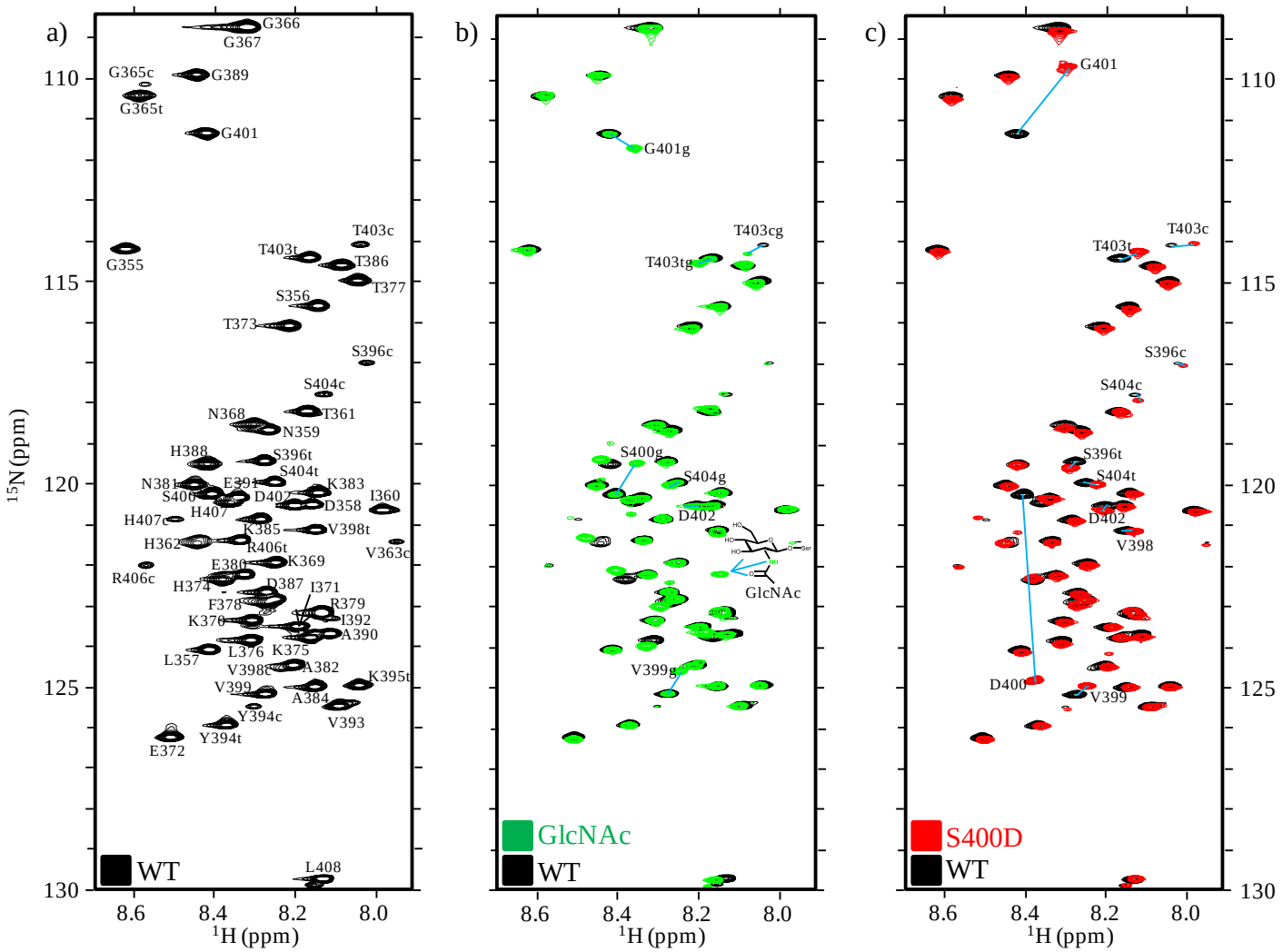

Figure 3.3: Assigned ^{1}H-^{15}N HSQC spectrum of a)wild-type tau$^{353\text{-}408}$ (black). Also shown are the spectra representing b)GlcNAc-modified tau$^{353\text{-}408}$ (green) and c) S400D tau$^{353\text{-}408}$ (red) overlaid on that of the wild-type species, with selected peaks labeled. Distinct signals from residues perturbed by the *trans* (t) and *cis* (c) isomers of Val363-Pro364, Ser396-Pro397, and Ser404-Pro405 are identified. Selected chemical shift changes due to modifications have also been indicated by blue lines.

3.2 *Cis*/*trans* conformational isomers of tau

The ^{1}H-^{15}N HSQC spectra of the three tau$^{353-408}$ species contained an extra set of weaker signals, attributable to the c*is* isomer of the three X-Pro (where X is any amino acid) peptide bonds. The *cis* and *trans* conformers exchange slowly, on the order of tens of seconds, that results in two distinct peaks in the ^{1}H-^{15}N HSQC spectra[67]. Usually *trans* conformers are favored energetically over *cis* conformers (typically 70%-90% *trans* population), and hence results in a larger peak relative to the *cis* conformer[67]. Importantly, these ratios only apply to X-Pro bonds in unstructured polypeptides or in the dynamic termini and loops of structured proteins. In contrast, usually one conformer is stabilized in the well-ordered regions of structured proteins.

The signals from residues perturbed by the *cis*/*trans* isomerization of Val363-Pro364, Ser396-Pro397, and Ser404-Pro405 in tau were assigned with two methods. First, each gave a separate set of $^{13}C^{\alpha}$, $^{13}C^{\beta}$ ^{15}N, and $^{1}H^{N}$ correlations in HNCACB and CBCA(CO)NH spectra. Second, to confirm experimentally which signals arose from which conformer, an (H)CC(CO)-TOCSY-NH spectrum was used to measure the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts of the three prolines. If the chemical shift difference between the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ is ~10ppm, the proline is in the *cis* conformation; however, if the chemical shift difference is ~5 ppm, the proline is in the *trans* conformation[68]. Indeed, as shown unambiguously in Figure 3.4, the weaker signals arose from sub-populations of tau$^{353-408}$ in a *cis* conformation.

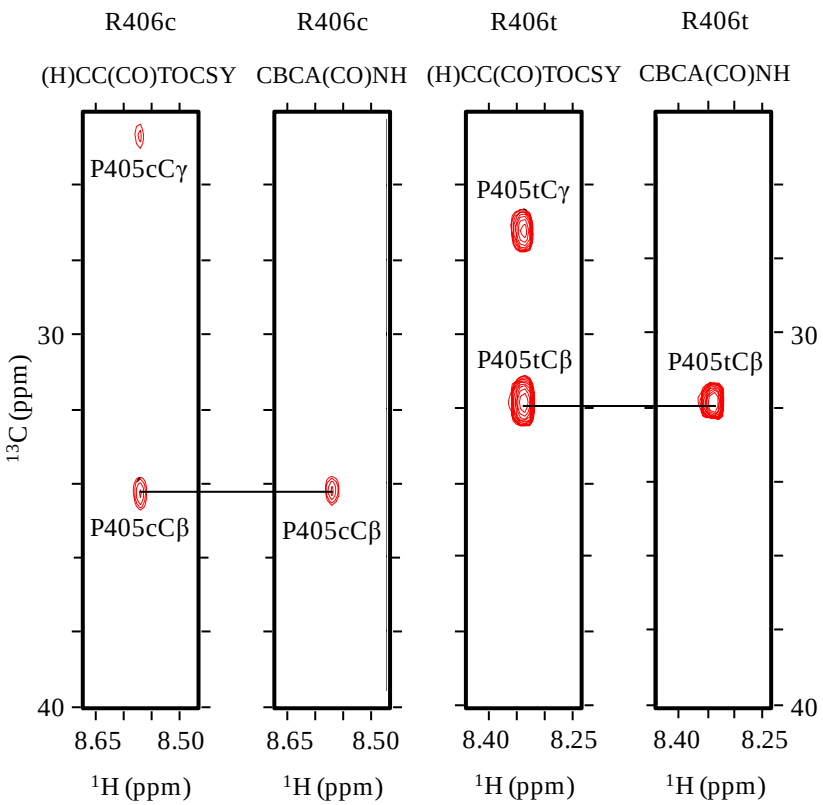

Figure 3.4: *Cis* versus *trans* X-Pro peptide bond assignment. *Cis* and *trans* X-Pro conformations are assigned based on the chemical shift differences between the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ observed in the (H)CC(CO)TOCSY. A chemical shift difference ~10 ppm indicates the *cis* conformer, whereas a difference of ~5 ppm indicates the *trans* conformer. Because $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of the *i-1* residue can be seen in both the (H)CC(CO)TOCSY and CBCA(CO)NH, the *cis* and *trans* peaks can be identified while assigning signals from sequential main-chain nuclei.

3.3 Assignment of *O*-GlcNAc NMR signals

In addition to characterizing the polypeptide, NMR can also provide insights into the covalently linked *O*-GlcNAc moiety. Signals from the $^{13}C/^{15}N$-labeled sugar were also assigned using ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNHA, HNCACB, and HCCH-TOCSY experiments, tabulated in Appendix B, and shown in Figure 3.5. Fortunately, *N*-acetylglucosamine has several easily distinguished signals in both the ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC as shown in Figure 3.3b and Figure 3.5a, respectively[53]. The HNHA experiments

correlated amide ^{1}H-^{15}N and $^{1}H_2$ signals, whereas the HNCACB experiments allowed the chemical shifts of C_1, C_2, and C_3, as well as the methyl carbon nuclei to be assigned (refer to Figure 3.5d for the numbering scheme) for *O*-GlcNAc. An HCCH-TOCSY spectrum correlates the signals from all protons within a ^{1}H-^{13}C spin system to a directly bonded ^{1}H-^{13}C pair. As such, a common pattern of proton chemical shifts appears at ^{1}H-^{1}H planes taken at each sugar ^{13}C chemical shift. The C_6 plane has two of these patterns, since there are two directly-bonded protons, denoted arbitrarily as $H_{6'}$ and $H_{6''}$ (Figure 3.5b). The C_4 and C_5 carbon chemical shifts were determined by a similar pattern as well, and subsequently mapped on the ^{1}H-^{13}C HSQC. These signals could not be identified unambiguously for C_4 or C_5 and were instead discriminated based on previously reported NMR data (HMDB ID: 00803)[69].

Of note, these NMR experiments did not provide a direct link between any nuclei in the *O*-GlcNAc and tau$^{353\text{-}408}$ residues as required to unambiguously identify Ser400 as the site of covalent attachment. However, the largest chemical shift difference between wild-type and *O*-GlcNAc-modified tau$^{353\text{-}408}$ occurred for Ser400 and its neighboring nuclei, as seen in the ^{1}H-^{15}N HSQC (Figure 3.3b). Furthermore, the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of Ser400 in the HNCACB and CBCA(CO)NH were altered by 2 ppm and 7 ppm, respectively, from the Ser400 chemical shifts in wild-type tau$^{353\text{-}408}$. This is entirely consistent with previous studies that used MS to identify Ser400 as the *O*-GlcNAc acceptor[53].

Finally, ^{15}N-resolved NOESY experiments that correlate the signals of ^{15}N-labeled amide protons with those of other protons through space, were performed (Figure 3.2). However, only intra-sugar NOE interactions between the *O*-GlcNAc amide and the $^{1}H_1$, $^{1}H_2$, $^{1}H_3$ and CH_3 were observed. Importantly, no other NOE correlations were observed between

GlcNAc and tau (Figure 3.5c). This reveals that the GlcNAc amide is not localized near the polypeptide chain.

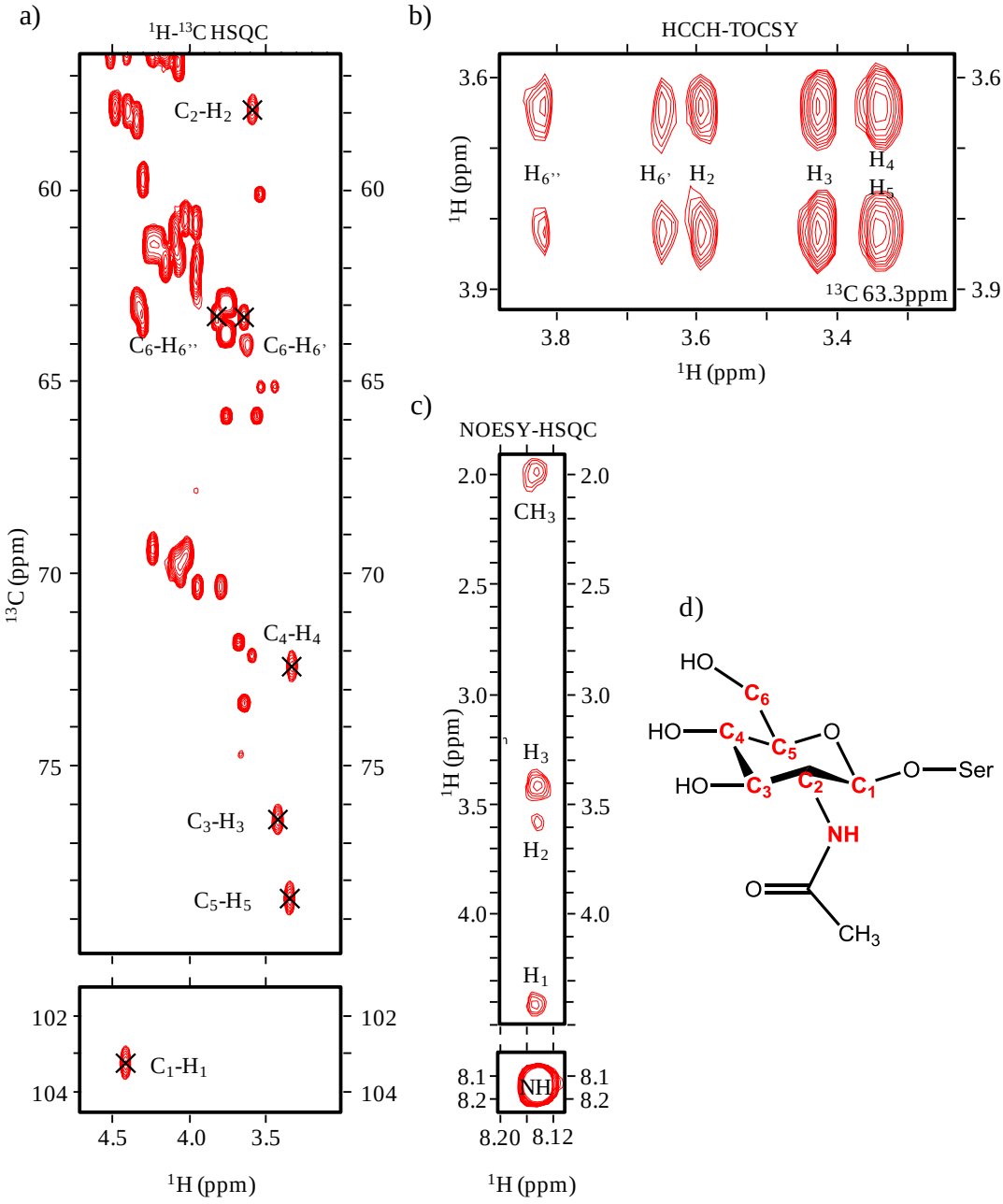

Figure 3.5: Assignment of *O*-GlcNAc NMR signals. a) ^{1}H-^{13}C HSQC with $^{13}C_1$ through $^{13}C_6$ assigned. The acetyl methyl signal at 24.8 and 1.9 ppm is not shown. Although many peaks are seen in this spectrum, the b) HCCH-TOCSY helped discern them from those of the tau polypeptide due to its distinct ^{1}H correlation pattern. c) The HNH-NOESY-HSQC showing only intra-sugar NOE interactions to the amide. d) Numbering scheme for *O*-GlcNAc.

Chapter 4: Structural, functional, and dynamic studies of tau

4.1 Chemical shift perturbations

The change in amide chemical shifts for the *O*-GlcNAc-modified and S400D tau$^{353-408}$ compared to wild-type tau$^{353-408}$ were all in the vicinity of Ser400 (Figure 4.1). However, this is difficult to interpret as amide chemical shifts are extremely sensitive to conformational changes and electrostatic effects. On the other hand, the lack of any significant chemical shift perturbations for the rest of tau$^{353-408}$ strongly indicates that the changes near position 400 do not affect distal residues or the global properties of tau$^{353-408}$.

4.2 Secondary structure propensities derived from chemical shifts

Chemical shift values are very indicative of secondary structure, as such, three different methods were used to interrogate and compare secondary structure in wild-type, *O*-GlcNAc-modified, and S400D tau$^{353-408}$.

4.2.1 δ2D

The main-chain ($^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{1}Ha, $^{1}H^{N}$, and ^{15}N) signals of a residue are strongly dependent upon its secondary structure within a protein[70]. Therefore, numerous algorithms have been developed to accurately predict secondary structure from chemical shifts alone. Of these, the program δ2D (http://www-vendruscolo.ch.cam.ac.uk/d2D/index.php) was specifically calibrated for intrinsically disordered proteins, and provides the normalized propensities of a residue to adopt an α-helical, β-sheet, polyproline II helical or random coil conformation, with scores totaling to one[71].

The results from δ2D analysis of the three forms of tau$^{353-408}$ considered herein are plotted in Figure 4.2. Most significantly, all the residues in each form show random coil

scores of larger than 0.7. This is consistent with the ^{1}H-^{15}N HSQC spectra (Figure 3.3), as well as numerous other studies indicating that soluble tau is an intrinsically disordered protein[72-74].

Upon closer inspection, some small differences are seen between the δ2D results for wild-type, GlcNAc-modified and S400D tau$^{353-408}$ (Figure 4.2). For example, residues around Ser404 have the highest β-sheet propensity in wild-type tau$^{353-408}$, whereas residues around Asp400 in S400D tau$^{353-408}$ have a slightly elevated polyproline II helical propensity. Also, the GlcNAc-modification is predicted to slightly increase the β-sheet propensity around Ser400, and the coil propensity of neighboring residues. However, the significance of these predictions is unclear as the differences in conformational propensities are small. It is important to note that the δ2D database does not contain reference chemical shift values for GlcNAc-modified serine, and treated Ser400 as unmodified. Furthermore, inspection of the assigned NMR chemical shifts for each species relative to those predicted for completely disordered reference polypeptides does not yield an obvious pattern of perturbations indicative of any secondary structure (Appendix C)[75]. Therefore, GlcNAc modification or mutation of Ser400 does not induce any predominant secondary structure.

4.2.2 $^{3}J_{HN\text{-}H\alpha}$ coupling

The three-bond $^{3}J_{HN\text{-}H\alpha}$ coupling constant is dependent upon the φ dihedral angle of a residue and thus is also a measure of secondary structure. For example, residues in helices have $^{3}J_{HN\text{-}H\alpha}$ values around 3-4 Hz, those in β-sheets have values around 9-10 Hz, and those in random coils have couplings around 7 Hz[62]. Using a HNHA experiment the $^{3}J_{HN\text{-}H\alpha}$ coupling constants of the three tau$^{353-408}$ species were measured[62]. As shown in Figure 4.3, the measured couplings were uniformly indicative of a conformational disorder in wild-type,

GlcNAc-modified and S400D tau. An increase in helical or sheet propensity should appear as a cluster of adjacent residues with $^{3}J_{HN\text{-}H\alpha}$ values less than or greater than 7 Hz, respectively, and this was not observed.

4.2.3 NOESY experiments

NOESY experiments correlate through-space dipolar interactions between protons within ~5 Å. In addition to providing tertiary structural information, regular patterns of NOE interactions between main-chain protons are also very diagnostic of secondary structure. These include NOE interactions between the amide $^{1}H^{N}$ of residue *i* with its neighbor's amide protons (*i+1* and *i-1*), as well as *i* to the $^{1}H^{\alpha}$ of itself (*i*) and the preceding residue (*i-1*). The latter are often denoted as $d\alpha_{N}(i,i)$ and $d\alpha_{N}(i\text{-}1,i)$, respectively[70]. Furthermore, a cluster of adjacent residues with $d\alpha_{N}(i,i)$ versus $d\alpha_{N}(i\text{-}1,i)$ NOE intensity ratios of ~6 are indicative of an α-helical structure, whereas ratios of ~0.25 arise from β-sheet structures. In the case of a random coil polpeptide, the $d\alpha_{N}(i,i)/d\alpha_{N}(i\text{-}1,i)$ ratio is in the range of 0.35[76].

HNH- and NNH-NOESY-HSQC spectra were recorded for the tau$^{353\text{-}408}$ species. These spectra further confirmed that tau$^{353\text{-}408}$ is globally unfolded, since many residues had $d\alpha_{N}(i,i)/d\alpha_{N}(i\text{-}1,i)$ ratios ~0.35 (not shown). However, focusing on the residues proximal to residue 400, a trend could not be found in tau$^{353\text{-}408}$ to discern whether there were any local structural changes due to modifications. This is attributed mainly to poor dispersion in the ^{1}H dimension. As a result of overlapping $^{1}H^{\alpha}$ and $^{1}H^{N}$ signals, the $d\alpha_{N}(i,i)/d\alpha_{N}(i\text{-}1,i)$ intensity ratios could not be calculated accurately. Furthermore, there are two proline residues (which lack amide protons required for this experiment) and a glycine residue in this region. Since, the two glycine $^{1}H^{\alpha}$ signals were not resolved, correct $d\alpha_{N}(i,i)/d\alpha_{N}(i\text{-}1,i)$ intensity ratios could not be determined.

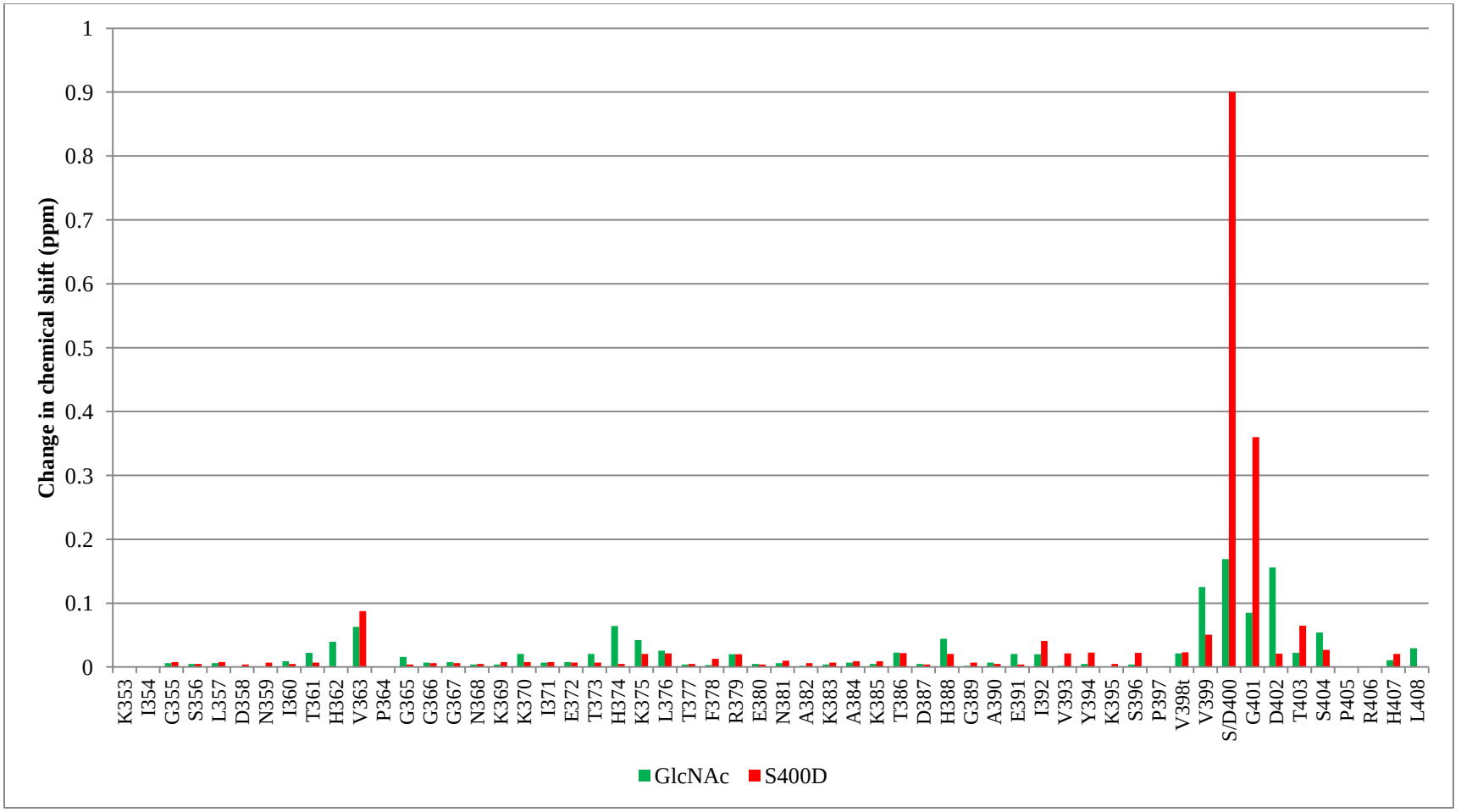

Figure 4.1: Change in amide chemical shifts ($\Delta\delta = [(\Delta\delta_{1H})^2 + (0.2\times\Delta\delta_{15N})^2]^{1/2}$) of the *trans* conformers of GlcNAc-modified and S400D tau$^{353\text{-}408}$ compared to wild-type tau$^{353\text{-}408}$. Significant perturbations occurred only near the 400 residue. Additional minor chemical shift perturbations clustered around histidine residues, and most likely reflect minor changes in sample pH value which was close to the pKa value of ~6 for a histidine side chain.

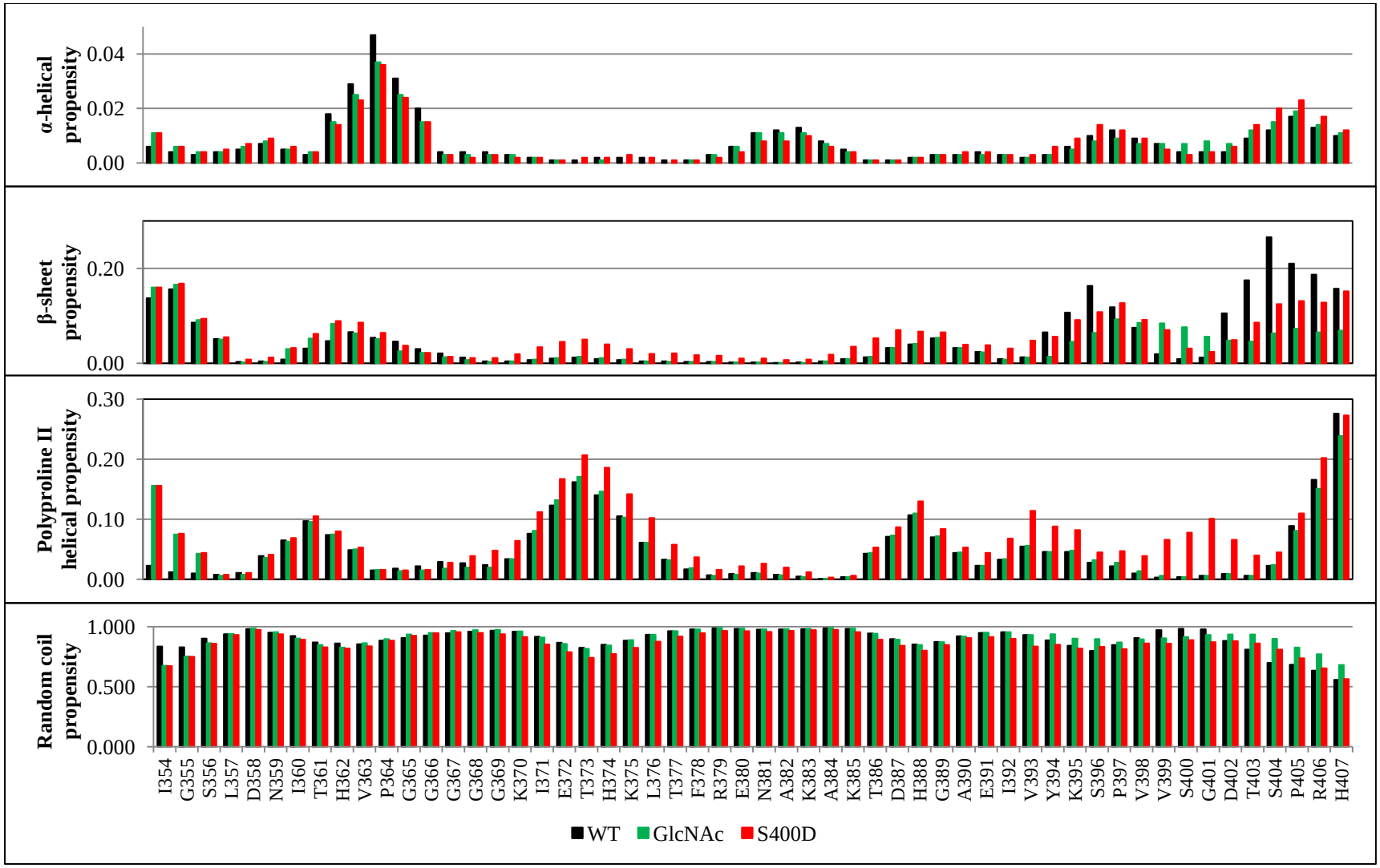

Figure 4.2: Secondary structure propensity calculated by ∂2D for each residue in wild-type (black), GlcNAc-modified (green), and S400D (red) tau$^{353\text{-}408}$. Note: the y-axes have different scales.

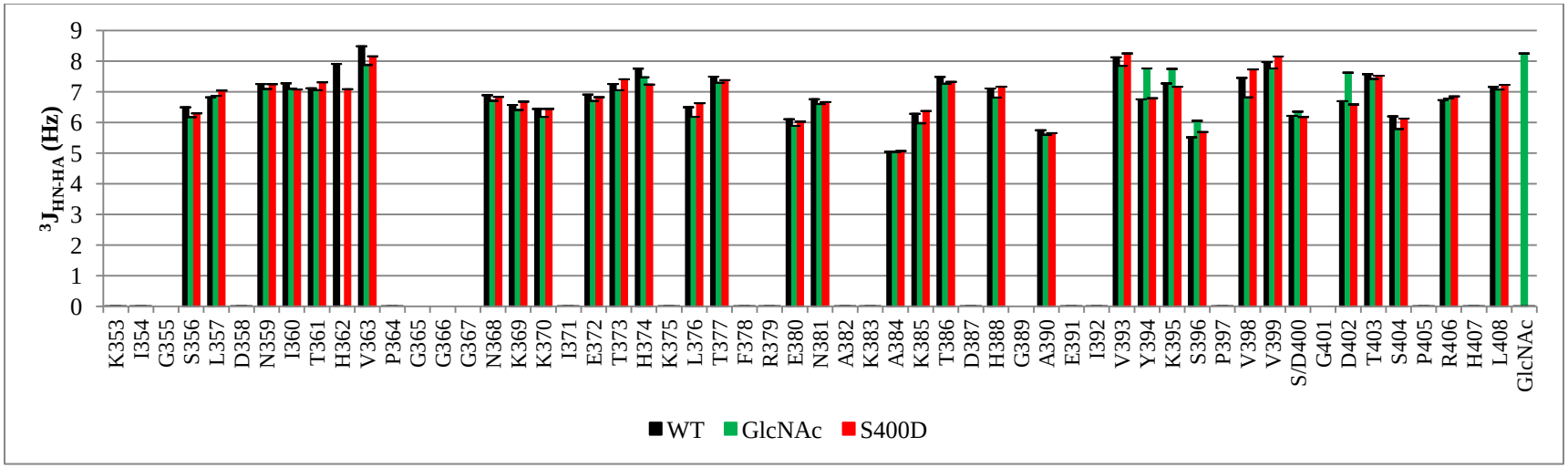

Figure 4.3: $^3J_{HN-HA}$ coupling values for non-glycine residues in tau$^{353-408}$. The three-bond $^3J_{HN-HA}$ coupling values around 3-4 Hz are typical for α-helices, whereas values around 9-10 Hz are typical for β-sheets. Finally, values around 7 Hz, which is observed above, is typical for random coils. Ambiguous values due to spectra overlap have not been shown. Error bars are based on spectral signal-to-noise ratios.

4.3 ^{15}N Relaxation

Based on chemical shifts and $^{3}J_{HN-H\alpha}$ coupling constants, wild-type tau is predominantly disordered and neither GlcNAc-modification nor S400D mutation induces any persistent secondary structure. To investigate this further, amide ^{15}N relaxation (T_2, ^{1}H-^{15}N-NOE) experiments were performed to gain insights on backbone dynamics. Note that the heteronuclear NOE values decrease from 0.8 to -0.3 with increasing mobility of the amide ^{15}N-$^{1}H^{N}$ bond vector on the nsec-psec time scale. Decreasing NOE values accompanied with increasing T_2 values indicate an increase in flexibility, and the reverse trend indicates an increase in rigidity[70].

The amide ^{15}N relaxation data for the three tau species are presented in Figure 4.3. Most notably, the central regions of each have roughly uniform heteronuclear NOE values ~0.35 and T_2 values ~300 msec, whereas the termini have reduced NOE values and increased T_2 values. This pattern is diagnostic of a conformationally disordered polypeptide. However, upon closer inspection, the NOE values of Ser400 and Gly 401 increase slightly relative to those of the wild-type species upon GlcNAc-modification, whereas the T_2 values decrease slightly relative to those of the wild-type species upon GlcNAc-modification, whereas the T_2 values decrease slightly. Similar effects are noticed for the S400D sample. This suggests that the presence of the GlcNAc or Asp400 slightly dampens the motions of these two amides; however, the effect is small. These patterns were also observed within the two populations found in the ~60% modified GlcNAc sample, indicating that these results are not due to differences in experimental conditions. The motions of the GlcNAc amide were also probed by these ^{15}N relaxation measurements. As shown on the very right hand side of Figure 4.4, the sugar amide has an unusually low NOE value of -0.1 and a high T_2 value of ~700 msec

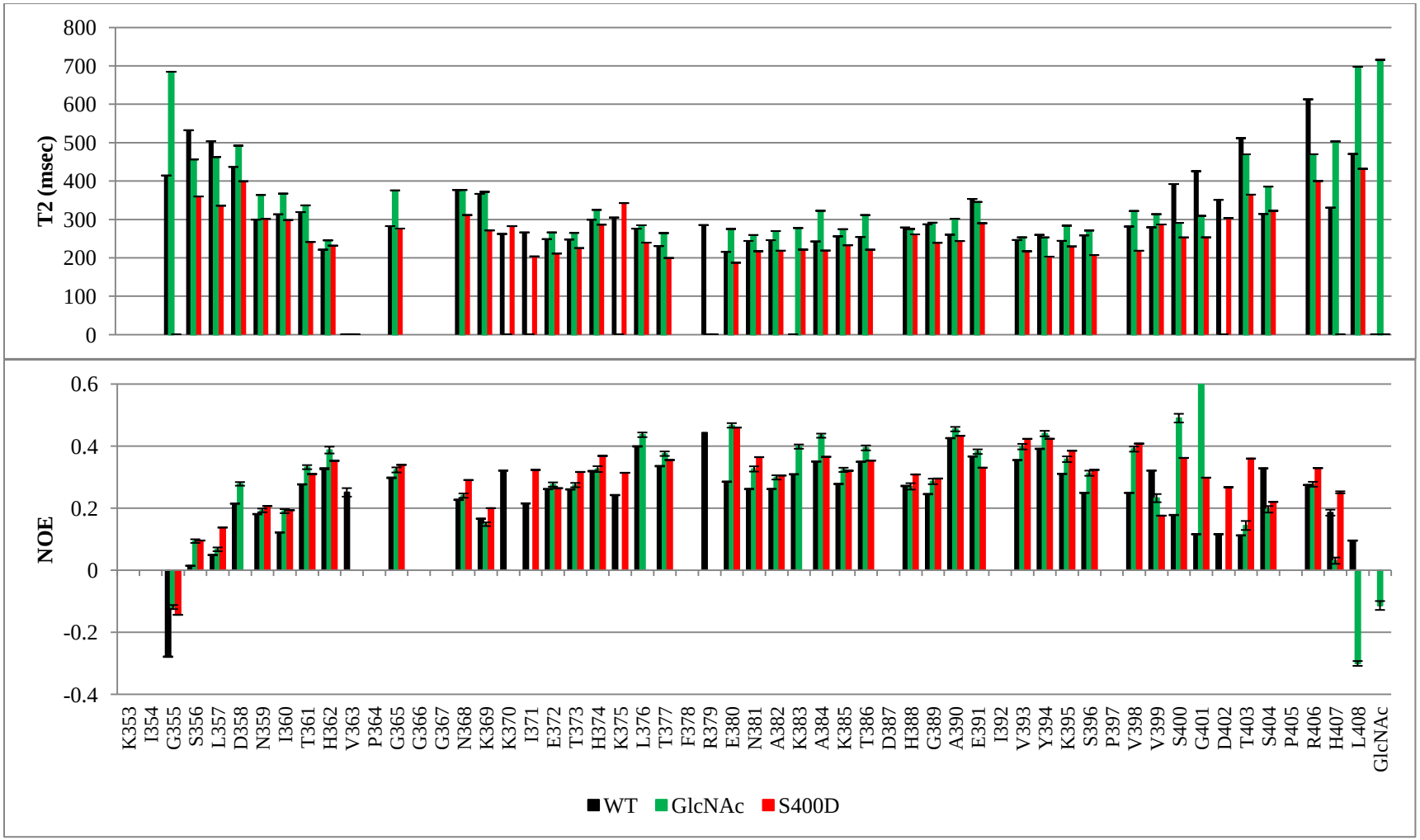

Figure 4.4: ^{15}N T_2, and heteronuclear NOE values for wild-type (black), GlcNAc-modified (green), and S400D (red) tau$^{353-408}$. These were measured on an 850 MHz Bruker AVANCE III NMR spectrometer. T_2 values and their errors were fit by SPARKY[60]. NOE values were calculated by the ratio of peak intensities in experimental and control spectra, and errors were calculated by using signal-to-noise ratios. Missing data correspond to proline residues or residues with signal overlap.

Therefore, it is highly mobile on the nsec-psec timescale, and is not conformationally restrained by the tau polypeptide.

4.4 *Cis*/*trans* isomerization

The possibility that the GlcNAc-modification or S400D mutation altered the *cis*/*trans* ratio of the adjacent to Ser396-Pro397 or Ser404-Pro405 was also examined. These ratios were determined from the relative ^{1}H-^{15}N HSQC peak intensities of amides giving resolved signals attributable to each X-Pro bond (Table 4.1). Note that this assumes similar NMR relaxation properties for the two conformers such that relative peak intensities reflect relative populations. The introduction of a negatively charged aspartic acid, in the place of serine (S400D), results in no significant change for either Ser396-Pro397 or Ser404-Pro405 (Table 4.2). In contrast, it appears that the GlcNAc modification on tau$^{353-408}$ has decreased the *cis* populations slightly for Pro405 and Pro397. However, it is difficult to say with certainty that these values are significant.

Table 4.1: Residues that were used to calculate relative *cis* populations, where y denotes a peak that was used.

Residue	**P364**	**P397**				**P405**			
	G365[a]	**Y394**	**K395[b]**	**S396**	**V398[c]**	**T403**	**S404**	**R406**	**H407**
Wild-type	y	y	-	y	-	y	y	y	y
GlcNAc	-	y	-	y	-	y	y	y	y
S400D	y	y	-	y	y	y	y	y	y

[a] The Gly365 *cis* peak was not observed for GlcNAc-modified tau$^{353-408}$, this may be the result of a minor chemical shift change in this region.

[b] Lys395 was not used for any of the constructs due to an unidentified *cis* peak.

[c] Val398c was only observed for wild-type and S400D tau$^{353-408}$, however, there were overlapping intensities observed in the wild-type spectrum.

Table 4.2: Relative *cis* populations calculated based on peak intensities of n number of residues about the X-Pro bonds. Errors were calculated based on signal-to-noise values for each of the peaks.

Construct	%*cis*					
	P364	**n**	**P397**	**n**	**P405**	**n**
Wild-type	8.0	1	8.7±0.13	2	10.6±0.13	4
GlcNAc	-	0	7.8±0.43	2	7.6±0.44	4
S400D	7.0	1	8.8±0.02	3	10.1±0.01	4

4.5 Aggregation assays

Valuable insight into tau's ability to aggregate *in vivo* can be obtained using polyanions, such as heparin to induce aggregation *in vitro*. Thioflavin S can bind to the resulting aggregates, increasing its relative fluorescence emission at 480 nm by 35-fold[77]. These assays are viewed somewhat qualitatively in terms of the rate and extent of indirectly-detected aggregation.

In this experiment, wild-type, S400D, and *O*-GlcNAc-modified tau$^{353\text{-}408}$ were tested alongside two additional phosphomimic constructs. The latter being S396D and S404D, thus representing all three serine residues that become phosphorylated in wild-type tau via a phosphorylation cascade (Figure 1.6). As shown in Figure 4.5, *O*-GlcNAc-modified tau$^{353\text{-}408}$ yielded the lowest amount of fluorescence, indicating the lowest of aggregation. S396D and S400D tau$^{353\text{-}408}$ mutants had increased aggregation relative to wild-type tau$^{353\text{-}408}$, whereas S404D had decreased aggregation compared to wild-type tau$^{353\text{-}408}$. Overall, it is striking that these single mutations and GlcNAc modifications have such a pronounced effect on the aggregation of this tau fragment, especially given that they are distal to the microtubule-binding repeat.

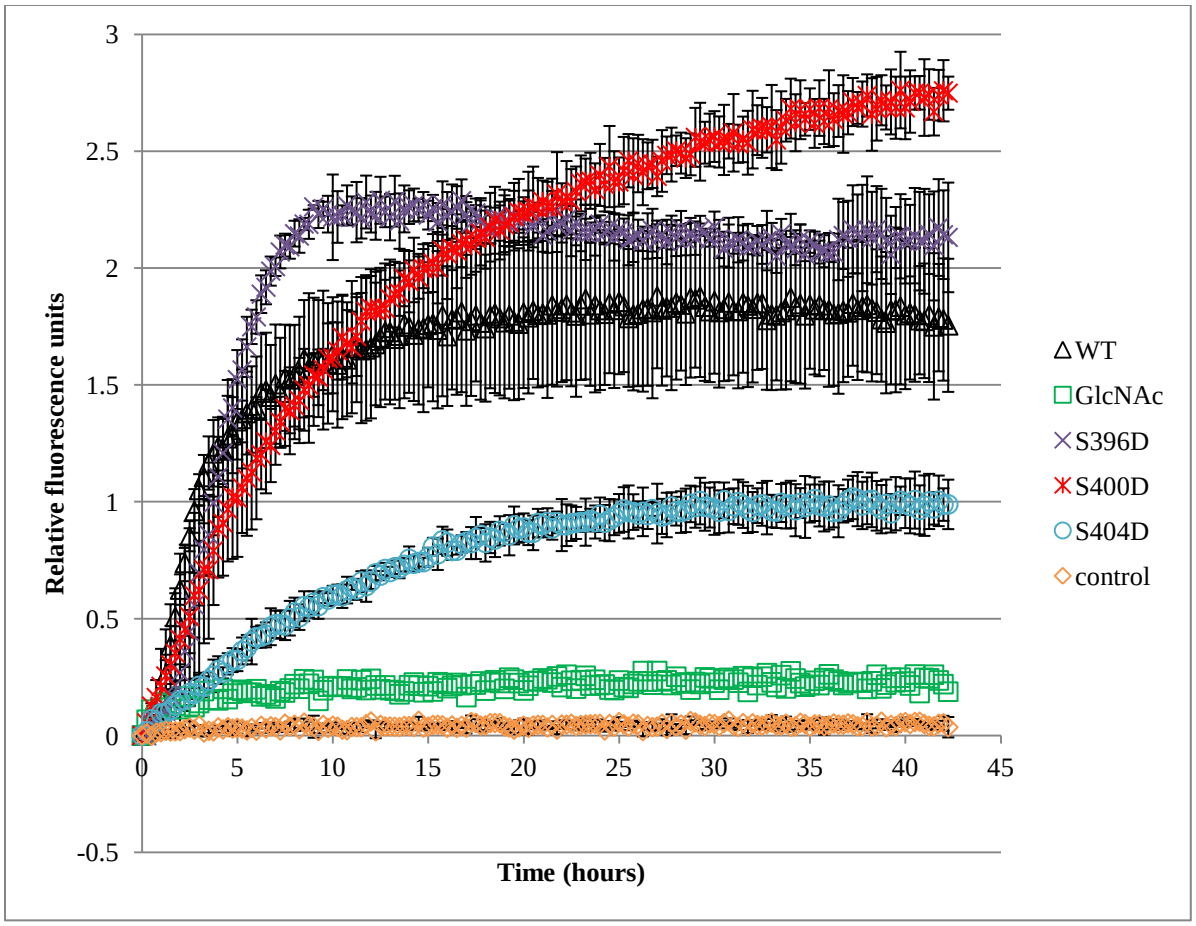

Figure 4.5: Aggregation assay results for tau$^{353\text{-}408}$. S400D (red snowflakes) and S396D (purple crosses) have higher amounts of aggregation relative to wild-type (black triangles), whereas S404D (blue circles) has lower aggregation. Most significantly, GlcNAc-modified tau$^{353\text{-}408}$ showed very little aggregation under these conditions. Average data for triplicates are shown, with error bars representing standard deviations.

Chapter 5: Discussion and conclusions

The goals of my thesis were to determine the effects of *O*-GlcNAc-modified Ser400 on the local structure and dynamics of tau$^{353\text{-}408}$ as well as its effects on heparin-induced aggregation. My initial hypothesis was that there would be a change in local structure which could correlate with reduced aggregation. Indeed, the *O*-GlcNAc modification did decrease the level of tau$^{353\text{-}408}$ aggregation, however, no significant structural changes were observed.

5.1 Strategies to overcome experimental challenges

In order to achieve the goals of my thesis, I had to overcome a number of experimental challenges that will be discussed in this section.

5.1.1 Optimizing protein expression and *O*-GlcNAcylation yield

The original pET28a vectors encoding full-length or C-terminal half constructs of tau, provided by our collaborator Dr. David Vocadlo (SFU) did not yield milligram levels of expressed protein required for typical NMR studies. Therefore, I developed a SUMO-fusion system in order to increase the expression of tau fragments in *E. coli*. Such fusions have proven useful for increasing protein expression and also have the advantage of leaving no extra amino acid acids after cleavage by a SUMO-specific hydrolase[58]. Furthermore, to simplify the resulting NMR spectra, a shorter construct spanning the fourth MT-binding repeat and including Ser400 was chosen. In the end, a plasmid encoding His_6SUMO-tau$^{353\text{-}408}$ was created and expressed high level of tau$^{353\text{-}408}$ required for structural and functional characterization. The next step was to create a GlcNAc-modified peptide.

To date, GlcNAc-modified peptides have mostly been produced synthetically with Fmoc chemistry[78]. However, a 56-residue $^{13}C/^{15}N$-labeled peptide would have been very

costly and difficult to synthesize. Alternatively, *in vitro* methods have been used to *O*-GlcNAcylate peptides, Smet-Nocca *et al*. achieved 8.2% *O*-GlcNAcylation after incubating their tau peptide (residues 392-411) for two days at 37 ºC with recombinant ncOGT and a 10-fold excess of UDP-GlcNAc[53]. Although a viable approach with a biosynthetically-labeled peptide, it would still be very costly to produce $^{13}C/^{15}N$-UDP-GlcNAc as required to label the monosaccharide. The Vocadlo group developed an *in vivo* system that resulted in ~10% *O*-GlcNAcylation tau$^{353-408}$. This was preferred to the *in vitro* method because of the slightly improved levels of *O*-GlcNAcylation without the need to produce and purify OGT, and more significantly, it allowed for uniform labeling of both the peptide and GlcNAc, as required for NMR studies. However, in order to study the local dynamic and structural effects of *O*-GlcNAcylation, spectral assignments were first required.

5.1.2 Bigger is better: assigning IDPs with a high field magnet

Tau$^{353-408}$ is an intrinsically disordered peptide (IDP), and thus has poor spectral dispersion in the ^{1}H dimension. Use of a recently installed high field magnet (850 MHz) tremendously helped in improving spectral resolution as well as sensitivity for tau. Standard main-chain directed 3D experiments became feasible for backbone assignment, which was further helped by the fact that most $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts matched the expected random coil values of each amino acid[79]. This allowed for the full assignment of the signals from backbone nuclei in wild-type, S400D, and *O*-GlcNAcylated tau$^{353-408}$. Furthermore, *cis* conformers signals were also assigned for many residues in the vicinity of the three X-Pro bonds. Since GlcNAc was also $^{13}C/^{15}N$-labeled, it was possible to assign the signals from all of the NMR-active nuclei as well.

In the ^{1}H-^{15}N HSQC spectra, residues adjacent to position 400 underwent chemical shift perturbations. There were no drastic changes in chemical shift, with the exception of Asp400 in the S400D tau$^{353-408}$ ^{1}H-^{15}N HSQC. There were minor chemical shift changes observed near histidine residues (Figure 4.1), as they titrate around pH 6.0, the pH at which NMR experiments were conducted. Since chemical shift perturbations were observed, further structural investigations were made.

5.1.3 Gleaning structural information from chemical shifts in the absence of NOESY-based distance restraints

Residues adjacent to position 400 showed amide $^{1}H^{N}$ and ^{15}N chemical shift changes upon mutation or *O*-GlcNAcylation of this site, indicating that any conformational perturbations were local rather than global. Accordingly, the secondary structures for all three species of tau$^{353-408}$ were interrogated by three approaches. The first involved the algorithm δ2D, which predicts structure from main-chain chemical shifts. The second involved measuring the three-bond $^{3}J_{HN-HA}$, scalar coupling values, which depend upon backbone φ angles. The third involved NOESY experiments, indicative of through-space proton-proton interactions, however, accurate values could not be calculated for residues in the 400 region. Based on each of these methods, I concluded that there was no obvious change in secondary structure induced upon *O*-GlcNAcylation or the mutation of Ser400. This is in agreement with a study performed on full-length tau, where residual β-sheet, α-helical, and polyproline II helical structure was found throughout tau, but not for residues spanning 345 to 430[72]. Although structural changes were not observed, probing the dynamic properties between all three species of tau$^{353-408}$ resulted in differences.

5.1.4 ^{15}N relaxation experiments report on local dynamics

Amide ^{15}N relaxation experiments were recorded for all three species of tau$^{353-408}$. For residues near GlcNAc-Ser400 and Asp400, there was a very small increase in NOE values and a decrease in ^{15}N T_2 values, indicating a slight dampening in the nsec-psec timescale motions of the amide ^{15}N-^{1}H^{N} bond vectors. Even more notable, the GlcNAc amide had exceptionally high T_2 and low NOE values, together with the lack of any detected saccharide-polypeptide NOESY correlations, it can be concluded that, although covalently linked to Ser400, GlcNAc does not interact with tau$^{353-408}$ and is very mobile on the nsec-psec timescale. This is consistent with earlier findings, where glycosylation results in a short-range decrease in backbone flexibility for highly flexible peptides[80-82]. With structural and dynamic properties assessed, functional aggregation assays were then performed.

5.1.5 Optimizing the aggregation assays

The aggregation assays were performed at a physiological temperature of 37 ºC, however, this quickly led to uneven sample evaporation due to heating edge effects. To circumvent this problem, optically clear crystallography tape was used to seal the wells, such that each well maintained the same volume throughout the experiment. In the heparin-induced aggregation assays, *O*-GlcNAc-modified tau$^{353-408}$ reproducibly aggregated much less than wild-type tau$^{353-408}$. In contrast, S396D and S400D exhibited increased aggregation compared to wild-type tau$^{353-408}$, whereas S404D aggregated somewhat less. Thus despite the lack of any structural perturbations, these modifications had a very pronounced effect on the aggregation of tau$^{353-408}$.

5.2 Changes in *cis*/*trans* population

In the ^{1}H-^{15}N HSQC spectra of tau$^{353-408}$, more than the expected number of peaks were observed. The subset of weak extra signals has been assigned to residues adjacent to the *cis* isomers of the three X-Pro bonds present in tau$^{353-408}$. Although full-length tau has 48 prolines and has been completely assigned by NMR, there has been no mention of the *cis* isoform[73,83]. This is surprising and may simply reflect their relatively low populations and difficulty to observe in the very crowded spectra of the full-length species. However, in a study on the same *O*-GlcNAc site on tau, in a construct spanning residues 392-411 (and thus overlapping with tau$^{353-408}$) there was no mention of a *cis* population either. Regardless, there are many important implications of *cis*/*trans* isomerization.

Several studies have reported a change in X-Pro *cis*/*trans* isomer ratio due to PTMs. In the conserved loop region of nicotinic acetylcholine receptor, Rickert and Imperiali noted that *cis* percentage decreased from 45% to 30% due to *N*-linked glycosylaion[84]. However, in disordered regions, such as the octapeptide studied in the mucin domain of MAdCAM-1, the *cis* population was found to be a more typical 8-10%. It was noted that for this octapeptide, *O*-GlcNAc alone caused no change in *cis*/*trans* isomerization, whereas the addition of a steric branched fucose resulted in a *cis* percentage decrease by two-fold, from 8% to 4%[85]. Finally, the introduction of a negative charge, such as a phosphate group, or a negatively charged amino acid, has been suggested to favor the *cis* isomer[86,87]. For tau$^{353-408}$, a small decrease in the *cis* to *trans* ratio of Ser396-Pro397 was observed as a result of the *O*-GlcNAc modification, and no effects were observed for the S400D mutation, relative to wild-type.

The small increase in relative *trans* population due to *O*-GlcNAcylation may have an interesting implication. A decrease in PP2A phosphatase activity has been shown to increase

hyperphosphorylation of tau in mice[88]. The levels of phosphorylation for residues Ser202, Thr205, Thr231, and Ser235 were affected by phosphatase activity of PP2A[89]. Interestingly, proline-directed phosphatases, such as PP2A, can only dephosphorylate pSer/Thr-Pro residues in the *trans* conformer[90]. PP2A activity is thus stimulated by Pin1, a *cis*/*trans* prolyl isomerase[91]. Pin1 has been observed to accelerate the isomerization of amyloid precursor protein by over 1000-fold[92]. Pin1 knockouts in mice results in an accumulation of phosphate at residues 202, 205, 231, and 235[93]. The level of soluble Pin1 in AD brains is reduced by five-fold relative to age-matched normal brains[94]. Recently, the development of *cis*- and *trans*-specific antibodies have shown that it is *cis*, but not *trans*, phosphorylated tau (p-tau) that appears early on in the brains of humans exhibiting mild cognitive impairment since the *cis* isomer cannot promote microtubule assembly. Using these antibodies, Lu and coworkers discovered that Pin1 catalyzes the *cis*/*trans* interconversion of p-tau to prevent AD tau pathology[95]. *O*-GlcNAcylation at Ser400 may be playing the same role by increasing the amount of tau in a phosphatase susceptible *trans* conformation, thereby reducing p-tau aggregation.

5.3 Aggregation of tau

In solution, tau is intrinsically disordered, and when hyperphosphorylated, will form paired helical filaments that lead to neurofibrillary tangles. However, *O*-GlcNacylation has been seen to reciprocally affect PHF formation[35]. Glutamic and aspartic acid residues are commonly used to mimic the effects of phosphorylation since they possess a negative charge. It was shown that a single phosphomimic was sufficient to cause a decrease in MT-binding, to induce conformational changes in the first and second MT-binding repeats, and most importantly, to increase tau aggregation[96-98]. In concordance with this observation, the S396D

and S400D mutations increased the aggregation of tau$^{353-408}$ as measured by thioflavin S fluorescence. In contrast, the S404D construct modestly reduced aggregation. Most dramatically, the *O*-GlcNAc-modified tau$^{353-408}$ showed very little aggregation. This indicates that *O*-GlcNAcylation can inhibit oligomerization of tau either by destabilizing PHF formation, or by stabilizing the monomer in solution.

The effect of GlcNAcylation may also arise from its interplay with tau phosphorylation. Assuming that aspartic acid is a functional mimic of phosphoserine, I surmise that the phosphorylation of Ser396 and Ser400 are more crucial in inducing paired helical formation than the phosphorylation of Ser404. Given that the *O*-GlcNAcylation at Ser400 is reported to block the phosphorylation cascade, it could further reduce aggregation by indirectly reducing phosphorylation[53].

To better understand the reciprocal relationship in tau$^{353-408}$, it would be valuable to test the different combinations of phosphomimics. Would two or three phosphomimics increase the amount of aggregation? Could one *O*-GlcNAc counteract the effects of the phosphomimics? Finally, this study would be more accurate if a phosphate group (bearing almost two negative charges at neutral pH) was added, instead of mutating it to an aspartic acid residue (which only has one negative charge).

5.4 The global fold of tau

Although the site of modification being studied is distal to the MT-binding repeats, it has the ability to change the amount of aggregation. What is surprising is that just one new negative charge or one monosaccharide can cause such an effect. Considering the global "paperclip" fold of tau (Figure 1.5), where both C-terminal truncations and phosphomimics

result in more accessible MT-binding repeats, it is clear that the C-terminus is involved in paired helical formation[47,48]. Furthermore, both Glu391 and Asp 421 truncations have been observed to enhance the rates of tau filament formation *in vitro* and have been found in NFTs[99,100]. Conceivably, C-terminally truncated tau$^{353\text{-}408}$ results in an accessible MT-binding repeat, and the phosphomimics further amplifies this effect. Although the effects of modifications on the global folding of tau have not been studied, presumably it results in a less accessible MT-binding repeat.

Now considering the global structure of aggregated tau, recently, Müller *et al*. proposed a tau paired helical filament model based on immunogold labeling transmission electron micrography experiments. They demonstrated that N- and C-terminal tails are not central to filament formation, but form a "two-layered polyelectrolyte brush"[101]. This suggests that, around the aggregated MT-binding domains, there is a rigid layer of fibrils with positive charges surrounded by a sparse and soft brush layer with negative charges (Figure 5.1). I hypothesize that the introduction of aspartic acid residues, beyond creating a more accessible MT-binding repeat, also forms strong interactions with positive charges found in the "brush", resulting in a sturdier framework around the already-formed filaments. Furthermore, the introduction of a bulky GlcNAc moiety may lead to a less densely packed core.

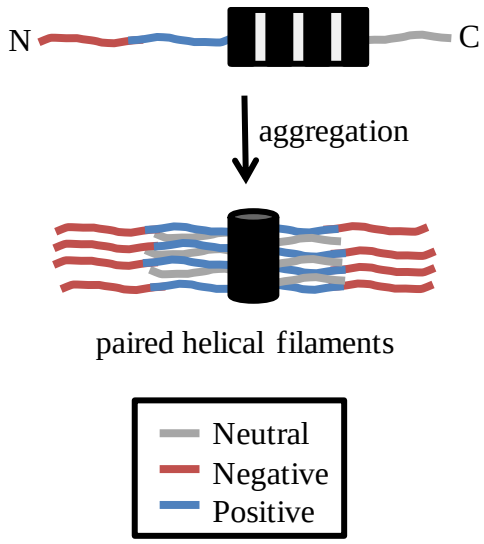

Figure 5.1: When tau fragments aggregate in to paired helical filaments, they form a two-layered polyelectrolyte brush.

5.5 Multiple GlcNAcylation sites

Although I was careful in avoiding multiple *O*-GlcNAc sites at the C-terminal end of the tau$^{353\text{-}408}$ construct, there was an *O*-GlcNAc modification site present at the N-terminal end of the construct. That is, Ser356 has been observed to be *O*-GlcNAcylated. However, this residue is part of the fourth MT-binding repeat, and considered to be required for aggregation studies. From NMR experiments, *O*-GlcNAcylation was never observed at Ser356, and only observed at Ser400. As such, I speculate that the TPR domain of OGT recognizes peptides sequences N-terminal to the site of modification.

5.6 GlcNAcylation in various systems

O-GlcNAc has been shown to induce turn-like structures in small peptide models of the C-terminal domain of RNA polymerase II and the N-terminus of murine estrogen receptor β[24]. This could result from hydrogen bonding between the N-acetyl moiety and the polypeptide backbone. The evidence for turn formation was based primarily on NOE

connectivities, and in the case of RNA polymerase II, further corroborated with $^{1}H/^{2}H$ exchange experiments[78]. Both peptides exhibited NOE patterns diagnostic for turns upon GlcNAcylation. In particular, Li and coworkers concluded that in a small peptide from murine estrogen receptor β, *O*-GlcNAc promotes turn formation based on a newly observed NOESY peak between the modified residue and its *i+1* neighbor. Conversely, they suggest that phosphorylation at this location creates a more extended conformation largely based on the absence of a NOESY signal otherwise seen in the unmodified peptide. In the case of Simanek *et al*. using $^{1}H/^{2}H$ exchange experiments, they found that amide protons belonging to the turn-like structure exchanged more than 20× slower than the remaining amide protons. In contrast to these examples, no clear evidence was found for a GlcNAc-induced turn in tau$^{353-408}$. In HNH-NOESY-HSQC spectra of tau$^{353-408}$ (Figure 3.5c), the only NOE correlations observed to the GlcNAc amide were intra-sugar connectivities. Even though *O*-GlcNAcylation at Ser400 resulted in minor local chemical shift perturbations, there were no indications of any local structural changes (Figure 4.1).

Carbohydrates can modulate proteins in many ways. *N*-linked glycans usually have a common core structure, but differ terminally, with complex, and sometimes, branched, structures[102]. Proteins undergo glycosylation in the lumen of the endoplasmic reticulum, as a cotranslational event. As such, the addition of large polar sugars is considered to play a role in their correct folding, as well as increasing their solubility. Furthermore, due to their polar nature, sugars are usually found on the exterior of a protein[102]. For example, *O*-linked glycans are often located in hinge or linker regions between folded globular domains[103]. Glycans can shield the protein surface to prevent aggregation, and increase thermal stability, [104,105]. However, tau is intrinsically disordered and extremely soluble, yet it still possesses

several glycosylation sites[72]. Thus, the link between increased *O*-GlcNAcylation and decreased tau aggregation must lie in the monosaccharide increasing the solubility of tau$^{353-408}$.

5.7 Summary

The goals of this thesis were to compare the local structure and dynamic effects of *O*-GlcNAcylation and phosphomimic on tau, and their functional effects on aggregation. As far as we know, this is the longest peptide to be studied with an *O*-GlcNAc modification. Using NMR experiments, I determined that neither *O*-GlcNAcylation nor aspartate mutations (considered to be a phosphomimic) induced a change in secondary structure. Although there were minimal changes in the local dynamics around position 400, no specific conformational perturbations were detected. Since there were no structural changes, the modifications effects on tau's aggregation must arise for other reasons. In performing functional aggregation assays, I confirmed that indeed aggregation propensity decreased for *O*-GlcNAc-modified tau$^{353-408}$ and increased for phosphomimics.

5.8 Future directions

The effects of *O*-GlcNAcylation and serine to aspartic acid mutations on tau were studied in this thesis. Although tau was an interesting model to investigate due to its link with Alzheimer's disease, there was only ~10% modification and the process of purifying the resulting *O*-GlcNAc modified peptide to yield milligram quantities was difficult. Studying a peptide for which OGT has a higher activity for modifying *in vivo* could result in more amenable samples. Furthermore, studying other peptides could provide more insights into the common and distinct structural and dynamic changes due to *O*-GlcNAcylation.

Selective *O*-GlcNAc labeling could circumvent the issue of low modification, and would help in further investigating the dynamics of the linked *O*-GlcNAc moiety. In the case of the 56-residue tau$^{353\text{-}408}$, the NMR signals from the sugar amide were easily identifiable. However, in order to monitor the dynamics of *O*-GlcNAc in larger proteins, selective labeling might be required. In principle, this can be achieved by using *in vivo* modification by providing exogenous labeled GlcNAc. However, to achieve high levels of GlcNAc labeling without also labeling the protein an *E. coli* strain that is auxotrophic for UDP-GlcNAc would be required.

UDP-GlcNAc is the end product of the hexosamine biosynthetic pathway, and thus it is important to understand this pathway in order to create the desired auxotrophic strain (Figure 5.2)[106]. In *E. coli* GlcNAc and GlcN are taken up into the cell by specific permesases of the phosphotransferase system. This system converts GlcNAc into GlcNAc-6-P. With a removal of the acetate group, it becomes GlcN-6-P, which is isomerized to GlcN-1-P. Both of these become re-acetylated and are then converted to UDP-GlcNAc. However, GlcN-6-P may also be synthesized from fructose-6-P (made during glycolysis) and glutamine by GlmS. In order to maximize the use of labeled GlcNAc, a GlmS knock-out strain must be created so that unlabeled GlcNAc (from glycolysis) will not be incorporated into the protein. Rather, all labeled GlcNAc or GlcN will be derived from exogenous sources. Although *E. coli* strains with the GlmS mutation have been isolated and characterized, they are not suitable with today's expression systems[107]. Thus, recombineering technology can be used to generate a GlmS knock-out in *E. coli* BL21(λDE3).

Finally, ^{15}N-labeled GlcNAc may be synthesized chemo-enzymatically. Fungi grown on ^{15}N-ammonia will result in ~60% labeled glucosamine in their cell wall. Subsequent

depolymerization and acetylation of their cell wall will yield ^{15}N-labeled GlcNAc. This in turn can be provided to the GlmS auxotrophic strain in order to generate GlcNAcylated proteins with only the monosaccharide labeled.

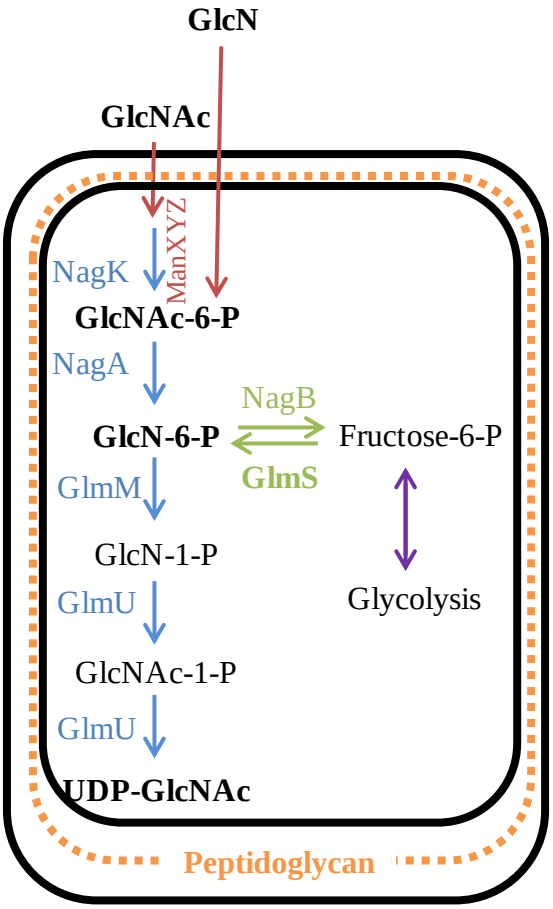

Figure 5.2: The hexosamine biosynthetic pathway. In order to create an auxotrophic strain for UDP-GlcNAc, it will be necessary to create a GlmS knockout.

Works cited

1. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).

2. Jensen, O.N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* **8**, 33-41 (2004).

3. 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).

4. Greis, K.D. & Hart, G.W. Analytical methods for the study of O-GlcNAc glycoproteins and glycopeptides. *Methods Mol Biol* **76**, 19-33 (1998).

5. Wang, Z. et al. Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **9**, 153-60 (2010).

6. 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).

7. Hart, G.W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem* **80**, 825-58 (2011).

8. Butkinaree, C., Park, K. & Hart, G.W. O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta* **1800**, 96-106 (2010).

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

10. Dias, W.B. & Hart, G.W. O-GlcNAc modification in diabetes and Alzheimer's disease. *Mol Biosyst* **3**, 766-72 (2007).

11. Shafi, R. et al. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci U S A* **97**, 5735-9 (2000).

12. Park, B.H., Karpinets, T.V., Syed, M.H., Leuze, M.R. & Uberbacher, E.C. CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20**, 1574-84 (2010).

13. Vocadlo, D.J. O-GlcNAc processing enzymes: catalytic mechanisms, substrate specificity, and enzyme regulation. *Curr Opin Chem Biol* **16**, 488-97 (2012).

14. Liu, Y. et al. Developmental regulation of protein O-GlcNAcylation, O-GlcNAc transferase, and O-GlcNAcase in mammalian brain. *PLoS One* **7**, e43724 (2012).

15. Jinek, M. et al. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* **11**, 1001-7 (2004).

16. Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-7 (2011).

17. Gorlich, D., Prehn, S., Laskey, R.A. & Hartmann, E. Isolation of a protein that is essential for the first step of nuclear protein import. *Cell* **79**, 767-78 (1994).

18. Butkinaree, C. et al. Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis. *J Biol Chem* **283**, 23557-66 (2008).

19. 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).

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

21. Dennis, R.J. et al. Structure and mechanism of a bacterial beta-glucosaminidase having O-GlcNAcase activity. *Nat Struct Mol Biol* **13**, 365-71 (2006).

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

23. 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).

24. Chen, Y.X. et al. Alternative O-GlcNAcylation/O-phosphorylation of Ser16 induce different conformational disturbances to the N terminus of murine estrogen receptor beta. *Chem Biol* **13**, 937-44 (2006).

25. 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).

26. 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).

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

28. Dias, W.B., Cheung, W.D., Wang, Z. & Hart, G.W. Regulation of calcium/calmodulin-dependent kinase IV by O-GlcNAc modification. *J Biol Chem* **284**, 21327-37 (2009).

29. Housley, M.P. et al. O-GlcNAc regulates FoxO activation in response to glucose. *J Biol Chem* **283**, 16283-92 (2008).

30. Wang, Z., Pandey, A. & Hart, G.W. Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics* **6**, 1365-79 (2007).

31. Wang, Z., Gucek, M. & Hart, G.W. Cross-talk between GlcNAcylation and phosphorylation: site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proc Natl Acad Sci U S A* **105**, 13793-8 (2008).

32. Whelan, S.A., Lane, M.D. & Hart, G.W. Regulation of the O-linked beta-N-acetylglucosamine transferase by insulin signaling. *J Biol Chem* **283**, 21411-7 (2008).

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

34. Liu, F. et al. Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease. *Brain* **132**, 1820-32 (2009).

35. Arnold, C.S. et al. The microtubule-associated protein tau is extensively modified with O-linked N-acetylglucosamine. *J Biol Chem* **271**, 28741-4 (1996).

36. 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).

37. Binder, L.I., Frankfurter, A. & Rebhun, L.I. The distribution of tau in the mammalian central nervous system. *The Journal of Cell Biology* **101**, 1371-1378 (1985).

38. Cleveland, D.W., Hwo, S.Y. & Kirschner, M.W. Purification of tau, a microtubule-associated protein that induces assembly of microtubules from purified tubulin. *J Mol Biol* **116**, 207-25 (1977).

39. Andreadis, A., Broderick, J.A. & Kosik, K.S. Relative exon affinities and suboptimal splice site signals lead to non-equivalence of two cassette exons. *Nucleic Acids Research* **23**, 3585-3593 (1995).

40. Goode, B.L. et al. Functional interactions between the proline-rich and repeat regions of tau enhance microtubule binding and assembly. *Mol Biol Cell* **8**, 353-65 (1997).

41. Ittner, A. et al. Brief update on different roles of tau in neurodegeneration. *IUBMB Life* **63**, 495-502 (2011).

42. Chen, J., Kanai, Y., Cowan, N.J. & Hirokawa, N. Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons. *Nature* **360**, 674-7 (1992).

43. Daebel, V. et al. β-Sheet Core of Tau Paired Helical Filaments Revealed by Solid-State NMR. *Journal of the American Chemical Society* **134**, 13982-13989 (2012).

44. Wischik, C.M. et al. Structural characterization of the core of the paired helical filament of Alzheimer disease. *Proc Natl Acad Sci U S A* **85**, 4884-8 (1988).

45. Friedhoff, P., Schneider, A., Mandelkow, E.M. & Mandelkow, E. Rapid assembly of Alzheimer-like paired helical filaments from microtubule-associated protein tau monitored by fluorescence in solution. *Biochemistry* **37**, 10223-30 (1998).

46. Rissman, R.A. et al. Caspase-cleavage of tau is an early event in Alzheimer disease tangle pathology. *J Clin Invest* **114**, 121-30 (2004).

47. Jeganathan, S., von Bergen, M., Brutlach, H., Steinhoff, H.J. & Mandelkow, E. Global hairpin folding of tau in solution. *Biochemistry* **45**, 2283-93 (2006).

48. Jeganathan, S., Chinnathambi, S., Mandelkow, E.M. & Mandelkow, E. Conformations of microtubule-associated protein Tau mapped by fluorescence resonance energy transfer. *Methods Mol Biol* **849**, 85-99 (2012).

49. 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).

50. Ledesma, M.D., Bonay, P., Colaço, C. & Avila, J. Analysis of microtubule-associated protein tau glycation in paired helical filaments. *Journal of Biological Chemistry* **269**, 21614-21619 (1994).

51. Yuzwa, S.A. et al. Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol* **8**, 393-9 (2012).

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

53. Smet-Nocca, C. et al. Identification of O-GlcNAc sites within peptides of the Tau protein and their impact on phosphorylation. *Mol Biosyst* **7**, 1420-9 (2011).

54. Yuzwa, S.A. et al. Mapping O-GlcNAc modification sites on tau and generation of a site-specific O-GlcNAc tau antibody. *Amino Acids* **40**, 857-68 (2011).

55. van den Ent, F. & Lowe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods* **67**, 67-74 (2006).

56. Marblestone, J.G. et al. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci* **15**, 182-9 (2006).

57. Seal, R.L., Gordon, S.M., Lush, M.J., Wright, M.W. & Bruford, E.A. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res* **39**, D514-9 (2011).

58. Weeks, S.D., Drinker, M. & Loll, P.J. Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr Purif* **53**, 40-50 (2007).

59. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).

60. Goddard, T.D., Kneller, D.G. (ed.) *SPARKY 3*, (University of California, San Francisco).

61. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93-158 (1999).

62. Kuboniwa, H., Grzesiek, S., Delaglio, F. & Bax, A. Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J Biomol NMR* **4**, 871-8 (1994).

63. Kay, L.E., Torchia, D.A. & Bax, A. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-8979 (1989).

64. Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (ed.) *Protein Identification and Analysis Tools on the ExPASy Server*, 571-607 (Humana Press, 2005).

65. Smith, P.K. et al. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150**, 76-85 (1985).

66. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99**, 201-207 (1992).

67. Schmid, F.X., Mayr, L.M., Mucke, M. & Schonbrunner, E.R. Prolyl isomerases: role in protein folding. *Adv Protein Chem* **44**, 25-66 (1993).

68. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J Biomol NMR* **24**, 149-54 (2002).

69. Wishart, D.S. et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **41**, D801-7 (2013).

70. Wuthrich, K. (ed.) *NMR of Proteins and Nucleic Acids*, (Wiley, New York, 1986).

71. Camilloni, C., De Simone, A., Vranken, W.F. & Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **51**, 2224-31 (2012).

72. Mukrasch, M.D. et al. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol* **7**, e34 (2009).

73. Narayanan, R.L. et al. Automatic assignment of the intrinsically disordered protein Tau with 441-residues. *J Am Chem Soc* **132**, 11906-7 (2010).

74. Peterson, D.W., Zhou, H., Dahlquist, F.W. & Lew, J. A soluble oligomer of tau associated with fiber formation analyzed by NMR. *Biochemistry* **47**, 7393-404 (2008).

75. Tamiola, K., Acar, B. & Mulder, F.A.A. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **132**, 18000-18003 (2010).

76. Maltsev, A.S., Ying, J. & Bax, A. Impact of N-terminal acetylation of alpha-synuclein on its random coil and lipid binding properties. *Biochemistry* **51**, 5004-13 (2012).

77. Friedhoff, P., Schneider, A., Mandelkow, E.-M. & Mandelkow, E. Rapid Assembly of Alzheimer-like Paired Helical Filaments from Microtubule-Associated Protein Tau Monitored by Fluorescence in Solution†. *Biochemistry* **37**, 10223-10230 (1998).

78. Simanek, E.E. et al. Glycosylation of Threonine of the Repeating Unit of RNA Polymerase II with β-Linked N-Acetylglucosame Leads to a Turnlike Structure. *Journal of the American Chemical Society* **120**, 11567-11575 (1998).

79. Ulrich, E.L. et al. BioMagResBank. *Nucleic Acids Research* **36**, D402-D408 (2008).

80. Wormald, M.R. & Dwek, R.A. Glycoproteins: glycan presentation and protein-fold stability. *Structure* **7**, R155-60 (1999).

81. Wormald, M.R. et al. Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling. *Chem Rev* **102**, 371-86 (2002).

82. Imperiali, B. & Rickert, K.W. Conformational implications of asparagine-linked glycosylation. *Proc Natl Acad Sci U S A* **92**, 97-101 (1995).

83. Harbison, N.W., Bhattacharya, S. & Eliezer, D. Assigning Backbone NMR Resonances for Full Length Tau Isoforms: Efficient Compromise between Manual Assignments and Reduced Dimensionality. *PLoS One* **7**, e34679 (2012).

84. Rickert, K.W. & Imperiali, B. Analysis of the conserved glycosylation site in the nicotinic acetylcholine receptor: potential roles in complex assembly. *Chem Biol* **2**, 751-9 (1995).

85. Wu, W.-g. et al. Structural Study on O-Glycopeptides: Glycosylation-Induced Conformational Changes of O-GlcNAc, O-LacNAc, O-Sialyl-LacNAc, and O-Sialyl-Lewis-X Peptides of the Mucin Domain of MAdCAM-1. *Journal of the American Chemical Society* **121**, 2409-2417 (1999).

86. Grathwohl, C. & Wüthrich, K. Nmr studies of the rates of proline cis–trans isomerization in oligopeptides. *Biopolymers* **20**, 2623-2633 (1981).

87. Evans, C.A. & Rabenstein, D.L. Nuclear magnetic resonance studies of the acid-base chemistry of amino acids and peptides. II. Dependence of the acidity of the C-terminal carboxyl group on the conformation of the C-terminal peptide bond. *Journal of the American Chemical Society* **96**, 7312-7317 (1974).

88. Kins, S. et al. Reduced protein phosphatase 2A activity induces hyperphosphorylation and altered compartmentalization of tau in transgenic mice. *J Biol Chem* **276**, 38193-200 (2001).

89. Landrieu, I. et al. Molecular Implication of PP2A and Pin1 in the Alzheimer's Disease Specific Hyperphosphorylation of Tau. *PLoS One* **6**, e21521 (2011).

90. Lu, K.P. & Zhou, X.Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol* **8**, 904-16 (2007).

91. Zhou, X.Z. et al. Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. *Mol Cell* **6**, 873-83 (2000).

92. Pastorino, L. et al. The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. *Nature* **440**, 528-34 (2006).

93. Liou, Y.C. et al. Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration. *Nature* **424**, 556-61 (2003).

94. Lu, P.J., Wulf, G., Zhou, X.Z., Davies, P. & Lu, K.P. The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature* **399**, 784-8 (1999).

95. Nakamura, K. et al. Proline isomer-specific antibodies reveal the early pathogenic tau conformation in Alzheimer's disease. *Cell* **149**, 232-44 (2012).

96. Kiris, E. et al. Combinatorial Tau Pseudophosphorylation: MARKEDLY DIFFERENT REGULATORY EFFECTS ON MICROTUBULE ASSEMBLY AND DYNAMIC INSTABILITY THAN THE SUM OF THE INDIVIDUAL PARTS. *Journal of Biological Chemistry* **286**, 14257-14270 (2011).

97. Fischer, D. et al. Conformational changes specific for pseudophosphorylation at serine 262 selectively impair binding of tau to microtubules. *Biochemistry* **48**, 10047-55 (2009).

98. Chang, E., Kim, S., Schafer, K.N. & Kuret, J. Pseudophosphorylation of tau protein directly modulates its aggregation kinetics. *Biochim Biophys Acta* **1814**, 388-95 (2011).

99. Novak, M., Kabat, J. & Wischik, C.M. Molecular characterization of the minimal protease resistant tau unit of the Alzheimer's disease paired helical filament. *EMBO J* **12**, 365-70 (1993).

100. Abraha, A. et al. C-terminal inhibition of tau assembly in vitro and in Alzheimer's disease. *J Cell Sci* **113 Pt 21**, 3737-45 (2000).

101. Wegmann, S., Medalsy, I.D., Mandelkow, E. & Muller, D.J. The fuzzy coat of pathological human Tau fibrils is a two-layered polyelectrolyte brush. *Proc Natl Acad Sci U S A* **110**, E313-21 (2013).

102. Taylor, M.E., and Drickamer, K. (ed.) *Introduction to Glycobiology*, (Oxford University Press, Oxford, 2003).

103. Petrescu, A.J., Milac, A.L., Petrescu, S.M., Dwek, R.A. & Wormald, M.R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103-14 (2004).

104. Rudd, P.M. & Dwek, R.A. Glycosylation: Heterogeneity and the 3D Structure of Proteins. *Critical Reviews in Biochemistry and Molecular Biology* **32**, 1-100 (1997).

105. Wang, C., Eufemi, M., Turano, C. & Giartosio, A. Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35**, 7299-307 (1996).

106. Park, J.T. & Uehara, T. How bacteria consume their own exoskeletons (turnover and recycling of cell wall peptidoglycan). *Microbiol Mol Biol Rev* **72**, 211-27, table of contents (2008).

107. Sarvas, M. Mutant of Escherichia coli K-12 defective in D-glucosamine biosynthesis. *J Bacteriol* **105**, 467-71 (1971).

Appendices

Appendix A: Chemical shift tables for tau$^{353\text{-}408}$

Table A.1: Chemical shifts for $^{15}N/^{13}C$ wild-type tau$^{353\text{-}408}$

Residue	$^{13}C'$ (ppm)	$^{13}C^{\alpha}$ (ppm)	$^{13}C^{\beta}$ (ppm)	$^{1}H^{N}$ (ppm)	^{15}N (ppm)
K353	-	-	-	-	-
I354	176.4	61.48	38.37	-	-
G355	173.9	44.98	-	8.624	114.2
S356	174.7	58.21	63.76	8.151	115.6
L357	177.1	54.91	41.91	8.419	124.1
D358	175.8	54.34	40.95	8.164	120.5
N359	175	53.18	38.52	8.274	118.7
I360	176.3	61.14	37.95	7.99	120.6
T361	174.2	61.92	69.57	8.177	118.2
H362	173.8	55.64	29.65	8.454	121.4
V363	174.4	59.18	33.97	7.994	121.8
P364	177.5	63.42	31.9	-	-
G365	174.8	45.21	-	8.587	110.4
G366	173.8	45.17	-	8.324	108.7
G367	174.2	45.08	-	8.324	108.7
N368	175.1	53.06	38.66	8.309	118.5
K369	176.3	56.23	32.83	8.255	121.9
K370	176.2	56.19	32.75	8.313	123.4
I371	176	60.82	38.41	8.2	123.5
E372	176.3	56.03	30.34	8.513	126.3
T373	174.1	61.94	69.55	8.222	116.1
H374	174.2	55.41	29.67	8.39	122.3
K375	175.6	57.39	32.73	8.173	123.8
L376	177.3	55.08	42.14	8.319	123.9
T377	174	61.54	69.86	8.056	115
F378	175.4	57.9	39.5	8.265	122.8
R379	175.8	55.86	30.76	8.142	123.2
E380	176.2	56.6	30.12	8.329	122.2
N381	175	53.04	38.61	8.458	120
A382	177.7	53.07	18.98	8.21	124.5
K383	176.3	56.12	32.75	8.149	120.2
A384	177.7	52.37	19.12	8.158	125

Table A.1: Chemical shifts for $^{15}N/^{13}C$ wild-type tau$^{353-408}$

Residue	$^{13}C'$ (ppm)	$^{13}C^{\alpha}$ (ppm)	$^{13}C^{\beta}$ (ppm)	$^{1}H^{N}$ (ppm)	^{15}N (ppm)
K385	176.6	56.29	32.85	8.292	120.9
T386	174	61.55	69.74	8.093	114.6
D387	176.1	54.08	41.05	8.275	122.7
H388	175.2	55.75	29.01	8.429	119.5
G389	173.7	45.24	-	8.451	109.9
A390	177.5	52.23	19.18	8.123	123.7
E391	176.2	56.34	29.97	8.345	120.3
I392	175.7	60.96	38.28	8.139	123.3
V393	175.5	61.81	32.67	8.1	125.5
Y394c	175.2	57.6	38.81	8.308	125.5
Y394t	175.1	57.98	38.81	8.377	126
K395t	175.4	55.3	33.36	8.05	125
S396c	172.8	-	64.09	8.032	117
S396t	172.7	56.42	63.02	8.282	119.4
P397c	176.2	62.85	32.12	-	-
P397t	176.5	63.03	31.99	-	-
V398c	175.8	61.7	32.79	8.242	124.6
V398t	176.2	62.21	32.46	8.157	121.1
V399	175.9	61.87	32.64	8.281	125.2
S400	174.8	58.23	63.72	8.414	120.3
G401	173.7	45.12	-	8.428	111.4
D402t	176.7	54.23	41.01	8.208	120.5
T403c	173.6	61.48	69.96	8.046	114.1
T403t	174.6	61.52	69.4	8.174	114.4
S404c	172.7	55.66	63.87	8.136	117.8
S404t	172.4	56.86	62.93	8.257	120
P405c	176	62.69	34.22	-	-
P405t	176.7	63.14	31.91	-	-
R406c	175.9	56.22	30.61	8.576	122
R406t	175.7	55.82	30.69	8.342	121.4
H407	173.5	55.08	29.43	8.507	120.9
L408	175.5	56.66	42.83	8.143	129.7

Table A.2: Chemical shifts for $^{15}N/^{13}C$ GlcNAc-modified tau$^{353-408}$

Residue	$^{13}C'$ (ppm)	$^{13}C^{\alpha}$ (ppm)	$^{13}C^{\beta}$ (ppm)	$^{1}H^{N}$ (ppm)	^{15}N (ppm)
GlcNAc	177.1	-	-	8.143	122.2
K353	-	-	-	-	-
I354	176.4	61.47	38.37	-	-
G355	173.9	44.99	-	8.618	114.2
S356	174.7	58.2	63.79	8.146	115.6
L357	177.1	55.06	41.96	8.413	124.1
D358	175.8	54.35	40.96	8.164	120.5
N359	175	53.32	38.53	8.273	118.7
I360	176.3	61.12	38.26	7.981	120.6
T361	174.3	61.88	69.59	8.167	118.1
H362	173.3	55.41	29.41	8.488	121.3
V363	-	-	33.97	7.975	121.5
P364	177.5	63.39	31.88	-	-
G365	174.7	45.21	-	8.571	110.4
G366	174.8	45.19	-	8.317	108.7
G367	173.8	45.07	-	8.316	108.7
N368	175.1	53.13	38.72	8.305	118.5
K369	176.3	56.17	32.85	8.251	121.9
K370	176.2	56.19	32.72	8.308	123.3
I371	176	60.77	38.42	8.193	123.5
E372	176.3	56.01	30.37	8.505	126.3
T373	174	61.92	69.54	8.217	116.2
H374	174.1	55.19	29.39	8.413	122
K375	176.2	56.21	32.83	8.187	123.6
L376	177.2	55.04	42.17	8.335	124
T377	174	61.48	69.86	8.06	115
F378	175.4	57.83	39.51	8.262	122.8
R379	175.8	55.87	30.76	8.142	123.1
E380	176.2	56.59	30.12	8.324	122.2
N381	175	53.28	38.64	8.452	120
A382	177.7	52.75	18.96	8.208	124.5
K383	176.3	56.14	32.79	8.145	120.2
A384	177.7	52.37	19.11	8.151	125
K385	176.8	56.33	32.89	8.287	120.9
T386	174	61.47	69.75	8.082	114.5
D387	176	54.05	41.06	8.27	122.7
H388	175.1	55.59	28.85	8.448	119.3

Table A.2: Chemical shifts for $^{15}N/^{13}C$ GlcNAc-modified tau$^{353-408}$

Residue	$^{13}C'$ (ppm)	$^{13}C^{\alpha}$ (ppm)	$^{13}C^{\beta}$ (ppm)	$^{1}H^{N}$ (ppm)	^{15}N (ppm)
G389	173.7	45.26	-	8.453	109.9
A390	177.5	52.22	19.18	8.13	123.7
E391	176.2	56.35	29.98	8.34	120.4
I392	175.7	60.9	38.29	8.137	123.2
V393	175.5	61.79	32.65	8.098	125.5
Y394c	-	-	38.89	8.304	125.5
Y394t	175.1	57.89	38.84	8.372	126
K395c	174.9	55.73	33.39	-	-
K395t	175.4	55.26	33.38	8.049	125
S396c	-	-	64.13	8.029	117
S396t	-	56.36	63	8.278	119.4
P397c	176	62.29	32.46	-	-
P397t	176.5	62.97	31.94	-	-
V398c	-	-	-	8.245	124.6
V398t	176.2	62.29	32.53	8.15	121.2
V399	175.9	61.85	32.68	8.275	125.2
V399g	175.8	61.72	32.93	-	-
S400	174.8	58.22	63.76	8.406	120.2
S400g	174.2	56.25	70.39	8.352	119.5
G401	173.7	45.13	-	8.42	111.3
G401g	-	45.16	-	8.359	111.7
D402t	176.7	56.05	46.7	8.204	120.5
D402tg	176.7	54.24	41.12	8.184	120.5
T403c	173.6	61.55	70.02	-	-
T403t	174.6	61.43	69.41	8.171	114.4
T403tg	-	61.39	69.36	8.198	114.5
S404c	-	-	63.97	8.135	117.8
S404t	-	56.88	62.88	8.267	120
P405c	176	62.69	34.24	-	-
P405t	176.7	63.09	31.89	-	-
R406c	175.9	56.24	30.65	8.57	122
R406t	175.7	55.77	30.71	8.336	121.4
H407c	-	55.44	29.32	8.516	120.8
H407t	173.4	54.96	29.32	8.376	120.4
L408	-	56.63	42.82	8.164	129.8

Table A.3: Chemical shifts for $^{15}N/^{13}C$ S400D tau$^{353-408}$

Residue	$^{13}C'$ (ppm)	$^{13}C^{\alpha}$ (ppm)	$^{13}C^{\beta}$ (ppm)	$^{1}H^{N}$ (ppm)	^{15}N (ppm)
K353	-	-	-	-	-
I354	176.4	61.46	38.37	-	-
G355	173.9	44.99	-	8.616	114.2
S356	174.7	58.21	63.8	8.146	115.6
L357	177.1	55.06	41.94	8.411	124.1
D358	175.8	54.35	40.97	8.16	120.5
N359	175	53.3	38.54	8.267	118.7
I360	176.3	61.12	38.47	7.985	120.6
T361	174.3	61.9	69.58	8.17	118.2
H362	173.4	55.58	29.63	8.453	121.4
V363c	-	59.05	34	7.958	121.4
P364c	176.9	58.85	32.66	-	-
P364t	177.5	63.38	31.89	-	-
G365c	-	-	-	8.579	110.3
G365t	174.7	45.22	-	8.583	110.4
G366	174.7	45.2	-	8.318	108.7
G367	173.8	45.08	-	8.318	108.7
N368	175.1	53.07	38.72	8.304	118.5
K369	176.2	56.18	32.86	8.247	121.9
K370	176.2	56.19	32.79	8.305	123.4
I371	176	60.78	38.41	8.192	123.5
E372	176.3	56.03	30.36	8.506	126.3
T373	174	61.92	69.54	8.215	116.1
H374	174.2	55.32	29.63	8.385	122.3
K375	176.2	57.11	32.83	8.168	123.7
L376	177.2	55.04	42.17	8.312	123.8
T377	174	61.54	69.93	8.051	115
F378	175.4	57.82	39.52	8.252	122.8
R379	175.8	55.89	30.75	8.14	123.1
E380	176.2	56.6	30.14	8.325	122.2
N381	175	53.16	38.64	8.448	120
A382	177.7	52.75	18.97	8.204	124.5
K383	176.3	56.15	32.81	8.142	120.2
A384	177.7	52.36	19.12	8.149	125
K385	176.7	56.31	32.9	8.283	120.9
T386	174	61.44	69.76	8.084	114.5
D387	176	54.05	41.07	8.271	122.7

Table A.3: Chemical shifts for $^{15}N/^{13}C$ S400D tau$^{353-408}$

Residue	$^{13}C'$ (ppm)	$^{13}C^{\alpha}$ (ppm)	$^{13}C^{\beta}$ (ppm)	$^{1}H^{N}$ (ppm)	^{15}N (ppm)
H388	175.2	55.71	28.98	8.424	119.4
G389	173.7	45.27	-	8.444	109.9
A390	177.5	52.22	19.2	8.118	123.7
E391	176.2	56.38	29.98	8.341	120.3
I392	175.7	60.9	38.3	8.13	123.1
V393	175.4	61.78	32.72	8.093	125.4
Y394c	-	57.51	38.89	8.3	125.5
Y394t	175.1	57.91	38.86	8.366	125.9
K395c	174.8	55.69	33.42	-	-
K395t	175.3	55.28	33.41	8.045	125
S396c	-	-	64.15	8.015	117
S396t	-	56.36	63.02	8.292	119.5
P397c	176.2	62.93	32.11	-	-
P397t	176.5	63.05	31.92	-	-
V398c	-	61.49	32.97	8.197	124.1
V398t	176.1	62.28	32.48	8.134	121.1
V399	175.6	61.8	32.76	8.253	124.9
D400	176.5	54.51	41.07	8.38	124.8
G401	173.9	45.27	-	8.301	109.6
D402	176.5	54.21	41.09	8.209	120.6
T403c	173.6	61.59	69.95	7.989	114
T403t	174.6	61.5	69.36	8.128	114.2
S404c	-	55.62	63.97	8.128	117.9
S404t	-	56.88	62.88	8.225	119.9
P405c	176	62.64	34.24	-	-
P405t	176.7	63.11	31.87	-	-
R406c	175.9	56.2	30.64	8.565	122
R406t	175.8	55.71	30.72	8.339	121.4
H407c	-	55.05	29.35	8.525	120.8
H407t	173.5	54.99	29.38	8.359	120.3
L408	-	56.61	42.87	8.138	129.7

Appendix B: Chemical shift table for *O*-GlcNAc

Table B.1: Chemical shifts for $^{15}N/^{13}C$ *O*-GlcNAc

Atom	^{13}C (ppm)	^{1}H (ppm)
C_1-H_1	103.269	4.418
C_2-H_2	57.93	3.587
C_3-H_3	76.424	3.421
C_4-H_4	72.432	3.329
C_5-H_5	78.48	3.344
C_6-$H_{6'}$	63.33	3.642
C_6-$H_{6''}$	63.31	3.82
CH_3	24.831	1.907

Appendix C: Change in chemical shifts compared to predicted values

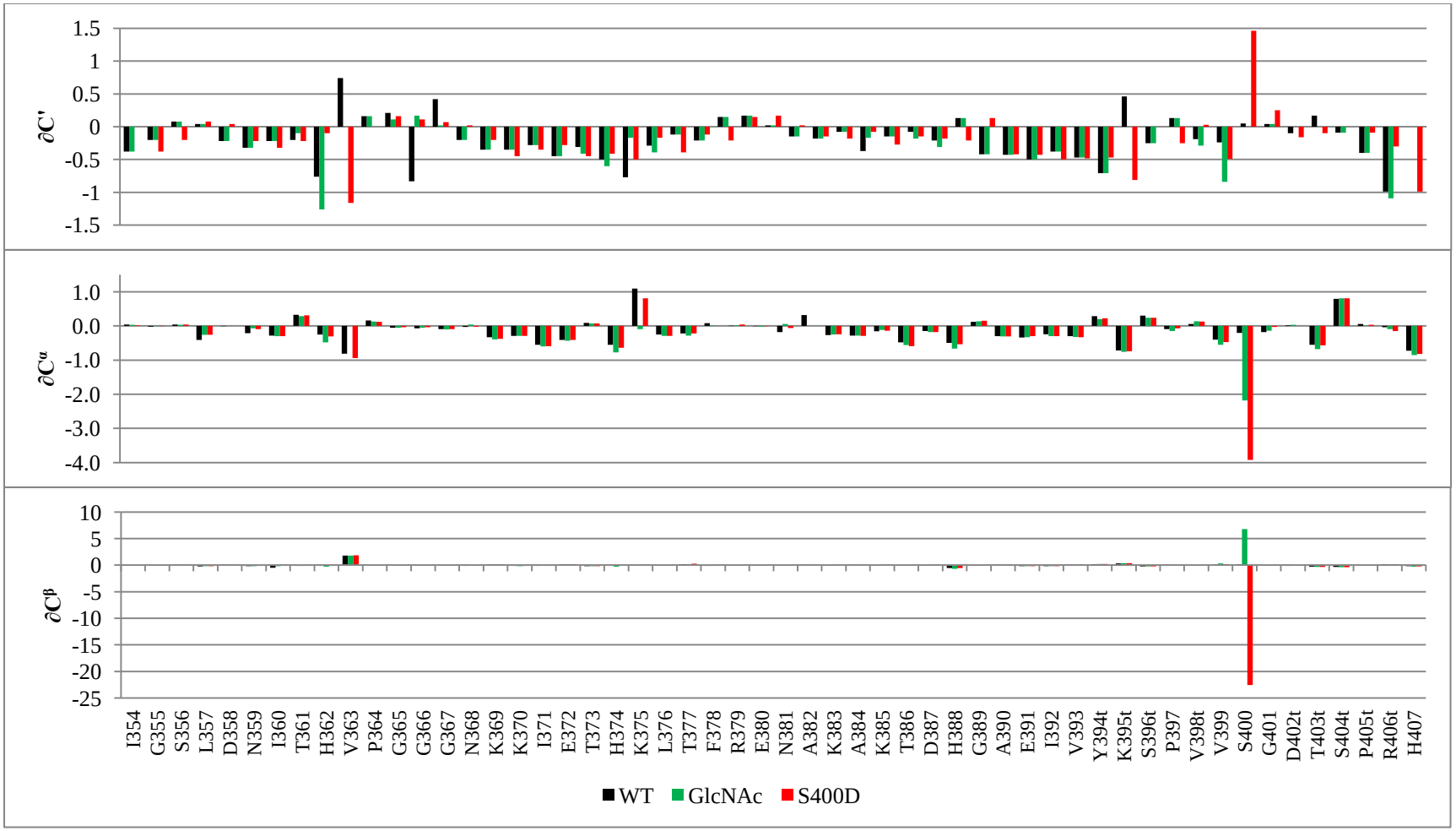

Figure C.1: Change in $^{13}C'$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ chemical shift values for wild-type, *O*-GlcNAc-modified, and S400D tau$^{353-408}$ compared to predicted chemical shift values [75].

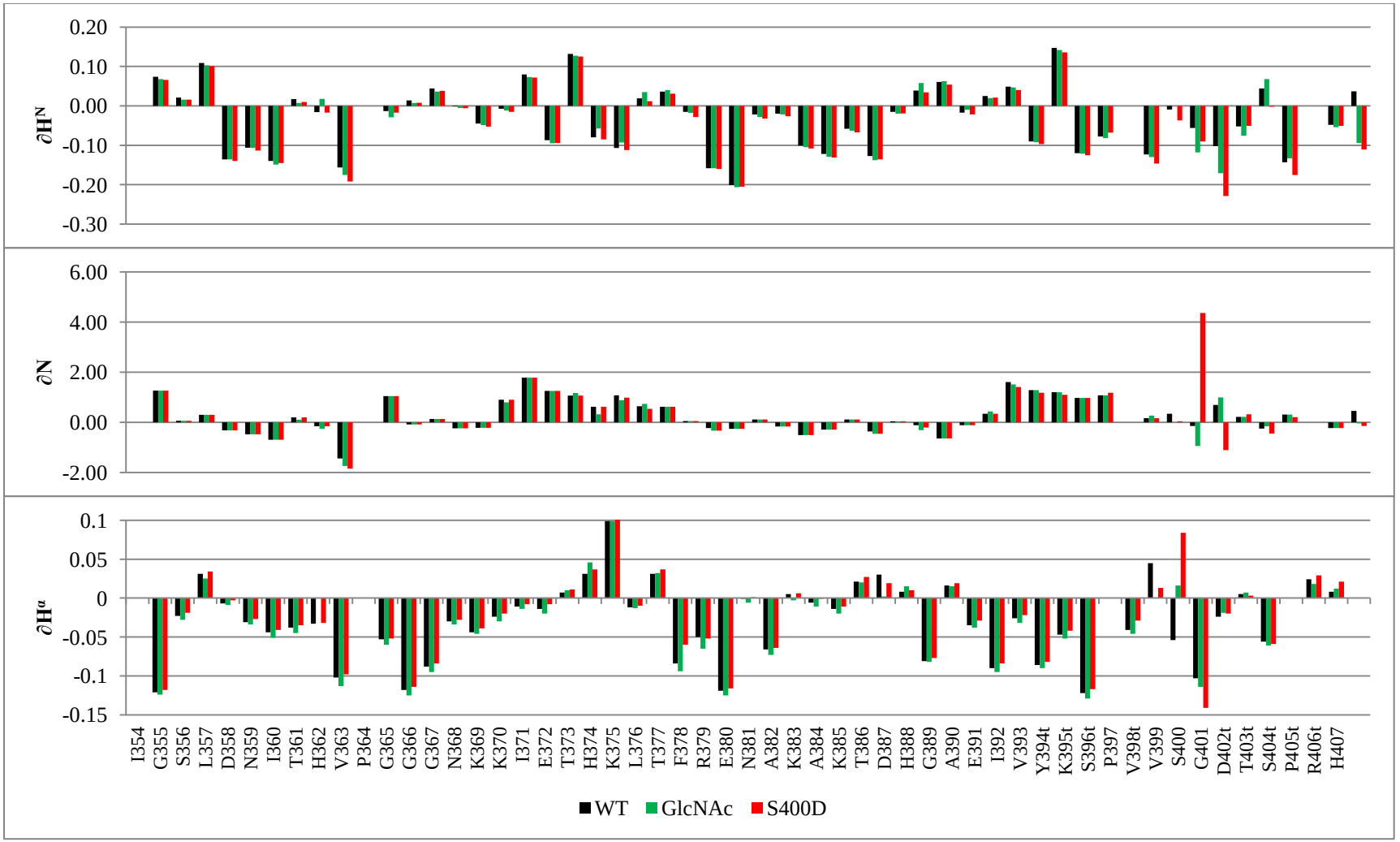

Figure C.2: Change in $^{1}H^{N}$, ^{15}N, and $^{1}H^{\alpha}$ chemical shift values for wild-type, *O*-GlcNAc-modified, and S400D tau$^{353-408}$ compared to predicted chemical shift values [75].