

# Tau protein misfolding and aggregation induced by abnormal N-glycosylation: Insights from Molecular dynamics simulation studies.

Alen Mathew T<sup>1</sup>, Anurag Baidya TK<sup>1</sup>, Bhanuranjan Das<sup>1</sup>, Bharti Devi<sup>1</sup>, and Rajnish Kumar<sup>1</sup>

<sup>1</sup>Indian Institute of Technology BHU Varanasi

July 4, 2022

## Abstract

Various post translational modifications like hyper phosphorylation, O-GlycNAcylation, and acetylation have been attributed to induce the abnormal folding in tau protein. Recent *in vitro* studies revealed the possible involvement of N-glycosylation of tau protein in the abnormal folding and tau aggregation. Hence in this study, we performed microsecond long all atom molecular dynamics simulation to gain insights into the effects of N-glycosylation on Asn-359 residue which forms part of the microtubule binding region. Trajectory analysis of the stimulations coupled with essential dynamics and free energy landscape analysis suggested that tau, in its N-glycosylated form tend to exist in a largely folded conformation having high beta sheet propensity as compared to unmodified tau which exists in a large extended form with very less beta sheet propensity. Residue interaction network analysis of the lowest energy conformations further revealed that Phe378 and Lys353 are the functionally important residues in the peptide which helped in initiating the folding process and Phe378, Lys347&Lys370 helped maintaining the stability of the protein in the folded state.

## Tau protein misfolding and aggregation induced by abnormal N-glycosylation: Insights from Molecular dynamics simulation studies.

Alen T Mathew, Anurag TK Baidya, Bhanuranjan Das, Bharti Devi, Rajnish Kumar\*

*Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (B.H.U.), Varanasi-221005 (U.P.) India*

\* Address correspondence and reprint requests to

Rajnish Kumar, M. Pharm., Ph. D.

Assistant Professor

Department of Pharmaceutical Engineering & Technology

Indian Institute of Technology (B.H.U.)

Varanasi-221005 (U.P.) India

Email: rajnish.phe@iitbhu.ac.in

ORCID: 0000-0002-7021-6996

## Acknowledgments

Alen T Mathew, Anurag TK Baidya & Bhanuranjan Das acknowledges the Indian Institute of Technology (BHU) Varanasi for providing teaching assistantship. Dr. Rajnish Kumar is grateful to the Indian Institute of Technology (BHU) Varanasi for the seed grant and Science Engineering & Research Board (SERB), India for providing start-up research grant (SRG/2021/000415). Bharti Devi is grateful to Science Engineering & Research Board (SERB) for providing Junior Research Fellowship. The support and the resources provided by ‘PARAM Shivay Facility’ under the National Supercomputing Mission, Government of India at the Indian Institute of Technology (BHU), Varanasi are gratefully acknowledged.

## Abstract

Various post translational modifications like hyper phosphorylation, O-GlycNAcylation, and acetylation have been attributed to induce the abnormal folding in tau protein. Recent *in vitro* studies revealed the possible involvement of N-glycosylation of tau protein in the abnormal folding and tau aggregation. Hence in this study, we performed microsecond long all atom molecular dynamics simulation to gain insights into the effects of N-glycosylation on Asn-359 residue which forms part of the microtubule binding region. Trajectory analysis of the stimulations coupled with essential dynamics and free energy landscape analysis suggested that tau, in its N-glycosylated form tend to exist in a largely folded conformation having high beta sheet propensity as compared to unmodified tau which exists in a large extended form with very less beta sheet propensity. Residue interaction network analysis of the lowest energy conformations further revealed that Phe378 and Lys353 are the functionally important residues in the peptide which helped in initiating the folding process and Phe378, Lys347&Lys370 helped maintaining the stability of the protein in the folded state.

**Keywords:** Tau protein; N-glycosylation; Molecular dynamics simulation; Glycosylation; Neurodegenerative disease; Protein folding

## Conflict of Interest:

The authors declare no conflict of interest.

## Introduction

A progressive neurological disease, which mainly affects the senile population, Alzheimer’s disease (AD) is one of the most prominent form of dementia [1]. The advances in molecular biology, structural biology and neurobiology in the recent past have enhanced our understanding of the pathological process which underlies this progressive neurodegenerative disease, but still we do not have a complete understanding of the molecular mechanisms associated with the disease. An estimated 55 million people across the globe suffer from Alzheimer’s and other related dementia, with no disease modifying drugs or therapy still available [2]. The current strategy of treatments revolves around providing symptomatic relief with cognition enhancement drugs and early diagnosis with biomarkers [3]. Central to the neuropathology of AD is the loss of synapse and synaptic plasticity, which reduce the ability of individuals to make and store new memories [4]. The two most prominent hypotheses which help explain this cognitive decline are the amyloid cascade hypothesis and the tau hyperphosphorylation hypothesis [5, 6].

Tau protein also called microtubule associated protein tau, is an intrinsically distorted protein (IDP), which does not have a complete secondary or tertiary structure when it’s in the unbound form. Encoded by the MAPT gene located on chromosome 17, with an mRNA of 16 exons, tau protein occurs in different isoforms in the human brain [7, 8]. Alternate splicing of the exons 2, 3 & 10 leads to the formation of six different isoforms with chain lengths ranging from 352-441 amino acids [9, 10]. The major role of tau protein in neurons is that they bind and stabilize the cytoskeletal fragments called microtubules. This in turn helps in the proper intracellular trafficking [11, 12]. Once bound to the microtubules, the tau protein will have residue specific secondary structure which stabilizes the binding [13, 14]. Hyperphosphorylation of tau protein have been attributed as one of the main causes of AD because of its ability to destabilize the tau-microtubule binding and induce the formation of neurofibrillary tangles [15, 16].

The structure of the longest tau isomer with 441 amino acids can be divided into three different domains,

the projection domain, proline rich domain and microtubule binding domain. The projection domain is the region that is close to the N terminal and has two sub regions called the N1 & N2 (1-165), which are not attached but projected away from the microtubules. The proline rich domain extends from amino acid 166 to 242 having a high density of proline residues. The microtubule binding region consists of four repeat regions (R1, R2, R3 & R4) extending from residues 243 to 367, which binds to the microtubules and have certain residue specific secondary structures, followed by the C terminal region [17, 18].

Post translational modifications (PTMs) have been associated with the pathophysiology of many intrinsically distorted proteins like tau [19]. Alquezer and coworkers have summarized various PTMs on tau protein where they have given a detailed review of all modifications along with the probable modification sites [20]. The most widely studied PTM associated with AD is tau hyperphosphorylation. Glycosylation is another important PTM on tau, which has been shown to be correlated with AD [21, 22]. Tau undergoes O-GlcNAcylation, which involves an attachment of the  $\beta$ -N-acetylglucosamine (GlcNAc) monosaccharide to either Serine or Threonine residues, and it has been found that the O-GlcNAcylation levels have been reduced in AD brains. Since both phosphorylation and O-GlcNAcylation takes place on serine/threonine residues, the presence of one is thought to hinder the other [23, 24]. Another interesting PTM is N-glycosylation of the tau protein. It was found that tau protein obtained from AD brains have been N-glycosylated but not those from the healthy brains [25]. N-glycosylation of tau was an unusual observation as the machinery needed for tau N-glycosylation, i.e., endoplasmic reticulum and Golgi are located on pathways for secreted proteins and usually not accessible to tau protein, which is an intracellular protein [26-28]. It was also observed that N-glycosylation takes place on asparagine residues having sequence N-X-S/T (where X can be any amino acid except proline) [29] and the three potential N-glycosylation sites of human tau protein are Asn167, Asn359, and Asn410; numbering based on longest tau isoform [26].

The fact that tau is N-glycosylated in AD brains and not in healthy brains, suggested the possible involvement of this PTM in AD pathology. It was also found that paired helical filament (PHF) tau when deglycosylated, changed the structure from the folded state to straight filaments [30]. A recent *in vitro* study conducted by Losev and coworkers using site specific mutation of putative N-glycosylation sites in tau protein expressed in *Drosophila* and SH-SY5Y cells, suggested the involvement of the three potential N-glycosylation sites in tau aggregation. SH-SY5Y cells and transgenic *Drosophila* expressing tau protein with asparagine to glutamine mutants at the putative N-glycosylation sites were generated, and it was found that N-glycosylation at Asn167 & Asn359 enhances the tau aggregation, while Asn410 reduces tau aggregation. They also found that expression of the N359Q hTau mutant leads to amelioration of AD-like symptoms, reduced neurodegeneration and importantly, N359Q is the only mutant that improved the lifespan of the flies [31].

Even though these *in vitro* results suggest that N glycosylation of tau protein at specific residues (Asn359) leads to tau aggregation, the exact mechanism by which it occurs is still not fully understood. Understanding the N-glycosylation induced effects on protein dynamics of tau can help us gain some insights into misfolding patterns and some of the important residue interactions. This in turn can help the scientific community to further explore the pathogenic role played by N-glycosylation in AD and identify new targets in the pathogenic N-glycosylation pathway. Here, in this study we used molecular dynamics (MD) simulations coupled with essential dynamics, free energy landscape analysis (FEL) & residue interaction network analysis to evaluate the effects caused by the N-glycosylation on Asn-359 residue, which forms part of the microtubule binding region of the tau protein. Through this study, we aimed to study whether N-glycosylation can induce tau protein folding and what are the important residue specific interactions which helps in the process. We hope that the information that we gain from this study can be used in future research to build new anti-tau aggregation compounds as disease modifying therapeutics against AD and other related tauopathies.

## Results & Discussion

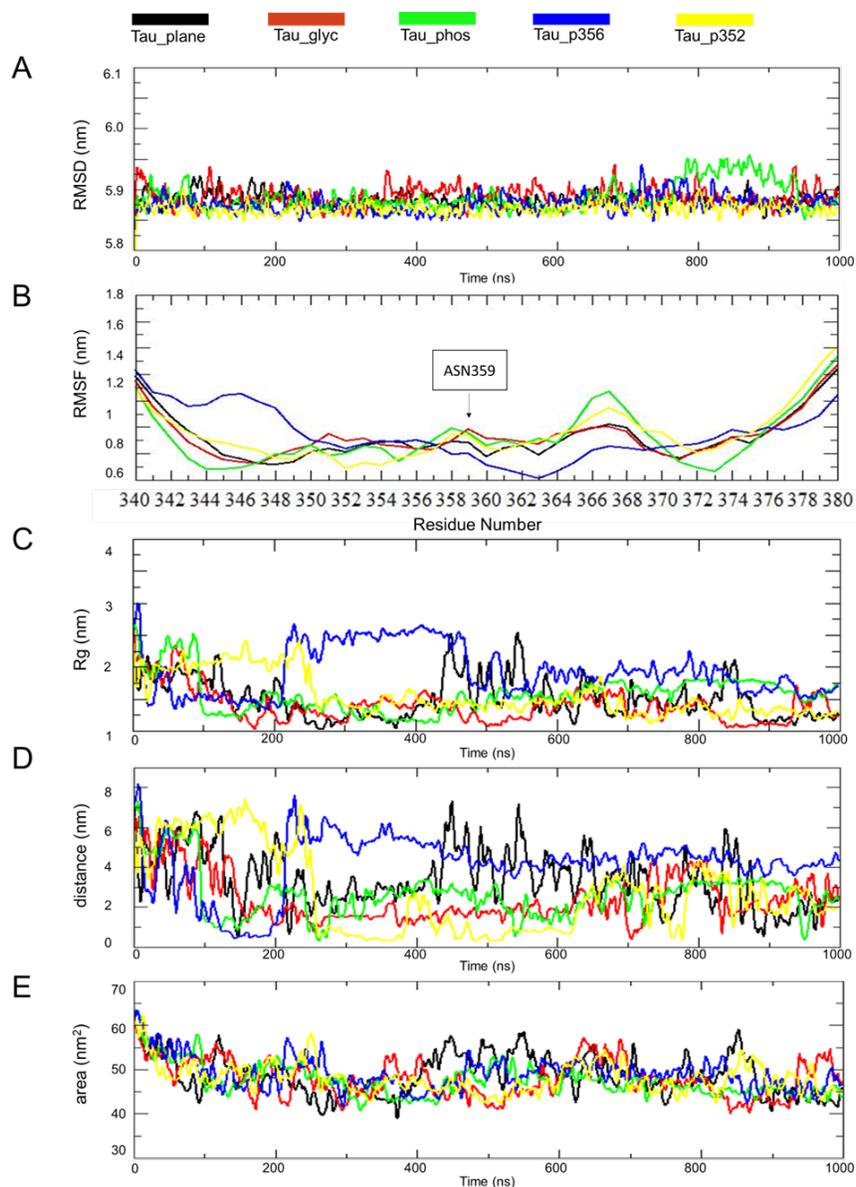
MD simulations are now recognized as reliable computational chemistry tool which helps in predicting atomistic movements of small biomolecules like proteins and help in understanding various biological phenomenon like protein folding, misfolding, protein stabilities, allosteric modulation and aggregation. MD simulations offer various advantages that help us model biomolecules in a predefined system having user-defined conditions

of pH, ionic concentration, and temperature and predict the molecular behavior using classical mechanics. At the same time, the reliability of the results depends on the timescale of the simulation, trajectory analysis and type of simulation performed [32, 33]. In the present study we tried to explore the effects that N-glycosylation of Asn359 residue have on the tau protein dynamics using one microsecond length all atom molecular dynamics simulation. The Asn359 residue, where N-glycosylation takes place, is located within the R4 domain (repeat domain) of the microtubule binding region of tau protein [34]. Therefore, the presence of sugars in this region can sterically hinder Tau ability to bind microtubules leading to enhanced misfolding and neurodegenerative effects. Moreover, Asn359 is part of the  $\beta 7$  (356SLDNITHV363) region of Tau, which is one of the eight sheets ( $\beta 1$ – $\beta 8$ ) which form the  $\beta$ -core (V306–F378) of the Tau filament [35]. We chose to study the region of tau from residue 340 to 380 which include the entire B7 region and R4 repeat domain (337-369). Subjecting only the 41 amino acid long chain of tau for microsecond long molecular dynamics simulation helps to capture most of the conformational ensemble the protein can occupy, which is not the case with simulating the whole protein. Two systems were generated, Tau\_plane – having only the selected residues with no modifications and Tau\_glyc – having N glycosylation at Asn- 359. Since the selected region of tau has two important phosphorylation sites, i.e. Ser356 and Ser352, we choose to model separate phosphorylated systems of tau in order to compare the effects induced by both phosphorylation and by N-glycosylation [36]. Details of the different systems generated for the study are given in the methods section.

**N-glycosylated protein exists in a folded /collapsed conformation while unmodified protein exists in a large extended conformation**

#### **RMSD & RMSF**

To analyze the structural stability of each of the systems throughout the simulation, we performed RMSD and RMSF calculations. The RMSD values were calculated for the protein backbone atoms (**Figure 1 A** ). Each of the systems attained equilibrium within the first 10 nanoseconds, and the RMSD values of all systems showed deviations within the range of 5.8 to 6 nm and remained within the range throughout the trajectory. Tau\_plane is relatively stable throughout the simulation, while Tau\_glyc is showing minor fluctuations, within the 0.1 nm range. Tau\_phous shows fluctuations from 800 to 900 ns. The analysis shows that each of the systems tends to remain stable throughout the trajectory without any major structural deformations and is converging.



**Figure 1: Analysis of MD trajectories.** (A) Root mean square deviation of the five simulated protein systems, (B) Root mean square fluctuations of the C-alpha atoms in the simulating proteins, (C) Radius of gyration of the protein systems indicating protein compactness, (D) End-to-end distance, showing mean distance between two extreme ends of the proteins, (E) Total solvent accessible surface area of the protein indicating the stability of the hydrophobic protein core.

Root mean square fluctuations of the C alpha atoms of each system were analyzed (**Figure 1 B**) to get an insight into the flexible regions of the protein throughout the simulations. The plot suggests that the two terminal ends of the protein are more fluctuating than the middle regions, as they are more exposed to the solvent surface with values in the range of 1.2 to 1.6 nm. The middle regions of Tau\_plane and Tau\_glyc showed similar fluctuations with values in the range of 0.7 to 1.0 nm while that of Tau\_phos and Tau\_p352 showed more fluctuations, while Tau\_p356 showed the least fluctuations among all systems throughout the middle region.

The atoms 409 to 429, which represents the three consecutive glycine (365, 366, 367) residues are highly fluctuating in all five systems. In Tau\_glyc the atoms of Asn359, where glycosylation takes place is fluctuating more as compared to Tau\_plane.

### Radius of Gyration & End to End Distance

The radius of gyration is a measure of the compactness of the protein structure throughout the simulation. It can also give hints about the conformational patterns occupied by the protein, as a stable, relatively low value indicates a probable compact folded conformation and vice versa [37]. From **figure 1 C** we can see that, the radius of gyration of Tau\_plane is not constant and is highly varying over the course of the simulation, with values reaching up to 2.75 nm, while Tau\_glyc and Tau\_phos show relatively stable and low Rog throughout the simulation, with values below 1.5 nm, indicating they have a stable compact and more likely a folded state. Tau\_p356 gets stabilized after 500 ns, and maintains a constant value, while Tau\_p352 gets stabilized by 250 ns and remains compact thereafter. The average values are given in **Table 1** and suggest that Tau\_glyc remains the most compact of all systems.

Like the Radius of gyration, end to end distance is another parameter that gives an idea about the mobility and confirmation of proteins during the simulation. It is the average distance between the first and last atoms in the peptide, with respect to time. A lower stable value might indicate a probable, largely folded / compact conformation and while a large value indicates extended conformation [38, 39]. From **Figure 1 D** we can see that initially, all systems had a very high end to end distance, as initially, proteins are in an unfolded extended confirmation. Within the first 100 ns, we can see that Tau\_glyc and Tau\_phos have attained an equilibrium value less than 2 nm, and remains constant throughout the simulation, except for a small spike in Tau\_glyc near the end of the simulation. Tau\_plane remains mostly in an extended form as indicated from the plot as the end-to-end distance remains higher and not stable throughout the simulation. Tau\_p352 have a higher end to end distance while Tau\_p356 gets stabilized after 230 ns with some fluctuations towards the end of the simulation. The average values, as given in **Table 1** suggest that Tau\_glyc have the least end to end distance, followed by Tau\_phos.

### Solvent accessible surface area

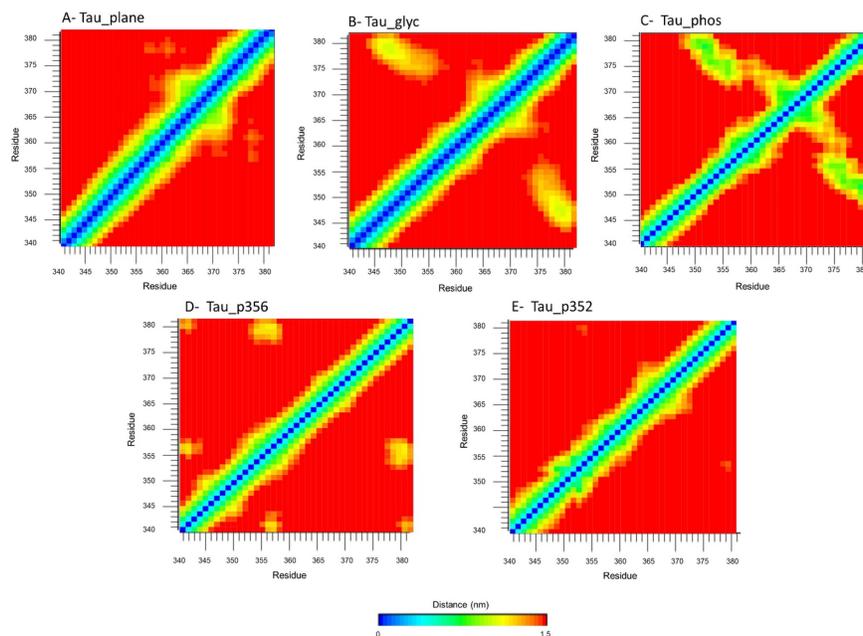
Hydrophobic interactions of amino acids at the core of the protein plays important role in the protein folding process. As the core hydrophobic interactions get buried inside the compact protein, the surface of the protein exposed to the solvent decrease since only polar amino acids at the surface makes contact with the solvent. Hence, solvent accessible surface area (SASA) analysis of proteins help to understand the surface area of proteins that are interacting with the solvent, and the stability of the hydrophobic core [14, 40]. The total SASA of each system were analyzed and plotted (**Figure 1 E**) . We can see that SASA of each system gets stabilized after the first 100 ns, while Tau\_plane shows maximum deviations and remains high for most of the simulation. Tau\_phos shows the least deviations while Tau\_glyc sasa is relatively stable and low as compared to Tau\_plane. The average SASA values indicate that Tau\_phos have the lowest SASA (38.9 nm<sup>2</sup>), followed by Tau\_glyc (44.9 nm<sup>2</sup>), while SASA for Tau\_plane (48.7 nm<sup>2</sup>) remains high (**Table 1**) . The probability distribution graphs for radius of gyration, end to end distance and solvent accessible surface have been incorporated in supporting information along with representative extended and compact conformations of the peptide in each system in **Figure S1 and S2** respectively.

**Table 1: Average (a) Solvent accessible surface area (b) Radius of gyration and (c) end-to-end distance of five systems**

System	(a) Avg. SASA (nm <sup>2</sup> )	(b) Avg. Rog(nm)	(c) Avg. End to end distance (nm)
Tau_plane	48.7	1.485	3.734
Tau_glyc	44.9	1.374	2.756
Tau_phos	38.9	1.590	2.941
Tau_p356	49.0	1.877	4.417
Tau_p352	48.0	1.542	3.455

## Close Contact Map Analysis

Analysis of residue-residue close contacts can help in identifying residues in protein that are making contact with each other during the simulation. This can give hints about the residues which are more likely to interact with each other during the protein folding process. A residue-residue close contact map for each of the systems was generated (**Figure 2**) using the gmx mdmat, with 1.5 nm cutoff distance, which calculates the inter-residue distance matrix and provides the average close contacts between the residues.



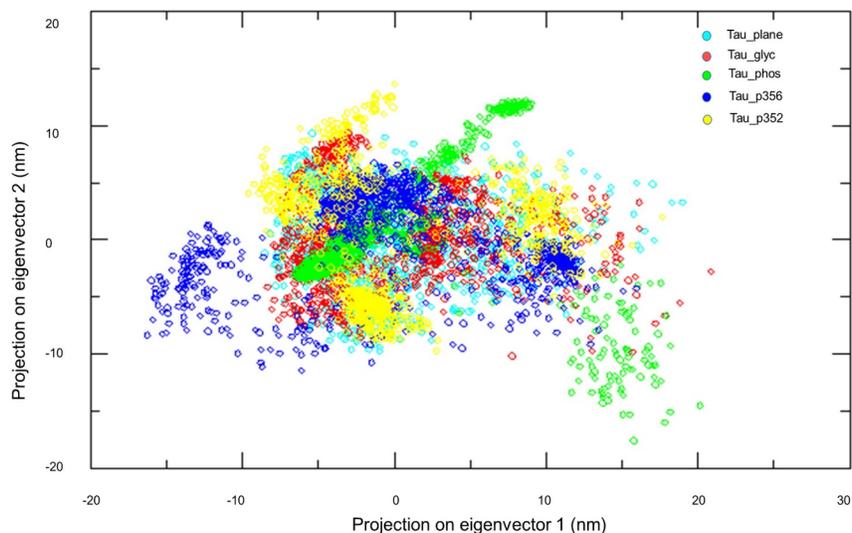
**Figure 2: Close contact maps of the five simulated systems, indicating average residue-residue close contacts across the trajectory.** Tau\_glyc and Tau\_phos show residue-residue close contacts, while the unmodified tau (Tau\_plane) shows negligible inter-residue close contacts.

We can see that in Tau\_plane, the average inter-residue close contacts are negligible, indicating the protein is making very few inter-residue close contacts, and most likely to exist in an extended conformation as a folded conformation is likely to have inter-residue interactions. But in Tau\_glyc, residue Lys343-Gly355 is making close contact with residues Glu372-Glu380. This shows that the two ends of the protein are more likely to be in close contact, and might be existing in a collapsed/folded conformation than an extended form. A similar trend is visible in Tau\_phos in which residues Lys347-Ser356 are in close contact with residues Ile371-Glu380. Tau\_p356 forms negligible close contacts between residues Ile354-Ser356 with Leu376-Glu380 and Lys340-Lys341 with Gly355-Ser356-17. Tau\_p352, like Tau\_plane form negligible close contacts.

## Principal Component Analysis

The principal component analysis (PCA), also called essential dynamics analysis is an important statistical analysis tool which helps in analyzing protein motions by dimensionality reduction. It filters the trajectories and helps identify the essential motions or the functionally relevant motions of the protein [41]. We performed PCA analysis of peptides in each system using the gmx cover & gmx ana eig, by diagonalizing the covariance matrix of atomic coordinates.

The plot showing the eigenvalues and eigenvectors of the five simulated systems are given in **Figure S3** in supporting information. The 2D projections of the first two PCs were plotted (**Figure 3**), with each dot representing a conformation space occupied by the protein.



**Figure 3: Projection of the first two principal components of the five simulated systems.** Unmodified Tau (Tau\_plane) showing wide distribution of conformational sampling, while Tau\_glyc & Tau\_phos occupies a more concentrated conformational space. Tau\_p356 and Tau\_p352 show the most widespread distribution of conformational space.

Large conformational changes in the protein cause the spreading of the distribution in the conformational space. **Figure 3** suggest that the conformational space occupied by Tau\_plane is widely distributed indicating the wide range of conformational states occupied by the protein over the course of the entire trajectory, while Tau\_glyc is having a relatively narrow conformational space as compared to Tau\_plane. Tau\_phos have the most concentrated conformational sampling, suggesting a stable conformational space occupied the phosphorylated system. Tau\_p356 and Tau\_p352 show concentrated conformational groupings, which are separated from each other, suggesting clustering of conformations.

In order to get deeper insights into the essential motions of tau in each of the systems, we extracted the movements of the protein across the first and second principal components. The functionally relevant motions of each of the tau systems as represented by the first two principal components are given in the supporting information (**Movie S1**) as movies. It is evident from the motions that unmodified tau (Tau\_plane) tend to exist in an extended conformation as we can observe the motions suggest the peptide chain is forming elongated conformation. But this is not the case with N-glycosylated system, as we can see that the glycan is inducing the peptide system to fold. The phosphorylated system (Tau\_phos) also follows a similar trend as we can see the peptide getting folded, rather than extending. Tau\_p356 and Tau\_p352 both do not show much folding tendency as the fundamental motions suggest they exist in an elongated form.

Results of the analysis of essential dynamics of the protein systems are in line with the previous analysis given above, suggesting that the glycosylated (Tau\_glyc) and phosphorylated (Tau\_phos) systems are actually folding and largely exist in a folded state whereas unmodified tau is largely existing in an open extended state.

### Higher beta sheet propensity in N-glycosylated tau as compared to unmodified tau

#### Secondary Structure

Secondary structure is an important analytical tool which help us understand the structural preference of a protein throughout the simulation. Proteins during the folding process take up different structural conformations like alpha helix, beta sheets, turns etc. which plays significant role in defining protein functions

in the underlying biological cascades [42]. Hence it is important to understand the secondary structure elements present in protein because, just as a particular structure can determine the protein function, any deviation from the normal structure, possibly due to misfolding or mutations, can lead to a pathological condition as well [43].

Analysis of PHF tau obtained from the brain of Alzheimer’s patients have revealed that tau is rich in beta sheet conformation in the microtubule binding region[35]. Indicating that tau protein when undergoing folding or aggregation tend to form secondary structure elements that are rich in beta sheet content [44, 45]. Hence, we performed secondary structure analysis of each of the systems using the DSSP [46] Algorithm and the time evolution of secondary structure throughout the trajectory is plotted in **Figure S5** in supporting information.

**Table 2: Average secondary structure across the trajectory for five systems.**

% Secondary Structure	Tau_plane	Tau_glyc	Tau_phos	Tau_p356	Tau_p352
coil	60	63	68	73	63
Beta sheet	1	4	3	0	0
Beta bridge	1	3	6	1	1
Bend	21	22	18	22	24
Turn	11	7	5	4	11
A-helix	3	0	0	0	0
5-helix & 3-helix	2	1	1	0	0

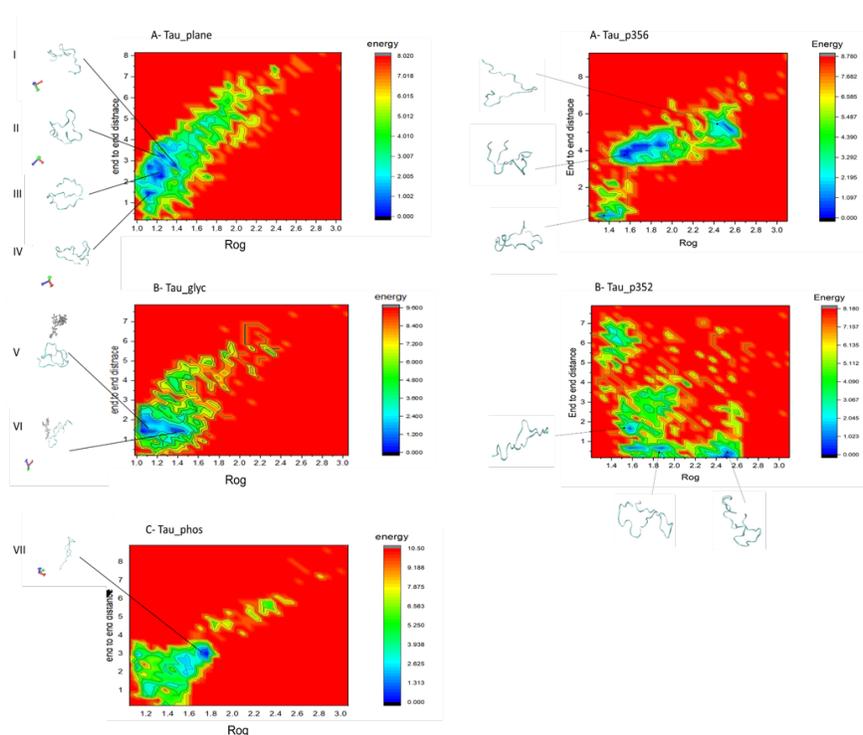
Secondary structural preference of tau in each of the systems (**Table 2**) shows random coil is the most preferred confirmation. All five system has more than sixty percent of the amino acids in a random coil state. The maximum number of beta sheets are found in the glycosylated tau (Tau\_glyc) form, i.e., 4 %, followed by phosphorylated (Tau\_phos) tau form (3%), while in Tau\_plane percent of beta sheet is as low as 1%. The percentage of beta bridges are more in phosphorylated tau (Tau\_phos), followed by glycosylated tau. Bend confirmation is the second most preferred state in all systems throughout the trajectory, with each system having close to 20 % amino acids in bend form. Turns and alpha helix are highest in Tau\_plane, other systems have fewer turns and negligible helix forms. The secondary structure analysis indicates that the glycosylated form of tau shows the maximum preference to beta sheet confirmations, followed by the phosphorylated tau (Tau\_phos). The unmodified tau protein shows a very low beta sheet propensity. This indicates that glycosylated tau have a high aggregation tendency as indicated by its high beta sheet content. It is also worth mentioning that, apart from the intra molecular beta sheets that we analyzed from the simulation trajectories, tau fibril formation & aggregation also depends on various inter molecular beta sheet formations which are formed as a result of the interaction of more than one tau protein during the aggregation process, which could not be sampled as we were studying only a single tau peptide.

### **Folding funnel coupled with residue interaction networks reveals functionally important residues for N-glycosylation induced tau folding process**

#### **Free Energy Landscape (FEL) analysis**

Folding funnel is a fundamental concept in the free energy landscape of ordered proteins. It helps in understanding the different conformational ensembles explored by the protein throughout the simulations and how they fold into an ensemble having a global energy minimum. Each funnel in the free energy landscape diagram represents a reaction path in the folding process, and ideally, a protein folds into its native structure, having the lowest energy, with a single narrow funnel, where the intermediate structures are represented within the funnel at higher energy levels [47, 48]. But the folding scenario in IDP is different, as they do not have proper tertiary structure and tend to fold once bound to a binding partner. This indicates that the free energy landscape of IDPs is rugged and contains multiple funnels [49-51]. But they can provide insights into the folding pattern and minimum energy conformations occupied by the protein. A FEL diagram was

generated for each of the simulated systems, with Radius of gyration and End to end distance as the two reaction coordinates using the gmx sham module [52] and 2D contour plots and 3D projections of the free energy are plotted in **figure 4** and supporting information **Figure S4** respectively.



**Figure 4: 2D free energy landscape diagram of Tau\_plane , Tau\_glyc , Tau\_phos , Tau\_p356 and Tau\_p352 along with the structure corresponding to the local energy minima.** The figure shows the wide distribution of structural ensembles in Tau\_plane with a large energy barrier, while in Tau\_glyc and Tau\_phos, we have comparatively narrow distribution and fewer local energy minima.

**Figure S4** shows Tau\_plane having four funnels each corresponding to a local minimum with an energy barrier separating them suggesting the folding process is rugged and not stable. The 2D contour plot, **Figure 5** , indicates the wide distribution of ensembles suggesting many of the conformations are trapped in high energy states. Tau\_glyc shows two local minima, without much energy barrier separating them and the conformational ensembles are not widely distributed as in Tau\_plane. This suggests a comparatively stable folding process in Tau\_glyc as compared to Tau\_plane. Tau\_phos system have a single energy minimum, suggesting the protein folds into a single native structure, but we can see that the funnel is broad and have many conformations trapped in high energy states. Five local energy minima observed in the FEL of Tau\_p356 indicate the rugged nature of the folding process. These local minima are separated by a large energy barrier, suggesting the possibility of the protein getting trapped in one local energy minima. Tau\_p352 FEL indicates two local minima, one at low energy and another at high energy, separated by a large energy barrier, suggesting the entrapment of protein in high energy states. Tau\_plane having four local energy minima metastable states have an energy barrier of about 3-4 KJ/mol separating them, while in tau\_glyc we have two low energy metastable states with  $\sim 2$  KJ/mol energy barrier. In Tau\_phos the we have a single prominent lower energy state separated by the next higher energy state by a transition barrier of  $\sim 3$  KJ/mol. In the case of tau\_p356 and tau\_p352 the lowest energy metastable states are separated by an energy barrier of  $\sim 6-7$  KJ/mol.

Structural analysis of the lowest energy metastable conformations from each of the systems obtained from

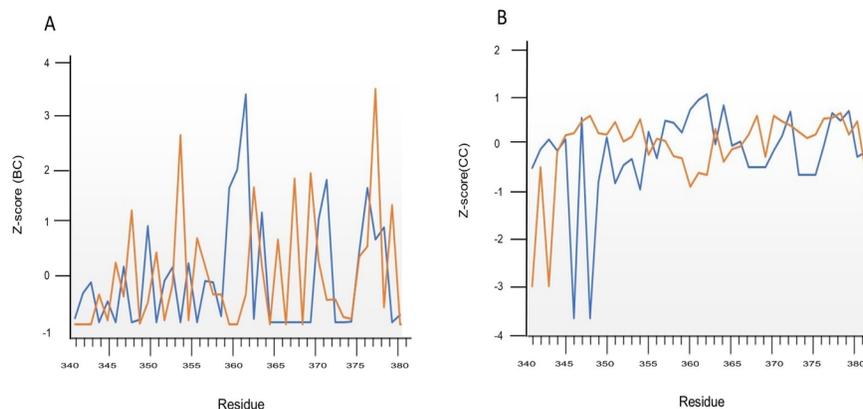
FEL were performed. The four local energy minima conformations of Tau\_plane (I-IV) was analyzed for secondary structure preference, especially beta sheet formation, and none of the four local energy minimal structures had beta sheet confirmations, while only III had a 5 percent isolated beta bridge. Both local minima conformations (V & VI) of Tau\_glyc had beta sheets with 10 percent beta sheets in each structure. Both V and VI were analyzed for salt bridges and we found that V had two salt bridges Asp345-Lys343 & Glu372- Lys375, while VI also had two salt bridges, Glu380-Lys347 & Glu380-Lys353. The lowest energy conformation of Tau\_phos VII had zero percent beta sheets but had around 10 percent residues with beta bridges and no salt bridges. The energy minimal metastable structures of Tau\_p356 and Tau\_p352 did not have any beta sheets but largely exist in random coils and helix. This suggests that the N-glycosylation induced folding of tau is able to form beta sheet rich regions in the protein, while the unmodified tau is having an unstable folding process with none of the lowest energy metastable structures having beta sheet propensity.

### Residue Interaction Network (RIN) Analysis

Structural analysis of the lowest energy conformation of Tau\_plane showed that it is having 1 salt bridge and 17 strong hydrogen bonds, while that of Tau\_glyc have 2 salt bridges and 23 strong hydrogen bonds. These interactions might be helping in the stabilization of the conformations, but we do not know which are the functionally important residues helping in folding /aggregation or maintaining the stability of the confirmation.

Residue interaction network (RIN) is an analysis tool that make use of the mathematical concept of graph theory, in which the 3D protein structures are represented as a set of nodes and edges. Here nodes represent the amino acids while the edges represent the interactions between the amino acids, creating a network [53, 54]. This network topology can be analyzed with help of various centrality measures and helps identify structurally and functionally important nodes and edges [55]. It has been used in studying protein stability and folding, effects of mutations on protein structure, protein dynamics, allosteric regulations, identification of functionally and biologically important residues, identification of ligand binding sites etc. [56-60]. Here we make use of RIN to study the functionally important residues. In order to make interaction networks, we made use of the lowest energy conformations of Tau\_plane and Tau\_glyc obtained from the free energy landscape analysis. The residue interaction networks are plotted in **Figure S6** in supporting information, & topological analysis of the networks is done using various centrality measures like betweenness centrality (BC) and closeness centrality (CC) [61]

Betweenness centrality measure how many times a particular node acts as a bridge between the shortest path between two other nodes. It is a measure of how important a residue is in the communication of signals within a protein. Residues which have high BC reveal locations which are important for controlling the inter domain communication [62]. Vendruscolo and coworkers have previously demonstrated that residues having high BC values are found in the native structure and/or the transition state ensemble of proteins, which they identified as key residues which can initiate the folding process [63]. BC of Tau\_plane and Tau\_glyc were calculated and normalized Z score-based plotted in **Figure 5 A**.



**Figure 5: (A) Z-score of betweenness centrality (Tau\_glyc-orange & Tau\_plane-blue), indicating Phe378, Lys353 are functionally important residues which helps in initiating folding process in glycosylated tau and (B) Z-score of closeness centrality showing Phe378, Lys347&Lys370 are the residues which helps stabilize the glycosylated tau energy minimum structure.**

In Tau\_glyc (orange) we can see that residue Phe378 and Lys353 have the two highest BC values, while Tau\_plane (blue) has one peak BC value corresponding to Thr361. Structural analysis of Tau\_glyc confirmation revealed that Phe378 is having a strong H bond with Thr377, both of which are involved in beta sheet formation at the region. Phe378 is also involved in a weak hydrogen bond with Asp345, aromatic hydrophobic interactions with Leu344 and Val363. Lys-353 is involved in weak hydrogen bond formation between Lys370 and Ile371. In Tau\_plane Thr361 was found to have a strong hydrogen bond between Phe378 and Leu376. Thr377 was also one of the two residues forming the beta bridge, the other one being Thr377 with which Thr361 is having a weak hydrogen bond.

Closeness centrality is a measure of the ability of a node in a network to spread information very efficiently. It is considered as the inverse of the average shortest path length. It provides global information, indicating how many steps are needed to reach every other node in the network [64]. Dokholyan and coworkers have demonstrated that the proteins in post transition state (i.e., those confirmations which are more likely to form native structures or close to native structures) have low average shortest path length [65]. Since CC is inverse of average shortest path length, The residues which have high CC tend to be important as they help stabilize the structure, and usually lie in the protein core and remain compact [66].

CC analysis of Tau\_glyc (**Figure 5 B**) showed that in Tau\_glyc the residues having the highest CC are Phe378, Lys347 &Lys370, while in Tau\_plane Thr361, Ile360 & Asn359. In Tau\_glyc Phe378 is involved in beta sheet formation as explained earlier, while Lys347 is also involved in the second beta sheet formation having a strong hydrogen bond between Thr377. It also has a weak hydrogen bond with Leu376. Lys370 is involved in an isolated beta bridge having a weak hydrogen bond with Lys353. In Tau\_plane (blue) we have Thr361 with an isolated beta bridge, having interactions as discussed earlier. Ile360 has a strong hydrogen bond with Leu357 & weak interaction with Leu376. Asn359 is involved in strong hydrogen bonds between Arg349, Ser356 & Leu344. The CC graph revealed that residues which showed high CC are those which were involved in secondary structure formation and helps in protein stabilization.

The above two centrality measures suggest that when tau protein is not glycosylated the central residues or the functionally important residues are those present in the middle region of the protein, close to Asn359 (N-glycosylation site), but these residues do not induce much folding or aggregation. while the tau is glycosylated the centrality of residues shifted away from Asn359. BC analysis showed that in the N glycosylated folded state, Phe378 & Lys353 are the functionally important residues, suggesting their role in initiating the folding process in transition state ensemble or stabilization of the native structure. Comparison of RIN analysis

results to Cryo Electron microscopic (EM) structure of PHF tau showed that the residues having high BC, i.e., Phe378 is part of the  $\beta$ 8 and Lys353 is part of  $\beta$ 6 region of PHF tau [67]. Also, residues having high CC values, i.e., Phe378&Lys370, part of the  $\beta$ 8 region, Lys347 part of the  $\beta$ 5 region of PHF tau. This indicates that the functionally important residues as indicated by the RIN analysis are actually key players in the beta core region of tau as predicted by the Cryo-EM structure of tau.

## Conclusion

Through this study, we tried to explore the effects of N-glycosylation on the structural dynamics of tau protein by glycosylating one of the PTM site, i.e., Asn-359 using a microsecond long all atom molecular dynamics simulation. Analysis of system stability showed that each tau system were quickly equilibrated, and remained stable throughout the trajectory without any major instabilities with the two extreme end regions of the protein being the most flexible compared to the middle regions. Protein compactness analysis and essential dynamics further showed that when the peptide was N-glycosylated it remained most compact and had a folded state, while in the unmodified state, the peptide was less compact and had an extended conformation. Further insights were obtained from the secondary structure analysis, which showed that the glycosylated tau, had a higher beta sheet propensity as compared to the unmodified peptide. Free energy landscape analysis revealed the folding process of the unmodified tau is rugged and unstable, while that of N-glycosylated and phosphorylated systems were comparatively stable and structural Analysis of the local energy minima structures revealed that glycosylated energy minimum structures had a higher amount of beta sheets while the unmodified tau had zero beta sheets, further justifying that N-glycosylation induces the tau protein to fold. Residue interaction network analysis revealed the locations in the protein which helps in the folding process. BC analysis revealed Phe378 & Lys353, as the functionally important residues for the folding process, while the Phe378, Lys370 & Lys343 helped stabilize the folded state of the protein.

Tau hyperphosphorylation is already an established factor which induces tau protein folding and aggregation. Comparative analysis of our results of the glycosylated system (Tau\_glyc) and phosphorylated system (Tau\_phos) suggest that both showed similar behavior, both existing in a folded conformation having comparatively high beta sheets regions. But Tau\_p356 & Tau\_p352 doesn't induce much folding. Our analysis and results are in line with the *in-vitro* studies conducted by Losev & coworkers as our study confirmed that N-glycosylation at Asn-359 is inducing tau protein folding. Also, the high beta sheet content found in the glycosylated tau shows it is having high aggregation propensity as compared to the unmodified non glycosylated tau. We believe that the results of this analysis can be further be used for the development of novel tau aggregation inhibitors which can target the functionally important residues and their residue specific interactions. Moreover, the process of N-glycosylation is initiated by oligosaccharyltransferase enzyme and the pathogenic role played by this enzyme in AD pathology and its druggability together with the exact mechanisms by which tau is N-glycosylated can be further explored. We also believe that, this study can help to explore the development of novel therapeutic molecules targeted at blocking the N-glycosylation induced tau misfolding and aggregation.

## Methods

### Selection of Glycans

Chemical analysis of the different glycans attached to the PHF tau and AD phosphorylated tau protein obtained from AD brain was carried out by Sato and coworkers and found that the glycans were rich in High mannose-type sugar chains and truncated N-glycans were also found in both tau in addition to a small amount of sialylated bi- and triantennary sugar chains. Proposed structures and their percent molar ratios of N-glycans obtained from AD P-tau and PHF-tau were reported by sato and coworkers [68]. For the purpose of the study, we have selected the glycan which is present in both PHF tau and AD P-tau, and at the highest proportion. The structure of the glycan is given in supporting information (**Figure S7**).

### Systems

The 41 amino acid long tau sequence from residue number 340 to 380, along with coordinates were extracted

from protein with PDB-ID 6TJX, chain A [69]. The structure was energy minimized, equilibrated and the random coil conformation was used for running the production simulations. The N-glycosylation, with selected glycan, was added to the asparagine amino acid at residue number 359 using CHARMMGUI web server [70]. Likewise, the phosphorylation was also added through the CHARMMGUI webserver.

**Table 3 – Different systems modeled for molecular dynamics stimulation**

System	PTM	PTM site	Total atoms	Ions(Na/Cl)
Tau_plane	Nil	nil	39997	37/40
Tau_glyc	N-glycosylation	Asn359	39979	37/40
Tau_phos	Phosphorylation	Ser356 & Ser352	39971	38/37
Tau_p356	Phosphorylation	Ser356	39953	37/38
Tau_p352	Phosphorylation	Ser352	39941	37/38

### Simulation Protocol

All the systems were simulated using the CHARMM-36M [71] forcefield, which is an improved forcefield for MD simulations for intrinsically distorted proteins and accurately define carbohydrates [71]. Each of the systems were solvated using the TIP3 water model in a rectangular box of dimensions 115 Å × 75 Å × 75 Å, and neutralized by adding Na<sup>+</sup> and Cl<sup>-</sup> ions using the Monte carlo [72] ion placing method. Each system was later energy minimized using the steepest descent algorithm for 10000 steps, with bonds having H constrained using the LINCS algorithm [73]. This was followed by a 1 ns equilibration step using Nose-Hoover [74] temperature coupling with the temperature kept at 303.15 K and pressure at 1 atm. Later each system was subjected to isobaric-isothermal (NPT) ensemble MD production stimulation for 1 microsecond length under periodic boundary conditions having isotropic Parrinello-Rahman pressure coupling [75], with pressure kept at 1 atm. The leapfrog algorithm was used for integrating the Newtonian equations of motions. Coulombic interactions were calculated using the particle-mesh Ewald [76] method having a 1.2 Å cutoff. Lennard-Jones or van der Waals interactions were calculated using 1.2 Å cutoff. All stimulations were performed using GROMACS 2020.3 [77, 78].

### Trajectory Analysis

#### Protein stability and compactness analysis

Root mean square deviation (RMSD) and root mean square fluctuations (RMSF) were analyzed using the GROMACS[77] in build analysis modules, `gmx rms` & `gmx rmsf` [79]. RMSD was calculated for the protein backbone atoms and RMSF were calculated for c-alpha atoms. The radius of gyration and end-to-end distance analysis were also performed using the inbuilt GROMACS modules like `gmx gyrate` and `gmx polystat` respectively. Also, `gmx sasa` was used to calculate the total solvent assessable surface area.

#### Contact map and Secondary structure

Contact maps for the c-alpha atoms were calculated using GROMACS [77] in build modules like `gmx mdmat` with 1.5 nm a distance cutoff. The secondary structure of the proteins was analyzed using the `dssp` algorithm, using GROMACS `gmx do_dssp`.

#### PCA and Free energy landscape analysis

PCA was also performed using GROMACS [77] modules like `gmx covar` & `gmx anaeg`, which create and diagonalize covariance matrix of atomic coordinates of atoms. Movies which shows the movement of protein along the first two principal components were created using VMD, with principal components calculated for the unmodified tau (Tau\_plane), tau-glycan complex (Tau\_glyc) and the phosphorylated systems (Tau\_phos, Tau\_p356, Tau\_p352) [80], after extracting the frames corresponding to the first two PCs using `gmx anaeg`. Free energy landscape was calculated with the radius of gyration and end to end distance as two reaction coordinates, using `gmx sham` module of GROMACS [77].

## Residue Interaction network analysis

The protein conformations were visualized in Chimera [81] and interaction networks were obtained using Structure Viz [82], a Cytoscape [83] plugin, having contacts with default parameters set as criteria for building the network which was visualized using Cytoscape. Topological analysis like betweenness centrality (BC), closeness centrality (CC) were calculated using CytoNCA [84], another cytoscape plugin, each giving inputs regarding the functionally relevant residues. Also Z-score for BC & CC were calculated using the formula,  $z = (x-\mu) / \sigma$ , where  $x$  is the actual value,  $\mu$  is the mean of the dataset, and  $\sigma$  is the standard deviation of the dataset.

## Data and Software Availability

GROMACS, Cytoscape, Chimera, VMD are freely available software, while CHARMM-GUI is a free web server. Origin Lab, software used for making figures were obtained through institutional license provided by Indian institute of Technology, (B.H.U) Varanasi. System modeling and analysis procedures are given in the methods section. The plots showing the principal components data, time evolution of secondary structure, residue interaction networks of lowest energy conformations of tau & glycan structure used for N-glycosylation of the tau protein is given in supporting information. The movies which shows the motion of tau peptides along the first two eigenvectors are also given in the supporting information.

## Declaration

The authors declare that this manuscript has been published in the form of pre-print at <https://chemrxiv.org/engage/chemrxiv/article-details/61e7ba534a603d5967353862> as a working paper with DOI as 10.26434/chemrxiv-2022-5bs5r.

## References

1. 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2021;17(3):327-406.
2. Gauthier S R-NP, Morais JA, & Webster C. World Alzheimer Report 2021:Journey through the diagnosis of dementia *Alzheimer's & Dementia*. 2021.
3. Yiannopoulou KG, Papageorgiou SG. Current and future treatments in Alzheimer disease: an update. *Journal of central nervous system disease*. 2020;12:1179573520907397.
4. Vaz M, Silvestre S. Alzheimer's disease: Recent treatment strategies. *European Journal of Pharmacology*. 2020:173554.
5. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine*. 2016;8(6):595-608.
6. Kametani F, Hasegawa M. Reconsideration of Amyloid Hypothesis and Tau Hypothesis in Alzheimer's Disease. *Front Neurosci*. 2018;12:25-.
7. Poorkaj P, Bird TD, Wijsman E, Nemens E, Garruto RM, Anderson L, et al. Tau is a candidate gene for chromosome 17 frontotemporal dementia. *Annals of neurology*. 1998;43(6):815-25.
8. Himmler A, Drechsel D, Kirschner MW, Martin Jr DW. Tau consists of a set of proteins with repeated C-terminal microtubule-binding domains and variable N-terminal domains. *Molecular and cellular biology*. 1989;9(4):1381-8.
9. Goedert M, Spillantini M, Jakes R, Rutherford D, Crowther R. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron*. 1989;3(4):519-26.
10. Panda D, Samuel JC, Massie M, Feinstein SC, Wilson L. Differential regulation of microtubule dynamics by three-and four-repeat tau: implications for the onset of neurodegenerative disease. *Proceedings of the National Academy of Sciences*. 2003;100(16):9548-53.

11. Wang J-Z, Liu F. Microtubule-associated protein tau in development, degeneration and protection of neurons. *Progress in neurobiology*. 2008;85(2):148-75.
12. Drubin DG, Kirschner MW. Tau protein function in living cells. *Journal of Cell Biology*. 1986;103(6):2739-46.
13. Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, et al. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS biology*. 2009;7(2):e1000034.
14. Lu S, Wagaman AS. On methods for determining solvent accessible surface area for proteins in their unfolded state. *BMC research notes*. 2014;7(1):1-7.
15. Alonso AdC, Zaidi T, Grundke-Iqbal I, Iqbal K. Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease. *Proceedings of the National Academy of Sciences*. 1994;91(12):5562-6.
16. LeBoeuf AC, Levy SF, Gaylord M, Bhattacharya A, Singh AK, Jordan MA, et al. FTDP-17 mutations in Tau alter the regulation of microtubule dynamics. *Journal of Biological Chemistry*. 2008;283(52):36406-15.
17. Gustke N, Trinczek B, Biernat J, Mandelkow E-M, Mandelkow E. Domains of tau protein and interactions with microtubules. *Biochemistry*. 1994;33(32):9511-22.
18. Barbier P, Zejneli O, Martinho M, Lasorsa A, Belle V, Smet-Nocca C, et al. Role of Tau as a Microtubule-Associated Protein: Structural and Functional Aspects. *Frontiers in Aging Neuroscience*. 2019;11(204).
19. Uversky VN. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Frontiers in aging neuroscience*. 2015;7:18.
20. Alquezar C, Arya S, Kao AW. Tau post-translational modifications: Dynamic transformers of tau function, degradation, and aggregation. *Frontiers in Neurology*. 2021:1826.
21. Alquezar C, Arya S, Kao AW. Tau Post-translational Modifications: Dynamic Transformers of Tau Function, Degradation, and Aggregation. *Frontiers in Neurology*. 2020;11.
22. Haukedal H, Freude KK. Implications of Glycosylation in Alzheimer's Disease. *Front Neurosci*. 2020;14:625348.
23. Ryan P, Xu M, Davey AK, Danon JJ, Mellick GD, Kassiou M, et al. O-GlcNAc modification protects against protein misfolding and aggregation in neurodegenerative disease. *ACS chemical neuroscience*. 2019;10(5):2209-21.
24. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annual review of biochemistry*. 2011;80:825-58.
25. Liu F, Zaidi T, Iqbal K, Grundke-Iqbal I, Merkle RK, Gong CX. Role of glycosylation in hyperphosphorylation of tau in Alzheimer's disease. *FEBS letters*. 2002;512(1-3):101-6.
26. Sato Y, Naito Y, Grundke-Iqbal I, Iqbal K, Endo T. Analysis of N-glycans of pathological tau: possible occurrence of aberrant processing of tau in Alzheimer's disease. *FEBS letters*. 2001;496(2-3):152-60.
27. Liu F, Zaidi T, Iqbal K, Grundke-Iqbal I, Merkle RK, Gong C-X. Role of glycosylation in hyperphosphorylation of tau in Alzheimer's disease. *FEBS letters*. 2002;512(1-3):101-6.
28. Haukedal H, Freude KK. Implications of glycosylation in Alzheimer's disease. *Front Neurosci*. 2021;14:1432.
29. Aebi M. N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2013;1833(11):2430-7.
30. Wang J-Z, Grundke-Iqbal I, Iqbal K. Glycosylation of microtubule-associated protein tau: An abnormal posttranslational modification in Alzheimer's disease. *Nature medicine*. 1996;2(8):871-5.

31. Losev Y, Frenkel-Pinter M, Abu-Hussien M, Viswanathan GK, Elyashiv-Revivo D, Gerles R, et al. Differential effects of putative N-glycosylation sites in human Tau on Alzheimer's disease-related neurodegeneration. *Cellular and Molecular Life Sciences*. 2021;78(5):2231-45.
32. Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC*. 2015;8:37.
33. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biology*. 2011;9(1):71.
34. Barbier P, Zejneli O, Martinho M, Lasorsa A, Belle V, Smet-Nocca C, et al. Role of tau as a microtubule-associated protein: structural and functional aspects. *Frontiers in Aging Neuroscience*. 2019;11:204.
35. Fitzpatrick AW, Falcon B, He S, Murzin AG, Murshudov G, Garringer HJ, et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature*. 2017;547(7662):185-90.
36. Wang JZ, Xia YY, Grundke-Iqbal I, Iqbal K. Abnormal hyperphosphorylation of tau: sites, regulation, and molecular mechanism of neurofibrillary degeneration. *Journal of Alzheimer's disease : JAD*. 2013;33 Suppl 1:S123-39.
37. Sneha P, George Priya Doss C. Chapter Seven - Molecular Dynamics: New Frontier in Personalized Medicine. In: Donev R, editor. *Advances in Protein Chemistry and Structural Biology*. 102: Academic Press; 2016. p. 181-224.
38. Prigozhin MB, Scott GE, Denos S. Mechanical modeling and computer simulation of protein folding. *Journal of Chemical Education*. 2014;91(11):1939-42.
39. Sandoval C. *Molecular Dynamics Simulation of Synthetic Polymers*. Molecular Dynamics-Studies of Synthetic and Biological Macromolecules: IntechOpen; 2012.
40. Ausaf Ali S, Hassan I, Islam A, Ahmad F. A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. *Current Protein and Peptide Science*. 2014;15(5):456-76.
41. David CC, Jacobs DJ. *Principal component analysis: a method for determining the essential dynamics of proteins*. Protein dynamics: Springer; 2014. p. 193-226.
42. Tramontano A, Cozzetto D. *The relationship between protein sequence, structure and function*. Supramolecular Structure and Function 8: Springer; 2005. p. 15-29.
43. Valastyan JS, Lindquist S. Mechanisms of protein-folding diseases at a glance. *Dis Model Mech*. 2014;7(1):9-14.
44. Soeda Y, Takashima A. New insights into drug discovery targeting tau protein. *Frontiers in Molecular Neuroscience*. 2020;13.
45. von Bergen M, Barghorn S, Biernat J, Mandelkow E-M, Mandelkow E. Tau aggregation is driven by a transition from random coil to beta sheet structure. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2005;1739(2):158-66.
46. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577-637.
47. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry*. 1997;48(1):545-600.
48. Mallamace F, Corsaro C, Mallamace D, Vasi S, Vasi C, Baglioni P, et al. Energy landscape in protein folding and unfolding. *Proceedings of the National Academy of Sciences*. 2016;113(12):3159-63.

49. Gianni S, Freiburger MI, Jemth P, Ferreiro DU, Wolynes PG, Fuxreiter M. Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Accounts of chemical research*. 2021;54(5):1251-9.
50. Chong S-H, Ham S. Folding free energy landscape of ordered and intrinsically disordered proteins. *Scientific reports*. 2019;9(1):1-9.
51. Strodel B. Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins. *Journal of Molecular Biology*. 2021:167182.
52. Jong K, Grisanti L, Hassanali A. Hydrogen Bond Networks and Hydrophobic Effects in the Amyloid  $\beta$ 30–35 Chain in Water: A Molecular Dynamics Study. *Journal of Chemical Information and Modeling*. 2017;57(7):1548-62.
53. Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A. Protein contact networks: an emerging paradigm in chemistry. *Chemical reviews*. 2013;113(3):1598-613.
54. O'Rourke KF, Gorman SD, Boehr DD. Biophysical and computational methods to analyze amino acid interaction networks in proteins. *Computational and structural biotechnology journal*. 2016;14:245-51.
55. Brysbaert G, Lensink M. Centrality Measures in Residue Interaction Networks to Highlight Amino Acids in Protein-Protein Binding. 2021.
56. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. *Proceedings of the National Academy of Sciences*. 2002;99(13):8637-41.
57. Soundararajan V, Raman R, Raguram S, Sasisekharan V, Sasisekharan R. Atomic interaction networks in the core of protein domains and their native folds. *PLoS One*. 2010;5(2):e9391.
58. Süel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature structural biology*. 2003;10(1):59-69.
59. Lopes TJS, Rios R, Nogueira T, Mello RF. Protein residue network analysis reveals fundamental properties of the human coagulation factor VIII. *Scientific Reports*. 2021;11(1):12625.
60. Yadav M, Igarashi M, Yamamoto N. Dynamic residue interaction network analysis of the oseltamivir binding site of N1 neuraminidase and its H274Y mutation site conferring drug resistance in influenza A virus. *PeerJ*. 2021;9:e11552.
61. Oldham S, Fulcher B, Parkes L, Arnatkevic Iūtė A, Suo C, Fornito A. Consistency and differences between centrality measures across distinct classes of networks. *PLoS One*. 2019;14(7):e0220061.
62. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol*. 2007;3(4):e59.
63. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. *Physical Review E*. 2002;65(6):061910.
64. Karain WI, Qaraeen NI. Weighted protein residue networks based on joint recurrences between residues. *BMC Bioinformatics*. 2015;16(1):173.
65. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. *Proceedings of the National Academy of Sciences*. 2002;99(13):8637.
66. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanel D, Venger I, et al. Network analysis of protein structures identifies functional residues. *J Mol Biol*. 2004;344(4):1135-46.
67. Fitzpatrick AWP, Falcon B, He S, Murzin AG, Murshudov G, Garringer HJ, et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature*. 2017;547(7662):185-90.
68. Sato Y, Naito Y, Grundke-Iqbal I, Iqbal K, Endo T. Analysis of N-glycans of pathological tau: possible occurrence of aberrant processing of tau in Alzheimer's disease. *FEBS Lett*. 2001;496(2-3):152-60.

69. Zhang W, Tarutani A, Newell KL, Murzin AG, Matsubara T, Falcon B, et al. Novel tau filament fold in corticobasal degeneration. *Nature*. 2020;580(7802):283-7.
70. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*. 2008;29(11):1859-65.
71. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*. 2017;14(1):71-3.
72. Bonate PL. A Brief Introduction to Monte Carlo Simulation. *Clinical Pharmacokinetics*. 2001;40(1):15-22.
73. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry*. 1997;18(12):1463-72.
74. Sidler D, Riniker S. Fast Nosé–Hoover thermostat: molecular dynamics in quasi-thermodynamic equilibrium. *Physical Chemistry Chemical Physics*. 2019;21(11):6059-70.
75. Parrinello M, Rahman A. Strain fluctuations and elastic constants. *The Journal of Chemical Physics*. 1982;76(5):2662-6.
76. Harvey MJ, De Fabritiis G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *Journal of chemical theory and computation*. 2009;5(9):2371-7.
77. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *Journal of computational chemistry*. 2005;26(16):1701-18.
78. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25.
79. Maiorov VN, Crippen GM. Size-independent comparison of protein three-dimensional structures. *Proteins*. 1995;22(3):273-83.
80. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996;14(1):33-8, 27-8.
81. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*. 2004;25(13):1605-12.
82. Morris JH, Huang CC, Babbitt PC, Ferrin TE. structureViz: linking Cytoscape and UCSF Chimera. *Bioinformatics (Oxford, England)*. 2007;23(17):2345-7.
83. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498-504.
84. Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Bio Systems*. 2015;127:67-72.