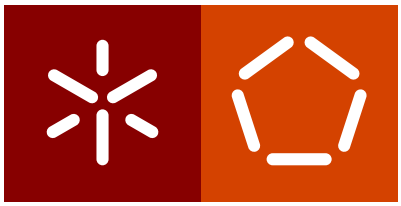
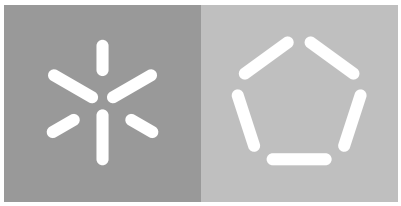


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Tiago Manuel Rocha Ferreira

Acetylation's role in Tau structure, electrostatics and dynamics

December 2020



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Tiago Manuel Rocha Ferreira

Acetylation's role in Tau structure, electrostatics and dynamics

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Artur Cavaco-Paulo
Tarsila Gabriel Castro

December 2020

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

AGRADECIMENTOS

Após um ano árduo de trabalho, algumas frustrações e acima de tudo aprendizagem, queria em primeiro lugar abrir esta secção com o meu voto de apreço e um agradecimento em especial à Prof. Dra. Tarsila Gabriel Castro, um pilar importantíssimo neste trabalho, que mais que uma orientadora, serviu também de professora, e com toda a simpatia garantiu com que durante esta jornada nunca me tivesse sentido sozinho/abandonado numa área em que tinha muito pouca, ou nenhuma experiência prévia. A sua disponibilidade e o modo como encara a profissão e o ensino, são e serão sempre referências de disciplina de trabalho e empenho.

Também deixo o meu profundo agradecimento ao Prof. Dr. Artur Cavaco-Paulo, pela oportunidade, pela orientação, por ter ajudado a proporcionar um ambiente de trabalho propício a que tudo o que esteja neste documento, se pudesse concretizar com a melhor qualidade possível, e por tudo o que fez e tem feito por este projeto.

Sem descurar do grupo, deixo também um pequeno espacinho neste documento, para os felicitar pela sua simpatia, e agradecer por me terem acolhido, naquela salinha, todos aqueles dias e aquelas horas! Os vossos sorrisos e pequenas conversas, aliados a alguns almoços/lanches foram um ótimo ingrediente para tornar a minha vida mais fácil. Caso este pedaço de papel fosse um copo de champanhe, saudava com um brinde! Ao BBRG!

A realização desta dissertação tem um peso fantástico também, de muitos fatores externos, que não me querendo alongar muito, sinto que lhes devo muito, pois a amizade, companheirismo e aquele um ou outro insulto típicos de amigos que se adoram imenso, fizeram-me muitas vezes descontrair e descomprimir dum documento, duma proteína que alugaram um espacinho cá no cérebro durante 1 ano. A vocês, os meus "bros", uma francesinha (vocês pagam).

E como chave de ouro deixo o agradecimento a quem mais me ama, e a quem eu mais devo, e devolvo esse amor: Em primeiro aos meus pais, sem eles, sem o apoio moral, financeiro, logístico e todos os etceteras nada disto estava aqui hoje concretizado, e a eles tenho muito a dever. À minha irmã - ter uma futura cientista em casa é também uma ajuda, poder discutir certos temas e ter alguém com quem falar de coisas não muito banais é bastante refrescante.

Por último, e em edição especial, à Bruna! Que é o meu refúgio há já uns anos, que tratas de me tranquilizar, de me motivar e de acima de tudo, recusares-te a duvidar de mim! Agradeço as tardes de trabalho mútuo, em que me impedes (ou tentas impedir) com tudo o que podes, de procrastinar, de tão facilmente me distrair! De mil e uma vezes que me dizes que tenho de dormir bem, que tenho de espairecer, que isto ou que aquilo. Acima de tudo obrigado por seres o meu grilo falante, por seres um bocadinho do meu cérebro e por muitas vezes fingires ouvir ou tentares perceber o que estou a dizer sobre o meu trabalho, o teu input nesta tese é de uma importância tremenda, e agradeço-te por isso!

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.
I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ABSTRACT

Tau is a Microtubule (MT) associated protein known to malfunction when it undergoes post-translational modifications. The present work implements molecular modelling techniques in order to assess the effects of Lys acetylation on Tau's overall conformation, electrostatics and interactions. Through a thorough assessment of reported harmful Lys sites, 8 analog Tau replicates were generated by mutation of those sites by their acetylated form (aTau). These replicates were evaluated in intracellular fluid and next to a MT model. Our findings demonstrated that the lack of positive Lys charges, through mutation, generate sufficient electrostatic changes in order to acquiesce very different final conformations, in comparison to native Tau. These results are consistent, even when a single site is altered. In both scenarios (Tau in intracellular fluid and near a MT), aTau is frequently folded over itself, although not in the "paperclip" conformation, as expected. Moreover, the acetylation process hinders the microtubule binding region (MTBR), thus undermining the association to the MT tubulins. This result is in agreement with the fact that these reported Lys positions have been found acetylated in Tau's deposits in Alzheimer's disease brains. Therefore, our *in silico* study elucidates about important molecular events behind multifactorial dementias, as Alzheimer's.

Keywords: tau, molecular dynamics simulation, acetylation, Alzheimer's disease.

RESUMO

Tau é uma proteína associada a microtúbulos, conhecida por ter a sua função normal prejudicada por eventos de modificações pós translacionais. O presente trabalho implementa técnicas de modelação molecular com o objetivo de avaliar os efeitos da acetilação de locais Lys, na conformação, eletroestática e interações gerais em Tau. Através de uma procura cuidada de locais Lys reportados como prejudiciais, foram gerados 8 réplicas análogas de Tau, através de mutação desses locais Lys, pela sua forma acetilada (aTau). Estas réplicas foram avaliadas em fluído intracelular e perto de um modelo de microtúbulo. Os resultados revelam que a falta de cargas Lys positivas, devido à mutação, gera alterações electroestáticas suficientes para gerar uma conformação final bastante diferente, comparando com Tau nativa. Estes resultados são consistentes, mesmo quando apenas um local é alterado. Em ambos os cenários (Tau em fluido intracelular e perto do MT), aTau apresenta-se frequentemente dobrada sobre si mesma, no entanto não na formação em "paperclip" como seria esperado. Mais ainda, o processo de acetilação prejudica a região de ligação do microtúbulo (MTBR), e portanto debilitando a sua associação às tubulinas do MT. Estes resultados estão em concordância com o facto de que estas posições Lys foram reportadas como acetiladas em depósitos de Tau em cérebros com doença de Alzheimer, portanto, o nosso estudo *in silico* elucida acerca de processos moleculares importantes em demências multifatorias, tais como Alzheimer.

Palavras-chave: tau, simulações de dinâmica molecular, acetilação, doença de Alzheimer.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AD	Alzheimer's disease
AI	Artificial Intelligence
aLys	Acetylated lysine
aTau	Acetylated tau
ATB	Automated Topology Builder
ATB	Automated Topology Builder
CG	Conjugate Gradient
CPU	Central Processing Unit
CTE	chronic traumatic encephalopathy
DI	Departamento de Informática
DSSP	Dictionary of Secondary Structure in Proteins
FF	force fields
FTD	frontotemporal dementia
GPU	Graphics Processing Unit
GROMACS	Groningen MAchine for Chemical Simulation
GSMM	genome scale metabolic model
IDP	intrinsically disordered protein
MAP	microtubule associated protein
MBIOINF	Mestrado em Bioinformática
MD	molecular dynamics
MM	Molecular Mechanics
MT	Microtubule
MTBR	microtubule binding region

Acronyms

NFTs	neurofibrillary tangles
ns	nanoseconds
PBC	Periodic Boundary Conditions
PHF	paired helical filaments
PME	particle-mesh Ewald
PSP	Protein Structure Predictions
pTau	phosphorylated tau
PTM	post translational modifications
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SD	Steepest Descent
SS	secondary structure
UM	Universidade do Minho

CHAPTER 1

Introduction

1.1 Motivation

Tau is a microtubule associated protein (MAP), usually present in the neuron's axonal region, to stabilise microtubules (MT), but not exclusively, as it is also present in the nucleolus, distal ends of growing neurons, oligodendrocytes and muscle [1, 2]. *In vitro* results have proven Tau to be an important protein in mammalian brains, promoting MT assembly and stability [3]. In abnormal conditions, Tau presents itself attached to MTs in a filamentous form, causing Tau-dependent diseases called Tauopathies. There are 6 known Tau isoforms: four of them have 4 MT-binding repeat amino acid sequences (4R Tau), and the other 3 lack the last repeat (3R Tau) [4]. The main isoform present in the adult human brain has 441 amino acids, two amino-terminal inserts, a proline rich region and the MTBR. This isoform is most commonly cited as 2N4RTau [5] and will be the main object of the following study.

Tau is an intrinsically disordered protein (IDP), however in certain Tauopathies such as Alzheimer's disease (AD), Dementia and Parkinson's Disease, Tau has been shown to aggregate into amorphous structures, called neurofibrillary tangles (NFTs) [6]. These structures are widely accepted to be the result of post-translational modifications, mainly hyperphosphorylation, creating these aberrant neural fibrils [7]. A novel MD study has shown that Tau has an elongated structure when bound to MT, with small patches presenting secondary structure (SS), and a paper-clip arrangement when in solution, agreeing with previous experimental evidences [5].

After translational processes complete, proteins are subjected to more transformations, as some molecules can be added to their structure and alter their overall conformation and function [8]. In Tau, some very important post-translational modifications occur, being phosphorylation the most widely cited and subsequently studied [6, 9–12]. A low degree of phosphorylation is usual and does not interfere with Tau optimal function. Only a hyperphosphorylation process is able to inhibit Tau-MT association [13]. However, other important modifications such as glycosylation and acetylation also occur during AD pathogenesis [12].

Acetylation is a modification that occurs in lysine residues, adding an acyl group to the NH moiety on the side chain [14] (fig. 1). This modification will alter Tau's electrostatics and thus the folding properties near that specific site [15], as neutralise the lysines, which are positive charged at physiological pH conditions. When this process wrongfully occurs, it can then affect the protein's interactions within itself, or by consequence, with other proteins or binding sites, as is the case in neurodegenerative diseases such as Huntington's [14].

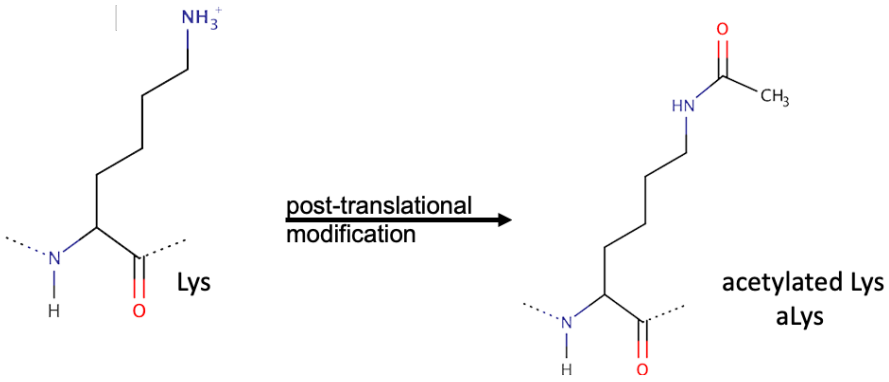


Figure 1: 2D representation of a lysine in its natural form, felting a physiological pH, and an acetylated lysine, after post-translational modification.

Molecular Modelling techniques, such as homology modelling or protein structure prediction (PSP) algorithms came as the *in silico* solution towards predicting structural conformation of molecules, batches of molecules or even larger, more complex structures such as proteins in solution, membrane embedded proteins or even nucleosomes and ribosomes [16]. MD simulations are able to stabilise and validate these conformational predictions yet their performance depend highly on empirical parameters for potential energy functions, or force fields (FF). FF parameters are calibrated taking into account well-structured proteins, due to the experimental amount of data and characterization available for this group of proteins [17]. Therefore, MD simulations that shed light into IDPs' structure have to be conducted carefully; exhaustively during many nanoseconds (ns), to guarantee a better sampling of conformational states. Aditionally, robust parameterisation of post-translationally modified amino acids must be made in order to guarantee these modifications stay true to experimental results.

1.2 Objectives

Depending on the specific sites, Acetylation of Tau has been proven to be a determining factor on MT-binding and Tau aggregation [18–21]. As such, this work will focus on the modifications of specific Lysines and subsequent structural and electrostatic analysis, according to an MD algorithm, provided by the GROMACS tool.

This work aims to understand and conclude on the effects of post translational modifications (PTM) by lysine acetylation on Tau, and it's impacts on MT-binding and possible formation of NFTs typical to those found on Alzheimer's Disease. These results will hopefully give rise to another setting stone to understand the molecular mechanisms behind these diseases, possibly given new detailed information which might contibute for their treatment.

More specifically, this work will address the following objectives:

- Review relevant literature about Tau protein and its modifications.
- Review relevant literature about MD techniques.
- Developing parametrization for acetylated Lys, in the scope of GROMOS FF.
- Learn how to work with PyMOL to generate aTau analogs and to analyse simulations' trajectories.
- Learn how to use GROMACS to perform MD simulations planned for aTau systems.
- Writing master thesis and a scientific publication with the main results.

References

[1] Ittner, L. M., & Götz, J. (2011). Amyloid- and tau — a toxic pas de deux in Alzheimer's disease. *Nature Reviews Neuroscience*, *12*(2), 67–72.

[2] Himmler, A., Drechsel, D., Kirschner, M. W., & Martin, D. W. (1989). Tau consists of a set of proteins with repeated C-terminal microtubule-binding domains and variable N-terminal domains. *Molecular and cellular biology*, *9*(4), 1381–8.

[3] Drubin, D., Kobayashi, S., & Kirschner, M. (1986). Association of Tau Protein with Microtubules in Living Cells. *Annals of the New York Academy of Sciences*, *466*(1 Dynamic Aspec), 257–268.

[4] Sato, C., Barthelemy, N. R., Mawuenyega, K. G., Patterson, B. W., Gordon, B. A., Jockel-Balsarotti, J., Sullivan, M., Crisp, M. J., Kasten, T., Kirmess, K. M., Kanaan, N. M., Yarasheski, K. E., Baker-Nigh, A., Benzinger, T. L. S., Miller, T. M., Karch, C. M., & Bateman, R. J. (2018). Tau Kinetics in Neurons and the Human Central Nervous System [WOS:000428235400014]. *Neuron*, *97*(6), 1284–+.

[5] Castro, T., Munteanu, F.-D., & Cavaco-Paulo, A. (2019). Electrostatics of Tau Protein by Molecular Dynamics. *Biomolecules*, *9*(3), 116.

[6] Lee, V. M.-Y., Goedert, M., & Trojanowski, J. Q. (2001). Neurodegenerative Tauopathies. *Annual Review of Neuroscience*, *24*(1), 1121–1159.

[7] Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's Disease. *New England Journal of Medicine*, *362*(4), 329–344.

[8] Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, *21*(3), 255–261.

[9] Hoover, B. R., Reed, M. N., Su, J., Penrod, R. D., Kotilinek, L. A., Grant, M. K., Pitstick, R., Carlson, G. A., Lanier, L. M., Yuan, L. L., Ashe, K. H., & Liao, D. (2010). Tau Mislocalization to Dendritic Spines Mediates Synaptic Dysfunction Independently of Neurodegeneration. *Neuron*.

[10] Stoothoff, W. H., & Johnson, G. V. (2005). Tau phosphorylation: physiological and pathological consequences. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1739*(2), 280–297.

[11] Thies, E., & Mandelkow, E. M. (2007). Missorting of tau in neurons causes degeneration of synapses that can be rescued by the kinase MARK2/Par-1. *Journal of Neuroscience*.

[12] Gong, C.-X., Liu, F., Grundke-Iqbal, I., & Iqbal, K. (2005). Post-translational modifications of tau protein in Alzheimer's disease. *Journal of Neural Transmission*, *112*(6), 813–838.

[13] Buée, L., & Delacourte, A. (2001). Tau Phosphorylation. *Functional neurobiology of aging* (pp. 315–332). Elsevier.

[14] Drazic, A., Myklebust, L. M., Ree, R., & Arnesen, T. (2016). The world of protein acetylation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1864*(10), 1372–1401.

[15] Narita, T., Weinert, B. T., & Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. *Nature Reviews Molecular Cell Biology*, *20*(3), 156–174.

[16] Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: Advances and applications.

[17] Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules.

[18] Cohen, T. J., Guo, J. L., Hurtado, D. E., Kwong, L. K., Mills, I. P., Trojanowski, J. Q., & Lee, V. M. (2011). The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nature Communications*, *2*(1), 252–259.

[19] Cohen, T. J., Constance, B. H., Hwang, A. W., James, M., & Yuan, C.-X. (2016). Intrinsic Tau Acetylation Is Coupled to Auto-Proteolytic Tau Fragmentation (C.-X. Gong, Ed.). *PLOS ONE*, *11*(7), e0158470.

[20] Min, S.-W., Cho, S.-H., Zhou, Y., Schroeder, S., Haroutunian, V., Seeley, W. W., Huang, E. J., Shen, Y., Masliah, E., Mukherjee, C., Meyers, D., Cole, P. A., Ott, M., & Gan, L. (2010). Acetylation of Tau Inhibits Its Degradation and Contributes to Tauopathy. *Neuron*, *67*(6), 953–966.

[21] Min, S.-W., Chen, X., Tracy, T. E., Li, Y., Zhou, Y., Wang, C., Shirakawa, K., Minami, S. S., Defensor, E., Mok, S. A., Sohn, P. D., Schilling, B., Cong, X., Ellerby, L., Gibson, B. W., Johnson, J., Krogan, N., Shamloo, M., Gestwicki, J., … Gan, L. (2015). Critical role of acetylation in tau-mediated neurodegeneration and cognitive deficits. *Nature Medicine*, *21*(10), 1154–1162.

CHAPTER 2

State of the art

2.1 Microtubules and Tubulin: Structure and role in cellular function

Microtubules are hollow rod-shaped tubular structures, approximately 25nm in size that are found inside every eukaryotic cell. They are the second most important constituent of the cytoskeleton, preceded only by actin filaments [1]. In normal cells these structures function as scaffolding elements, helping to determine cell shape, some sorts of locomotion, organelle transport and also help to separate chromosomes during mitosis [1, 2].

Structurally, microtubules are composed by dimers of a globular protein, by the name of tubulin. Tubulin polypeptides α and β organise themselves in protofilaments assembled around a hollow core, forming the microtubule. These dimers can depolymerise and polymerise fairly quickly and thus microtubules are highly unstable [3]. This process of rapid growth followed by disassembly is called dynamic instability [2] (Figure 2).

Thus, this creates a need in which the instability of microtubules needs to be regulated. MAPs are a set of proteins important in this regulation, having roles from MT initiation, polymerases that help elongation and depolymerases that promote MT shrinkage, CLASP proteins for end protection and overall rescue of microtubule catastrophe and stabilisers such as MAPs 1, 2 and 4 and Tau. Many MAPs are cell-type specific, as is the case of Tau [1].

In nerve cells, MTs present themselves organised slightly different along the axon and dendrites. Normally MTs anchor around the centrosome, but as nerve cells are more dynamic and elongate in different areas, MTs are structurally required in the cell processes and regions characteristic of nerve cells: axons and dendrites [4]. Microtubules in these areas are different both in orientation and in associated MAPs. In dendrites, MTs present themselves oriented both with the plus-end towards the cell body and against it, having MAP2 attached to them. In axons, the associated peptides are Tau proteins and not MAP2, and the Microtubule plus-ends are oriented away from the cell body (similar to normal cells). The intricacy of axonal microtubules is that, instead of being attached to the centrosome, minus ends are capped and attached to the cytoplasm [5]. These different conformations in neural MTs seem to have an important role in their stable disposition along these cell structures [4, 5].

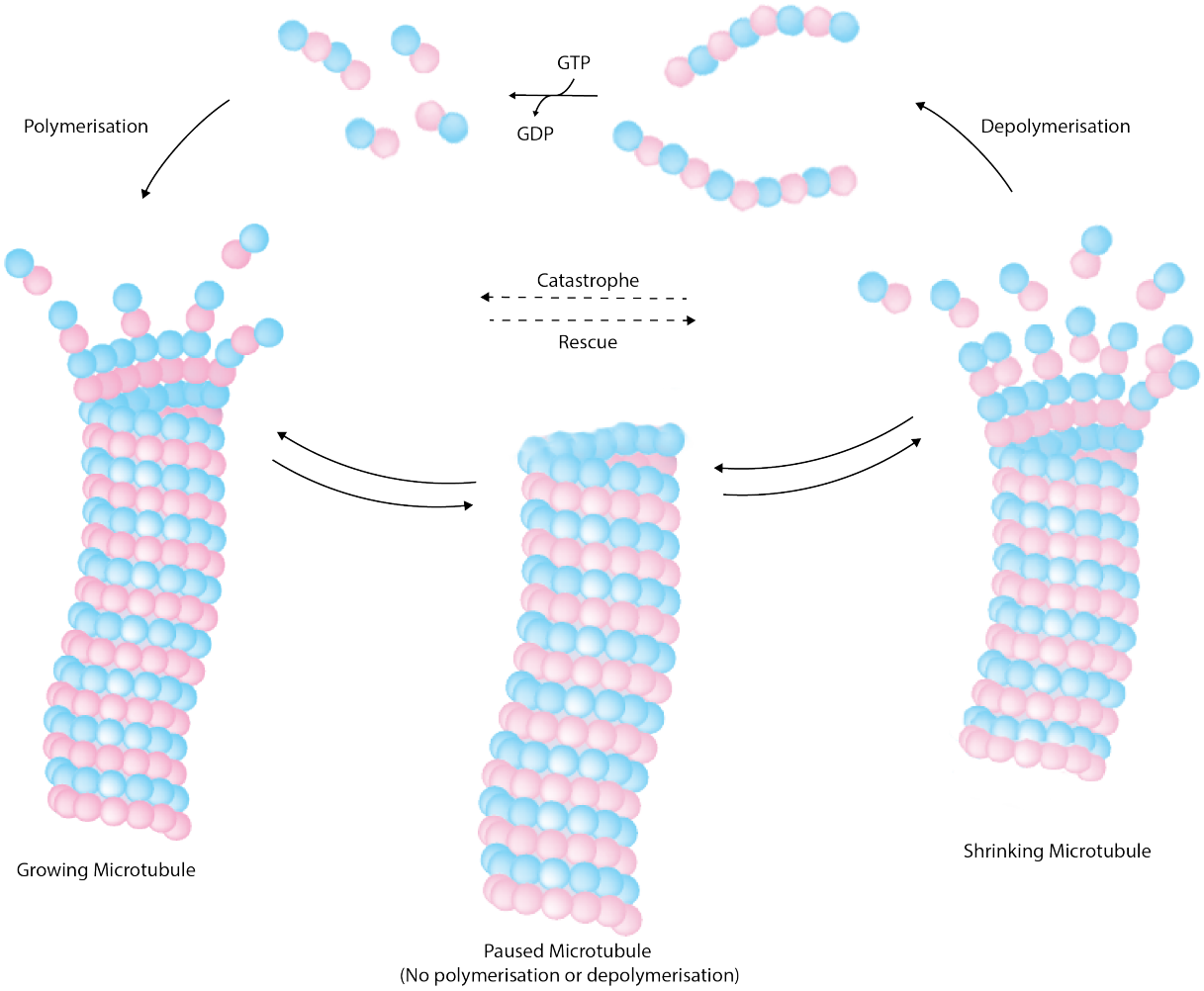


Figure 2: Mechanism underlying the dynamic instability of MT. α and β-tubulin heterodimers are represented in blue and pink, respectively and their assembly/disassembly is regulated by GTP. Adapted from Conde and colleagues (2009) [2].

2.2 Tau Protein

2.2.1 Tau: Characteristics and importance in mental health

In Alzheimer's disease, there are two main disease markers: neuritic plaques, and neurofibrillary lesions. [6, 7] These NFTs are a bundle of a mostly hyperphosphorylated filamentous protein called Tau. These proteins are heavily folded within themselves and therefore are called tangles.

Tau is a microtubule-associated protein, discovered in 1975 [8]. Through studies about Tau and its isoforms, it is now known that Tau is predominantly expressed mostly in mature and

growing neurons as well as astrocytes and oligodendrocytes. It is found widely in association with tubulin, a major component of MT.

Under normal circumstances, Tau appears in nerve cells either associated with microtubules or freely loose in the cytoplasm. Loose Tau is more commonly seen in a so called "paperclip" structure, where its C terminus rotates over the MTBR placing itself above it, whereas the N-terminus folds behind the latter, having both termini in close distance from each other (Figure 3) [9].

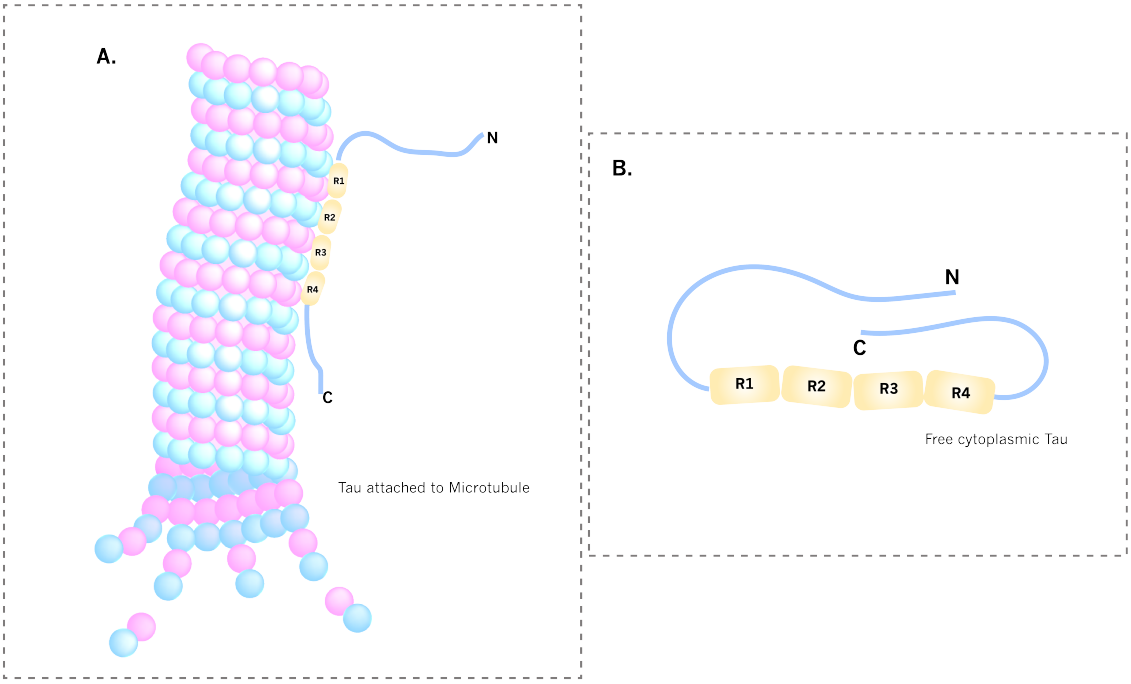


Figure 3: Tau structure and orientation both attached to microtubules and in free form in the cytoplasm. In **A**, Tau bonded to MT with the N-terminus facing away from the microtubule. In **B** is the illustration of Tau predicted conformation in the cytoplasm, folded within itself, with both termini over the MTBR.

2.2.2 Tau structure

Structurally, Tau is characterised as an IDP [8, 10, 11], meaning it has an undefined tertiary structure. However recent work has revealed Tau to possess different small chunks of what appears to be α-helix sections along the entire structure [11]. This protein encompasses 441

amino acids in its largest isoform, out of a total of 6 different ones, ranging from 352 to 441 amino acids with molecular masses ranging from 45 to 65 kDa [12]. Along the amino acidic structure, 4 different sections can be observed: a projection domain (encompassing the N-terminal), a proline-rich region, MTBR and the C-terminal (Figure 4). The different isoforms differ around the repeat number in the MTBR. Four of the isoforms have 4 MT-binding repeat amino acid sequences (4R Tau), whilst the remaining ones lack the last repeat (3R Tau) [13].

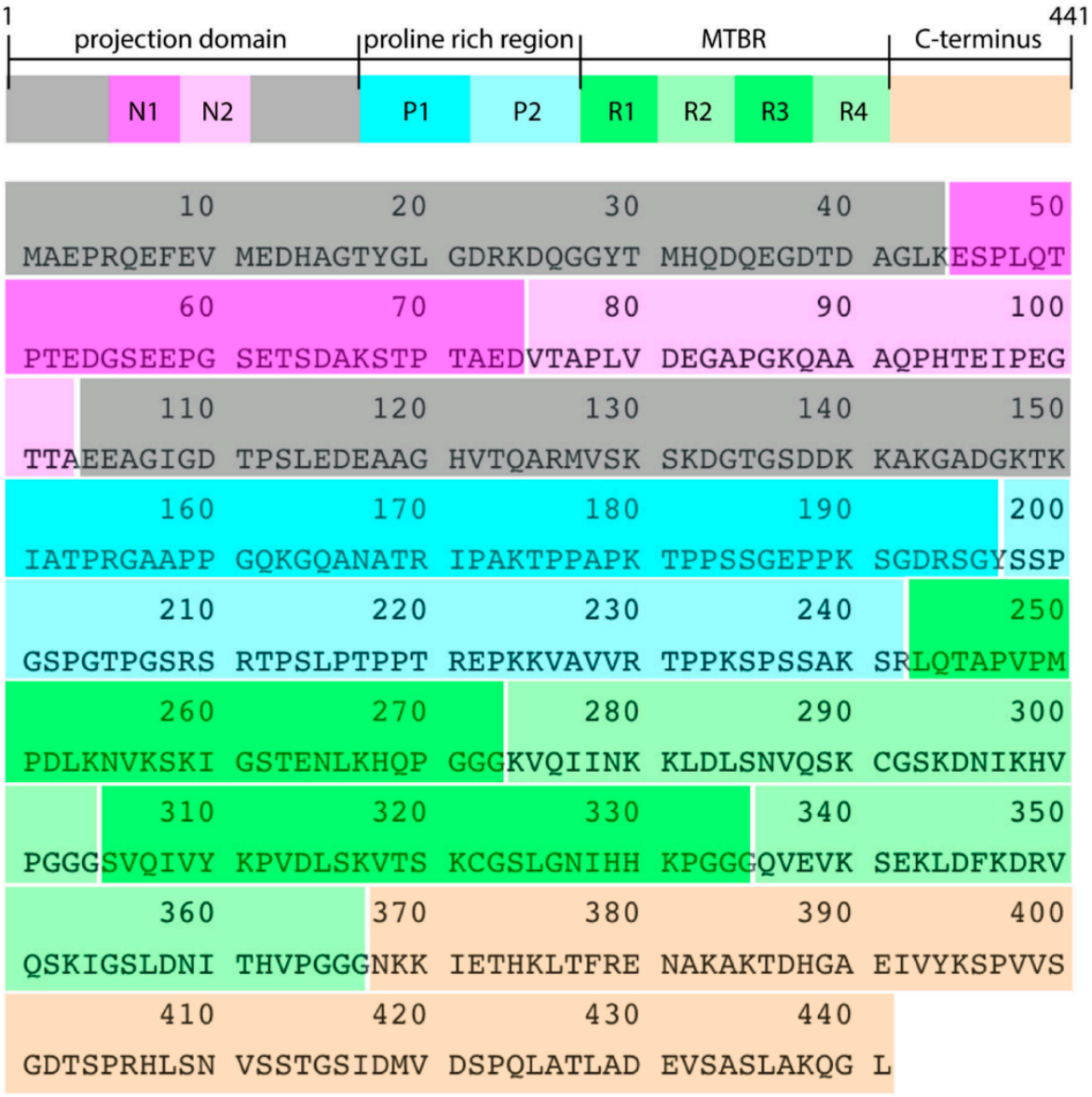


Figure 4: Tau aminoacid structure. A representation of the aminoacidic structure of 2N4RTau and the different regions in itself: P1 and P2 comprising the projection domains 1 and 2, respectively, N1 and N2 the proline rich region n-terminal inserts and the 4 different repeats (R1 to R4) of the MTBR. Adapted from Castro et al (2018) [11].

A study by Kanaan and co-workers (2011) demonstrated that a small region of Tau (amino acids) in proximity to the N-terminus in mutated forms of filamentous but not soluble Tau, would impact axonal transport, showing this region to be somewhat involved in a signaling cascade of axonal transport in neurons [14]. Another region that has a functional impact in nerve cell function is the proline-rich region. Several works point roles such as signaling modulators for Tau, interaction sites for nucleic acids as well as an important protein interaction site (typical of proline-rich regions in other proteins). But most importantly for neural function, proline-rich regions of Tau have been shown to be important in microtubular assembly regulation [15] and in actin (a cytoskeletal protein constituent of microfilaments [1]) binding [16]. Thus, Tau proline-rich region seems to be important in nerve cell motility and overall structure.

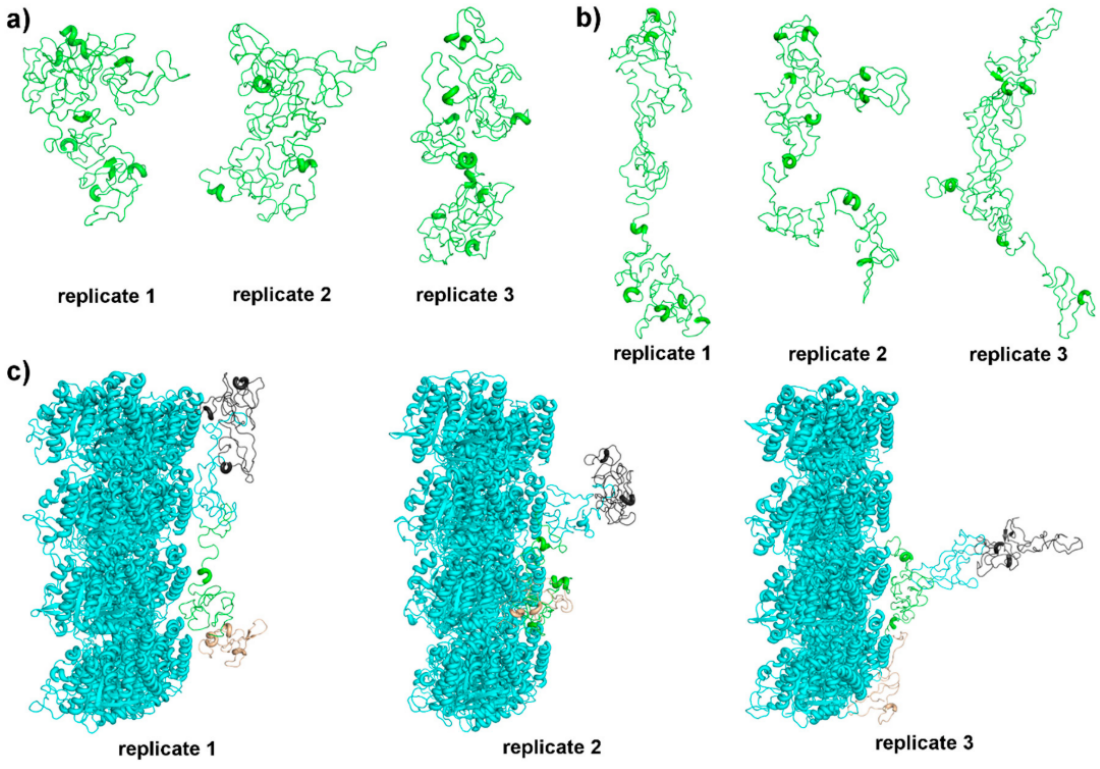


Figure 5: Tau's predicted 2N4R middle structures. Results are taken from the work of Castro et al (2018) [11] and represent **a)** three replicates of Tau simulated middle structures of Tau when associated with microtubules, with N-termini oriented to the top, and **b)** Tau::MT middle structures, with a color scheme where: grey is for N-terminal, cyan for proline-rich domain, green representing the MTBR, and beige for C-terminal. **(c)** Microtubule wall is represented in cyan cartoon and **(a,b)** Tau is represented in green.

Later, the first tau's full structure prediction and MD simulations has surged with the work of Castro and colleagues [11]. Much in the way of what is intended to do with the present work, through *in silico* methods for protein simulations, the overall electrostatics of Tau was analysed, and structure evaluations were set for both intracellular fluid and in interaction with the MT. This work came to confirm the previous remarks that Tau in fluid takes a paperclip-like structure, as mentioned above, and neat a MT this protein presents itself in a more extended form, binding through the MTBR (fig. 5).

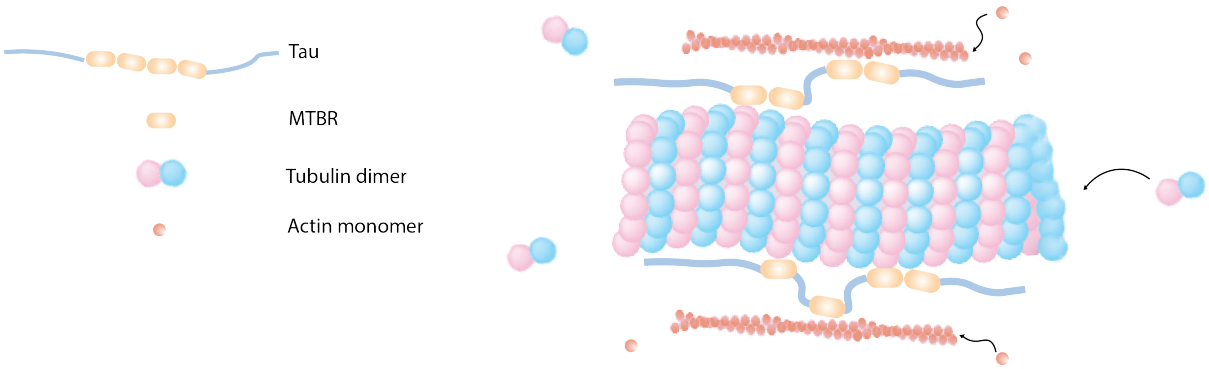


Figure 6: Model for Tau interaction mechanism for both actin filaments and microtubules. The cytoskeletal elements interact mutually in a system where both MT (blue and pink) and actin filaments (red) are bound to Tau, promoting their elongation and structural integrity. Adapted from Elie et al (2015) [17].

The MTBR is the most variable region amongst different Tau isoforms. Both 3R-Tau and 4R-Tau can coexist in the cell, but they are typically found in different structures as 4R-Tau has previously been studied in a brain region called the dentate gyrus and has not been detected [18], suggesting that different domains might be involved in different areas and thus, different physiological functions. It has also been proven that the 4R-Tau domain has a stronger bond to microtubules than 3R-Tau [19].

Additional to interactions with tubulin dimers and therefore microtubules, the MTBR is also reported to interact with actin filaments in a way that allows for both microtubule and filamentous actin to bind [17] (Figure 6). This tethering of both cytoskeletal elements can be a critical factor during neural development and the maintenance of healthy synapses.

2.2.3 Post translational modifications in Tau

Tau is subjected to a wide number of PTMs. From glycosylation to ubiquitination, glycation, nitration, phosphorylation and the subject of the present work, acetylation [20]. The majority of these modifications occur naturally as regulatory mechanisms act upon Tau. However, a cascade of events can cause neurological pathologies such as the ones discussed further in this work. The widest studied field, with numerous works around it, addresses the hyperphosphorylation process that occurs on Tau, misshaping the protein and causing deformities such as the paired helical filaments (PHF) that accumulate into proteic tangles (NFTs), affecting the neural cell's structural integrity and neurodegeneration [16, 21–25].

Phosphorylation

Tau has 85 putative phosphorylation sites in 2N4RTau, namely in serine (Ser), threonine (Thr) or tyrosine (Tyr) binding residues, even though more discreetily in Tyr sites rather than Ser or Thr. Although phosphorylation mostly occurs on these sites, a number of other phosphorylation sites have also been found in Tau [12]. Tau phosphorylation is required for a normal functioning of this protein. It is a phosphoprotein, in which around 2-3 moles of phosphates per molecule of Tau are found. This PTM is important for microtubule binding and stabilisation, hence it's an important regulation mechanism.

Whilst phosphorylated Tau is ideal for optimal function, failure in this regulatory mechanism often means dire effects for Tau activity and interaction with MTs and neuronal function. An immunocytochemical study by Bancher and coworkers in 1989 [26], found abnormally phosphorylated Tau in early stage NFTs, and later chemical work on Alzheimer's disease brains found not only an elevated number of hyperphosphorylated Tau in the cytoplasm of affected neurons, but also a large number of Tau sediment was encountered on AD brains, an uncommon feature in normal brains, suggesting an altered Tau biophysical state in AD [27]. These findings, along with several studies and literary reviews along the following decades helped to lay a path that connects hyperphosphorylation and consequent development of dense PHF, and the pathogenesis of AD [5, 28–33].

Glycosylation

Another quite common PTM in most proteins is glycosylation. It is characterised as an addition of an oligossacharide to a protein or lipid and as phosphorylation, when out of control, this mechanism can lead to a series of diseases [34]. There are two main types of glycosylation: N-glycosylation and O-glycosylation, either from the addition of sugar to the amino radical of Asparagine or by adding the oligossacharide to the hydroxyl terminal of either Serine or Threonine. The latter mechanism is seen as a protector against abnormal phosphorylation, since it acts along mostly the same residues. As such, it has been largely assumed that glycosylation acts as an antagonist to phosphorylation, protecting against hyperphosphorylation [20, 34].

A study on murine brains, under the premise that brain glucose uptake is impaired right before AD symptomatic onset, was conducted in order to evaluate how phosphorylation mechanisms evolve after fasting, in order to reduce glucose uptakes. Results demonstrated that most Tau was hyperphosphorylated on the target sites that normally should be the subject of O-glycosylation [35]. These results came to confirm that O-glycosylation in Tau can serve as a prevention mechanism against abnormal phosphorylation. However, Tau has also proven to be heavily glycosylated in AD and the precise relation between phosphorylation and glycosylation remains difficult to understand [36]. This *in vitro* study has also found Tau to reorganise from PHFs to straight filaments after enzymatic deglycolysation, suggesting that glycosylation has somewhat of a role on maintaining PHF structure.

Acetylation

The study of acetylation as a PTM of Tau, along with the aforementioned ones, is very important for the understanding of Tau's behavior under normal and pathological conditions. Contrary to phosphorylation and glycosylation, which combined share a wide number of sites, Tau acetylation occurs under 31 Lysine sites, a relatively small number which turns its investigation into a necessary ordeal [37].

Since lysine residues have the capacity to intervene in electrostatics and hydrophobic interactions [4] and are the binding site for acetyl groups onto proteins, research has been coming forward with some clarifications on the role of Tau lysine acetylation, and how the breakage in these mechanisms can affect protein structure.

Cook and colleagues [38] found, by means of a site-specific antibody that in AD brains, Tau was discovered to be hypoacetylated and thus vulnerale to pathogenic events. Under normal conditions, Tau acetylation is conducted under some enzymatic events, in order to ensure normal function, and since Tau is mainly unfolded, several modifications compete with each other. In the case of lysine acetylation, methylation and ubiquitination also present interaction with lysine residues [39]. This competitiveness ensures Tau to function properly, mainly in the KXGS motif, which regulates neurite extention and MT-binding, which presents competitive regulation by both phosphorylation and acetylation. Tau acetylation thus presents a complex mechanism, differently mediated in different sites of Tau and by different enzymes [38].

SIRT1 and p300 are two types of these enzymes, with completely oposing action. Whilst SIRT1 proteins enhance Tau aggregation, p300 inhibits it. This mechanism is best described by Min and his peers [40] when proving Tau to be acetylated and also promoting the formation of phosphorylated tau (pTau), used both these enzymes to respectively deacetylate and acetylate this protein. This work has also been important on finding a correlation between Tau acetylation and phosphorylation, and mark these events as potential hallmarks of neurodegeneration and AD onset and progression.

Another study on acetylation points towards impaired Tau function after finding insoluble Tau on cell cultures treated with heparin which induces fibrillization. aTau on some sites was found to fibrillize, unlike native samples of Tau, indicating a promoting effect on Tau fibrillization. This work also sheds light on how PHFs only form after acetylation, further indicating the pathogeinicity of these events. Potential harmful acetylation sites were identified as lysines K163, K280, K281 and K369 [41]. Other acetylation positions were reported and associated to pathogenic process on Tau, such as K174, making Tau insoluble and more disposed to aggregation [42, 43]. This position was also found acetylated in AD brains together

with K274 and K281[44]. A very recent review summarizes the principal acetylation sites related to tauopathies [45], in which indicates also the positions K180 and K240, as risky sites for aggregation and conformational abnormalities. These sites will pose a great deal of importance for further progression of the present work.

Earlier in 2019, Kim and his colleagues [46], through computational tools that help predict protein sequence and thus mimic the effects on real case studies, demonstrated Tau to be the subject of putative pathogenic acetylation on some sites. This work had in consideration previous literature regarding Tau acetylation and specific acetylation sites, and probable pathogenic targets were identified. This, besides further confirming the importance of Tau acetylation on the contribution to disease onset and progression, also shows the robustness and importance of *in silico* studies on biochemical research.

2.3 Computational Chemistry: Brief history and impact on scientific research

With the help of computational tools, the understanding of some biological processes and chemical reactions among living organisms has been increased. Important steps were achieved in these fields, serving as important tools in various areas, including drug development, genetic profiling, pathway analysis, structural prediction, simulation and many more [47]. *Nature* even released a special issue when the first Human genome scale metabolic model (GSMM) was first described [48], marking an important feat for bioinformatics tools and since then new tools have been surfacing, helping research on a wide number of areas in biological science.

Computational chemistry is widely defined by Lipkowitz and Boyd, in 1990 [47] as the aspects of chemical research that are expedited or rendered practical by computers. Otherwise, in the very same book, computational chemistry can also be described as quantitative modeling of chemical phenomena by computer-implemented techniques. Practically, computational chemistry takes advantage of computer algorithms and tools to process available experimental data, simulate, predict, design or otherwise modulate entire

molecules or systems of molecules [47]. Most importantly, these processes take advantage of findings in chemistry, physics and informatics to compute results from user inputs into the system.

2.3.1 Protein modeling's importance in research

As previously mentioned, the three-dimensional structure of a protein will be what ultimately defines its function and thus, research requires means to determine a certain protein's structure in order to study it or probable alterations of it [49]. Protein structure determination by means of laboratory work turns out to be a time consuming process, that consequently turns expensive. These data, aligned with the abundance of sequences both genetic and amino acidic present in modern databases largely exceeding the number of available structural data, creates a necessity for a cheap, quicker system for determining sequence and structure [50].

Computerised strategies for structure determination and molecular modelling came as the solution for this issue. For decades, the study of *in silico* (computer generated) methods has improved and novel tools became a constant in research.

In particular, the field of Protein Structure Predictions (PSP) is one of the biology's significant challenges, as allow to assign a 3D structure to proteins not resolved experimentally due to diverse factors. More recently, beyond the traditional homology modelling techniques, Artificial Intelligence (AI) has been used to determine a protein's 3D shape. The ability to accurately predict protein structures from their amino acid sequence is of extreme importance to life sciences and medicine, because it unveils molecular processes and features behind diseases [51].

2.4 Molecular Dynamics studies' impact and progress in computational chemistry

Another important subset of the computational chemistry field is Molecular Dynamics. MD is a field in development since the 1970's. These techniques are capable of representing the molecular interactions between all the system's components at an atomic scale [52]. Generally speaking, MD methods involve solving Newton's equations of motion in small time steps, based on cartesian coordinates of the particles and using a conservative FF [50].

The MD base algorithm, solves classical equations of motion for a given set of particles in a certain time period. The resulting output, called a trajectory can thus be read by further tools in computer simulation and the final result is able to be visualised and analysed [52, 53]. Following these lines of thought, investigators have released through the years, some simulation tools, namely AMBER [54], CHARMM [55], NAMD [56], OpenMM [57], LAMMPS [58], ESPResSo [59] and, finally, and the tool chosen for the present work, GROMACS [60, 61].

MD simulations take advantage of complex equations of forces acting on every atom in the system, or otherwise, FF. FF equations, though complex, are fairly easy to calculate, since these FFs represent molecular features such as springs for bond length and angles, periodic functions for bond rotations, Lennard–Jones potentials for Van deer Waals, and the Coulomb's law for electrostatic interactions [53]. This also assures that energy and force calculations can be a fast process, even in large systems. There is currently a wide range of force fields available for MD simulations, organised into FF families. A few examples are CHARMM, AMBER, Slipids, OPLS and GROMOS.

References

[1] Cooper, G. M., Hausman, R. E., & Hausman, R. E. (2000). *The cell: A molecular approach* (Vol. 10). ASM press Washington, DC.

[2] Conde, C., & Cáceres, A. (2009). Microtubule assembly, organization and dynamics in axons and dendrites. *Nature Reviews Neuroscience*, *10*(5), 319–332.

[3] Mitchison, T., & Kirschner, M. (1984). Dynamic instability of microtubule growth. *Nature*, *312*(5991), 237–242.

[4] Cook, C., Stankowski, J. N., Carlomagno, Y., Stetler, C., & Petrucelli, L. (2014). Acetylation: a new key to unlock tau's role in neurodegeneration. *Alzheimer's Research & Therapy*, *6*(3), 29.

[5] Iqbal, K., Del C. Alonso, A., Chen, S., Chohan, M. O., El-Akkad, E., Gong, C. X., Khatoon, S., Li, B., Liu, F., Rahman, A., Tanimukai, H., & Grundke-Iqbal, I. (2005). Tau pathology in Alzheimer disease and other tauopathies. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1739*(2), 198–210.

[6] Ittner, L. M., & Götz, J. (2011). Amyloid- and tau — a toxic pas de deux in Alzheimer's disease. *Nature Reviews Neuroscience*, *12*(2), 67–72.

[7] Kellogg, E. H., Hejab, N. M., Poepsel, S., Downing, K. H., DiMaio, F., & Nogales, E. (2018). Near-atomic model of microtubule-tau interactions. *Science*, *360*(6394), 1242–1246.

[8] Weingarten, M. D., Lockwood, A. H., Hwo, S. Y., & Kirschner, M. W. (1975). A protein factor essential for microtubule assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(5), 1858–1862.

[9] Jeganathan, S., Von Bergen, M., Brutlach, H., Steinhoff, H. J., & Mandelkow, E. (2006). Global hairpin folding of tau in solution. *Biochemistry*.

[10] Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., & Zweckstetter, M. (2017). Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature Communications*, *8*(1).

[11] Castro, T., Munteanu, F.-D., & Cavaco-Paulo, A. (2019). Electrostatics of Tau Protein by Molecular Dynamics. *Biomolecules*, *9*(3), 116.

[12] Buée, L., & Delacourte, A. (2001). Tau Phosphorylation. *Functional neurobiology of aging* (pp. 315–332). Elsevier.

[13] Sato, C., Barthelemy, N. R., Mawuenyega, K. G., Patterson, B. W., Gordon, B. A., Jockel-Balsarotti, J., Sullivan, M., Crisp, M. J., Kasten, T., Kirmess, K. M., Kanaan, N. M., Yarasheski, K. E., Baker-Nigh, A., Benzinger, T. L. S., Miller, T. M., Karch,

C. M., & Bateman, R. J. (2018). Tau Kinetics in Neurons and the Human Central Nervous System [WOS:000428235400014]. *Neuron*, *97*(6), 1284–+.

[14] Kanaan, N. M., Morfini, G. A., LaPointe, N. E., Pigino, G. F., Patterson, K. R., Song, Y., Andreadis, A., Fu, Y., Brady, S. T., & Binder, L. I. (2011). Pathogenic forms of tau inhibit kinesin-dependent axonal transport through a mechanism involving activation of axonal phosphotransferases. *Journal of Neuroscience*, *31*(27), 9858–9868.

[15] Eidenmüller, J., Fath, T., Maas, T., Pool, M., Sontag, E., & Brandt, R. (2001). Phosphorylation-mimicking glutamate clusters in the proline-rich region are sufficient to simulate the functional deficiencies of hyperphosphorylated tau protein. *Biochemical Journal*, *357*(3), 759–767.

[16] Gong, C. X., Lidsky, T., Wegiel, J., Zuck, L., Grundke-Iqbal, I., & Iqbal, K. (2000). Phosphorylation of microtubule-associated protein tau is regulated by protein phosphatase 2A in mammalian brain. Implications for neurofibrillary degeneration in Alzheimer's disease. *Journal of Biological Chemistry*, *275*(8), 5535–5544.

[17] Elie, A., Prezel, E., Guérin, C., Denarier, E., Ramirez-Rios, S., Serre, L., Andrieux, A., Fourest-Lieuvin, A., Blanchoin, L., & Arnal, I. (2015). Tau co-organizes dynamic microtubule and actin networks. *Scientific Reports*, *5*.

[18] Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J., & Crowther, R. A. (1989). Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *The EMBO Journal*, *8*(2), 393–399.

[19] Goode, B. L., Denis, P. E., Panda, D., Radeke, M. J., Miller, H. P., Wilson, L., & Feinstein, S. C. (1997). Functional interactions between the proline-rich and repeat regions of tau enhance microtubule binding and assembly. *Molecular Biology of the Cell*, *8*(2), 353–365.

[20] Gong, C.-X., Liu, F., Grundke-Iqbal, I., & Iqbal, K. (2005). Post-translational modifications of tau protein in Alzheimer's disease. *Journal of Neural Transmission*, *112*(6), 813–838.

[21] Lee, V. M.-Y., Goedert, M., & Trojanowski, J. Q. (2001). Neurodegenerative Tauopathies. *Annual Review of Neuroscience*, *24*(1), 1121–1159.

[22] Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's Disease. *New England Journal of Medicine*, *362*(4), 329–344.

[23] Götz, J., Chen, F., Van Dorpe, J., & Nitsch, R. M. (2001). Formation of neurofibrillary tangles in P301L tau transgenic mice induced by Aβ42 fibrils. *Science*, *293*(5534), 1491–1495.

[24] Council, M. R., Road, H., & Kingdom, U. (2001). Eurodegenerative auopathies. *Genetics*.

[25] Sharma, P., Srivastava, P., Seth, A., Tripathi, P. N., Banerjee, A. G., & Shrivastava, S. K. (2019). Comprehensive review of mechanisms of pathogenesis involved in Alzheimer's disease and potential therapeutic strategies. *Progress in Neurobiology*, *174*(September 2018), 53–89.

[26] Bancher, C., Brunner, C., Lassmann, H., Budka, H., Jellinger, K., Wiche, G., Seitelberger, F., Grundke-Iqbal, I., Iqbal, K., & Wisniewski, H. M. (1989). Accumulation of abnormally phosphorylated τ precedes the formation of neurofibrillary tangles in Alzheimer's disease. *Brain Research*, *477*(1-2), 90–99.

[27] Kopke, E., Tung, Y. C., Shaikh, S., Del Alonso, C. A., Iqbal, K., & Grundke-Iqbal, I. (1993). Microtubule-associated protein tau. Abnormal phosphorylation of a non- paired helical filament pool in Alzheimer disease. *Journal of Biological Chemistry*, *268*(32), 24374–24384.

[28] Gong, C.-X., & Iqbal, K. (2008). Hyperphosphorylation of Microtubule-Associated Protein Tau: A Promising Therapeutic Target for Alzheimer Disease. *Current Medicinal Chemistry*, *15*(23), 2321–2328.

[29] Duka, V., Lee, J. H., Credle, J., Wills, J., Oaks, A., Smolinsky, C., Shah, K., Mash, D. C., Masliah, E., & Sidhu, A. (2013). Identification of the Sites of Tau Hyperphosphorylation and Activation of Tau Kinases in Synucleinopathies and Alzheimer's Diseases. *PLoS ONE*, *8*(9), e75025.

[30] Ksiezak-Reding, H., Liu, W. K., & Yen, S. H. (1992). Phosphate analysis and dephosphorylation of modified tau associated with paired helical filaments. *Brain Research*, *597*(2), 209–219.

[31] Martin, L., Latypova, X., & Terro, F. (2011). Post-translational modifications of tau protein: Implications for Alzheimer's disease. *Neurochemistry International*, *58*(4), 458–471.

[32] Šimić, G., Babić Leko, M., Wray, S., Harrington, C., Delalle, I., Jovanov-Milošević, N., Bažadona, D., Buée, L., de Silva, R., Giovanni, G. D., Wischik, C., & Hof, P. R. (2016). Tau Protein Hyperphosphorylation and Aggregation in Alzheimer's Disease and Other Tauopathies, and Possible Neuroprotective Strategies. *Biomolecules*, *6*(1), 2–28.

[33] Stoothoff, W. H., & Johnson, G. V. (2005). Tau phosphorylation: physiological and pathological consequences. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1739*(2), 280–297.

[34] Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease.

[35] Li, X., Lu, F., Wang, J.-Z., & Gong, C.-X. (2006). Concurrent alterations of O-GlcNAcylation and phosphorylation of tau in mouse brains during fasting. *European Journal of Neuroscience*, *23*(8), 2078–2086.

[36] Wang, J. Z., Grundke-Iqbal, I., & Iqbal, K. (1996). Glycosylation of microtubule-associated protein tau: An abnormal posttranslational modification in Alzheimer's disease. *Nature Medicine*.

[37] Kontaxi, C., Piccardo, P., & Gill, A. C. (2017). Lysine-directed post-translational modifications of tau protein in Alzheimer's disease and related tauopathies.

[38] Cook, C., Carlomagno, Y., Gendron, T. F., Dunmore, J., Scheffel, K., Stetler, C., Davis, M., Dickson, D., Jarpe, M., DeTure, M., & Petrucelli, L. (2014). Acetylation of the KXGS motifs in tau is a critical determinant in modulation of tau aggregation and clearance. *Human Molecular Genetics*, *23*(1), 104–116.

[39] Morris, M., Knudsen, G. M., Maeda, S., Trinidad, J. C., Ioanoviciu, A., Blinkingame, A. L., & Mucke, L. (2015). Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nature Neuroscience*, *18*(8), 1183–1189.

[40] Min, S.-W., Cho, S.-H., Zhou, Y., Schroeder, S., Haroutunian, V., Seeley, W. W., Huang, E. J., Shen, Y., Masliah, E., Mukherjee, C., Meyers, D., Cole, P. A., Ott, M., & Gan, L. (2010). Acetylation of Tau Inhibits Its Degradation and Contributes to Tauopathy. *Neuron*, *67*(6), 953–966.

[41] Cohen, T. J., Guo, J. L., Hurtado, D. E., Kwong, L. K., Mills, I. P., Trojanowski, J. Q., & Lee, V. M. (2011). The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nature Communications*, *2*(1), 252–259.

[42] Drazic, A., Myklebust, L. M., Ree, R., & Arnesen, T. (2016). The world of protein acetylation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1864*(10), 1372–1401.

[43] Min, S.-W., Chen, X., Tracy, T. E., Li, Y., Zhou, Y., Wang, C., Shirakawa, K., Minami, S. S., Defensor, E., Mok, S. A., Sohn, P. D., Schilling, B., Cong, X., Ellerby, L., Gibson, B. W., Johnson, J., Krogan, N., Shamloo, M., Gestwicki, J., … Gan, L. (2015). Critical role of acetylation in tau-mediated neurodegeneration and cognitive deficits. *Nature Medicine*, *21*(10), 1154–1162.

[44] Tracy, T. E., Sohn, P. D., Minami, S. S., Wang, C., Min, S. W., Li, Y., Zhou, Y., Le, D., Lo, I., Ponnusamy, R., Cong, X., Schilling, B., Ellerby, L. M., Huganir, R. L., & Gan, L. (2016). Acetylated Tau Obstructs KIBRA-Mediated Signaling in Synaptic Plasticity and Promotes Tauopathy-Related Memory Loss. *Neuron*.

[45] Silva, M. C., & Haggarty, S. J. (2020). Tauopathies: Deciphering Disease Mechanisms to Develop Effective Therapies. *International Journal of Molecular Sciences*, *21*(23), 8948.

[46] Kim, Y.-C., & Jeong, B.-H. (2019). In Silico Evaluation of Acetylation Mimics in the 27 Lysine Residues of Human Tau Protein. *Current Alzheimer Research*, *16*(5), 379–387.

[47] Lipkowitz, K. B., & Boyd, D. B. (2007). *Reviews in Computational Chemistry*.

[48] Abbas, A. E., & Holmes, S. P. (2004). Bioinformatics and Management Science: Some Common Tools and Techniques. *Operations Research*, *52*(2), 165–190.

[49] Pitman, M. R., & Menz, R. I. (2006). Methods for protein homology modelling. *Applied mycology and biotechnology* (pp. 37–59). Elsevier.

[50] Leach, A. R. (2006). Ligand-based approaches: Core molecular modeling. *Comprehensive medicinal chemistry ii* (pp. 87–118). Elsevier Ltd.

[51] Callaway, E. (2020). 'it will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature*.

[52] Marrink, S. J., Corradi, V., Souza, P. C., Ingólfsson, H. I., Tieleman, D. P., & Sansom, M. S. (2019). Computational Modeling of Realistic Cell Membranes.

[53] Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: Advances and applications.

[54] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs.

[55] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., … Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*.

[56] Nelson, M. T., Humphrey, W., Gursoy, A., Dalke, A., Kale, L. V., Skeel, R. D., & Schulten, K. (1996). NAMD: A parallel, object-oriented molecular dynamics program. *International Journal of High Performance Computing Applications*.

[57] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*.

[58] Plimpton, S. (1995). LAMMPS.

[59] Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., Dal Corso, A., De Gironcoli, S., Fabris, S., Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., … Wentzcovitch, R. M. (2009). QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials. *Journal of Physics Condensed Matter*.

[60] Berendsen, H. J., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*.

[61] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: Fast, flexible, and free.

CHAPTER 3

Methods

3.1 Molecular Dynamics Simulations

The task of studying matter at a molecular level was a near impossible challenge until the development of modern computers. In 1957, Alder and Wainwright [1] paved the path for the birth of MD when they studied a system of rigid spheres. In this work, they have designed a relatively simple method that accurately calculates the simultaneous classical equations of motion for several hundred particles, using what they described back then as "fast electronic computers". Two years had gone by and the same pair of colleagues then presented a more complete review on these so called MD methods [2] with a few more practical examples and a more complete description of the model for MD. At this point, they had a computational barrier that limited them to a maximum number of 500 molecules. Down the line, more works in MD started to surge, namely Vineyard et al. [3] and Rahman's [4] works, being the pioneers studying materials and systems under continuous potentials.

Accompanied with the evolution of technology and subsequently, more modern computers, Rahman and Stilinger[5] simulated the behaviour of liquid water, marking in history the first study on a molecular liquid. In this work, a system comprising of 216 molecules of water was simulated with minor corrections, which further proved this dynamical model can realistically represent liquid water.

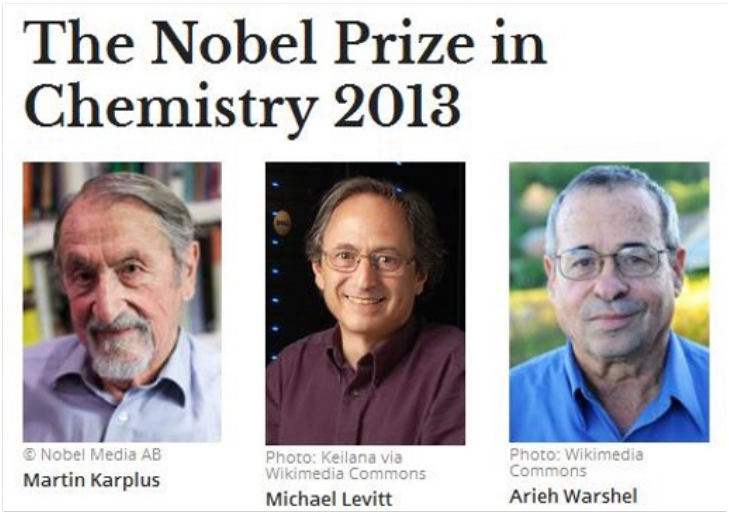


Figure 7: The developers of CHARMM (Molecular modelling Molecular Dynamics), nobel laureates in chemistry in 2013.

A number of other studies were carried throughout the subsequent years and new forms of biological models have surfaced since then, with a special mention to the work of Martin Karplus, Michael Levitt and Ariel Warshel who were recognised with a Nobel Prize in chemistry in 2013 for their efforts on the development of multiscale models for complex chemical systems[6] (Fig. 7). The development and research in this field has been an extremely important step for the fields of systems biology and biotechnology, namely their important contribution, in 1972, of powerful computational tools that help predict and study chemical processes[7].

This can largely be considered the work that perhaps pioneered the field of MD, since it steered away from the more common algorithms of molecular mechanics that were being used, and took advantage of modern technologies in order to fulfil some drawbacks, such as having no way of providing a quantitative value of energy per each particular structure[9]. A brief schematic of the history and some key events that took part in the last decades in computatinal chemistry and biology are illustrated above (Fig. 8).

The findings presented in this work depend largely on much more complex MD methods than the ones cited above. Along with the continuous growth in technology, the advent of clustered Central Processing Unit (CPU)s and increasingly better Graphics Processing Unit (GPU)s, the systems under simulation are now capable of more iterations of calculations, and thus larger, and more complex systems can be simulated[9–11].

In a quick interpretation of Leach's work[9], modern MD algorithms comprise of three basic principles: First and foremost, a molecular system must be described; Next comes an algorithm that better uses that molecular descriptor and better manipulates it against another measurable property of the system and/or the system itself; And lastly, the hardware/software and all the equipment used in the process, that must gather enough capacities to better combine the two latter factors.

These days, MD simulations are essential tools in geoengineering for their remarkable capacity to describe with great detail and accuracy, the dynamical properties of biological systems, somewhat better or at the very least much more quickly and cheaper than experimental methods[9, 11].

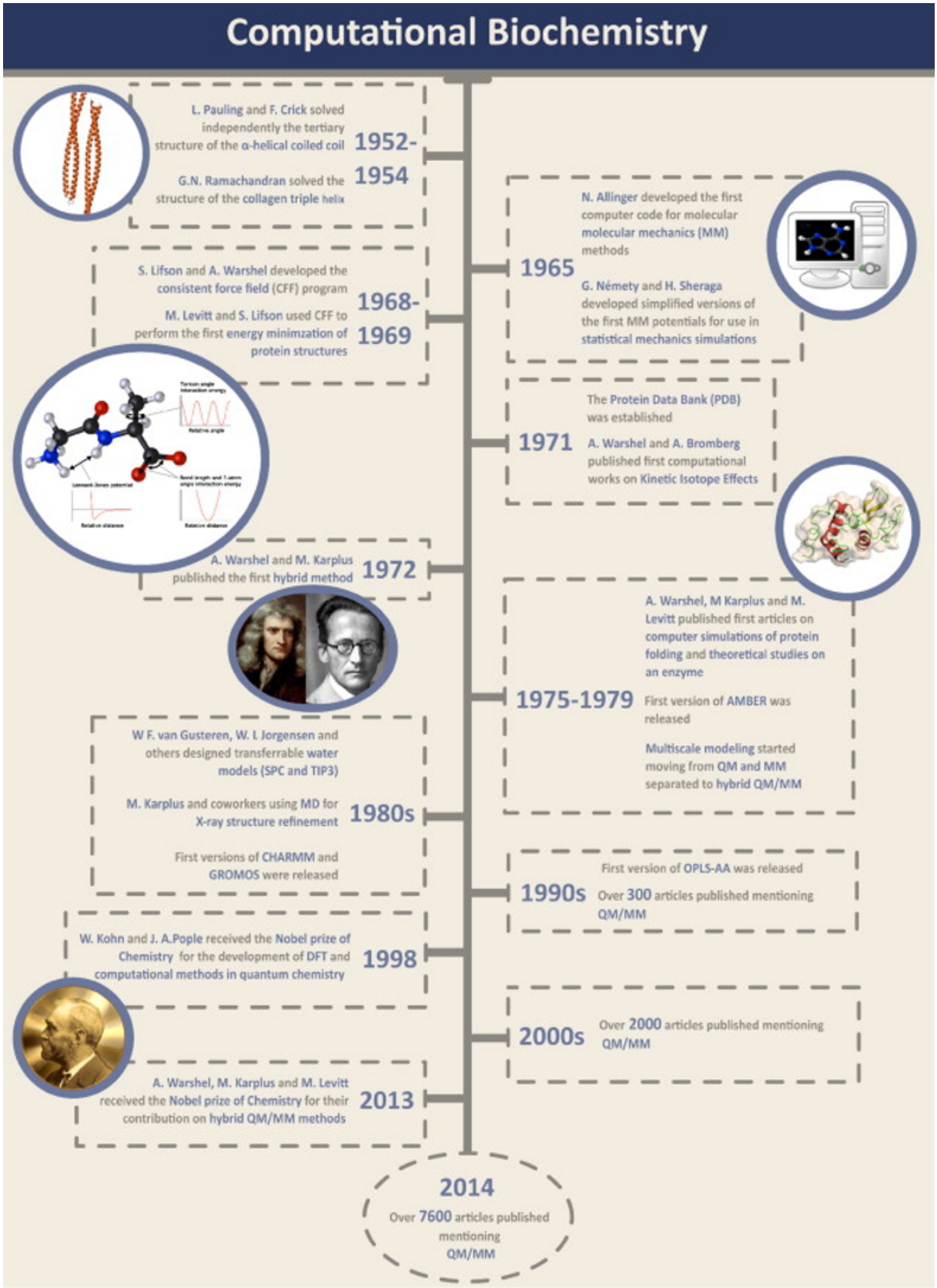


Figure 8: Illustrative timeline of some key events in computational chemistry and biology. Image extracted from Carvalho et al. (2014) [8]

3.2 Force Fields

A biomolecular FF is a set of parameters commonly used to calculate the potential energy of a system. As such, a FF has to describe a multitude of classes of molecules, as many as possible, with reasonable accuracy. And so it becomes of extreme importance to select an adequate FF in a project, for it will be the limiting factor and dictate the quality of the results, in case some parameters are missing [12].

Molecular Mechanics (MM) is a field that defines a set of widely accepted strategies to calculate molecular geometries and energies [13]. These methods consider the composition of a molecule to be a collection of masses interacting with each other via harmonic forces. This simplified way of approach to calculations is what ultimately gives its advantage regarding classical quantum mechanics, and funcions as an alternative approach to obtaining a FF. However, FF parameters are derived from a number of processes, from experimental data, to quantum mechanics in complex systems [9].

In a MM system, atom types are interpreted as rubber balls of different types and bonds come in the form of springs of varying length, so Hooke's law is used to assess the potential energy of the whole assembly, where bonds and angles describe harmonic potentials, torsional parameters describe how the energy changes when a bond rotates and the non-bonded parameters are describing the long-range electrostatic and van der Waals (vdW) forces between molecules or atoms that are linked to a distance of more than three covalent bonds [14]. These parameters are also schematically described below in figure 9 [9, 13, 15] according to the following equation:

$$E_{pot} = E_{bonds} + E_{angles} + E_{torsions} + E_{vdW} + E_{electrostatic} \tag{1}$$

$$m_j a_j = F_i = -\nabla_i E_{pot} \tag{2}$$

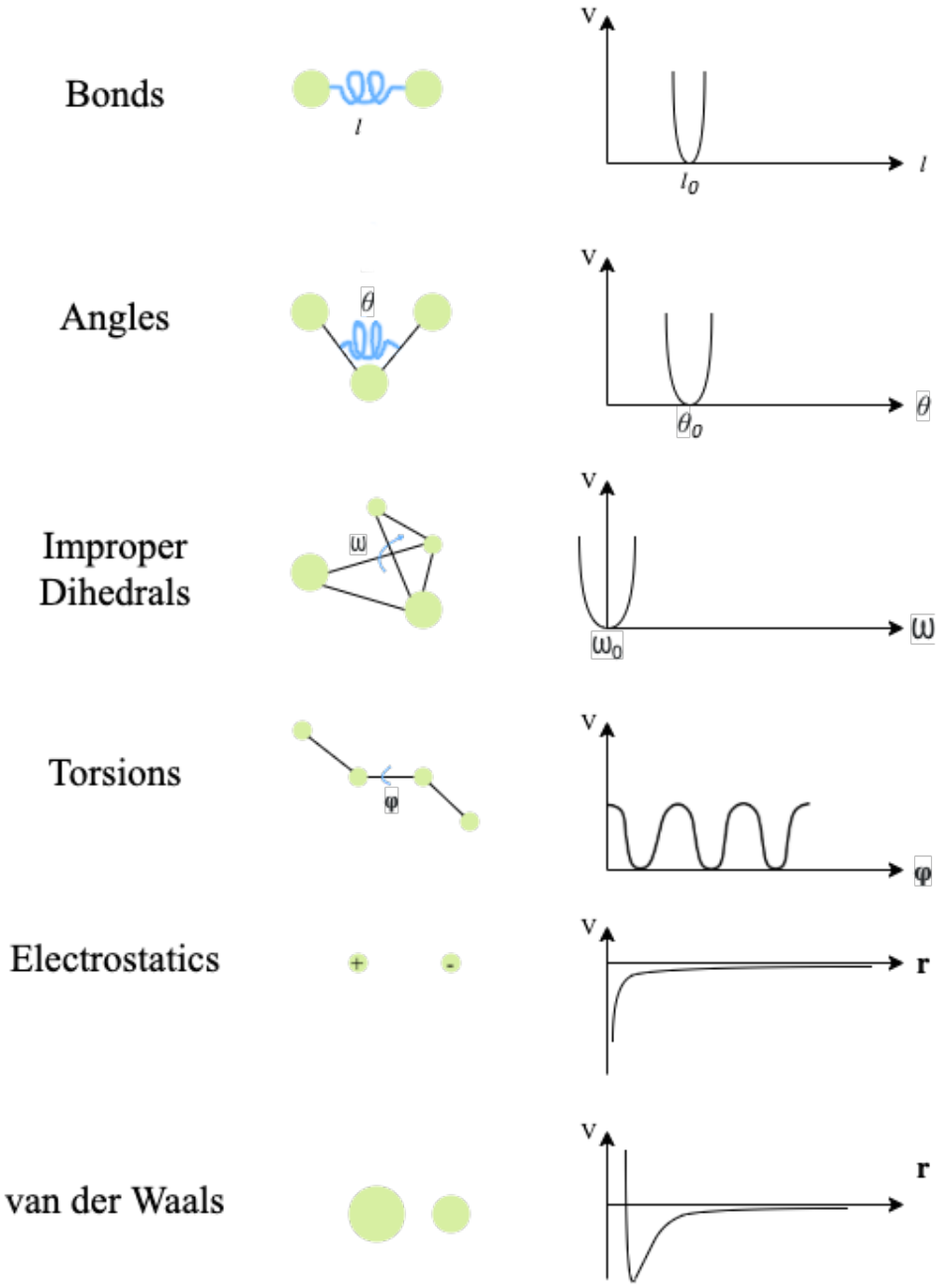


Figure 9: Illustration of the potential energy functions (V) described for molecular Force Fields

3.2.1 GROMOS FF and GROMACS package

Groningen MAchine for Chemical Simulation (GROMACS) is a C language based software, initially developed in 1995 by Berendsen and his peers for a 32-processor machine, with capability to run on any parallel system of processors [16]. The aim of GROMACS is to

provide an efficient and versatile MD program with source code, mainly turned towards simulation of biological macromolecules in aqueous and membrane environments, able to run either in single processor units, or in large scale tandem computer systems [16, 17].

Although GROMACS is independent from the FF used for intramolecular interactions, the organisation of force and energy evaluations used does cause some limitations. Thus, GROMACS is compatible with GROMOS, Encad, OPLS and AMBER FF. These forces and interactions, according to Van der Spoel and colleagues [17], take into consideration three different types of interaction: *bonded interactions*; *nonbonded interactions* and other *special interactions*.

In the same work, these researchers also identify ideal simulation applications suited for GROMACS: membrane lipidic simulations; membrane protein simulations; molecule interactions with inonizing *X*-rays; combined quantum mechanics and classical mechanics and protein folding-related simulations [17].

MD simulations highly depend on the quality of representation of the FF. The interaction between the atoms in a given system must be as accurate as possible, in order to correctly represent the properties and mechanisms underlying the desired processes.

The GROningen MOlecular Simulation (GROMOS) package was first developed back in 1978, by a team of investigators of the same institute that developed GROMACS, initially written in Fortran (1996) [18], as commonly done for these programs, and C, since 2011 [19].

This FF has many iterations over the years, with new additions coming through in each version. The proposed work will take advantage of GROMOS 54a7, an improvement on the previous version GROMOS 53A6 with modifications in some parameters having to do with proteic torsion angles and some Van der Waals interactions parameters [18, 19].

GROMOS is, therefore, one of the most cited and suitable FF for the study of biomolecules, from small molecules to large macromolecules, due to its proven reliability and robustness.

3.3 Molecular Dynamics Protocol

In order for an MD simulation to occur, a strict set of procedures must be partaken and will be generally described in the subsections below. Steps are carefully done and in separate phases, as the system must be prepared in a way that guarantees an optimal convergence towards equilibrium and thus a better closeness to what happens in a real cell. These steps will be further elaborated and their sequence is represented in figure 10.

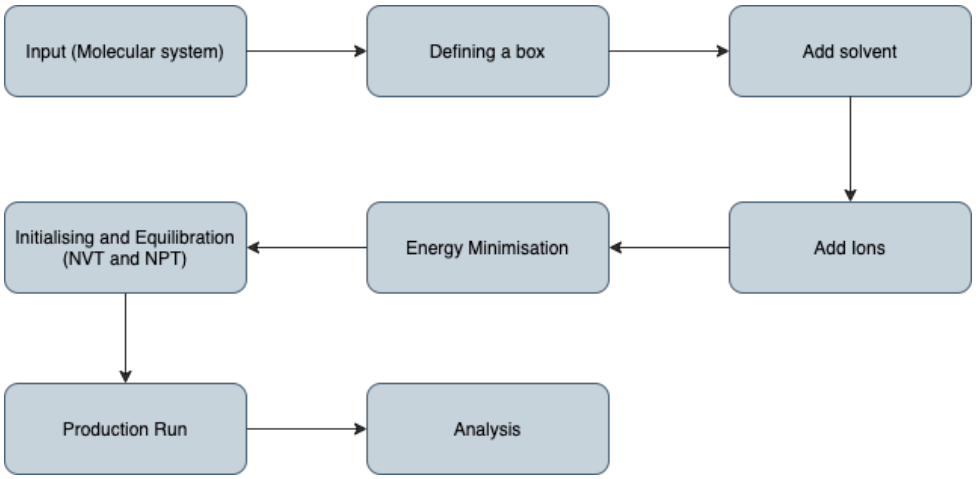


Figure 10: The basic steps regarding an MD protocol.

3.3.1 Initial Steps

At the start of any simulation, an initial 3D model structure must be given upfront. This system will be either deduced experimentally through x-RAY or NMR techniques, or by theoretical assessments, such as homology modelling. The initial structure must be inserted into a solvent (water, ethanol, organic solvents, membranes, etc) first, in order to mostly reproduce physiological conditions, thus giving a high degree of realism in the simulation of biomolecules. Furthermore, all the molecules in the media, must be equilibrated [9].

After solvating, it is important to set a scenario where the central cell containing the system must be surrounded by replicas of itself, allowing for a simulation with a small number of particles and minimise surface effects that would likely result, should the system interact

with a void. These are defined as Periodic Boundary Conditions (PBC), and to set them, one must deal with the non-bonded interactions between atoms of the central cell and the atoms of the surrounding images. The key role is setting a box size such that the system remains compact, but not interacting with mirror images of itself [20].

3.3.2 Energy Minimisation

As suggested by the name itself, this next stage encompasses a set of steps necessary to minimise the system's energy. In a series of relaxation steps, the aim of this stage is ensuring that there are no steric clashes or inappropriate geometry. The process in which a lower energy state is achieved in the following series of steps is schematised in figure 11 and is briefly defined as finding a local minimum, in regards to a certain starting point.

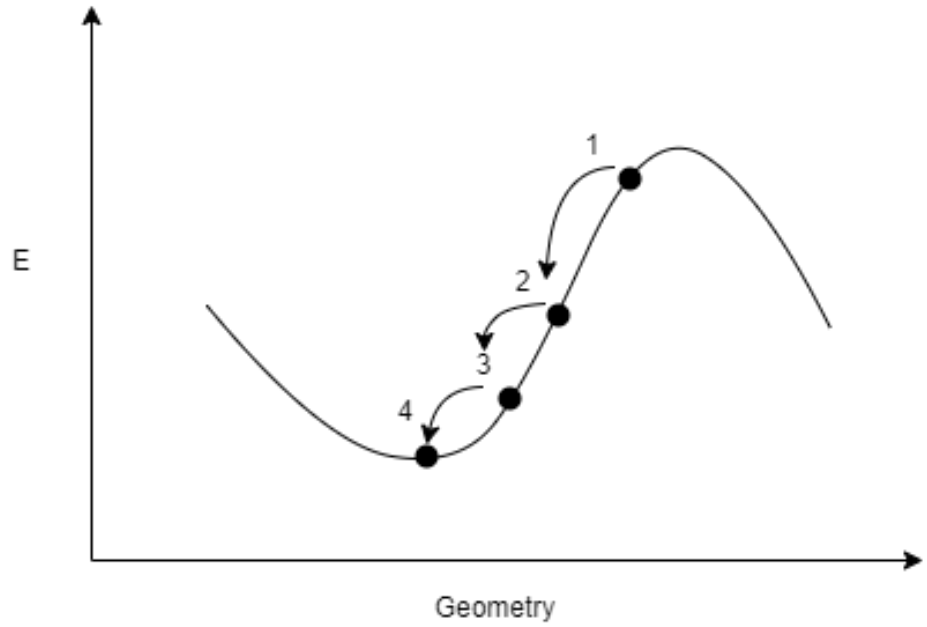


Figure 11: Description of a simplified version of an energy minimisation process. This graphical representation heightens the transitions of energy towards a local minimum according to succeeding changes in its geometry.

There are many algorithms used for geometrical optimisation, however, and for MD simulations the best known algorithms are three: the Steepest Descent (SD), the Conjugate Gradient (CG) and the Newton-Raphson. The SD method is the one used in this simulation

and therefore it is briefly described first. This algorithm, as the name suggests, at its starting point, begins from the direction that points towards the largest decrease, hence the steepest descent. Following steps occur iteratively in vastly the same way as before, until a stopping criterion is eventually reached [17].

The CG method in itself is a very similar method to the previous one, however, after the initial search for the steepest descent, the algorithm then proceeds to search for a direction perpendicular, or conjugate, to the previous. This method is quite slow in the beginning, but as iterations continue, the efficacy of this method starts working, as this slight caveat in the algorithm where it starts to search for the conjugate direction of the last, avoids oscillations typically observed in SD. Thus, this method tends to find the minimum faster [21].

Another approach to minimisation in MD is the Newton-Raphson method. This algorithm is based on a Taylor series expansion of the potential energy surface at the current geometry. This process is concurrently iterated until parameter values are stabilised. This is a complex method and depends on an Hessian estimation of parameters, and thus becomes much more computationally expensive [22].

As for long-range estimation methods for electrostatic interactions, there are two main methods. The Reaction Field assumes a continuous environment (beyond a certain cutoff value, that is usually 1.4 nm) and takes into consideration a dielectric constant that best describes the solvent in the simulation. Another long-range interaction estimation method is the Ewald summation, and is the one used in the present work. This method was initially developed to compute long-range interactions on crystals due to the summation being on over an infinite number of periodic images [23]. But, with the surge of PBC, particle-mesh Ewald (PME) is now widely used in MD, although not advised on complex systems, as the increase in number of particles will be proportional to the increase in the sum, which will turn the whole process computationally expensive

3.3.3 Initialising and Equilibration

In the current phase, initial velocities for the atoms in the system are assigned. As they're not know, they are generated according to temperature. These initial velocities are random, but they do follow a mathematical distribution, a Maxwell-Boltzman distribution [9] more specifically. The initial velocities v_i for each atom in the system are:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}}\,(-\frac{m_i v_i^2}{2kT}) \qquad i = 1, ..., 3N \tag{3}$$

where k represents the Boltzman constant, whereas m_i is the mass of each atom i, and T is the absolute temperature. Having the initial velocities been given, the potential energy of the system is assessed and Newton's equations of motion (1) start being integrated for each particle.

This step will determine the trajectory of each atom in the system and its execution can be acquired utilising some algorithms. The present work takes advantage of an alteration of the Verlet algorithm [14], called leap-frog integration [24].

Once a certain set of properties become stable, the system reaches a state of equilibrium. The process can be monitored by analysis, such as following the root mean square deviation.

A strategy applied to some systems, in order either to avoid disastrous deviations, or to include knowledge from experimental data, is applying position restraints, which in turn, and according to the name itself, restrain some atoms in certain positions. This procedure is used here in the initialization steps to relax the structure slowly, first by restraining all heavy atoms and then, restraining the main chain atoms, under two distinct ensembles: canonical (NVT) and isothermic-isobaric ensembles (NPT), respectively [25, 26].

The NVT ensemble conserves three variables, being the amount of substance (N), the volume (V) and the temperature (T). This canonical ensemble, will represent the possible states of the system in thermal equilibrium under a heat bath with fixed temperature [25].

There are a number of methods used to control temperature (thermostats), one of such is the Berendsen thermostat method [26], that was used in this work.

NPT ensembles, in a much similar way as the NPT previously, uses the same techniques with an added barostat, in order to control pressure. This ensemble keeps number of substance (N), pressure (P) and temperature (T) conserved. Instead of a heat bath, this system is more representative of an ambient temperature and pressure system, as before, Berendsen thermostat and barostat are used in the present work [26]. Another method, the LINCS algorithm [27, 28] is also applied in order to constrain bonds and angles and eliminate some vibrational modes of higher frequency, thus allowing larger integration time steps.

3.3.4 Production MD

After the initialization step is concluded, the system is now ready to run for a certain time and withou position restraints, in order to guarantee a good sampling of conformational states the system can acquire. This total time should be superior to the relaxation time of the properties to be analysed. In the particular case, for Tau in solution the simulation will be made under 50 nanoseconds, and the interaction with MT will be made under 30ns.

3.3.5 Analysis: Tools and functions

Obtained trajectories are put through some GROMACS's [16, 17, 21] package tools and components. These analysis are important in order to assess deviations from the original structure, number and type of SS, number of hydrogen bonds, clusters of middle structures, among others.

The most common tools used in MD and used in this work consist of RMSD, RMSF, SS analysis and clustering analysis. RMSD is typically used in order to assess structural stability of peptides in the system. An RMSD curve tracks the changes/deviations from the initial starting structure, by fitted peptides along the simulation. RMSF is similar to RMSD but instead of

deviations from the starting structure through time, this analysis tracks fluctuations in peptide movement along the entirety of the structure. This is useful to track/identify changes in specific domains or sites [17].

It is also possible to quantify SS through a DSSP, a method that tells the SSs taken by the system in simulation. Clustering analysis, is also an important tool, as it groups mathematically similar through a given distance metric. This one method permits guessing the middle structures of the system, predicting its most probable conformation for most of the time in physiological conditions [17, 21].

PyMOL consists of a graphical interface for molecular structures, allowing for user interaction and visualisation of structural properties of molecules or molecular systems [29]. It aims to provide most of the required implementatations of outputs from common MD simulation packages written either in Fortran or C [30].

It consists of a free open-source platform, allowing investigators to use and publish work without costs or limitations to it. This software was written in the python language.

PyMOL can reproduce high-quality videos and images of macromolecules in different representations such as ribbons, cartoons, dots, surfaces, spheres, sticks, and lines. The versatility of this program means it can be used in a wide variety of molecular modeling disciplines, with the most common usage for it being drug design and discovery, where it is used in combination with a MD package [29].

References

[1] Alder, B. J., & Wainwright, T. E. (1957). Phase transition for a hard sphere system.

[2] Alder, B. J., & Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*.

[3] Gibson, J. B., Goland, A. N., Milgram, M., & Vineyard, G. H. (1960). Dynamics of radiation damage. *Physical Review*.

[4] Rahman, A. (1964). Correlations in the motion of atoms in liquid argon. *Physical Review*.

[5] Rahman, A., & Stillinger, F. H. (1971). Molecular dynamics study of liquid water. *The Journal of Chemical Physics*.

[6] Walpole, J., Papin, J. A., & Peirce, S. M. (2013). Multiscale Computational Models of Complex Biological Systems Keywords.

[7] Warshel, A., & Karplus, M. (1972). Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules.1 I. Formulation and Parametrization. *Journal of the American Chemical Society*, *94*(16), 5612–5625.

[8] Carvalho, A. T., Barrozo, A., Doron, D., Vardi Kilshtain, A., Major, D. T., & Kamerlin, S. C. L. (2014). Challenges in computational studies of enzyme structure, function and dynamics. *Journal of Molecular Graphics and Modelling*.

[9] Leach, A. R. (2006). Ligand-based approaches: Core molecular modeling. *Comprehensive medicinal chemistry ii* (pp. 87–118). Elsevier Ltd.

[10] Lipkowitz, K. B., & Boyd, D. B. (2007). *Reviews in Computational Chemistry*.

[11] Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules.

[12] González, M. (2011). Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, *12*, 169–200.

[13] Ramachandran, K. I., Deepa, G., & Namboori, K. (2008). *Computational chemistry and molecular modeling: Principles and applications*.

[14] Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*.

[15] Höltje, H., Folkers, G., Mannhold, R., Kubinyi, H., & Timmerman, H. (2008). *Molecular modeling: Basic principles and applications*. Wiley.

[16] Berendsen, H. J., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*.

[17] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: Fast, flexible, and free.

[18] Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., & van Gunsteren, W. F. (1999). The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A*, *103*(19), 3596–3607.

[19] Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., & Van Gunsteren, W. F. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*.

[20] Van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hünenberger, P. H., Kastenholz, M. A., Oostenbrink, C., Schenk, M., Trzesniak, D., Van Der Vegt, N. F., & Yu, H. B. (2006). Biomolecular modeling: Goals, problems, perspectives.

[21] Abraham, M., Van Der Spoel, D., Lindahl, E., Hess, B., van der Spoel, D., & Lindahl, E. (2014). GROMACS User Manual version 5.0.4. *www.Gromacs.org*.

[22] Verbeke, J., & Cools, R. (1995). The newton-raphson method. *International Journal of Mathematical Education in Science and Technology*.

[23] Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*.

[24] Roberts, K. V. (1979). Computational methods in physics. In Riazuddin (Ed.), *Physics and contemporary needs: Volume 3* (pp. 385–472). Springer US.

[25] BRYAN, G. H. (1902). Elementary Principles in Statistical Mechanics. *Nature*.

[26] Berendsen, H. J., Postma, J. P., Van Gunsteren, W. F., Dinola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*.

[27] Hess, B. (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*.

[28] Hess, B., Bekker, H., Berendsen, H. J., & Fraaije, J. G. (1997). LINCS: A Linear Constraint Solver for molecular simulations. *Journal of Computational Chemistry*.

[29] Schrödinger, LLC. (2015). *The PyMOL molecular graphics system, version 1.8*
PyMOL The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC.

[30] DeLano, W. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*.

CHAPTER 4

Results and Discussion

4.1 Methodology for aTau Simulations

4.1.1 aLys Parameterisation

Before any of the processes and events of the MD protocol itself, it was necessary to generate the topological coordinates file for Gromos FF. As such, a lysine with an acetyl group was prepared earlier in PyMOL [1] and then submitted to the online platform Automated Topology Builder (ATB)[2], generating a file with all classical FF for the submitted protein, in the wanted gromos 54a7 FF. This lysine system is illustrated bellow (figure 12).

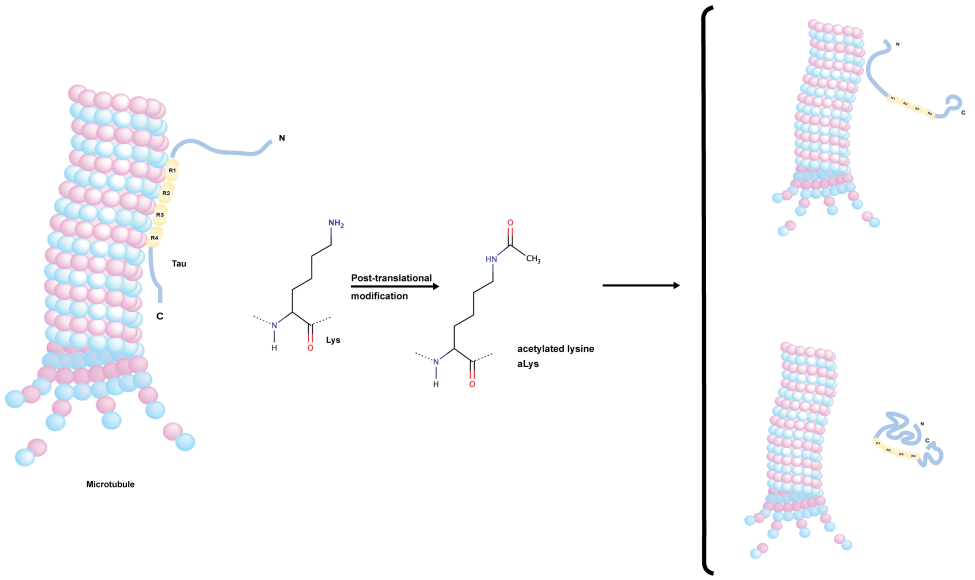


Figure 12: Illustrative representation of the main components taken into consideration for this work: a) Tau's domains and lysine positions, in which each acetylation site to be altered is marked in pink; b) representation of each replicate's modifications, with a description of where lysines are being modified and c) a computational model displaying a normal lysine and its morphology following a post-translational modification, adding an acetyl group.

After this step, it is important to incorporate this new topology into the gromos54a7 FF, name it properly and then start modifying our pre-existent model, the predicted 2N4RTau's 441 aminoacid structure from Castro and colleagues [3], on targeted sites as shown in figure 12.

4.1.2 Systems Setup

Tau modifications were made under a precise scrutiny of literature. These findings were previously discussed in chapter 2, and as such a brief elaboration will take place: some works report potentially harmful lysines to be K163, K280, K281 and K369 [4] whereas other findings report an association of modified K174 to some neuropathologies [5, 6] where Tau becomes insoluble and consequently tends to aggregate. More authors also have published works where AD brains reportedly present acetylated Tau on K180, K274, K240 and K281 [7, 8]. Thus, these positions, along with some others, regarding their spatial disposition along some critical domains in Tau, were targeted in 8 different replicates. The more detailed descriptions of each acetylated position and their respective domains is illustrated above (fig. 12b.) and a visual 3D representation of each replicate is in figure 13. Each modification has been made with either intent to verify some findings in literature that point towards significant changes in Tau behaviour, Tau protection or by targeting positions in the protein structure, responsible for Tau regulation.

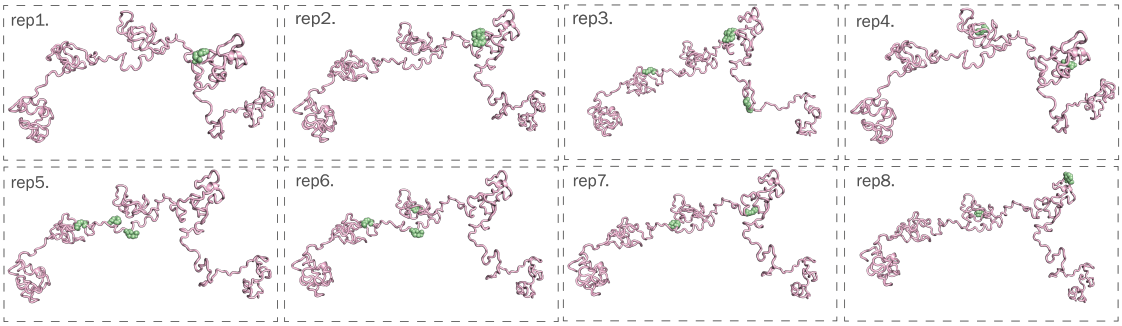


Figure 13: Representation of Acetylated positions of each replicate of Tau (1-8). Light green spheres represented aLys residues into the Tau's pink cartoon.

Two types of systems were considered for simulation in order to study Tau post-translational acetylation. The first system consists of a simulation box in which each replicate interacts alone, in fluid and the second is a simulation box containing aTau next to a tubulin section (PDB ID: 5JCO), simulating interactions with a MT. The systems will thus be named aTau and aTau::MT, respectively. The first system is centered in the simulation box, with water and ions surrounding the aTau protein and in the case of the second system,

the tubulin wall was placed next to the box's wall and aTau near it, in a random position. These systems are illustrated in figure 14.

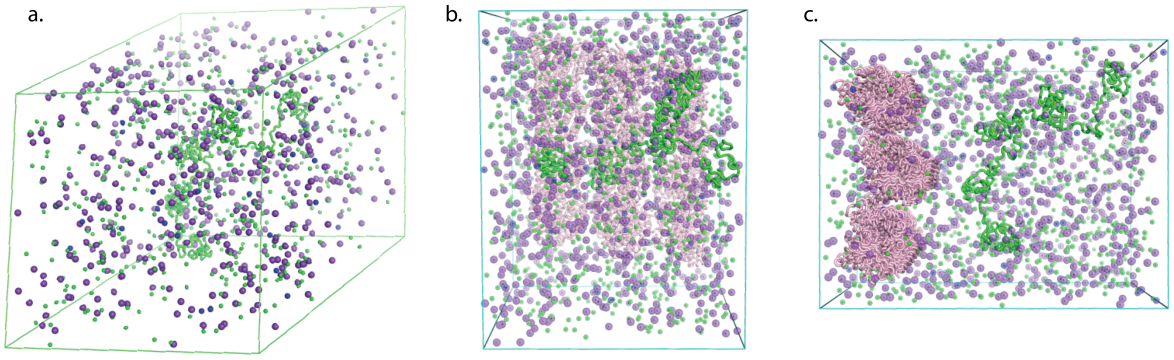


Figure 14: Front view snapshots of systems under simulation: **a)** represents aTau in fluid whereas **b)** shows an aTau::MT system and **c)** presents a perspective side-view of the aTau::MT system. Systems are shown with no water molecules in order to provide better visualisation.aTau proteins are herein represented in lime green, the MT is in pink ad for K^+ in purple, Na^+ in blue, and Cl^- in green spheres.

As previously mentioned, the next step in the simulation protocol is solvation. This step is quickly done and remains the same for both systems. The addition of ions follows, and in a replication of the fluid inside a common neuron, the system comprises a high concentration of potassium ions, and a fixed concentration of K^+ and Na^+ was established, in line to what Castro et al. [3] also reported in their paper, and also in accordance to references therein. These ions are heavily involved in the transmission of the electric signals in neurons, and there is a high chance that they also play a big role, affecting the hugely negative MT wall and Tau proteins. These values were respectively set at 140 nM for K^+ and 10 nM for Na^+.

The highly negative balance that nullifies these cations in neurons is usually obtained from several negatively charged proteins inside neurons. However, computational limitations prevent simulating such highly complex systems in a feasible time, plus, the final balance of charges will be the same with the contribution of Cl^-. This ion addition to the MD box is a common practice in many MD simulations. This is mainly due to the utilisation of PBC and PME, for electostactic treatment. As such, the simulation only converges if the net electric charge is zero, and these ions are therefore calibrated in a way that its final concentration guarantees this.

4.1.3 MD simulations options

Keeping in line with the MD protocol and simulation norms, energy minimisation was performed in two different fashions: for the aTau in fluid system, and since it is a relatively small one, a single step of energy minimisation was performed, with the steepest descent method. For the aTau::MT system, as a more demanding simulation box, in order to guarantee that the system converges properly and avoids some abnormalities that may surge by trying to minimise such a heavy system in one go, the simulation occured in several stages, with iterations of steps using interpolations of both steepest descent and conjugate gradient methods. For both systems, minimisation was constrained to a maximum of 50,000 steps.

Initialisation steps were performed using canonical (NVT, constant number of particles, volume and temperature) and isothermal-isobaric (NPT, constant number of particles, pressure and temperature) ensembles, where position restraints (with force constant of 1000 kJ · mol-1 · nm-2) were applied to all heavy atoms in the NVT procedure, and to the main chain at the NPT initialisation step, during 100 ps each. Systems were constrained at the temperature, where it was kept at a constant 310 K with V-rescale algorithm [9] and pressure kept at 1 atm with the Parrinello-Rahman barostat [10]. Additionaly, coupling constants were set as T_t = 0.10 ps and T_p = 2.0 ps.

The NPT ensamble was then utilised in the MD run. The system comprising aTau in fluid was run for 50 ns whereas aTau::MT was run for 30 ns. These systems were not subject to any position constraints, and all 8 replicates of each condition went through simulation. All simulations were conducted using the GROMACS 5.1.4 [11] on the GROMOS 54a7 FF.

4.1.4 Analysis

aTau in fluid

Once the MD simulation is complete, the GROMACS package offers a number of post-processing analytical tools, which were previously explored in chapter 3. RMSD analysis

was performed aTau in fluid, in order to observe how the overall structure evolves through each time step following acetylation, when compared to the wild-type, ergo unaltered protein.

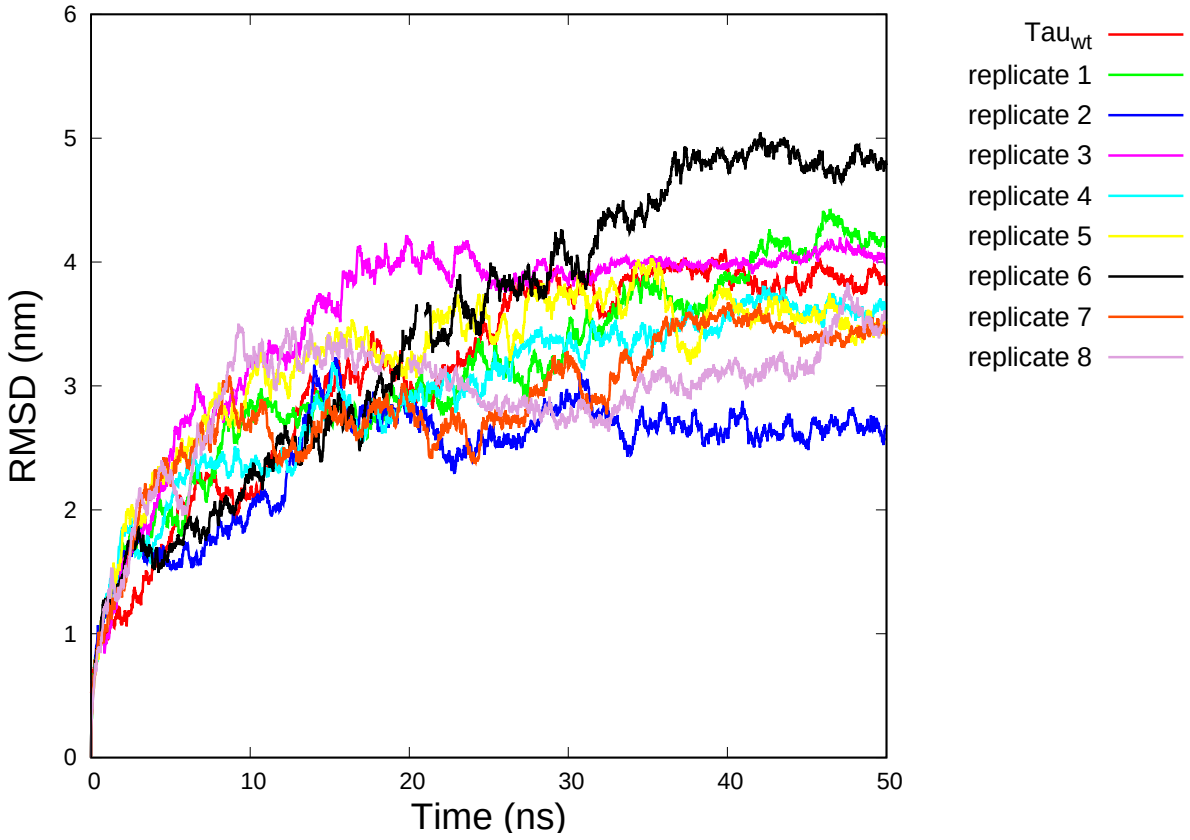


Figure 15: RMSD of each replicate of aTau in fluid, compared against a native Tau protein ($aTau_{wt}$), fitting the backbone for intracellular fluid simulations.

A carefull look into figure 15 reveals that all aTau replicates change substantially throughout time. This is an expected attitude for an IDP like Tau, as the lack of well defined secondary and tertiary structures results in consequent lack of stability and structural cohesion. All RMSD curves appear high in value, mainly replicates 1, 3 and 6, which all surpass Tau_{wt} at a certain point, whilst other replicates express a lower RMSD pattern when compared to Tau_{wt}. Equilibration appears to happen somewhat around 20 ns. These results suggest significant structural changes in aTau.

RMSF data analyses fluctuations amongst residues and in this step, data was collected for the last 20 ns of simulation, as from the aforementioned analysis, it was shown these are more representative of a final conformation. Much like RMSD beforehand, RMSF data is also very inconstant, with high variations, for much the same reasons as before (fig. 16). Additionally,

data seems to suggest residue behavior to not keep up with that of Tau_{wt}, in some cases with less fluctuations within the terminals, such as the case with replicates 3 and 5. Some replicates have higher movement within these regions, as replicates 1 and 8, but what remains constant for all situations is that there are clear deviations from the expected normal attitude in intracellular fluid in a wild-type protein.

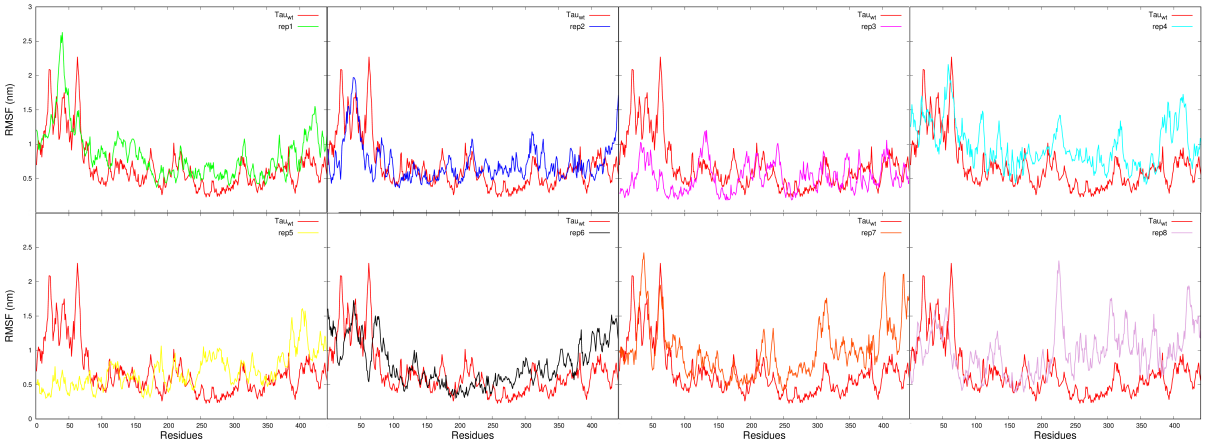


Figure 16: RMSF of each replicate of aTau in fluid, each compared against a native Tau protein ($aTau_{wt}$), fitting the backbone for intracellular fluid simulations.

Clustering analysis is also a feature of GROMACS computational package. It consists of grouping closely related data by a chosen distance criteria. This clustering generates middle structures for each replicate, where each structure will represent the most likely state of conformation that this protein will exhibit. Analysis were conducted, much like for RMSF, within the last 20 ns of MD simulation in this condition (aTau in solution).

These results show that only replicates 2, 5 and 8 appear more extended (fig. 17). On all other cases aTau tends to fold around itself and the MTBR proteins, hindering access to that particular area. In all cases, there seems to be little to no signs of the common "paperclip" structure Tau acquires in intracellular fluid [12], where both N- and C- termini fold in the vicinity of eachother. The only cases that may be happening is for replicate 3, and a slight tendency towards that type of folding can be observed in replicates 1 and 4 (fig. 17).

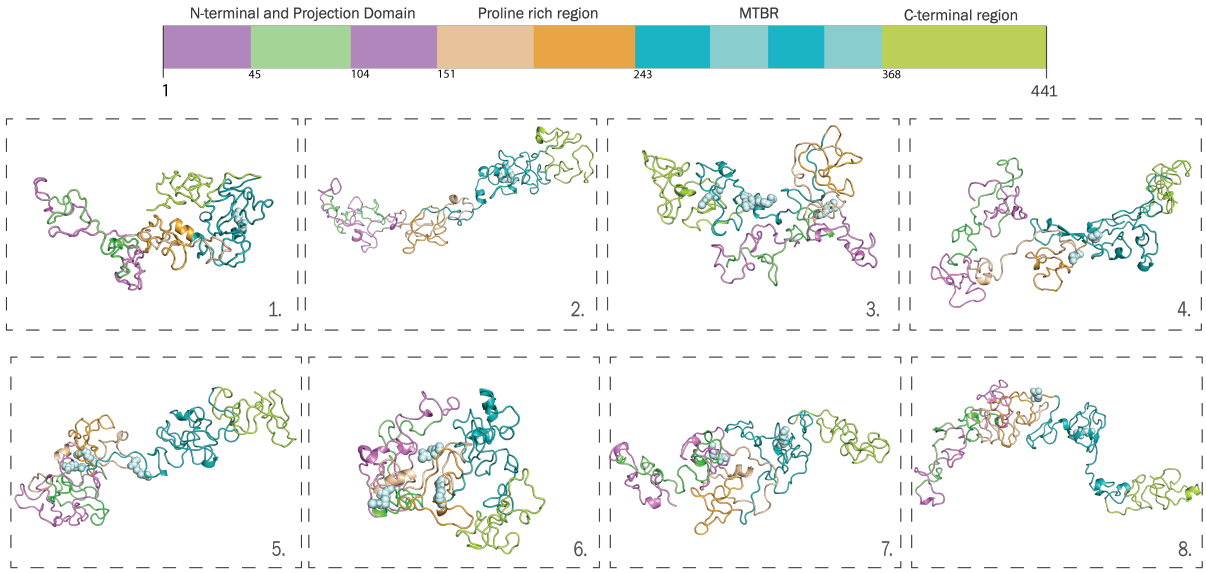


Figure 17: Middle structures generated for the last 20 ns of MD simulation for intracellular fluid. Color blocking follows the colors of each domain illustrated in the bar above. Acetylated lysines are represented in the 3D structures as light blue spheres.

aTau::MT

For the systems comprising aTau simulated against a tubulin wall, backbone RMSD data reveals a very similar behavior as before. RMSD curves remain highly variable for all replicates, as expected, with replicates 1 and 3 distancing from the apparent norm and replicate 4 starting with very high RMSD, but stabilising later on in the run (fig. 18 **a)**). Although lower than values observed before for aTau in fluid, these deviations are still generally higher than those of Tau under the same condition [3]. The time step that equilibrium seems to be generally acquired is around 20 ns.

RMSF data for aTau::MT was conducted, much in the fashion of the previous RMSF analysis, for data in the last 10 ns of simulation. Fluctuations are much higher now than before (fig. 18 **b)**). This is also quite an expected result. Near the MT, the electrostatics of all charged elements in solution, the elements within aTau itself and the highly negative tubulin dimer conglomerate that constitutes the MT, added with the fact that Tau is an IDP, are expected to cause high motility within residues. Results show that in this case, high fluctuations occurred within all residues of most replicates, whereas before, movements were

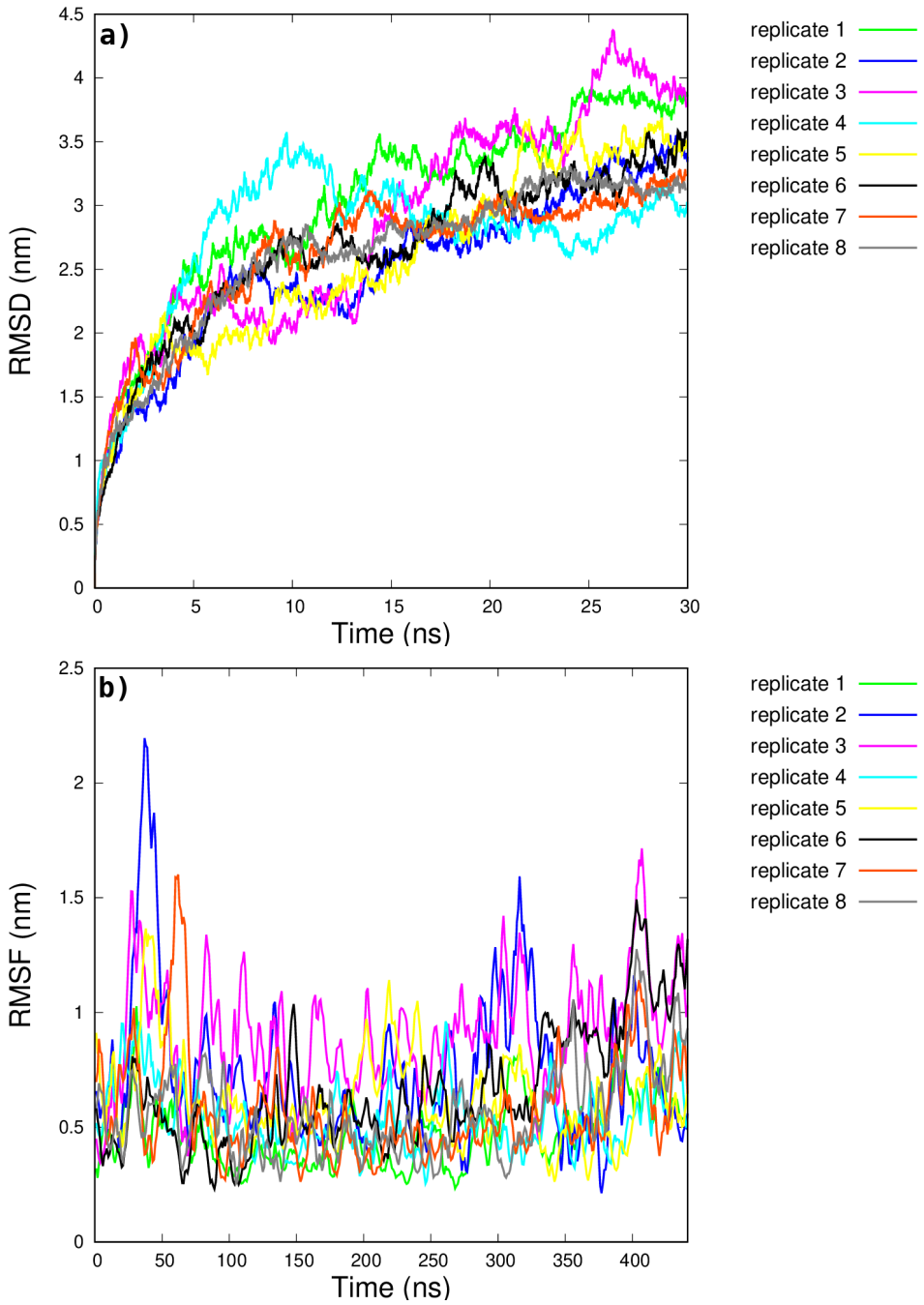


Figure 18: **a)** Backbone RMSD of all Tau replicates, fitting the backbone near the MT. **b)** Residues RMSF for all replicates of aTau near MT

more or less concentrated within terminal regions (fig. 16). This result hints towards lack of stability in mutant replicates, probably resulting from failure to bind to MT.

Clustering is thus, the next logical step, as 3D models can provide a better visual perception of conformational behaviors of mutant replicates after MT simulation. In this scenario, and for better view of the systems, aTau was also captured more closely, without the MT in sight, for ease in interpreting middle structures generated by the simulation. Results are shown in figure 19

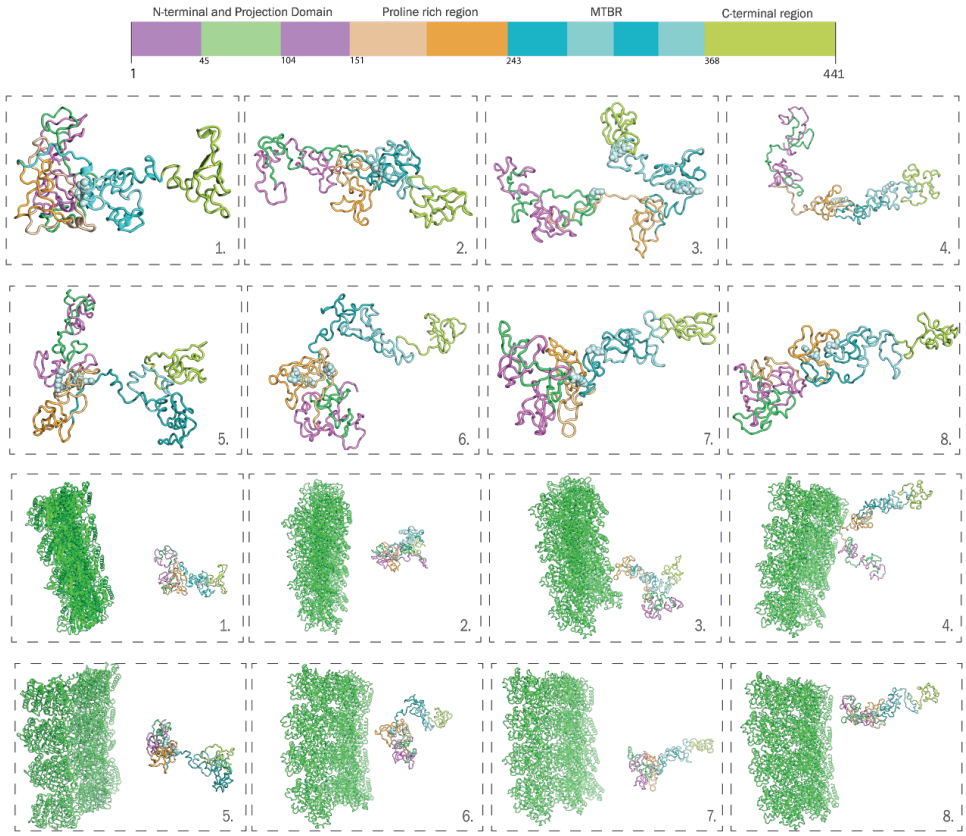


Figure 19: Middle structures generated for the last 10 ns of MD simulation for aTau::MT systems. Color blocking follows the colors of each domain illustrated in the bar at the top, and the proeminent structure in bright green refers to the tubulin wall. Acetylated lysines are represented in the above 3D structures as light blue spheres.

From each middle structure the most obvious result is gotten from the verification that in this condition, aTau barely shows any binding, and if so, this only happens in replicates 3 and 4 (fig. 19), but these interactions are within the proline-rich region, which is not the binding site for Tau. Furthermore, and reminiscing to the previous RMSF analysis (fig. 18 **b)**),

and acetylation sites (fig. 12 **b)**), one can observe that: replicate 1, with only K280 modified presents a highly globular structure, and consequent failure to bind to the MT. The same is happening for replicate 2 where a very fluctuant RMSF curve, takes a tremendously "ball"-like structure and logically fails to interact with the MT. These mutants have K280, for rep1 and K280, K281 modifications cited as detrimental for Tau structure integrity. These results, along with those of replicate 3, which also has these sites altered, suggest an important structural role for these lysines.

Similarly to what has happened for intracellular aTau, replicates 3, 5 and 6 still present somewhat of a tendency to approximate termini. For all others, with the exception of rep4, structures remain either too tangled or globular, hampering the odds of the MTBR binding to the MT, as the binding of Tau to MT is reported to occur through the MTBR in an extended form [13].

Secondary structure quantification

Table 1: Secondary structure percentages calculated using the DSSP method, for aTau replicates middle structures in intracelular fluid and near the MT.

Replicates	% of Secondary Structure (%SS)									
	β-sheet		β-bridge		turn		α-helix		3_{10}-helix	
	fluid	MT	fluid	MT	fluid	MT	fluid	MT	fluid	MT
rep1	4	4	5	5	14	10	5	2	3	2
rep2	1	4	2	6	29	12	1	4	1	1
rep3	1	-	6	5	17	8	2	5	2	3
rep4	1	4	4	2	15	15	1	3	1	2
rep5	2	-	5	3	19	16	1	2	1	3
rep6	2	2	3	2	20	15	7	3	1	1
rep7	1	6	5	5	17	8	3	2	1	3
rep8	3	2	2	2	17	15	3	2	1	1

The DSSP method was finally utilised in order to assign SS in both aTau systems. This method, through intra-backbone tracking of hydrogen bonds' energy and orientations, and settling these to a SS, predicts SS in a given protein. This method was used in order to

understand if after simulation, aTau acquires any well-defined SS, since previous work by Castro et al. [3] has shown Tau to have only small patches of SS.

The results evidenced in table 1 show that for all replicates, in both conditions, a very small number of helical SS is acquired. The same, although in slightly higher patches of β-bridges, seems to happen in β-sheets and bridges. A higher occurence in turns is registered. This can be evidenced by the increase of entanglements within Tau's own filaments, creating these structures.

4.1.5 Discussion

The results shown by aTau in fluid point towards a significant alteration of its structure when acetylated. Tau's condition as an IDP dictates a high instability for Tau under normal conditions [3, 12, 14], and as such the hypothesis that any small alteration of the electrostatics of certain areas may deeply influence the whole structure was confirmed, as critical sites such as K280, previously identified in other works to be highly likely of causing harm if tampered with [12], have proven to be detrimental when it comes to dictating Tau's structure malfunctioning if acetylated. This case is not noted only for K280, as other sites in all test cases caused no "paperclip" shape in intracellular fluid and no MT-binding by the MTBR. Even slight cases of tendency to approximate both N- and C- termini, as for some interactions with the tubulin wall observed by replicates 3, 4 and a tendency by rep6, still corroborate the initial hypothesis that these post-translational modifications will have some kind of effect on final conformation.

To further support this work's main objective, RMSD and RMSF data both skew away from those of previously reported wild-type Tau [3], and DSSP data also indicates an increase of some types of SS patches, creating a stronger, more well defined structure, and possibly contributing for the resulting difficulties in Tau stretching itself to its MT-binding position.

In sum, Tau is an IDP, crucial for MT's structural integrity, stabilising them and regulating the dynamic instability resulting from the concurrent polymerisations and depolymerisations of tubulin dimers within their structures [15]. This malfunctioning of Tau is being regarded as a potential cause for neurological cell decay in some neurological diseases called Tauopathies

[4, 16]. Post-translational phosphorylation has been heavily regarded as the main culprit for Tau's structural failure and the target for many studies [4, 17–21]. However more recently, Tau acetylation as moved some investigations into its potential harm towards the malfunctioning of this protein [22–24]. The present work, through computational chemistry protocols and tools such as GROMACS, successfully gathered proof of Tau erratic behavior when acetylated on potentially harmful lysines or strategic lysine sites on some proteic domains, altering its structure, and thus its cellular behavior.

4.1.6 Conclusions

Modelling Tau and its behaviour under some post-translational modifications have proven to be a fruitful task. In this work, the suspected harmful role of acetylation on Tau function has been tested and confirmed under MD simulations. This work has successfully predicted Tau's conformational behavior for aTau both in intracellular fluid and in the vicinity of a microtubule. All replicates have shown behaviors close to the ones reported by literature as harmful and potential causes of neurodegenerative events and long term effects such as mental diseases in the likes of Huntington's, Parkinson's and Alzheimer's diseases. In several studies, Tau is reportedly prone to, when damaged, become filamentous, creating entanglements and failing to bind MTs, ultimately contributing for its failure and neuronal damage [4–7, 22–29]. In sum, Tau malfunction has been highly associated with structural damage following some harmful post-translational modifications. This study on the molecular dynamics and electrostatics confirms these associations and provides in silico predictions for aTau structural profiles under such alterations. This work is yet another setting stone on understanding the mechanisms underlying Tau behaviour, structure and a more detailed view in the broader picture of neurodegenerative diseases.

References

[1] Schrödinger, LLC. (2015). *The PyMOL molecular graphics system, version 1.8* PyMOL The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC.

[2] Warshel, A., & Karplus, M. (1972). Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules.1 I. Formulation and Parametrization. *Journal of the American Chemical Society*, *94*(16), 5612–5625.

[3] Castro, T., Munteanu, F.-D., & Cavaco-Paulo, A. (2019). Electrostatics of Tau Protein by Molecular Dynamics. *Biomolecules*, *9*(3), 116.

[4] Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's Disease. *New England Journal of Medicine*, *362*(4), 329–344.

[5] Drazic, A., Myklebust, L. M., Ree, R., & Arnesen, T. (2016). The world of protein acetylation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1864*(10), 1372–1401.

[6] Min, S.-W., Chen, X., Tracy, T. E., Li, Y., Zhou, Y., Wang, C., Shirakawa, K., Minami, S. S., Defensor, E., Mok, S. A., Sohn, P. D., Schilling, B., Cong, X., Ellerby, L., Gibson, B. W., Johnson, J., Krogan, N., Shamloo, M., Gestwicki, J., ... Gan, L. (2015). Critical role of acetylation in tau-mediated neurodegeneration and cognitive deficits. *Nature Medicine*, *21*(10), 1154–1162.

[7] Tracy, T. E., Sohn, P. D., Minami, S. S., Wang, C., Min, S. W., Li, Y., Zhou, Y., Le, D., Lo, I., Ponnusamy, R., Cong, X., Schilling, B., Ellerby, L. M., Huganir, R. L., & Gan, L. (2016). Acetylated Tau Obstructs KIBRA-Mediated Signaling in Synaptic Plasticity and Promotes Tauopathy-Related Memory Loss. *Neuron*.

[8] Silva, M. C., & Haggarty, S. J. (2020). Tauopathies: Deciphering Disease Mechanisms to Develop Effective Therapies. *International Journal of Molecular Sciences*, *21*(23), 8948.

[9] Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *Journal of Chemical Physics*.

[10] Martoňák, R., Laio, A., & Parrinello, M. (2003). Predicting Crystal Structures: The Parrinello-Rahman Method Revisited. *Physical Review Letters*.

[11] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: Fast, flexible, and free.

[12] Jeganathan, S., Von Bergen, M., Brutlach, H., Steinhoff, H. J., & Mandelkow, E. (2006). Global hairpin folding of tau in solution. *Biochemistry*.

[13] Kellogg, E. H., Hejab, N. M., Poepsel, S., Downing, K. H., DiMaio, F., & Nogales, E. (2018). Near-atomic model of microtubule-tau interactions. *Science*, *360*(6394), 1242–1246.

[14] Guo, T., Noble, W., & Hanger, D. P. (2017). Roles of tau protein in health and disease.

[15] Conde, C., & Cáceres, A. (2009). Microtubule assembly, organization and dynamics in axons and dendrites. *Nature Reviews Neuroscience*, *10*(5), 319–332.

[16] Ittner, L. M., & Götz, J. (2011). Amyloid- and tau — a toxic pas de deux in Alzheimer's disease. *Nature Reviews Neuroscience*, *12*(2), 67–72.

[17] Lee, V. M.-Y., Goedert, M., & Trojanowski, J. Q. (2001). Neurodegenerative Tauopathies. *Annual Review of Neuroscience*, *24*(1), 1121–1159.

[18] Götz, J., Chen, F., Van Dorpe, J., & Nitsch, R. M. (2001). Formation of neurofibrillary tangles in P301L tau transgenic mice induced by Aβ42 fibrils. *Science*, *293*(5534), 1491–1495.

[19] Council, M. R., Road, H., & Kingdom, U. (2001). Eurodegenerative auopathies. *Genetics*.

[20] Sharma, P., Srivastava, P., Seth, A., Tripathi, P. N., Banerjee, A. G., & Shrivastava, S. K. (2019). Comprehensive review of mechanisms of pathogenesis involved in Alzheimer's disease and potential therapeutic strategies. *Progress in Neurobiology*, *174*(September 2018), 53–89.

[21] Gong, C. X., Lidsky, T., Wegiel, J., Zuck, L., Grundke-Iqbal, I., & Iqbal, K. (2000). Phosphorylation of microtubule-associated protein tau is regulated by protein phosphatase 2A in mammalian brain. Implications for neurofibrillary degeneration in Alzheimer's disease. *Journal of Biological Chemistry*, *275*(8), 5535–5544.

[22] Cohen, T. J., Guo, J. L., Hurtado, D. E., Kwong, L. K., Mills, I. P., Trojanowski, J. Q., & Lee, V. M. (2011). The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nature Communications*, *2*(1), 252–259.

[23] Cohen, T. J., Constance, B. H., Hwang, A. W., James, M., & Yuan, C.-X. (2016). Intrinsic Tau Acetylation Is Coupled to Auto-Proteolytic Tau Fragmentation (C.-X. Gong, Ed.). *PLOS ONE*, *11*(7), e0158470.

[24] Min, S.-W., Cho, S.-H., Zhou, Y., Schroeder, S., Haroutunian, V., Seeley, W. W., Huang, E. J., Shen, Y., Masliah, E., Mukherjee, C., Meyers, D., Cole, P. A., Ott, M., & Gan, L. (2010). Acetylation of Tau Inhibits Its Degradation and Contributes to Tauopathy. *Neuron*, *67*(6), 953–966.

[25] Drubin, D., Kobayashi, S., & Kirschner, M. (1986). Association of Tau Protein with Microtubules in Living Cells. *Annals of the New York Academy of Sciences*, *466*(1 Dynamic Aspec), 257–268.

[26] Lee, V. M. Y., Goedert, M., & Trojanowski, J. Q. (2001). Neurodegenerative tauopathies [WOS:000170109100037]. *Annual Review of Neuroscience*, *24*, 1121–1159.

[27] Narita, T., Weinert, B. T., & Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. *Nature Reviews Molecular Cell Biology*, *20*(3), 156–174.

[28] Cook, C., Carlomagno, Y., Gendron, T. F., Dunmore, J., Scheffel, K., Stetler, C., Davis, M., Dickson, D., Jarpe, M., DeTure, M., & Petrucelli, L. (2014). Acetylation of the KXGS motifs in tau is a critical determinant in modulation of tau aggregation and clearance. *Human Molecular Genetics*, *23*(1), 104–116.

[29] Kontaxi, C., Piccardo, P., & Gill, A. C. (2017). Lysine-directed post-translational modifications of tau protein in Alzheimer's disease and related tauopathies.

CHAPTER 5

Final Remarks

The dissertation at hands has taken the job of evaluating an extremely important IDP such as Tau, modelling its electrostatic and dynamic behavior when acetylated at specific lysines. The main objectives of this work, as it was mentioned in chapter 1, were all met. Studying Tau and its behaviour under some post-translational modifications has proven to be a fruitful task.

MD simulations have proven to be able to surpass the obstacles of modelling an IDP, as results prove favourable to what previous research cites. Furthermore, the skills taught throughout the development of this work and all the tools used, from parameterisation to simulations to analysis, has given tools to develop the work at hands as well as provide future developments in this area, helping in other such projects as the present one.

In conclusion, through *in silico* studies, aTau behavioral profile was assessed for both test conditions (solution and near the MT), suggesting the potential harm of post-translational acetylation of Tau in its structure and the direct effect on MTs and neurons. It has been clearly observed that electrostatics has a key role in mediating Tau's interaction with MT, with evident influence of the positive Lys charges on MT binding. Computational chemistry tools and procedures were successfully applied and results have been favourable. Future work may follow the same pattern of research, underlining other post-translational modifications and possibly finding a potential Tau-protective effect that prevents these conformational changes, either in Lys residues, or other modification sites on Tau.

The remarkable findings of this work were written in the form of a paper, that has been already submitted and awaits review.