

# **Computer Simulations of Post-Translationally Modified Microtubules**

**Christopher Robert Field**

**201261258**

**Submitted in accordance with the requirements of  
the degree of Master of Research in Physics**

**University of Leeds  
Faculty of Engineering and Physical Sciences**

**September 2022**

## **Intellectual Property and Publication Statements**

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

## **Acknowledgements**

I would like to personally thank my project supervisors Dr. Sarah Harris and Prof. Michelle Peckham for their endless support, knowledge, and guidance over the course of this project. This project would not have been possible without them. I would also like to thank Dr. Geoff Wells for his expert help parameterising the modifiable glutamate residues used throughout the project.

This work was undertaken on ARC4, part of the High-Performance Computing facilities at the University of Leeds, UK.

## Abstract

Microtubules are large, multimeric, hollow tubes made of tubulin found in all eukaryotic cells. They are integral to DNA segregation, organelle localisation and intracellular cargo transport. Microtubules are rich in post-translational modifications (PTMs), particularly along tubulin C-terminal tails, which modulate dynamics, protein recruitment and motor protein processivity. These modifications make up the *Tubulin Code*, which encodes a set of signals that result in specific and consistent changes in factors such as microtubule stability and levels of protein recruitment. The current understanding of each modification's effect is limited, due to limitations in both *in vitro* and *in silico* methods. Atomistic molecular dynamics (MD) simulations of microtubules have only been possible for the last 10 years, meaning aspects such as the effect of PTMs on microtubule dynamics are yet to be fully explored. This project presents a method for building and parameterising post-translationally modified microtubules for atomistic MD simulations. This was used to generate 100 ns simulations of a 7-protofilament model (*PF Sheet*) with either 10 residue poly(glycine) or poly(glutamate) chains. These were compared against an unmodified 13-protofilament microtubule simulation. These simulations showed the ends of both models curve away from the centre, caused by the plus end GDP-cap. Negatively charged poly(glutamate) chains positioned between protofilaments were found to interfere with inter-protofilament salt bridges, causing model to break apart. The smaller model flattened out during each simulation, which was not observed in the *Short MT* simulation, adding to existing evidence that microtubule fragment models do not accurately represent complete microtubules in atomistic simulations.

## List of Abbreviations

<b>Abbreviation</b>	<b>Meaning</b>
Å	Angstrom
ATAT	α-Tubulin acetyltransferase
CDK1	Cyclin-dependent kinase 1
CPP	Cytosolic carboxypeptidase
CPU	Central Processing Unit
<i>cryo-EM</i>	Cryogenic electron microscopy
CTT	C-terminal tail
Da	Dalton
DNA	Deoxyribonucleic acid
DYRK1A	Dual-specificity tyrosine-regulated kinase
EB1	End-binding protein 1
EMDB	Electron Microscopy Data Bank
<i>frmod</i>	Force field modification
GBSA	Generalised Born surface area
GDP	Guanosine-5'-diphosphate
GPU	Graphics Processing Unit
GTP	Guanosine-5'-triphosphate
HDAC6	Histone deacetylase 6
K	Kelvin
KCl	Potassium chloride
M	molar
MAP	Microtubule-associated protein
MD	molecular dynamics
Mg <sup>2+</sup>	Magnesium ion
MMGBSA	molecular mechanics generalized Born surface area
MMPBSA	molecular mechanics Poison-Boltzmann surface area
MT	Microtubule
MTOC	microtubule-organising centre
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PF	Protofilament
<i>PF Sheet</i>	Protofilament Sheet Model
PME	Particle Mesh Ewald
PMEMD	Particle Mesh Ewald Molecular Dynamics
<i>poly(E)</i>	poly(glutamate)
<i>poly(G)</i>	poly(glycine)
PRC1	Protein Regulator of Cytokinesis 1
PTM	Post-translational modification
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
SETD2	SET domain-containing protein 2
<i>Short MT</i>	Short Microtubule Model
SIRT2	Sirtuin 2
SUMO	Small ubiquitin-like Modifier
SVBP	Small vasohibin-binding protein
SYK	Spleen tyrosine kinase

<i>TIP3P</i>	Transferable Intermolecular Potential with 3 Points)
<i>TTL</i>	Tubulin-tyrosine ligase
<i>TLL</i>	Tubulin-tyrosine ligase-like protein
<i>TUBA</i>	Tubulin alpha gene
<i>TUBB</i>	Tubulin beta gene
<i>VASH</i>	Vasohibin

# Table of Contents

<i>Intellectual Property and Publication Statements</i>	2
<i>Acknowledgements</i>	3
<i>Abstract</i>	4
<i>List of Abbreviations</i>	5
<i>Introduction</i>	9
Microtubule Structure	9
Tubulin Diversity	12
Microtubule Dynamics	15
Post-translational modifications and the Tubulin Code	17
<i>In vitro vs. in silico</i>	21
Project Aims	25
<i>Materials and Methods</i>	26
Hardware, Software and Additional Parameters	26
Simulation Systems and Conditions	26
Initial Model Reconstruction	27
Parameterisation of Modifiable Glutamate Residue	28
Modifying Tubulin with poly(glycine) and poly(glutamate) chains	29
Building the <i>PF Sheet</i> and <i>Short MT</i> models	30
Building PTMs on multi-heterodimer models	30
<i>Results</i>	33
Structure of the <i>Short MT</i> and <i>PF Sheet</i> simulation models	33
Plus- and minus- end tubulin heterodimers bow away from the microtubule centre	38
<i>PF Sheet</i> and <i>Short MT</i> simulations exhibit differing model flexibility	41
$\beta$ -tubulin C-Terminal tails are the most mobile regions	45
Trapped poly( $\gamma$ E) chains sequester inter-protofilament interactions	48
GDP molecules diffuse from plus end $\beta$ -tubulins	50
<i>Discussion</i>	52
Changes to the current simulation systems	55
Future Experiments	59

<b>Improvements for Current Software</b>	<b>61</b>
<b>Conclusions</b>	<b>64</b>
<b>Bibliography</b>	<b>65</b>
<b>Appendices</b>	<b>73</b>
<b>Appendix A: Optimising AMBER input files for efficient simulations</b>	<b>73</b>
<b>Appendix B: Scripts/Input Files</b>	<b>75</b>
<b>Appendix C: PDB file formatting</b>	<b>83</b>

## Introduction

Microtubules are a class of filament found in all eukaryotic cells that play a key role in intracellular transport, localisation of organelles and segregating DNA strands during mitosis. Microtubules are made of  $\alpha$ - $\beta$  tubulin heterodimers, which polymerise in an alternating fashion to form protofilaments. These protofilaments are stacked next to each other in a staggered fashion, forming offset  $\alpha$ - $\alpha$  and  $\beta$ - $\beta$  inter-protofilament interactions. The resulting quaternary structure is a hollow tube, along which molecular motors like kinesin and dynein can travel (Fig. 1A and 1B). This review will focus on the structure of microtubules, how they assemble and disassemble, what kinds of post-translational modifications (PTMs) exist and how these modifications can alter interactions with binding partners.

### Microtubule Structure

The building blocks for all microtubules is the  $\alpha$ - $\beta$  tubulin heterodimer. The human genome contains at least 8  $\alpha$ -tubulin encoding genes and 10  $\beta$ -tubulin encoding genes, scattered across 10 different chromosomes (HUGO Gene Nomenclature Committee (HGNC), 2021; Tweedie et al., 2021). Each tubulin isoform has a slightly different mass, but all are close to 50 kDa (Margolis and Wilson, 1981). These isoforms are closely related in overall structure and sequence. Both bind GTP, mediated by a 7-residue nucleotide binding site (GGGTGSG) that occurs at residues 142-148 in  $\alpha$ -tubulin and 140-146 in  $\beta$ -tubulin (Cowan et al., 1983; Ranganathan et al., 1998). GTP bound to  $\alpha$ -tubulin is non-exchangeable, whereas GTP bound to  $\beta$ -tubulin is readily hydrolysed to GDP in the microtubule and exchanged with free GTP on non-polymerised tubulin heterodimers (Löwe et al., 2001).

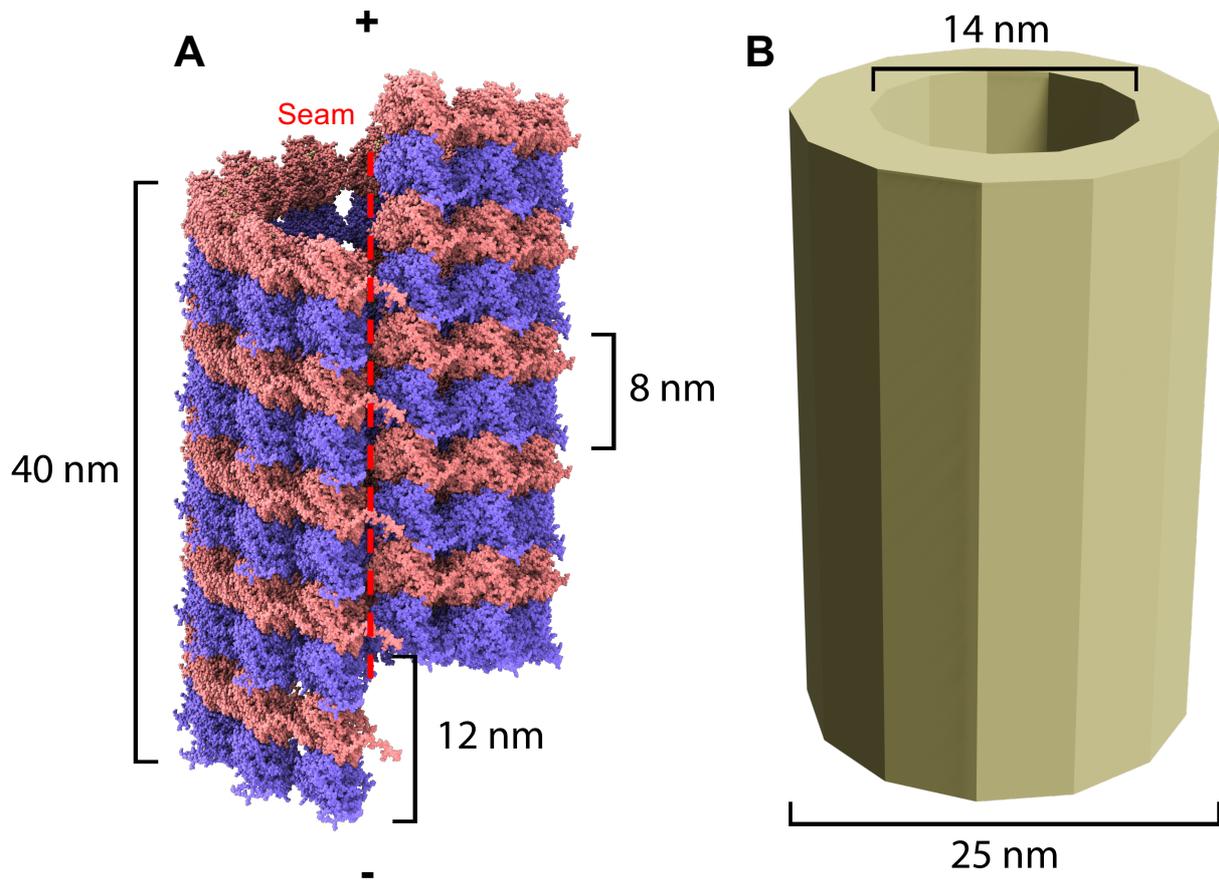
Individual tubulin heterodimers are stacked end-to-end to form long strands called protofilaments, which have an alternating  $\alpha$ -  $\beta$ - tubulin pattern. This asymmetry imbues protofilaments an inherent polarity, with  $\beta$ -tubulin at the growing plus end and  $\alpha$ -tubulin found at the opposite minus end. Protofilaments are stacked next to each other in parallel, forming

a hollow cylinder with  $\alpha$ - $\alpha$  and  $\beta$ - $\beta$  interactions between protofilaments. Lateral protofilament-protofilament interactions are mediated by salt bridges (a combination of ionic and hydrogen bonding), formed between the side chains of positively charged residues on the M-loops on one protofilament (e.g.,  $\alpha$ -tubulin Lys 280 or  $\beta$ -tubulin Arg 284) and negatively charged residues on the H1'-S2 (e.g.,  $\beta$ -tubulin Glu 55) and H2-S3 (e.g.,  $\alpha$ -tubulin Glu 90) loops on the neighbouring protofilament (Nogales et al., 1999; Löwe et al., 2001; Sui and Downing, 2010). These loop-loop salt bridges are further stabilised in  $\alpha$ -tubulin through stabilisation of the M-loop itself by H6 and the S9-S10 loop (Sui and Downing, 2010). Similar stabilisation of  $\beta$ -tubulin can be achieved through the use of paclitaxel (Taxol), an antitumour drug that binds to a taxane pocket formed by helix H7 and the H6-H7, S9-S10 and M loops (Löwe et al., 2001; Prota, Bargsten, et al., 2013).

Protofilament stacking is slightly staggered, allowing them to pack together more closely, giving the overall structure a helical arrangement. This creates a seam along the entire length of a microtubule, along which the  $\alpha$ -tubulin of one protofilament directly interacts with  $\beta$ -tubulin of the adjacent protofilament (Fig. 1A). A consensus on whether the lateral  $\alpha$ - $\beta$  interactions at the seam are weaker than in the rest of the lattice has not been reached, as evidence has been published that both supports (Katsuki et al., 2014; Alushin et al., 2014) and counters (Sui and Downing, 2010; Harris et al., 2018) this claim. Overall, the non-covalent linkages between protofilaments imbue microtubules with both structural rigidity and enough flexibility to resist mechanical breakage (Sui and Downing, 2010).

Microtubule polymerisation is a process that involves the addition of free tubulin heterodimers to the plus end of a microtubule. *In vitro* polymerisation occurs in three discrete steps: the slow nucleation phase in which many tubulin heterodimers form a stable aggregate; a rapid elongation phase in which heterodimers are rapidly added to the plus end of the new microtubule; and a steady state in which the rate of assembly (addition of heterodimers to the plus end) and disassembly (loss of heterodimers from the minus end) reaches equilibrium

(Alberts et al., 2002). In the steady state, tubulin heterodimers appear to travel along the microtubule, resulting in a phenomenon known as “treadmilling” (Margolis and Wilson, 1981).



**Figure 1. Dimensions of a microtubule.** (A) 3D representation of a microtubule, highlighting the approximate height of a tubulin heterodimer (8 nm), and a 5-heterodimer protofilament (40 nm). The helical pitch of the resulting microtubule is 12 nm.  $\alpha$ - and  $\beta$ - tubulin heterodimers are shown in purple and pink, respectively. The seam (dotted red line) occurs where the  $\alpha$ -tubulin along one protofilament is adjacent to the  $\beta$ -tubulin of the neighbouring protofilament. (B) Open tridecagonal (13-sided) prism representative of the simplified shape of a 13-protofilament microtubule, highlighting the inner and outer diameters of the molecule (14 nm and 25 nm, respectively).

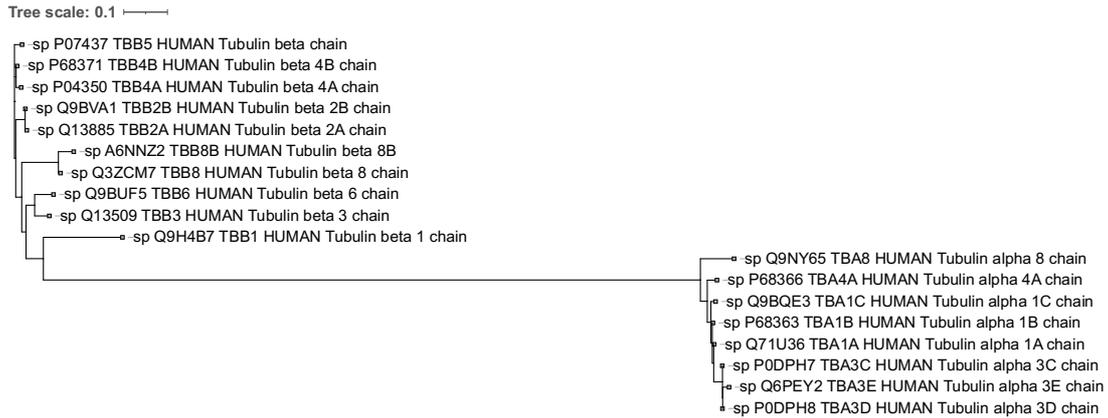
## Tubulin Diversity

Different variations of the main tubulin isoforms (e.g.,  $\alpha$ - and  $\beta$ - tubulin) are known as isotypes (e.g.,  $\alpha$ -tubulin 1A, 1B, 1C, etc.). Tubulin isotypes are encoded in different genes and have small variations in amino acid sequence but are still considered variants of a tubulin isoform (Fig. 2A). For example, TUBB and TUBB1 both encode  $\beta$ -tubulin isotypes, however TUBB is located on chromosome 6 and encodes a 444-residue  $\beta$ -tubulin isotype, whereas TUBB1 is located on chromosome 20 and encodes a 451-residue  $\beta$ -tubulin isotype (Deloukas et al., 2001; UniProt, 2001; Mungall et al., 2003; UniProt, 2004). Homology between the 8  $\alpha$ -tubulin and 10  $\beta$ -tubulin isotypes is relatively high. Within each isoform, several stretches of residues are identical between all isotypes, with many of the differences between isotypes arising from substitutions of similarly charged amino acids. The least similar pair  $\alpha$ -tubulin isotypes ( $\alpha$ 8 and  $\alpha$ 3E) share 89.51% of their amino acid sequences, whereas the least similar pair of  $\beta$ -tubulin isotypes ( $\beta$ 1 and  $\beta$ 8B) share 75.45% of their amino acid sequences (Fig. 2B). The sequence of the nucleotide binding site (GGGTGSG) is completely conserved across all isotypes of both isoforms of the protein.

Microtubules can be made up of any combination of tubulin isotypes, occurring at any point along a protofilament. Expression patterns for each isotype differ between cell types. For example, the  $\beta$ -tubulin 1 (TUBB1) gene is expressed specifically in platelets. Point mutations in this gene, leading to replacement (F260S, R307H or R318W) or nonsense (Q423\*) mutations, are known to result in disordered microtubule formation, and have the potential to lead to defective platelet production and macrothrombocytopenia (Kunishima et al., 2009; Kunishima et al., 2014; Fiore et al., 2017). Isotype composition can also influence the number of protofilaments which comprise a microtubule. Protofilaments made of  $\alpha$ 1B- $\beta$ 2B heterodimers tend to assemble into 14-protofilament microtubules, whereas ones made of  $\alpha$ 1B- $\beta$ 3 heterodimers show a preference for 13-protofilament microtubule formation (Ti et al., 2018). The  $\beta$ 3 tubulin isotype is expressed primarily in brain cells (NIH National Library of Medicine, 2023; Gene ID: 10381). Overexpression of this isotype is common in

certain types of cancers, which can cause issues as it is resistant to common taxane-based antitumor drugs, such as paclitaxel and docetaxel (Stengel et al., 2010; Ploussard et al., 2010; Maahs et al., 2019).

The C-terminal tail (CTT) is a highly disordered region of tubulin with higher variability in amino acid sequence (Fig. 3).  $\alpha$ -tubulin CTTs are between 10-13 residues in length, and often feature a terminal tyrosine residue, preceded by two glutamic acid residues.  $\beta$ -tubulin isotypes feature longer CTTs of between 18 and 24 residues and are typically capped with an alanine residue.  $\alpha$ - and  $\beta$ - tubulin CTTs are rich in glutamic acid residues, giving this region of the protein an overall negative charge. These glutamic acid residues are the sites of two important post-translational modifications: mono- or poly- glutamylation and glycylation (Redeker et al., 1994; Mary et al., 1997). Three different types of microtubules occur naturally. The most common is the singlet microtubule, generally comprising of 13 protofilaments, which is involved in the transport of large cargoes such as organelles in cells and the axonal transport of synaptic vesicles from neuron cell bodies to the synaptic cleft (Lasser et al., 2018). Microtubule doublets are found at the base of cilia and flagella and are composed of a complete 13 protofilament microtubule with an additional 9 protofilament segment fused to the outside, forming a second incomplete ring (Cavalier-Smith, 1974). Microtubule triplets are found in basal bodies and centrioles, and have two additional 10 protofilaments segments attached (Li et al., 2012).



### α-Tubulin Isootypes

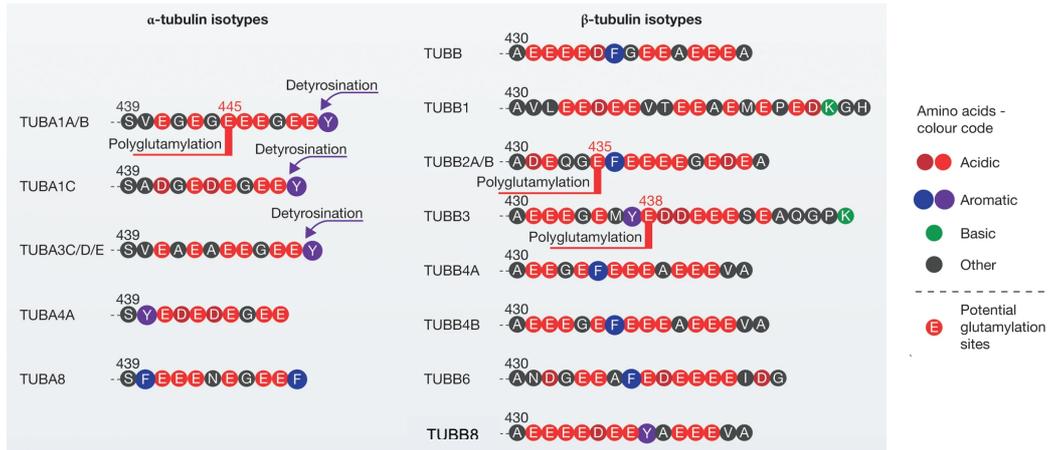
	TUBA8	TUBA3E	TUBA3D	TUBA3C	TUBA4A	TUBA1C	TUBA1B	TUBA1A
TUBA8	100.00	<b>89.51</b>	90.85	90.85	89.69	89.93	90.20	90.20
TUBA3E	<b>89.51</b>	100.00	98.67	98.67	93.06	94.87	95.78	96.22
TUBA3D	90.85	98.67	100.00	100.00	94.18	96.21	97.11	97.56
TUBA3C	90.85	98.67	100.00	100.00	94.18	96.21	97.11	97.56
TUBA4A	89.69	93.06	94.18	94.18	100.00	95.74	96.43	95.98
TUBA1C	89.93	94.87	96.21	96.21	95.74	100.00	98.44	98.44
TUBA1B	90.20	95.78	97.11	97.11	96.43	98.44	100.00	99.56
TUBA1A	90.20	96.22	97.56	97.56	95.98	98.44	99.56	100.00

### β-Tubulin Isootypes

	TUBB1	TUBB8B	TUBB8	TUBB6	TUBB3	TUBB2B	TUBB2A	TUBB	TUBB4B	TUBB4A
TUBB1	100.00	<b>75.45</b>	77.25	78.03	77.33	78.65	78.43	78.60	78.43	78.60
TUBB8B	<b>75.45</b>	100.00	96.17	83.56	84.23	86.49	86.26	86.94	87.84	86.91
TUBB8	77.25	96.17	100.00	85.36	86.04	88.29	88.06	88.74	89.64	88.71
TUBB6	78.03	83.56	85.36	100.00	92.38	90.56	90.34	90.77	90.79	91.67
TUBB3	77.33	84.23	86.04	92.38	100.00	91.69	91.46	92.57	92.81	92.12
TUBB2B	78.65	86.49	88.29	90.56	91.69	100.00	99.55	95.50	96.85	95.50
TUBB2A	78.43	86.26	88.06	90.34	91.46	99.55	100.00	95.05	96.40	95.05
TUBB	78.60	86.94	88.74	90.77	92.57	95.50	95.05	100.00	97.52	96.61
TUBB4B	78.43	87.84	89.64	90.79	92.81	96.85	96.40	97.52	100.00	98.65
TUBB4A	78.60	86.91	88.71	91.67	92.12	95.50	95.05	96.61	98.65	100.00



**Figure 2. Tubulin isotypes are closely related.** (A) Phylogenetic tree for 18 tubulin encoding genes (8 α-tubulin, 10 β-tubulin), generated using NGPhylogeny.fr (Lemoine et al., 2019). Branch length indicates genetic diversity, with longer branches indicating a higher degree of genetic change. (B) Percentage protein sequence identity matrices for 8 α-tubulin and 10 β-tubulin subunits, generated with Clustal 2.1 (Larkin et al., 2007). The lowest sequence identity percentages are highlighted with bold text.



**Figure 3. C-terminal tail sequences are variable between tubulin isotypes.** Adapted from Wehenkel and Janke (2014). Aligned sequences of the C-terminal tails of 8  $\alpha$ - and 10  $\beta$ - tubulin isotypes, highlighting the properties of each residue and potential modification sites. Residues are represented by their single-letter codes.

## Microtubule Dynamics

Microtubules nucleate from microtubule-organising centres (MTOC), found at different stages throughout the cell cycle as well as at the bases of flagella and cilia (Stearns et al., 1976; Brinkley, 1985). In interphase cells a single centrosome is positioned near the nucleus, from which microtubules sprout in all directions towards the cell periphery. The minus ends of MTOC-originating microtubules are anchored within the MTOC, with the plus ends extending outwards (Mitchison and Kirschner, 1984b). Neuronal microtubules are released from centrosomes (Yu et al., 1993). Axonal microtubule, originating from centrosomes, segments all have the same polarity, with plus ends pointed towards the presynaptic membrane, whereas in dendritic segments, originating from Golgi outposts, are more likely to exist in mixed orientations (Baas et al., 1988; Burton, 1988; Yu et al., 1993; Delandre et al., 2016). Basal bodies are the origin of doublet and triplet microtubules extending toward the tip of the protrusions (Li et al., 2012).

The length of microtubules *in vivo* is highly variable, stemming from a process called dynamic instability (Mitchison and Kirschner, 1984a). Microtubules experience events known as catastrophe and rescue. Catastrophe involves a growing microtubule suddenly shrinking at a rate of approximately 50 heterodimer lengths (8 nm) per second, caused by the hydrolysis of  $\beta$ -tubulin-bound GTP at the growing end (Margolin et al., 2012). *In vitro* studies have shown protofilaments capped with GDP- $\beta$ -tubulin at the plus end begin to peel away, causing short protofilament strands to be released and resulting in a swift decrease in length (Mandelkow et al., 1991).  $\beta$ -tubulin-bound GDP can dissociate from the nucleotide binding site, once depolymerised, allowing GTP to repopulate the site. Regenerated GTP- $\beta$ -tubulin can then be reincorporated into another microtubule, through polymerisation at the plus end, in a process known as rescue. Interactions between several loop regions on neighbouring protofilaments promote a more stable microtubule structure, improving resistance to catastrophe (Sui and Downing, 2010). Whilst growing microtubules are capped with GTP- $\beta$ -tubulin,  $\beta$ -tubulin subunits in the body of the microtubule contain GDP. The energy released from GTP hydrolysis by  $\beta$ -tubulin subunits is stored as structural strain within the microtubule lattice, which helps maintain the rigidity of the structure and can be harnessed to transport chromosomes towards the mitotic spindle during mitosis (Koshland et al., 1988; Igaev and Grubmüller, 2020).

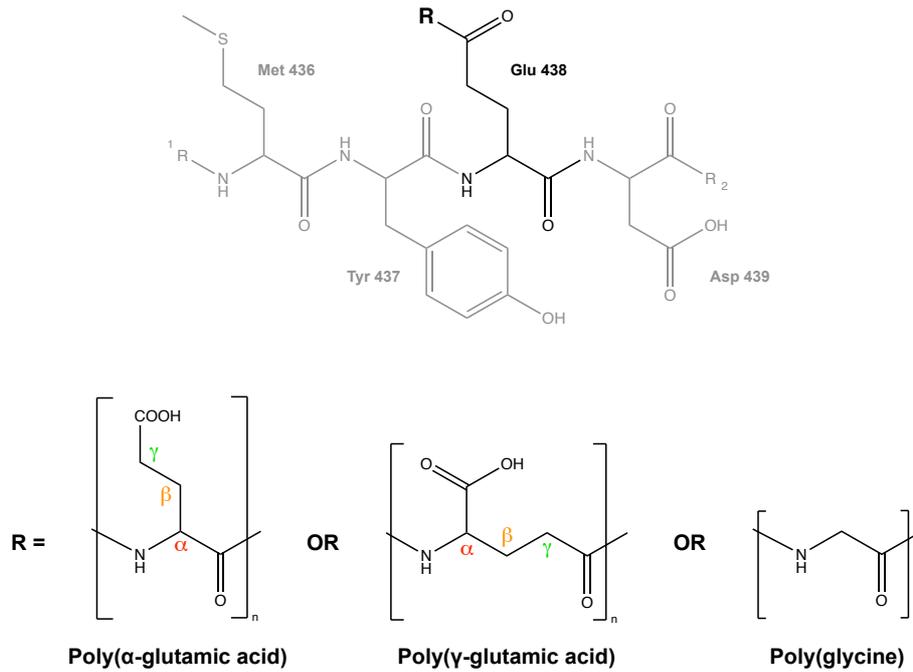
### Post-translational modifications and the Tubulin Code

Microtubules are highly post-translationally modified *in vivo* (Table 1). Polyglutamylation (poly(E)) and polyglycylation (poly(G)) can occur on any CTT glutamate residue, on both  $\alpha$ - and  $\beta$ - tubulin. Poly(E) tubulin helps upregulate kinesin binding and processivity, as well as modulating spastin (microtubule severing protein) recruitment (Sirajuddin et al., 2014; Valenstein and Roll-Mecak, 2016). Poly(G) tubulin is heavily associated with axonemal stability (Redeker et al., 1994; Grau et al., 2013), primary cilia formation in mammals and has been shown suppress colon tumourigenesis (Rocha et al., 2014). The length of these modifications, in some cases, can be longer than the CTTs themselves. Poly(E) chains of up to 21 residues and poly(G) chains of up to 40 residues have been observed (Schneider et al., 1998; Wall et al., 2016).

The N-terminal Glu or Gly residues of poly(E) and poly(G) PTMs is linked to the  $\gamma$ -carboxyl group of a CTT Glu residue via a  $\gamma$ -linkage (Wolff et al., 1994; Zhuang et al., 2022) (Fig. 4). Subsequent Gly residues in poly(G) PTMs are linked via the main chain of the residues ( $\alpha$ -linkage). The additional Glu residues in poly(E) PTMs can either be linked via the main chain (poly( $\alpha$ E)), or via the  $\gamma$ -carboxyl group of the previous Glu residue in the PTM (poly( $\gamma$ E)). poly( $\alpha$ E) is the most common isomer of the poly(E) PTM, however poly( $\gamma$ E) has been observed on neuronal tubulin CTTs (Wolff et al., 1994). The exact length of poly(E) chains modulates spastin-mediated microtubule severing through a graded, biphasic response curve (Valenstein and Roll-Mecak, 2016). Short poly(E) chains (1-8 residues) promote spastin activity, with activity improving as chain length tends towards 8 residues. Past 8 residues, spastin activity is inhibited, with longer poly(E) chains having a greater inhibitory effect (Valenstein and Roll-Mecak, 2016). It is even possible for tubulin heterodimers lacking polyglutamylation to be shielded from spastin-mediated severing through heavy glutamylation on neighbouring heterodimers (Valenstein and Roll-Mecak, 2016).

Modification sites	Forward enzymes	Reverse enzymes
<b>Acetylation</b>		
$\alpha$ -Tubulin Lys40 (LeDizet and Piperno, 1987)	$\alpha$ -Tubulin acetyltransferase 1 (ATAT1) (Akella et al., 2010; Shida et al., 2010)	Histone deacetylase 6 (HDAC6); sirtuin 2 (SIRT2) (Hubbert et al., 2002; North et al., 2003)
$\beta$ -Tubulin Lys252 (Chu et al., 2011)	San acetyltransferase (Chu et al., 2011)	Not known
<b>Methylation</b>		
$\alpha$ -Tubulin Lys40 (Park et al., 2016)	SET domain-containing protein 2 (SETD2) (Park et al., 2016)	Not known
<b>Detyrosination</b>		
$\alpha$ -Tubulin C-terminal Tyr residue	Vasohibin (VASH) proteins in complex with small vasohibin-binding protein (encoded by SVBP) (Aillaud et al., 2017; Nieuwenhuis et al., 2017; Adamopoulos et al., 2019; Liao et al., 2019; Li et al., 2019; Wang et al., 2019; Zhou et al., 2019)	Tubulin-tyrosine ligase (TTL) (Ersfeld et al., 1993)
<b>Generation of <math>\Delta</math>2-tubulin and <math>\Delta</math>3-tubulin (removal of C-terminal Glu residues from detyrosinated <math>\alpha</math>-tubulin)</b>		
$\alpha$ -Tubulin penultimate C-terminal Glu residues	Cytosolic carboxypeptidases (CCPs) (encoded by AGTPBP1, AGBL1, AGBL2, AGBL3, AGBL4 and AGBL5) (Kimura et al., 2010; Rogowski et al., 2010; Tort et al., 2014)	No reverse reaction known to date. Tyrosination of $\Delta$ 2-tubulin by TTL is not possible. (Paturle-Lafanechere et al., 1991; Prota et al., 2013)
<b>Glutamylolation or polyglutamylolation (addition of Glu to <math>\gamma</math>-carboxy group of Glu side chains and chain elongation by further addition of Glu residues)</b>		
$\alpha$ -Tubulin and $\beta$ -tubulin C-terminal tails (multiple Glu residues can be modified) (Eddé et al., 1990; Alexander et al., 1991; Rüdiger et al., 1992)	Tubulin-tyrosine ligase-like (TTLL) proteins, multiple members in most organisms (9 glutamylases in mammals) (Janke et al., 2005; Ikegami et al., 2006; van Dijk et al., 2007)	CCPs, multiple members in most organisms (6 deglutamylases in mammals) (Kimura et al., 2010; Rogowski et al., 2010; Tort et al., 2014)
<b>Glycylation or polyglycylation (addition of Gly to <math>\gamma</math>-carboxy group of Glu side chains and chain elongation by further addition of Gly residues)</b>		
$\alpha$ -Tubulin and $\beta$ -tubulin C-terminal tails (multiple Glu residues can be modified) (Redeker et al., 1994; Bre et al., 1996)	TTLL proteins, multiple members in most organisms (3 glycylation in mammals) (Wloga et al., 2009; Rogowski et al., 2009; Ikegami and Setou, 2009)	No reverse reaction or enzymes known
<b>Polyamination (addition of polyamines to the <math>\gamma</math>-carboxamide group of Gln side chains)</b>		
$\alpha$ -Tubulin and $\beta$ -tubulin, major modification site $\beta$ -tubulin Gln15 (Song et al., 2013)	Transglutaminases (Song et al., 2013)	No reverse reaction or enzymes known
<b>Phosphorylation (addition of phosphate group to Ser, Thr or Tyr)</b>		
$\beta$ -Tubulin Ser172 (Fourest-Lieuvin et al., 2006; Ori-McKenney et al., 2016)	Cyclin-dependent kinase 1 (CDK1); Dual-specificity tyrosine-regulated kinases (DYRK1A, Minibrain) (Fourest-Lieuvin et al., 2006; Ori-McKenney et al., 2016)	Not known
$\beta$ 3-Tubulin Ser444 (Ludueña et al., 1988)	Not known	Not known
$\alpha$ -Tubulin Tyr432 (Peters et al., 1996)	Spleen tyrosine kinase (SYK) (Peters et al., 1996)	Not known
$\alpha$ -Tubulin and $\beta$ -tubulin Tyr residues (not identified) (Akiyama et al., 1986; Matten et al., 1990)	Neuronal proto-oncogene tyrosine-protein kinase SRC (Akiyama et al., 1986; Matten et al., 1990)	Not known
<b>Ubiquitinylation (addition of ubiquitin to Lys residues of tubulin)</b>		
$\alpha$ -Tubulin, major modification site Lys304 (Ren et al., 2003; Huang et al., 2009)	Not known	No reverse reaction or enzymes known
<b>SUMOylation (addition of SUMO to Lys residues of tubulin)</b>		
$\alpha$ -Tubulin (modification site unknown) (Rosas-Acosta et al., 2005)	Not known	No reverse reaction or enzymes known
<b>Palmitoylation (addition of long-chain fatty acid palmitate to Lys residues of tubulin)</b>		
$\alpha$ -Tubulin, major modification site Lys376 (Ozols and Caron, 1997)	Not known	No reverse reaction or enzymes known

**Table 1. Tubulin post-translational modifications and associated enzymes. Taken from Janke and Magiera (2020).**



**Figure 4. Poly( $\alpha$ -) and poly( $\gamma$ -) glutamic acid side chains.** Skeletal structure of a segment of tubulin  $\beta 3$  CTT, showing a potential polyglutamylation site, and the possible isoforms of this PTM.  $R_1$  and  $R_2$  represent the rest of the protein sequence in the N- and C- terminal directions, respectively. Poly( $\alpha$ E) chains have peptide bonds connecting adjacent glutamate residues via the main chain. Poly( $\gamma$ E) chains have isopeptide bonds connecting their side chains to adjacent glutamates.

An important microtubule acetylation site is  $\alpha$ -tubulin K40, which is located inside the microtubule lumen. Acetylation at this site causes changes the conformation of the K40 loop, moving it 8 Å away from the neighbouring  $\alpha$ -tubulin M-loop and weakening any stabilising electrostatic interactions between the two regions (Eshun-Wilson et al., 2019). The result of these weaker interactions is a more flexible microtubule that is less likely to break due to mechanical stress (Eshun-Wilson et al., 2019).

Most  $\alpha$ -tubulin-associated genes encode for a C-terminal tyrosine residue. The presence or absence of this residue can affect which motor proteins bind to and walk along a given microtubule. Detyrosination is catalysed by vasohibins (VASH) in complex with a small vasohibin-binding protein (SVBP) (Nieuwenhuis et al., 2017). Kinesin-1, responsible for

anterograde cargo transport (minus end to plus end), has been shown to preferentially bind detyrosinated microtubules, and have increased velocity along detyrosinated microtubules when compared to tyrosinated and acetylated microtubules (Dunn et al., 2008; Kaul et al., 2014). The dynein/dynactin/BICD2 complex, responsible for retrograde cargo transport (plus end to minus end), has been shown to require tyrosination in order to initially bind to a microtubule, but is able to walk along microtubules that are detyrosinated after binding (McKenney et al., 2016). Complete detyrosination of a microtubule has been shown to protect it from kinesin-13 mediated depolymerisation, which recognises tyrosinated tubulin as its substrate (Peris et al., 2009). Detyrosination has recently been linked to the polyglutamylation of  $\alpha$ -tubulin, with the increase in negative charges localised to the CTT enhancing VASH/SVBP complex activity (Ebberink et al., 2022). Longer poly(E) chains have also been linked with increased VASH/SVBP efficiency (Ebberink et al., 2022).

The 'tubulin code' was first hypothesised by Verhey and Gaertig (2007). This theory states that a unique code is formed through variations in the isotype composition of a microtubule and the presence or absence of different PTMs along a microtubule. Differences in these two factors can directly regulate the structure and dynamics of a microtubule, by modulating stability and resistance to mechanical strain, and indirectly through the recruitment of microtubule-associated proteins (MAPs) that either stabilise, destabilise, or depolymerise the lattice (Verhey and Gaertig, 2007).

***In vitro vs. in silico***

Microtubules are often studied *in vitro*, where exact cellular conditions cannot be perfectly replicated for several reasons. *In vitro* experiments may not include one or more adapter/helper proteins, that may have not even been discovered yet, normal be present *in vivo* (Janke and Magiera, 2020). Little is known about how combinations of PTMs interact with each other, whether constructively or destructively. In a few cases, specific tubulin isoforms have been identified as important in certain cell types, however we lack a complete understanding of why each isoform is present in what amount across microtubule populations and between different cell types. Finally, it can be difficult to control the location and incidence of PTMs along a single microtubule, and across a population of microtubules experimentally. Whilst the number of forward and reverse enzymes discovered to interact with tubulin is ever increasing, we still do not have a full list. One or more of these issues have the potential to cause the results of *in vivo* and *in vitro* studying the same PTM to produce contradicting results (Janke and Magiera, 2020).

There are some examples of MD simulations being used to investigate microtubules, however due to technical limitations, it has not been possible to produce atomistic simulations of microtubules until the last 5-10 years, meaning this technique has been used less frequently than *in vivo* or *in vitro* techniques. In one of the few examples of studying the effects of tubulin C-terminal tail PTM composition, Bigman and Levy (2020) showed that End-Binding Protein 1 (EB1) and Protein Regulator of Cytokinesis 1 (PRC1) diffuse more slowly along polyglutamylated microtubules when compared to unmodified microtubules, and more quickly along polyglycylated microtubules. These modifications have been shown to have a preference for protruding away from the tail and collapsing in on themselves, respectively, though the diffusion of EB1 along microtubules has since been shown to not be affected by the length or position of the corresponding PTMs (Bigman and Levy, 2021).

Recent multi- $\mu\text{s}$  length, atomistic simulations of GTP- and GDP- capped microtubules have provided evidence that a GDP-cap prevents further microtubule growth by increasing the activation energy of straight lattice formation (Igaev and Grubmüller, 2022). MD simulations have also been used in conjunction with other *in silico* and *in vivo* techniques to determine the mechanism by which C-terminal tails are glutamylated, the effect of  $\alpha$ -tubulin tail  $\Delta$  modifications (detyrosination,  $\Delta 2$  and  $\Delta 3$ ) on microtubule stability and assembly, and how the properties of the tubulin heterodimer affect microtubule self-assembly (Natarajan et al., 2017; Mani and Subramanian, 2021; Nasedkin et al., 2021).

Using software packages such as AMBER (Case et al., 2021b) or GROMACS (Bekker et al., 1993), it is possible to simulate and visualise sections of microtubules composed of specific tubulin isotypes at an atomic level, with complete control over the number and location of PTMs. This allows us to study the effect of adding PTMs in predefined locations, without the need for additional enzymes to add modifications, and the controls necessary to restrict onto which residue a PTM is added. All-atom molecular dynamics simulations use Newtonian physics to approximate the behaviour of atoms (Dror et al., 2012). The atoms within a given system are represented by particles, to which Newton's second law is applied (equation 1). The forces applied to these particles are dictated by a force field (equations 2 and 3) (Fig. 5) designed using a combination of computational and experimental data (MacKerell et al., 1998; Dror et al., 2012). These force field equations ensure that the molecular geometry of a simulated protein (bond length, bond angle and torsion angles) agrees with the experimentally determined values for each type of bond. This is to prevent atoms from getting too close to each other or overlapping. Dihedral angles along the backbone of a protein are also considered, so that stabilising secondary structural features present at the start of a simulation are maintained, if left undisturbed. These equations also factor in non-bonded forces, like van der Waals forces, to ensure electrostatic attraction between pairs of atoms is possible, whilst preventing the formation of steric clashes.

$$F = ma$$

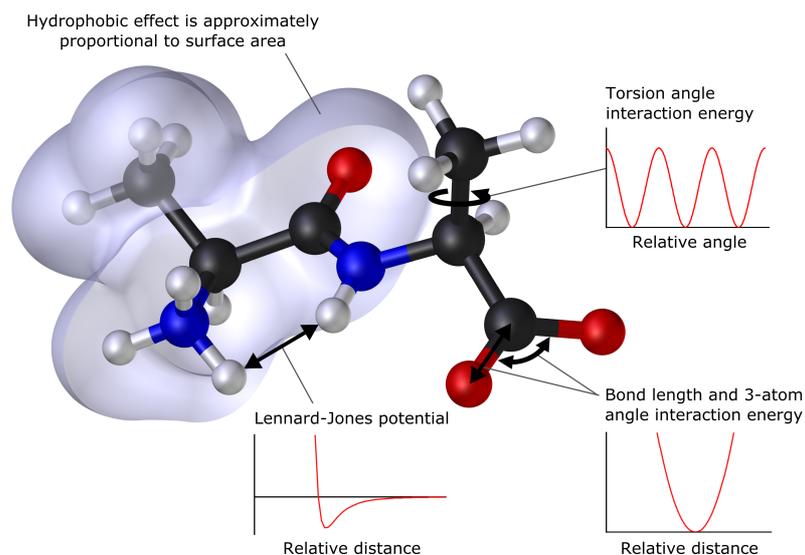
**Equation 1. Newton's Second Law.**  $F$  = force (N);  $m$  = mass of an object (kg);  $a$  = acceleration ( $m/s^2$ ).

$$\text{Total Force} = \text{Bonded forces} + \text{van der Waals forces} + \text{Electrostatic forces}$$

**Equation 2. General Force Field Equation.** 'Bonded forces' are interactions between groups of atoms connected by covalent bonds; 'van der Waals forces' are weak, short-range interactions which can induce temporary dipole-dipole interactions; 'Electrostatic forces' occur between pairs of atoms, with a much larger range than van der Waals.

$$V = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)] \\ + \sum_{\text{nonbonded}} \left( \epsilon \left[ \left( \frac{R_{\text{min}ij}}{r_{ij}} \right)^{12} - \left( \frac{R_{\text{min}ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}} \right)$$

**Equation 3. Force Field Energy Function. Adapted from MacKerell et al. (1998).**  $V$  is total force;  $b$  is bond length;  $\theta$  is bond angle;  $\phi$  is the dihedral angle; 0 subscripted versions of these parameters represent their equilibrium values;  $K_b$ ,  $K_\theta$  and  $K_\phi$  are the respective force constants. Nonbonded forces are calculated by combining the Coulomb and Lennard-Jones terms.  $\epsilon$  is Lennard-Jones well depth;  $\epsilon_1$  is the effective dielectric constant;  $R_{\text{min}}$  is distance at the Lennard-Jones minimum.  $q$  represents partial atomic charges for atoms  $i$  and  $j$ ;  $r$  is the distance between the atoms.



**Figure 5. Physics on a molecular scale. Taken from Dhatfield (2007).** Graphical the location and ranges of the different physical properties of a molecule.

Molecular dynamics (MD) simulations work by calculating the movement of particles over very short time increments called timesteps (typically no larger than 2 fs). At each timestep, the computer will alternate between calculating the forces exerted on each particle in the system, using the chosen force field, and calculating the resulting change in each particles velocity and position. This information is written to a trajectories file which can then be opened in viewing software, such as VMD (Humphrey et al., 1996), as an animated 3D model of the protein.

The goal of using molecular dynamics simulations is to visualise tubulin and microtubules (modified and unmodified) in motion and in atomistic detail, something that is impossible with current experimental techniques. This gives us the best chance at seeing any patterns emerge linking PTM or isoform composition to specific structural or dynamic behaviours, enabling us to better predict how microtubules might behave under the same conditions *in vivo* or *in vitro*.

## Project Aims

The aim of this project was to learn more about, and potentially contribute to, the idea of the tubulin code by investigating the effects that different PTMs have on microtubule structure and dynamics, particularly focussing on modifications to tubulin C-terminal tails. This was achieved using AMBER (Case et al., 2005) to run atomistic MD simulations of two representative microtubule models: the *short microtubule (Short MT)* model, a 13 protofilament microtubule model with a height of 2 tubulin dimers; and the *protofilament sheet (PF Sheet)* model, a cut-down version of the *Short MT* model containing 7 protofilaments in a semicylindrical arrangement. Studies into the effects of different PTMs were done using the *PF Sheet* model. A 7 protofilament model was chosen to provide a balance between improving simulation performance by significantly reducing the number of atoms per system, whilst maintaining the curvature of a complete 13 protofilament microtubule. These simulations were compared to an unmodified *Short MT* simulation to validate whether the *PF Sheet* model was an accurate representation of a complete microtubule.

The structure used as the basis for these models was PDB ID: 5ij0, a 3.80 Å cryo-EM resolved structure of a human tubulin heterodimer ( $\alpha$ 1B,  $\beta$ 3), stabilised with paclitaxel (not present in the deposited structure) and decorated with kinesin 1 $\Delta$ C (not present in the deposited structure) (Ti et al., 2016). It was selected as  $\beta$ 3-tubulin is a notable isotype that is overexpressed in malignant cells. Gaining a better understanding of how PTMs affect microtubules predominantly containing this isotype could eventually lead to new anti-cancer drugs that can stabilise them, providing an alternative to taxane-based drugs.

The PTMs chosen for this project were poly(glycine), poly( $\alpha$ -glutamate) and poly( $\gamma$ -glutamate) (poly(G), poly( $\alpha$ E) and poly( $\gamma$ E), respectively). Glutamate has a net charge of -1 at physiological pH (PubChem, 2022), so we would expect the increase in negative charges around the CTTs to affect microtubule structure and dynamics more than a glycine chain of similar length. To enable us to build additional polypeptide chains onto tubulin, we

parameterised a modifiable glutamate residue for use with AMBER, onto which poly(G) and poly(E) chains of any length can be built. This also helped to create a workflow for generating other modifiable amino acids, for PTMs such as SUMOylation of lysine residues, in the future (Feng et al., 2021). To study the effect of adding poly(G) or poly(E) chains to microtubules, MD simulations were run using a representative model made of 7 protofilaments, each comprised of 2 tubulin heterodimers, organised into a hollow semicylinder. This provided an approximation of microtubule behaviour in an atomistic MD simulation, whilst reducing the computing resources and time required to complete each simulation.

## Materials and Methods

### Hardware, Software and Additional Parameters

AMBER MD simulations were performed on ARC4 (Leeds). ARC4 GPU nodes are equipped with an Intel® Xeon® Gold 6138 CPU, 192 GB of DDR4 RAM and 4 NVIDIA® V100 GPUs. MD simulations were run using the Particle Mesh Ewald Molecular Dynamics engine (PMEMD) implementation in AMBER 20 (Case et al., 2005; Case et al., 2020). Systems were prepared using the ff14SB force field in *LEaP* and *Parmed* from the AmberTools21 software package (Maier et al., 2015; Case et al., 2021b). Root-Mean-Square Deviation (RMSD) and Root-Mean-Square Fluctuation (RMSF) calculations, and inter-protofilament measurements were carried out in *cpptraj* (AmberTools21). MD data was visualised using VMD 1.9.3 (Humphrey et al., 1996) and ChimeraX (Goddard et al., 2018). AMBER force field parameters for GTP and GDP were generated by Meagher et al. (2003). Parameters for  $Mg^{2+}$  were generated by Allnér et al., (2012). Python scripts for optimising the workflow were written in Python 3.9.13 (van Rossum and L. Drake, 2009).

### Simulation Systems and Conditions

7 simulations were run using the *Protofilament Sheet* (PF *Sheet*) model, a semicylindrical, 7 protofilament model with a height of 2 tubulin heterodimers. 10 residue PTMs of poly(G), poly( $\alpha$ E) or poly( $\gamma$ E) were added to either Glu 446 ( $\alpha$ -tubulin) or Glu 445 ( $\beta$ -tubulin) on each

tubulin heterodimer in the model. These were compared against simulations using an unmodified version of the *PF Sheet* model, and a larger, 13 protofilament model, the *Short Microtubule (Short MT)* model. Each system was solvated with water using the TIP3P (transferable intermolecular potential with 3 points) water model implementation (Jorgensen et al., 1983), specifying a cuboid periodic box shape to best match the semicylindrical and cylindrical shapes of the *PF Sheet* and *Short MT* models, respectively. The minimum solute-to-box-edge distance was set to 10 Å. Bounding box volume varied from  $1.218 \times 10^6 \text{ Å}^3$  (99.530 Å x 97.951 Å x 124.933 Å) to  $1.490 \times 10^6 \text{ Å}^3$  (121.742 Å x 97.951 Å x 124.933 Å) for *PF Sheet* simulations, depending on the length and angle of any additional PTM chains. The bounding box volume for the *Short MT* simulation was  $3.109 \times 10^7 \text{ Å}^3$  (306.349 Å x 310.762 Å x 326.581 Å). K<sup>+</sup> and Cl<sup>-</sup> ions were added to a final concentration of 150 mM using the *addionsrand* function in LEaP, to approximate *in vivo* conditions. Each system was heated to 300 K and equilibrated with any protein atoms restrained to minimise the energy in the system. Production MD was run using a 2 fs timestep for a total of 100 ns, following preliminary tests to improve the performance and efficiency of the simulations (appendix A).

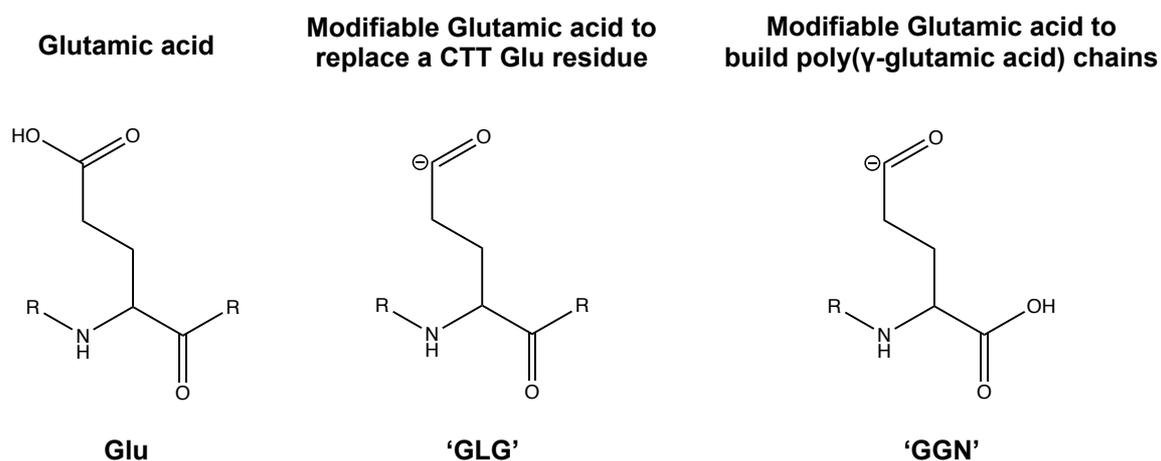
### **Initial Model Reconstruction**

The PDB model selected as a base for this project was PDB ID: 5ij0 (Ti et al., 2016). The model contains GTP and Mg<sup>2+</sup> in the  $\alpha$ -tubulin nucleotide binding site and GDP in the  $\beta$ -tubulin site. The 5ij0 PDB file was obtained from the RCSB Protein Data Bank (Berman et al., 2000). The model was missing 3 segments of its structure: residues 38-44 ( $\alpha$ -tubulin K40 loop), corresponding with a short, unstructured loop found between helices H1 and H2, 438-451 (CTT) of the  $\alpha$ -tubulin subunit and residues 427-450 (CTT) of the  $\beta$ -tubulin subunit. These three regions are positioned at either the outer (CTTs) or inner (K40 loop) solvent interface, causing them to be highly mobile. This makes it near impossible to assign the residues to any low-resolution density in these areas. The missing segments of the polypeptide chain were added back in using Clustal Omega (Goujon et al., 2010) and Modeller (Šali and Blundell, 1993).

The sequence for 5ij0 (i.e., the incomplete sequence) was extracted from the PDB file and the complete sequences for human TUBA1B and TUBB3 were taken from UniProt (Accession numbers: P68363; Q13509) and combined to create a single sequence for the heterodimer. The incomplete and complete sequences were added to a *fasta* file and aligned using Clustal Omega. This created an output file containing the incomplete sequence with dashes added for missing residues. A total of 4 PDB files were generated using *LoopModel.py* in *Modeller*, which were then ran through to *MolProbity* (Chen et al., 2010).

### Parameterisation of Modifiable Glutamate Residue

The branched amino acid structure of tubulin CTTs modified with poly(G) and poly(E) is non-standard and required a bespoke AMBER modification. This was carried out by Sarah Harris and Geoff Wells using the *Simulating a Solvated Protein that Contains Non-Standard Residues* procedure by Walker (2008). Two versions of a modifiable Glu residue, 'GLG' and 'GGN', were produced (Fig. 6).



**Figure 6. Modifiable Glutamate Residues with Unsatisfied Valences.** Glu is the built-in residue for glutamate in AMBER; 'GLG' is missing the OH group from its side chain carboxyl group (now a carbonyl group); 'GGN' lacks the OH group in its side chain carboxyl group but adds an OH group to the main chain carboxyl group, terminating its main chain.

'GLG' was generated to allow us to build an additional Gly or Glu residue onto the side chain of 'GLG' in the N to C direction. It featured an unsatisfied valence on the  $\gamma$ -carbonyl group that could be used to form a peptide bond with an additional amino acid, enabling the construction of 10 residue poly(G) and poly(E) chains onto tubulin C-terminal tails. 'GGN' was generated to allow us to build  $\gamma$ -linked poly(E) chains by forming a peptide bond between the unsatisfied valence and the amide group of another 'GGN' molecule. It features an unsatisfied valence on the  $\gamma$ -carbonyl group, and a C-terminal carboxylic acid group along the main chain instead of a non-terminal carbonyl group.

### **Modifying Tubulin with poly(glycine) and poly(glutamate) chains**

The completed PDB file containing the three missing segments was edited so that the chosen glutamate residue could be modified in XLEaP (example input scripts are printed in appendix B<sub>1</sub> and specific changes to PDB files are printed in appendix C). Glu 446 was changed to 'GLG' in  $\alpha$ -tubulins, and Glu 445 (Glu 896 in the PDB file) was changed to 'GLG' in  $\beta$ -tubulins. The force field modification (frcmod) and prep files for 'GLG' and 'GGN' were loaded into XLEaP, followed by the newly modified PDB file. The *sequence* command was used to create a separate polypeptide for the PTM. The carbon at the end of 'GLG' side chain was then bonded to the N-terminal nitrogen of the PTM chain using the *bond command*. In the *edit* window, the PTM was positioned so that the bond between it and the protein was of an appropriate length. The *relax selection* function was then used on the PTM and CTT atoms to ensure the length and angles of the bond were correct. Sets of coordinates and topology files exported for tubulin modified with poly( $\alpha$ E), poly( $\gamma$ E), and poly(G).

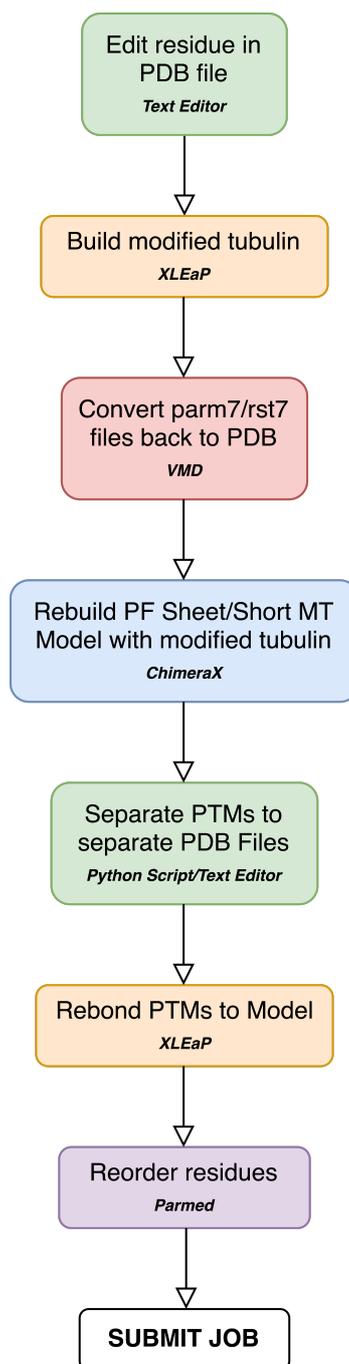
### **Building the *PF Sheet* and *Short MT* models**

The completed PDB file was fitted into the electron density map associated with PDB ID 5ij0 (Ti et al., 2016) with the *Fit in Map* function within *ChimeraX*. The density map was obtained from the Electron Microtubule Data Bank (EMDB code: EMD-8094) (Lawson et al., 2016). A total of 65 copies of the heterodimer structure were fitted, forming a 13-protofilament microtubule with a height of 5 heterodimers (~40 nm). Two models with reduced sizes were then created to minimise the number of atoms within simulation systems, improving simulation rate in the process. The *Short MT* model was made from the bottom two heterodimers from all 13 protofilaments, forming a ~16 nm tall microtubule. The *PF Sheet* model is a cut-down version of the *Short MT* model, made of 7 protofilaments that form a semicylindrical sheet. Both models have GTP populating each  $\alpha$ -tubulin nucleotide binding site and GDP populating each  $\beta$ -tubulin site, mimicking the nucleotide state of a microtubule experiencing catastrophe.

### **Building PTMs on multi-heterodimer models**

Coordinates and topology files for modified tubulin were converted to PDB files using *VMD*. The PDB file was then cleaned using a text editor, so that the file interpretable in XLEaP. Either the *PF Sheet* or *Short MT* model was then loaded into *ChimeraX* as a scaffold, along with the corresponding number of copies of the modified tubulin PDB files (14 for the *PF Sheet* model, 27 for the *Short MT* model). Each of the modified tubulins was fitted to a different tubulin heterodimer in the scaffold model using the *MatchMaker* command. A new modifiable PDB file was saved containing only the modified tubulins with *Save relative to model* enabled, using the scaffold as a reference (see example ChimeraX batch script in appendix B<sub>ii</sub>). All hydrogen atoms were removed from the modifiable PDB file using the *-trim* command line flag for the *Reduce* (Word et al., 1999). Each PTM chain was separated into its own PDB file using the *ptm-split.py* script (appendix B<sub>iii</sub>). The position of each PTM relative to the corresponding tubulin heterodimer was retained.

The modifiable PDB was imported into XLEaP. The PTM sequence was then added as a unit before importing each tail PDB with the *loadpdbusingseq* function, using the previously defined sequence as the input sequence. The tails were then re-bonded to the corresponding tubulin heterodimers, the system was solvated, salt ions were added, and topology and coordinates files were generated (see example XLEaP input script in appendix B<sub>IV</sub>). The topology and coordinates files were loaded into *Parsed* (AmberTools21) to reorder the atoms so that they are contiguous using the *CheckValidity* function. The reordered topology and coordinates files were then used as the input for AMBER simulations. The full workflow for preparing and building a multi-heterodimer model with modifications is shown in figure 7. PMEMD input files and ARC4 submission file information is detailed in appendix B<sub>V</sub> and B<sub>VI</sub>, respectively.



**Figure 7. Summarised method for modifying homooligomeric systems.** Steps that use a text editor or *Python* script are shown in green; *XLEaP* steps are shown in orange; *VMD* steps are shown in red; *ChimeraX* steps are shown in blue; *Parmed* steps are shown in purple.

## Results

### Structure of the *Short MT* and *PF Sheet* simulation models

A 901 residue  $\alpha$ 1B- $\beta$ 3 tubulin heterodimer model was generated after adding three missing segments of protein using *Modeller* (Fig. 8). The  $\alpha$ -tubulin subunit was made of 451 residues and had a GTP molecule bound in its nucleotide binding site, with an accompanying  $Mg^{2+}$  positioned to stabilise the  $\beta$ - and  $\gamma$ - phosphate groups. The  $\beta$ -tubulin subunit was made of 450 residues and had a GDP molecule bound in its nucleotide binding site (Fig. 9A). The restored  $\alpha$ -tubulin K40 loop (residues 38-44) underwent 4 iterations of loop structure refinement (Fig. 9B), whereas the structures of the restored CTTs were not refined past the initial model (Fig. 9C). The structure of the K40 loop was equally plausible in each of the iterations since this region is unstructured and no improper torsion angles were occupied. Any of the 4 structures could have been used to produce simulations of a similar level of accuracy. Ultimately, iteration 4 (Fig. 9B; cyan) was selected as the constituent model for the multi-tubulin models. This completed heterodimer model was then run through the '*Asn/Gln/His flips*' function in *MolProbity* to correct for steric overlaps originating from Asn, Gln or His residues. A total of 19 residues were flipped.

**$\alpha$ -tubulin 1B**

```

5ij0_incomplete  MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMP-----GDDSFNTFFSETGAGK 53 /
5ij0_complete    MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK 60 /
*****

5ij0_incomplete  AREDMAALEKDYEEVGV----- 430
5ij0_complete    AREDMAALEKDYEEVGVDSVEGEGEEGEEY 451
*****

```

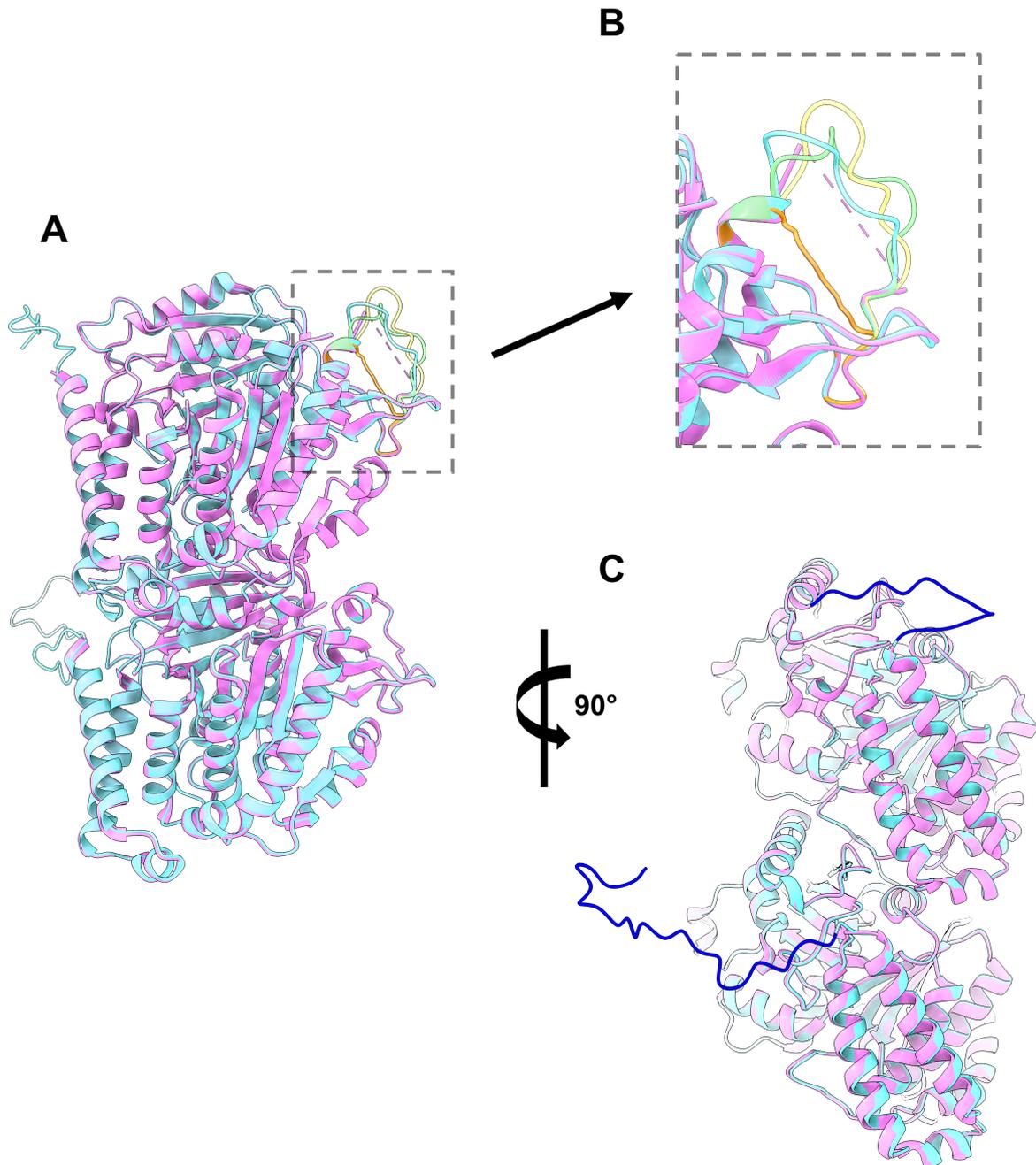
 **$\beta$ -tubulin 3**

```

5ij0_incomplete  RKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQ----- 856
5ij0_complete    RKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAEEGEMYEDDEESEAQGPK 901
*****

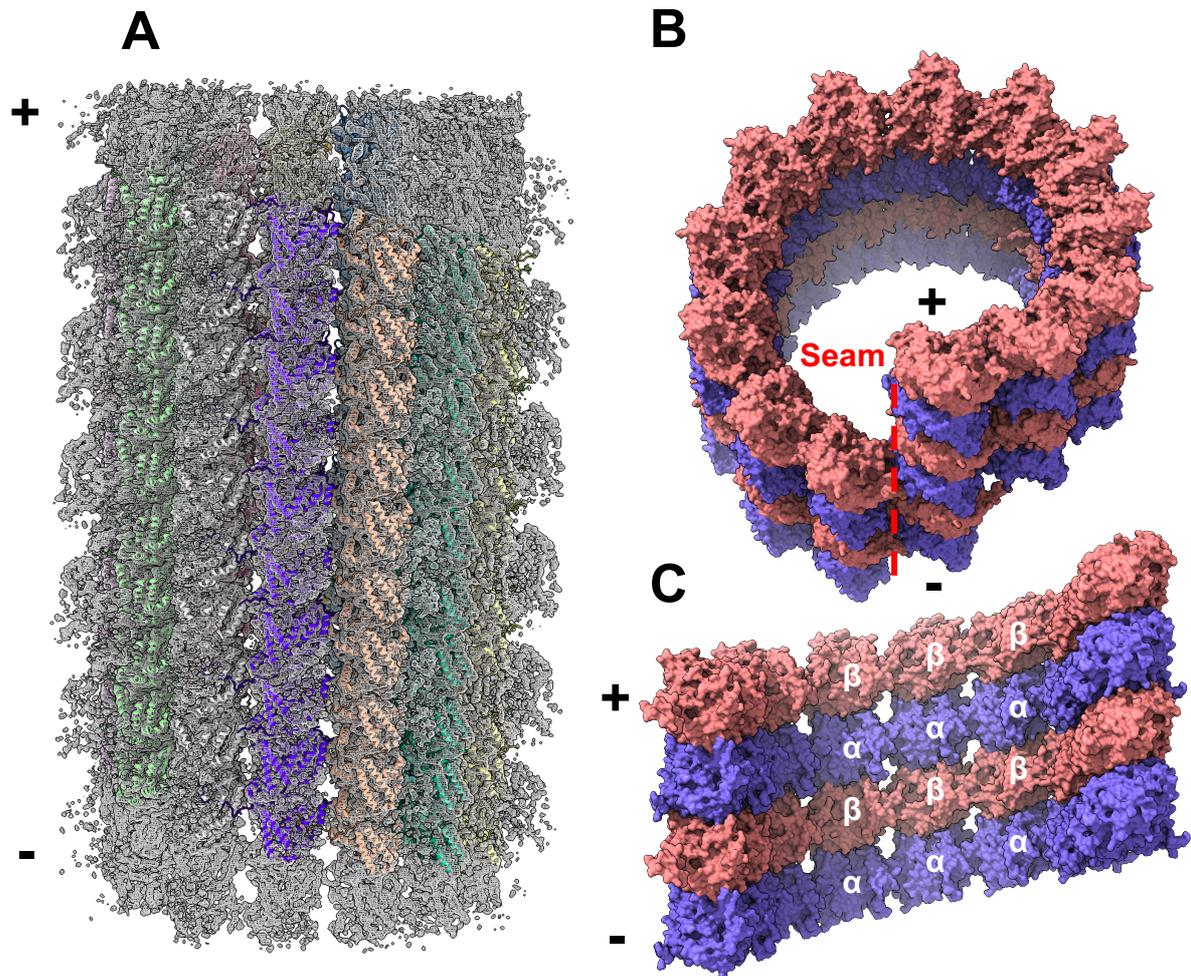
```

**Figure 8. Sequence alignment of 5ij0 (PDB structure for the  $\alpha$ 1B- $\beta$ 3 heterodimer, obtained by cryo-EM; Ti et al., 2016) and the known complete sequence for the  $\alpha$ 1B- $\beta$ 3 tubulin heterodimer. The missing K40 loop ( $\alpha$ -tubulin residues 38-44) and CTTs ( $\alpha$ -tubulin residues 438-451;  $\beta$ -tubulin residues 427-450) in 5ij0 and the corresponding amino acids in the complete sequence are highlighted in cyan. *5ij0\_incomplete* is the protein sequence for the 5ij0 PDB structure. *5ij0\_complete* is the known complete sequence for the  $\alpha$ 1B- $\beta$ 3 heterodimer. \* Indicates identical residues; - indicates missing residues; / indicates a break in sequence contiguity.**



**Figure 9. Comparing the 5ij0 PDB structure (Ti et al., 2016) to the completed *Modeller* structures.** (A) 5ij0 (magenta) overlaid with the *Modeller* generated PDB file (cyan), aligned in *ChimeraX* using the MatchMaker function. (B)  $\alpha$ -tubulin residues 38-44 from the 4 loop refinement steps, generated using *Modeller* (cyan, green, yellow, orange). The dashed line in the magenta structure represents residues 38-44, which are not present in the deposited PDB structure (5ij0). The position of segment differs greatly between models. It is unlikely any of these conformations are representative of this segments position in vivo, due to its high mobility/flexibility. (C) Rotated view of (A), highlighting the added CTTs dark blue for  $\alpha$ - (439-451) and  $\beta$ - (430-450) tubulin.

The completed heterodimer model was fitted into the 5ij0 electron density map to create a microtubule model with 13 protofilaments, each with a height of 5 heterodimers (65 total heterodimers) (Fig. 10A). This model was then reduced in size by removing heterodimers to create 2 models: the *Short MT* model (Fig. 10B) and the *PF Sheet* model (Fig. 10C). The *Short MT* model consisted of a 13 protofilament microtubule with a height of 2 heterodimers. An additional heterodimer was left at the minus end of PF13 to improve stability at the seam, leaving a total of 27 heterodimers in the model. The *PF Sheet* model was made by further reducing the *Short MT* model. It contained 7 protofilaments with a height of 2 heterodimers, for a total of 14 heterodimers, arranged into a semicylinder to emulate the curvature of a microtubule.



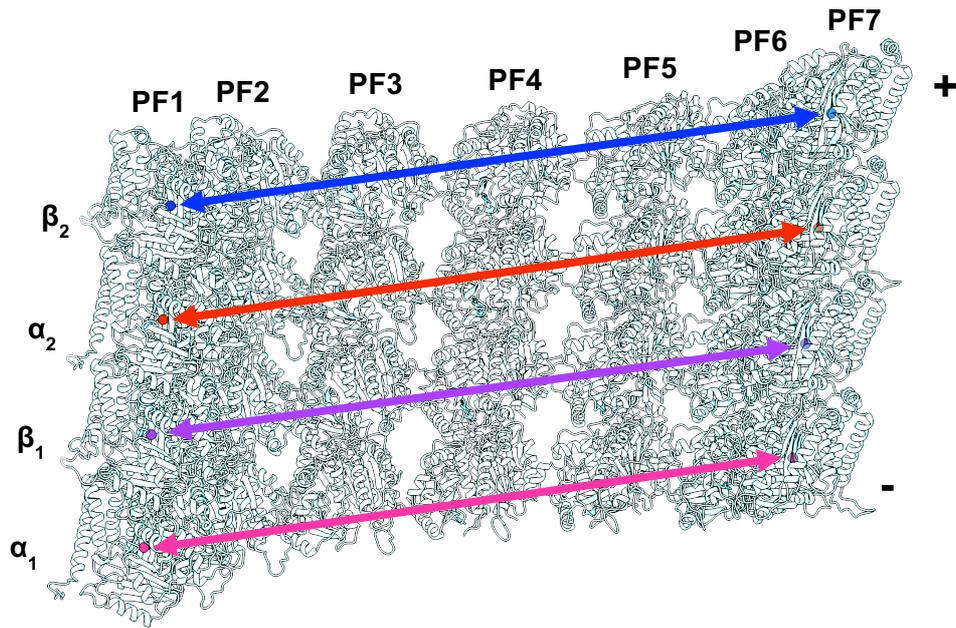
**Figure 10. The *Short MT* and *PF Sheet* models: representative microtubule models intended for simulation:**  
 (A) Reconstructed tubulin heterodimers (cartoon; multiple colours) fitted into the electron density map (grey mesh) deposited with PDB ID 5ij0 (Ti et al., 2016) (EMDB code: EMD-8094). (B) The *Short MT* model.  $\alpha$ -tubulins are in purple;  $\beta$ -tubulins are pink; The plus end is at the top of the model. The seam (red dotted line) occurs to the left of the three-dimer protofilament. (C) The *PF Sheet* model. The colour scheme is identical to (B).

### **Plus- and minus- end tubulin heterodimers bow away from the microtubule centre**

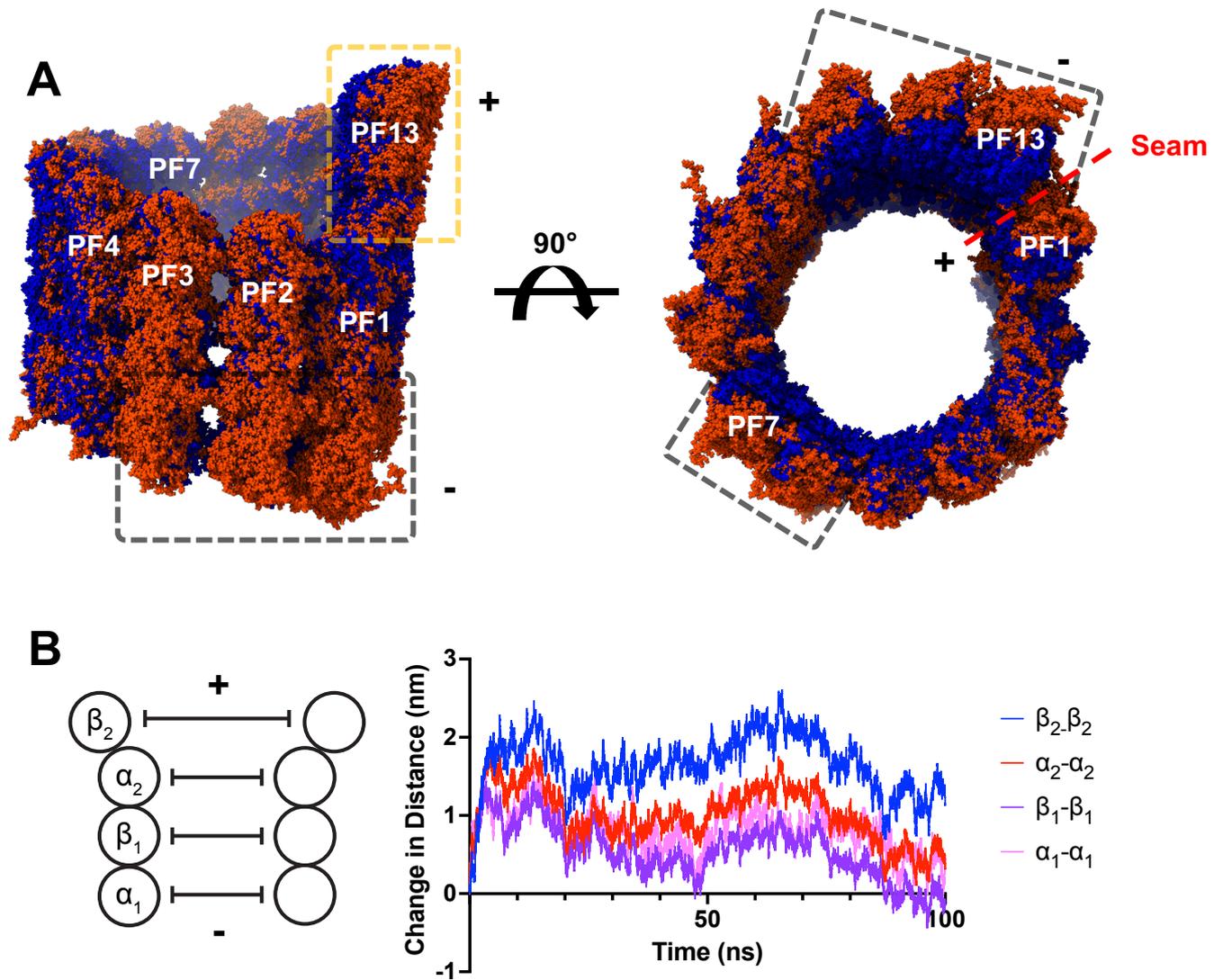
The distance between heterodimer subunits on protofilaments 1 and 7 of the *PF Sheet* and *Short MT* models was measured to track the overall model shape over the course of simulations (Fig. 11). The centre of mass for each subunit of PF1 and PF7 was calculated. Then, distance measurements were taken between corresponding centres of mass for each frame in the output trajectories file. GTP, GDP and PTM chains were omitted from subunit centre of mass calculations to ensure the centres of mass were in the same starting location for all simulations.

Over the course of the unmodified *Short MT* simulation, the plus- and minus- ends of the model bent away from the centre (Fig. 12A). The average distance between plus end  $\beta$  subunits on PF1 and PF7 (23.90 nm) was 0.778 nm higher than between other subunits in these protofilaments (Fig. 12B). This bowing effect was most prominent around the seam, where the protofilaments on either side of the seam both contain tubulin heterodimers that are missing a neighbouring heterodimer to interact with. Outwards bowing at the minus end was also observed.

Outwards bowing was also observed in most of the modified *PF Sheet* simulations. The  $\alpha$ - and  $\beta$ - modified tubulin with poly( $\alpha$ E) simulations exhibited outwards bowing at the plus end. The  $\alpha$ - and  $\beta$ - modified tubulin with poly(G) simulations, as well as  $\alpha$ -modified tubulin with poly( $\gamma$ E) simulation exhibited outwards bowing at the minus end. The  $\beta$ -modified tubulin with trapped poly( $\gamma$ E) tails simulation exhibited bowing at both the plus- and minus- end. Unlike the other *PF Sheet* simulations, the unmodified and  $\beta$ -modified tubulin with free poly( $\gamma$ E) simulations showed no significant signs of bowing.



**Figure 11. Internal centre of mass distances.** The 4 inter-prot filament measurements for  $\alpha_1$ - $\alpha_1$  (minus end) (pink)  $\beta_1$ - $\beta_1$  (purple),  $\alpha_2$ - $\alpha_2$  (red) and  $\beta_2$ - $\beta_2$  (plus end) (blue) used to track distances in *PF Sheet* simulations. Centres of mass for the measured tubulin subunits are shown as coloured spheres.



**Figure 12. Plus- and minus- end tubulin subunits sit further away from the model's centre than the non-terminal subunits.** (A) Comparison of the bowing present at the plus end of the unmodified *Short MT* simulation, sampled at 0 ns (blue) and 100 ns (orange). Outwards bowing at the plus end is highlighted in yellow; outwards bowing at the minus end is highlighted in black. (B) The change in distance between PF1 and PF7 subunits for the unmodified *Short MT* simulation. Each circle represents a tubulin subunit ( $\alpha$  or  $\beta$ ). The bars between subunits show relative distances. Distance at 0 ns is 22.5 nm;  $\alpha_1$  (minus end) is in red;  $\beta_1$  is in blue;  $\alpha_2$  is in pink;  $\beta_2$  (plus end) is in dark blue.

### ***PF Sheet and Short MT simulations exhibit differing model flexibility***

Root-mean-square deviation (RMSD) was used to determine the structural fluctuations of a model during simulation. First, a reference structure of the simulated model was generated by averaging the positions of each atom in the system, across the simulation run time. The average structure was chosen as the reference structure as it better represented the relaxed conformation seen during the majority of each simulation, as opposed to the contracted starting model conformation. The general position of the simulated structure was fitted to the position of the reference structure at each timestep. This was to prevent translation of the entire model, within the periodic box, from affecting RMSD calculations. Mass-weighted RMSD of the atomic positions (equation 4) was then calculated by comparing the position of each backbone atom to its position in the reference structure, at each timestep. A higher RMSD value indicates a larger average deviation in atom position in the simulation model from their positions in the reference structure.

$$RMSD = \sqrt{\frac{\sum_{i=0}^N [m_i (X_i - Y_i)^2]}{M}}$$

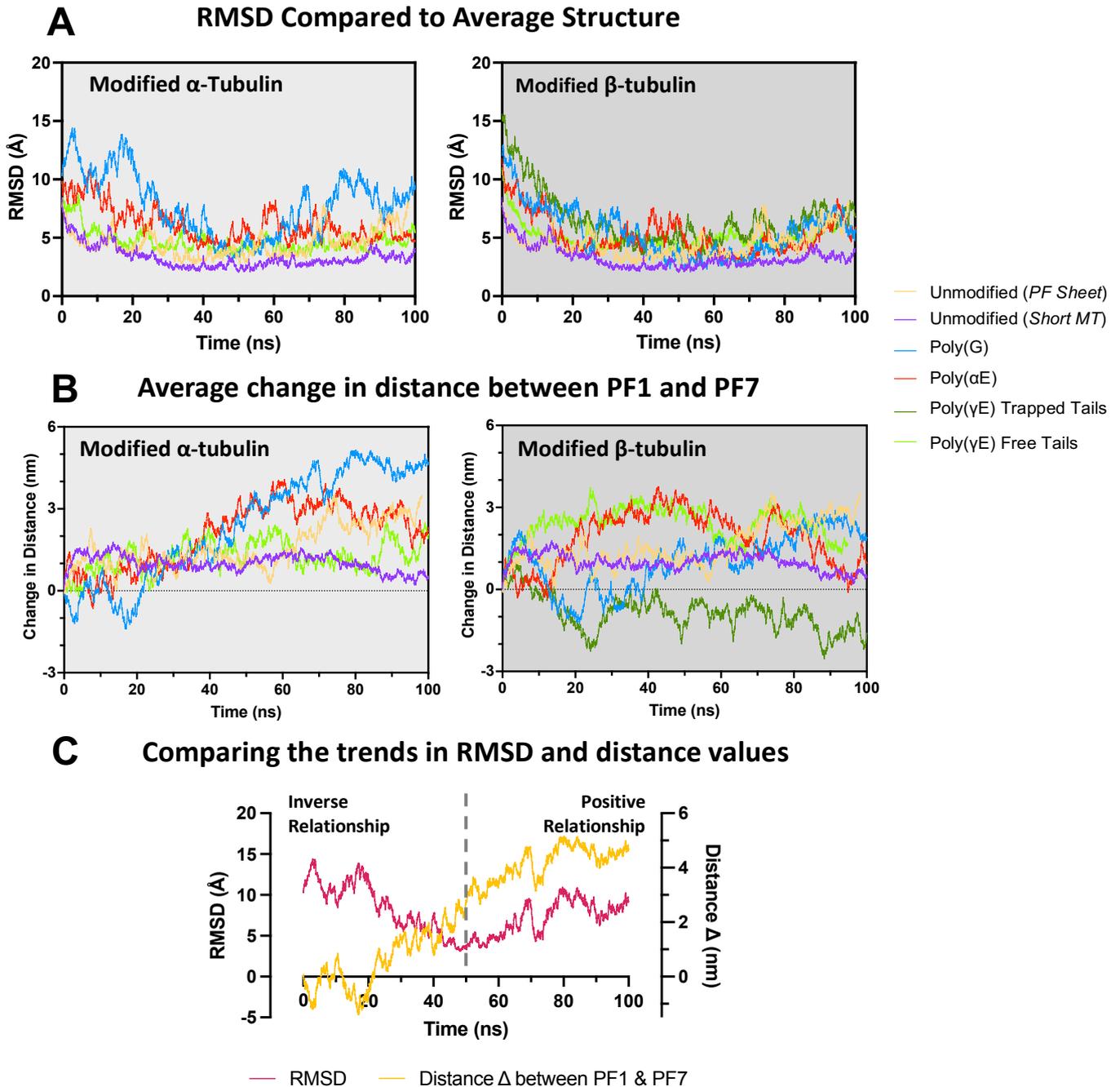
**Equation 4. Mass-Weighted Root-Mean-Square Deviation. Taken from Case et al. (2021a).**  $N$  is the number of atoms;  $m_i$  is the mass of atom  $i$ ;  $X_i$  is the coordinate vector for target atom  $i$ ;  $Y_i$  is the coordinate vector for reference atom  $i$ ;  $M$  is the total mass.

RMSD analysis shows that the starting structure was not the most occupied conformation for either the *PF Sheet* and *Short MT* models, and that a lower energy equilibrium state was reached about halfway through the simulation (Fig. 13A). During the first 20 ns RMSD values decreased, which was caused by an initial expansion of the model, as the distance between neighbouring protofilament increased slightly. An equilibrium state was reached between 40-60 ns, where RMSD values were at their lowest. Past 80 ns the RMSD values for the *PF Sheet* simulations began to increase above their equilibrium values. This contrasts with the

RMSD values for the unmodified *Short MT* simulation, which remained at an average of  $2.95 \pm 0.46 \text{ \AA}$  from 20-100 ns (Table 2).

RMSD values for the *PF Sheet* simulations had higher variances across 100 ns than the unmodified *Short MT* simulation. The  $\alpha$ -tubulin modified with poly( $\gamma$ E) simulation had marginally lower RMSD standard deviation when considering the entire timescale, however the unmodified *Short MT* simulation had lower RMSD standard deviation when excluding the first 20 ns of each simulation. RMSD values from 0-20 ns were excluded to reduce the impact of the initial model expansion on mean RMSD values.

The average distance between PF1 and PF7 was 22.5 nm at the first timestep. This distance increased over the course of each *PF Sheet* simulation, excluding the  $\beta$ -modified with trapped poly( $\gamma$ E) chains simulation (Fig. 13B). This trend resulted in an overall flattening of the semicylindrical shape of the *PF Sheet* model. This is in direct contrast to the unmodified *Short MT* simulation. An initial burst phase, resulting from a slight increase in distance between adjacent protofilaments, caused the distance to peak to 23.8 nm, before fluctuating around  $23.3 \pm 0.3 \text{ nm}$  for the remainder of the simulation.



**Figure 13.** *PF Sheet* simulations do not act like the unmodified *Short MT* simulation. **(A)** Best-fit RMSD calculated for the unmodified *Short MT*,  $\alpha$ -modified and  $\beta$ -modified *PF Sheet* simulations **(B)** The average change in distance between PF1 and PF7, over the course of each simulation. Measurements between each subunit (4 total) of these PFs were averaged. Average distance at 0 ns was 22.5 nm. Data above 0 indicates an increase in distance, and below 0 indicates a decrease in distance. **(C)** Comparison of the trends in RMSD (red) and  $\Delta$  inter-PF distance (orange) values for the *PF Sheet* simulation with poly(G) PTMs on  $\alpha$ -tubulin subunits. Between 0-50 ns there is an inverse relationship between the two measurements. Equilibrium is reached at 50 ns. Between 50-100 ns there is a positive relationship.

<i>Simulation</i>	<i>Modified subunit</i>	<i>0-100 ns</i>				<i>20-100 ns</i>			
		<i>Mean RMSD (Å)</i>	$\sigma$	$\sigma/\sqrt{n}$	<i>n</i>	<i>Mean RMSD (Å)</i>	$\sigma$	$\sigma/\sqrt{n}$	<i>n</i>
<b>Short MT</b>	<i>Unmodified</i>	<b>3.3106</b>	0.9339	<b>0.0042</b>	50000	<b>2.9491</b>	<b>0.4610</b>	<b>0.0023</b>	40000
<b>PF Sheet</b>	<i>Unmodified</i>	4.4874	1.1219	0.0051	49127	4.4221	1.1523	0.0058	39127
<b>Poly(G)</b>	$\alpha$	7.8254	2.6929	0.0120	50000	6.8974	2.0735	0.0104	40000
	$\beta$	5.4647	2.1090	0.0094	50000	4.7288	1.4768	0.0074	40000
<b>Poly(<math>\alpha</math>E)</b>	$\alpha$	6.1182	1.4934	0.0067	50000	5.5617	0.9391	0.0047	40000
	$\beta$	5.1985	1.6129	0.0072	50000	4.6961	1.0719	0.0054	40000
<b>poly(<math>\gamma</math>E)</b>	$\alpha$	4.7306	<b>0.9305</b>	<b>0.0042</b>	50000	4.4630	0.5620	0.0028	40000
	$\beta$ (trap)	6.5213	2.2391	0.0100	50000	5.6548	1.0670	0.0053	40000
	B (free)	5.0200	0.9487	0.0067	47337	4.7740	0.5451	0.0073	37337

**Table 2. Unmodified Short MT simulation is less variable than the PF Sheet Simulations.** The lowest values in each category are shown in green.  $\sigma$  is standard deviation;  $\sigma/\sqrt{n}$  is standard error; *n* is number of frames, where each frame is 2 ps.

Large peaks in RMSD values correspond with similarly large inter-protofilament distance changes (Fig. 13C). The initial spike in RMSD, from 10.5 to 14.5 Å, at 4 ns aligns with a decrease in distance between PF1 and PF7 of 1 nm from the starting distance of 22.5 nm. A wider peak of similar amplitude occurred between 10-20 ns, corresponding with another decline in inter-protofilament distance. Between 66-90 nm, there are two major peaks in inter-protofilament distance up to 5.2 nm above the starting distance. This is mirrored in the RMSD data, which has two near identical 10 Å peaks. This demonstrates an inverse correlation between RMSD and average inter-protofilament distances when inter-protofilament distances are shorter than in the reference model used to calculate RMSD, and a positive correlation when distances are longer than in the reference model. For the poly(G) on  $\alpha$ -tubulin simulation, this correlation flips from inverse to positive at around 50 ns. Factors such as plus- and minus-end bowing, and the initial model expansion to make total interaction energy within the system less negative, also contributed to increased RMSD values, as they cause the model to deviate more from the average structure.

### **$\beta$ -tubulin C-Terminal tails are the most mobile regions**

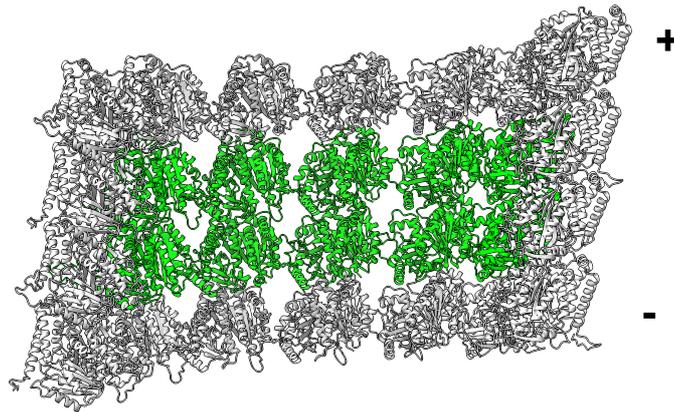
Mass-weighted root-mean-square fluctuation (RMSF) (equations 5 and 6) was calculated for each residue in each *PF Sheet* simulation, to quantify the mobility of different tubulin regions. RMSF values were only calculated for the central subunits of PF2 to PF6 (Fig. 14), to limit the impact of protofilament bowing and model flexibility. Values were averaged on a per-subunit basis (i.e., all  $\alpha$ -tubulins and all  $\beta$ -tubulins). RMSF values for PTM chains were calculated independently from tubulin heterodimers.

$$RMSF_i = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle}$$

**Equation 5. Root-mean-square fluctuation.** Taken from Case et al. (2021a).  $x_i$  is the position of atom  $i$ . RMSF values were averaged over the 100 ns duration of the simulations.

$$\langle Fluct \rangle = \frac{\sum AtomFluct_i \times Mass_i}{\sum Mass_i}$$

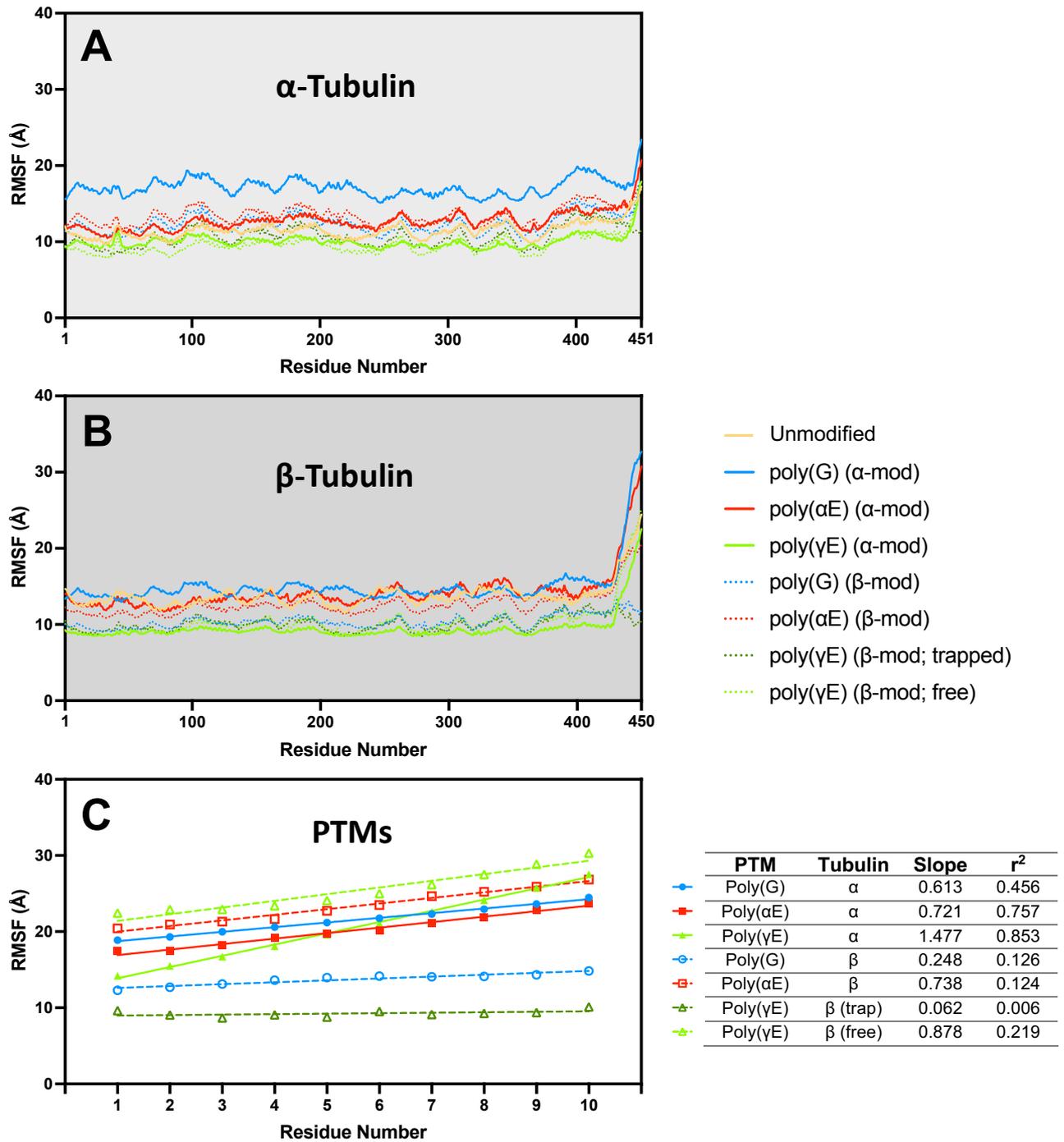
**Equation 6. Mass-weighted average of atomic fluctuations.** Taken from Case et al. (2021a).



**Figure 14. RMSF values were calculated for central tubulin subunits only.** *PF Sheet* model with central subunits coloured green, indicating that RMSF values were calculated for these subunits only. Central heterodimers include  $\beta_1$  and  $\alpha_2$  subunits from PF2 to PF6.

RMSF analysis shows the disordered C-terminal tails of both  $\alpha$ - and  $\beta$ - tubulin were by far the most mobile regions of the protein (Fig. 15A and 15B). The CTTs for  $\beta$ -tubulins were more dynamic than  $\alpha$ -tubulin CTTs, resulting in a more pronounced spike in RMSF values. This could be due to the  $\beta$ -tubulins being longer than the  $\alpha$ -tubulin CTTs (21 and 13 residues, respectively). Both poly( $\alpha$ E) simulations behaved similarly to the unmodified *PF Sheet* simulation. All 3 poly( $\gamma$ E) simulations had the lowest average RMSF. The poly( $\gamma$ E) simulation with trapped tails did not exhibit RMSF peaks for the CTTs on either tubulin subunit. This stemmed from an error when rebuilding the *PF Sheet* model that led to  $\beta$ -tubulin CTTs being positioned between protofilaments, restricting their movement during the simulation. A significant RMSF peak was present for both CTTs for the poly( $\gamma$ E) simulations with free CTTs. The  $\alpha$ -tubulin modified with poly(G) simulation saw the highest average RMSF values across each residue, when compared to other *PF Sheet* simulations. This is consistent with the RMSD values and inter-PF distances, which show the shape of the model was more variable than in other simulations.  $\beta$ -tubulin modified with poly(G) did not produce an RMSF peak for the  $\beta$ -tubulin CTT. This is, again, a result of these CTTs having their movement restricted by being positioned between two protofilament. RMSF peaks, like those seen for CTTs in other modified *PF Sheet* simulations, would be expected if the simulation was repeated with unobstructed CTTs and PTMs.

RMSF values for the unobstructed PTMs were higher at the C-terminus of the chain than at the N-terminus (Fig 15C). Poly( $\gamma$ E) modified onto  $\alpha$ -tubulin had the highest increase in RMSF from N- to C-terminus (gradient of 1.477 Å/residue) and the data best fit with the associated regression line ( $r^2$ : 0.853). Free poly( $\gamma$ E) on  $\beta$ -tubulin had the most flexible of the PTM chains, with the C-terminal Glu residue having the highest RMSF value of any PTM residue (30.31 Å). Values for poly(G) and trapped poly( $\gamma$ E) on  $\beta$ -tubulin were lower than the other PTMs, and their regression lines had the lowest gradients (0.248 and 0.062, respectively). This is caused by these PTMs being trapped between protofilaments, which restricts their movement.



**Figure 15.  $\beta$ -tubulin CTTs are more dynamic than  $\alpha$ -tubulin CTTs.** (A) Mass-weighted RMSF of atoms associated with  $\alpha$ -tubulin subunits, averaged on per-residue basis. (B) Mass-weighted RMSF of atoms associated with  $\beta$ -tubulin subunits. RMSF values for subunit type ( $\alpha$  or  $\beta$ ) were averaged. (C) Average mass-weighted RMSF per residue of PTM chains. Linear regression for each modification is shown as a straight line.

**Trapped poly( $\gamma$ E) chains sequester inter-protofilament interactions**

Over the course of the  $\beta$ -modified poly( $\gamma$ E) simulation with trapped tails, tears occurred between PF5-6 and PF6-7 (Fig. 16), resulting from the CTTs and poly( $\gamma$ E) chains being wedged between protofilaments. The breaks began at the plus end of the model and worked their way down to the minus end. The break between PF5 to PF6 first appeared at around 30 ns, occurring between just the plus end  $\beta$ -tubulins. It reached to the minus end  $\beta$ -tubulin after 70 ns. A similar break formed between PF6 and PF7 at 40 ns. These tears caused PF6 and PF7 to fold inwards, towards the centre of the model, resulting in the decrease in inter-protofilament distance between PF1 and PF7 (Fig. 13B). Similar breaks did not occur when poly(G) chains were positioned between protofilaments, or when repeating the simulation with the CTT and PTM positioned away from the inter-protofilament interface. The increase in local negative charge in these regions, caused by the presence of poly(E) PTMs, appears to sequester the inter-PF interactions that maintain the cylindrical macrostructure of the microtubule.

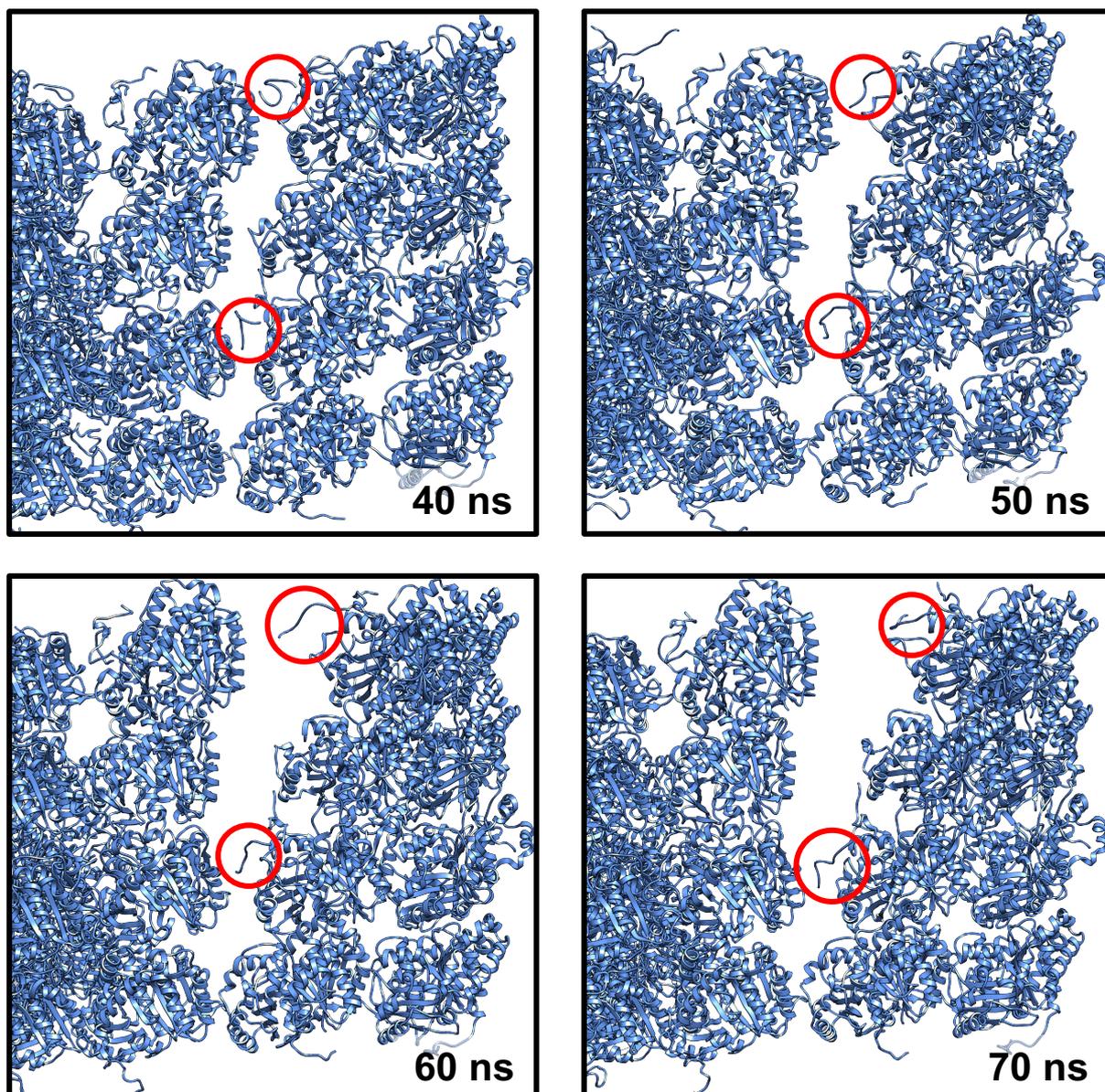
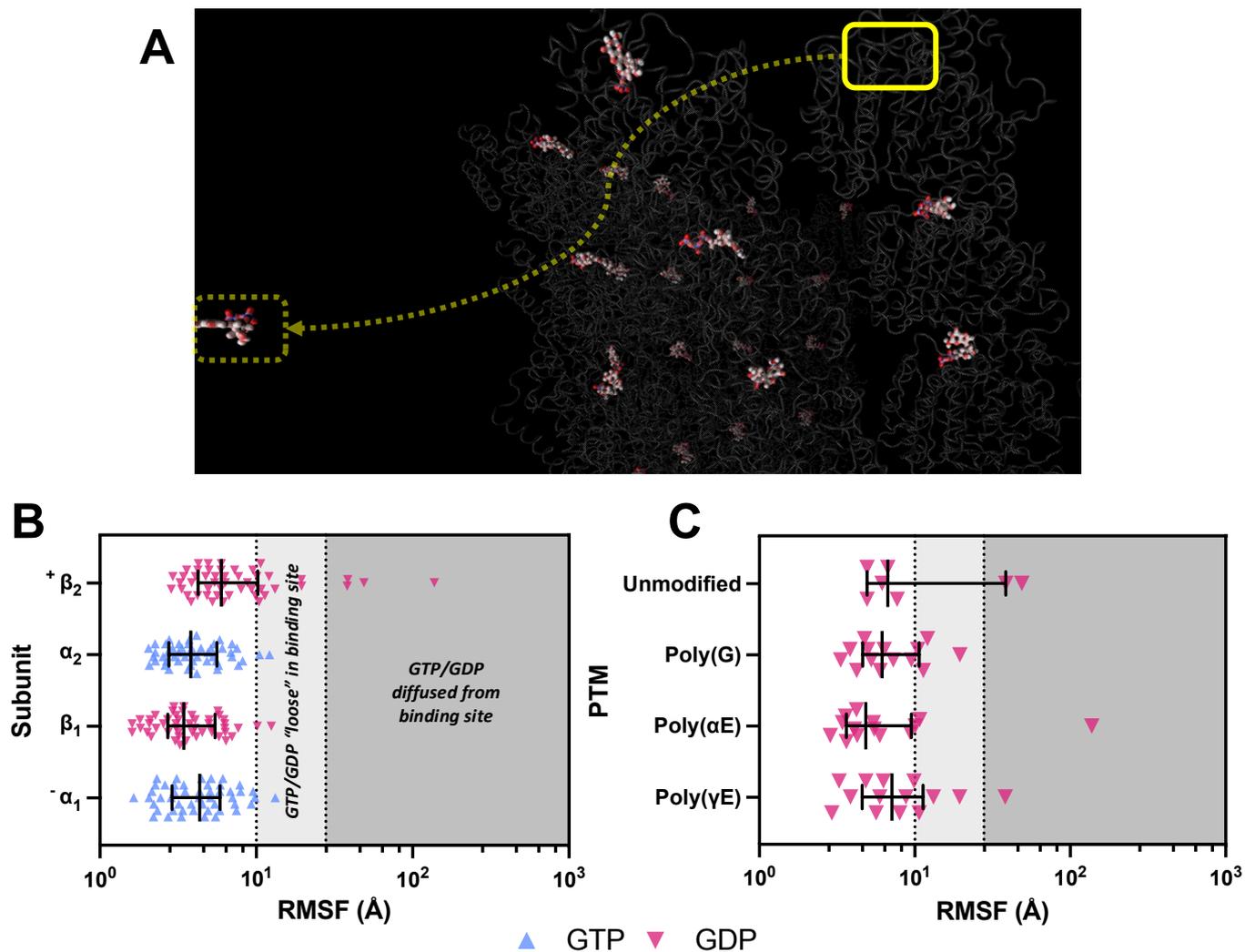


Figure 16. Poly( $\gamma$ E) chains cause a tear between PF5 and PF6.  $\beta$ -modified Poly( $\gamma$ E) PF Sheet simulation, showing the progression of the tear between PF5 and PF6. Poly( $\gamma$ E) chains are circled in red.

### **GDP molecules diffuse from plus end $\beta$ -tubulins**

Mass-weighted RMSF was calculated for the bound GTP and GDP molecules in each *PF Sheet* simulation, to quantify the mobility of these ligands. Values for the backbone atoms of each molecule were averaged, providing a single RMSF value for each molecule. Higher RMSF values indicate that a ligand is less strongly associated with its nucleotide binding site, or in extreme cases, completely dissociated from the nucleotide binding site and floating in solution (Fig. 17A).

RMSF values for GDP bound to plus end  $\beta$ -tubulin subunits were on average  $\sim 1.5$  to  $2.5$  Å higher than for GDP/GTP bound to other tubulin subunits (Fig. 17B). This is consistent with GDP molecules that are more mobile within the nucleotide binding site, as a result of higher solvent accessibility to the site. The increased movement seen at the plus- and minus- ends in *PF Sheet* simulations is also likely contributing to the increased RMSF values for the GDP and GTP molecules bound in these subunits. During 3 of the 7 *PF Sheet* simulations (unmodified, poly( $\alpha$ E) bound to  $\alpha$ -tubulin and poly( $\gamma$ E) bound to  $\beta$ -tubulin) one or more GDP molecules bound to a plus end  $\beta$ -tubulin diffused away from the nucleotide binding site and into the surrounding solvent (Fig. 17A). GDP bound to the exposed  $\beta$ -tubulin subunit at the plus end has little protection from the solvent, which can enter the binding site and promote removal of GDP. RMSF values for only the plus end-bound GDP molecules (Fig 17C) across *PF Sheet* simulations do not show a correlation between the addition of a particular PTM and GDP diffusion. No GTP or GDP molecules diffused away from the corresponding nucleotide binding sites in the unmodified *Short MT* simulation.



**Figure 17. GDP molecules with higher RMSF values are more likely to diffuse away from their nucleotide binding sites. (A)** A GDP molecule diffusing from PF1 during the unmodified *PF Sheet* simulation. Diffusion occurred at 22 ns. **(B)** Mass-weighted RMSF for GTP and GDP molecules during all *PF Sheet* simulations, organised by tubulin subunit. Above 10 Å, the molecules begin to vibrate more and detach from the binding sites. Above 30 Å, the molecules escape from the binding sites completely and diffuse into the solvent around the model. The centre bars show median values; error bars show interquartile range. **(C)** Mass-weighted RMSF for GDP molecules bound to  $\beta_2$  (plus end) tubulin heterodimers for all *PF Sheet* simulations. The centre bars show median values; error bars show interquartile range.

## Discussion

I have generated the *PF Sheet* and *Short MT* models to represent a microtubule in atomistic molecular dynamics simulations. The *PF Sheet* model was the basis for a series of simulations testing the effect adding different PTMs (poly(G), poly( $\alpha$ E) and poly( $\gamma$ E)) to  $\alpha$ - and  $\beta$ - tubulin subunits had on the model's structure and dynamics. The *Short MT* model was used to validate the accuracy of the *PF Sheet* simulations as a suitable representative model for a full microtubule. The results of these simulations revealed that tubulin CTTs and PTMs were the most mobile regions, with  $\beta$ -tubulin CTTs being, on average, more mobile than  $\alpha$ -tubulin CTTs. Tubulin subunits at the plus- and minus- ends pulled away from the centre of both models, causing the protofilaments to bow. The diameter of the *PF Sheet* and *Short MT* models rapidly increased over the first 2 ns of each simulation, as the distance between neighbouring protofilaments increased, before stabilising. The distance between protofilaments 1 and 7 in *PF Sheet* simulations tended to increase past this initial expansion distance over the 100 ns timescale. However, for the *Short MT* simulation, this distance stabilised at around 1 nm above its initial distance after 20 ns. Positioning negatively charged poly( $\gamma$ E) PTMs between protofilaments caused a tear to form in the *PF Sheet* model, where uncharged poly(G) PTMs did not. Finally, GDP was more likely to diffuse away from the plus end  $\beta$ -tubulin nucleotide binding site than from other positions in the model.

There are currently only a small number of examples of atomistic microtubule or protofilament sheet simulations being used to study microtubule dynamics, likely due to the difficulty associated with simulating multi-million atom systems. Nasedkin et al. (2021) used a 3x3 heterodimer model (3 adjacent protofilaments with a height of 3 tubulin heterodimers), in conjunction with a 13 protofilament microtubule model, to investigate the differences in dynamics between microtubules with GTP and GDP bound to plus end  $\beta$ -tubulins. Protofilament bowing was present in both the *PF Sheet* and microtubule simulations, as in the simulations presented in this project, likely enhanced by the plus end-bound GDP, which

resembles earlier computational work studying single protofilaments as well as complete microtubules (Grafmüller and Voth, 2011; Nasedkin et al., 2021; Igaev and Grubmüller, 2022). RMSD showed the smaller 3x3 model *PF Sheet* had higher average RMSD values than the full microtubule model, in accordance with my results (Nasedkin et al., 2021). Differences between the behaviour of the *PF Sheet* and full microtubule models during simulations were also observed, further implying *PF Sheet*-like models are not an accurate substitute for full microtubule models in atomistic simulations (Nasedkin et al., 2021).

Bigman and Levy (2021) used a 3x2 heterodimer model to investigate the effect of different length poly(G) and poly(E) PTMs on tubulin CTTs, observing poly(G) PTMs to collapse in on themselves and poly(E) tails to protrude from the CTTs more noticeably. This does not fully agree with the results of this project, which show unobstructed poly(G) and poly(E) tails to fluctuate at similar levels, however this may be due to differences in simulation conditions. Bigman and Levy (2021) used 0.125 mM NaCl to neutralise the system and each simulation was heated to 400K for the first 10 ns, whereas the simulations in this project used 150 mM KCl to approximate *in vivo* conditions and maintained the system temperature at 300 K for the entire 100 ns duration.

Neither Bigman and Levy (2021) nor Nasedkin et al. (2021) report diffusion of the plus end-bound nucleotide from the nucleotide binding site. Exchange of GDP bound to  $\beta$ -tubulin for GTP is thought to only occur on unpolymerised tubulin heterodimers (Desai and Mitchison, 1997). Kinetic Monte Carlo models have recently shown that GDP-to-GTP exchange at the plus end of microtubules is theoretically possible (Piedra et al., 2016). It is possible that the conformation of the  $\beta$ -tubulin nucleotide binding site is a GTP-accepting conformation. This would increase the chances of GDP dissociation, allowing a new GTP molecule to diffuse into the site ready for hydrolysis. Further simulations with excess GTP present in the solvent may provide a simulation where this GDP-GTP exchange occurs. Alternatively, a preliminary energy minimisation and structure optimisation simulation of a single tubulin heterodimer could

provide a structure with the  $\beta$ -tubulin nucleotide binding site is the GDP-bound conformation, which could be used to create variations of the *PF Sheet* and *Short MT* models. This would allow us to determine whether the binding site conformation is the cause of the observed GDP diffusion.

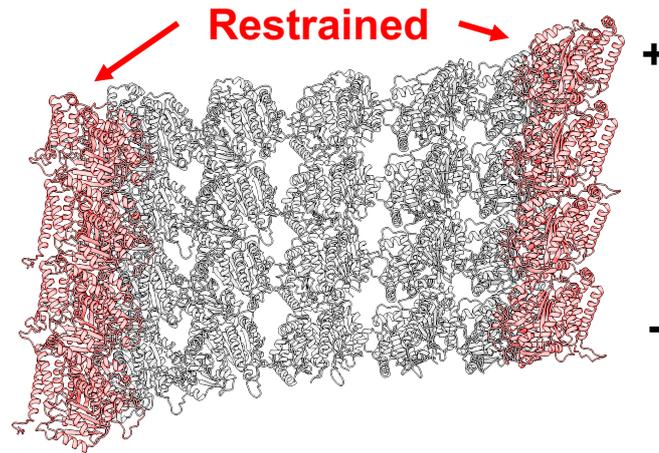
There are no other examples of tears resulting from negatively charged PTMs being positioned between protofilaments, further confirming that this resulted from an error during model building, and that the likelihood of the CTTs being positioned between protofilaments in a growing microtubule, as in the trapped poly( $\gamma$ E) and poly(G) simulations, is low. The additional negatively charged side chains afforded by the poly( $\gamma$ E) chains interfered with the inter-protofilament salt bridges. PF 5-6 and 6-7 were the only protofilament pairs to exhibit a breakage during the trapped poly( $\gamma$ E) simulation. This results from imperfect dimer fitting during the *Short MT* and *PF Sheet* model generation process, resulting in a slightly larger distance between these protofilaments. This would weaken the strength of the stabilising salt bridge interactions, making it easier of the misplaced poly( $\gamma$ E) chains to disrupt them and cause a tear to form.

There is not enough evidence to show that poly(E) can initiate microtubule catastrophe alone. However, it is possible that CTTs could position between PFs if a hole or breakage in the microtubule lattice was already formed. In this scenario, the poly(E) PTM acts as a catastrophe-accelerating factor, rather than a catastrophe-instigating factor. It is also possible for a microtubule-destabilising protein recognise and position longer poly(E) chains between protofilaments, causing a break to form between PFs. It is important to emphasise that these are purely hypothetical scenarios and mechanisms, and the results of this project do not provide enough evidence to present a concrete mechanism. Further simulations and, crucially, evidence from *in vitro* and/or *in vivo* studies is required.

### Changes to the current simulation systems

All future simulations should be carried out at 310.15 K (37 °C) as opposed to 300 K (26.85 °C). Not only will this produce simulations that more accurately represent microtubule dynamics at biologically relevant temperatures, but it should speed up the rate at which different conformations are sampled. Crucially, I would expect this temperature increase to increase the rate at which breakages between protofilaments form, caused by negatively charged PTMs. This would make further simulations investigating this phenomenon easier to analyse as a shorter simulation would be needed, either resulting in a smaller trajectories file or allowing longer simulations to be run ( $\geq 1 \mu\text{s}$ ).

For situations where atomistic detail is important, such as studying specific residue interactions, improvements can be made to *PF Sheet* model simulations so that they better represent the dynamics of a full microtubule. One such improvement could be to restrain the outer protofilaments (PF1 and PF7 for my *PF Sheet* model) during a simulation (Fig. 18). This would fix the atoms of those protofilaments in place, while leaving the remaining protofilaments free to move. I would expect this to prevent the model from flattening out, as observed in all the unrestrained *PF Sheet* simulations (Fig. 13B), whilst still leaving most tubulin heterodimers free to move and be influenced by PTMs. To improve similarity between *PF Sheet* and *Short MT* simulations, I would remake the *PF Sheet* model using coordinates from the unmodified *Short MT* simulation to account for the initial increase in model diameter seen during the first 1-2 ns. This would involve exporting a frame at around 5-10 ns as a PDB file, using a tool such as VMD (Humphrey et al., 1996). Then, tubulin heterodimers could be removed from the PDB file until the equivalent of the *PF Sheet* model remains (i.e., 7 protofilaments, 2 heterodimers tall, not including the seam).



**Figure 18. Restraining PF1 and PF7 should prevent model flattening.** PF1 and PF7 (red) can be held in place to prevent the model from flattening during a simulation.

Transitioning from the *PF Sheet* model to the *Short MT* model would likely result in simulations more accurate to the *in vivo* and *in vitro* dynamics of a microtubule. The increase in interprotofilament interactions across the structure, and the fact there are no terminal protofilaments, would eliminate the flattening seen in *PF Sheet* simulations. However, this change would significantly increase the number of atoms in the simulation. Not only does the *Short MT* model contain almost double the atoms of the *PF Sheet* model (192,375 and 99,750, respectively), but solvent can occupy the empty space inside the microtubule lumen, increasing the number of solvent atoms present in the simulation. Solvated *PF Sheet* simulations contained ~1,700,000 atoms, whereas the *Short MT* simulation contained 2,700,000 atoms once solvated, which significantly reduced the rate at which simulation ran. The *PF Sheet* simulations ran at an average rate of 12.25 ns/day, requiring 8 days to complete 100 ns, whereas the unmodified *Short MT* simulation ran at 7.55 ns/day and took 13 days to complete. Building PTMs onto the *Short MT* model would increase the periodic box size, as the new chains protrude further from the lattice. This further increases the number of water molecules in the system, driving simulation rate down and increasing the environmental impact of each simulation, through more energy being required to run HPC nodes for longer. It is,

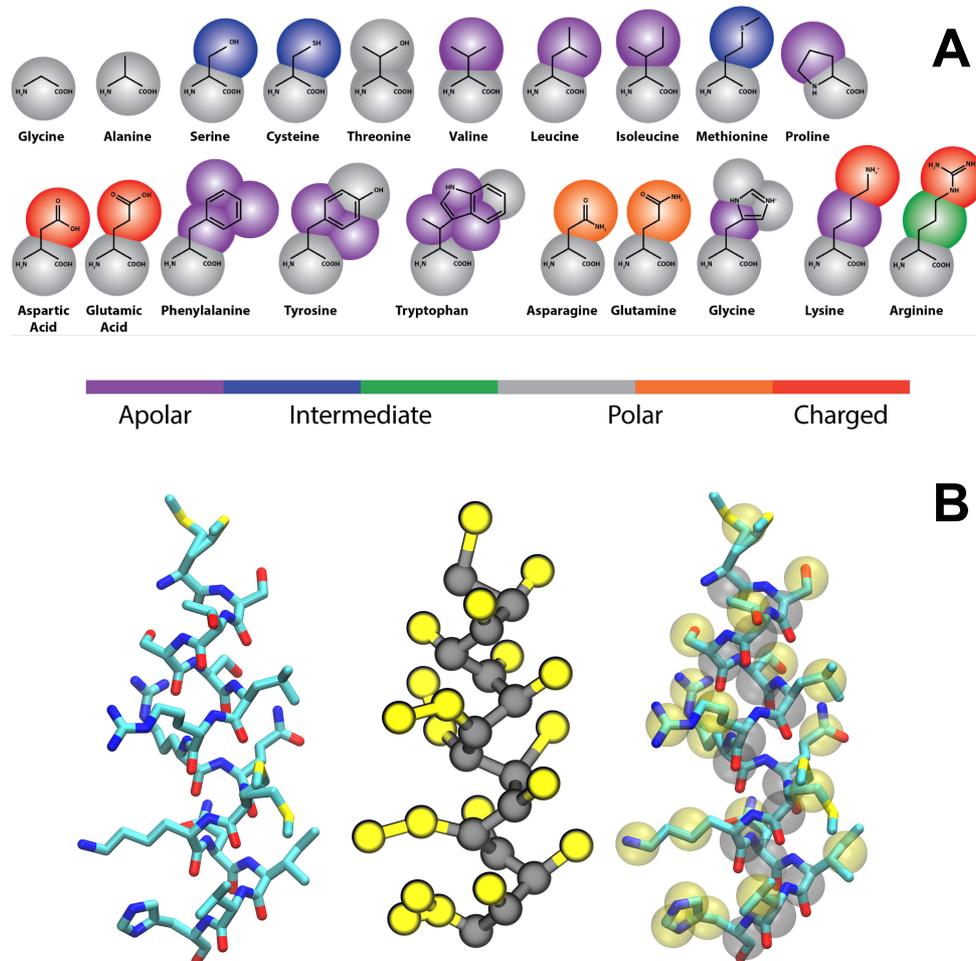
therefore, not recommended to completely switch to a full microtubule model, such as the *Short MT* model, for atomistic simulations using an explicit solvent model.

The larger *Short MT* model could be used in coarse-grained simulations, which summarise the constituent atoms of an amino acid down to a handful of particles. By reducing the total number of particles in the system through increasing the granularity of the system, we could study the effects of different combinations of PTMs on a much larger microtubule model, and how these PTMs might affect MAP binding or processivity. Coarse-grained simulations are a more widely used tool for studying microtubules *in silico* than atomistic simulations. They have already been used to show: defects in the microtubule lattice impair resistance to mechanical breakages (Jiang et al., 2017); clusters of acetylated tubulins provide better resistance to microtubule disassembly than more regularly spaced patterns (Aparna et al., 2020); poly(E) PTMs reduce the rate of diffusion of EB1 and PRC1 along microtubules and poly(G) PTMs enhance diffusion rates (Bigman and Levy, 2020); differences in the mechanical properties of microtubules up to 12  $\mu\text{m}$  in length, using an ultra-coarse-grained model (Zha et al., 2021).

The Martini force field is a commonly used coarse-grained model which uses an elastic network to connect backbone particles via springs that follow Hooke's law (equation 7) (Marrink et al., 2004; Monticelli et al., 2008). This model mostly uses a four-to-one atom mapping (four heavy atoms and their associated hydrogens summarised as one particle), though ring-like structures present in amino acids such as His and Phe use a higher resolution, two-to-one mapping to better represent their geometry (Monticelli et al., 2008) (Fig. 19). Coarse graining adds a layer of abstraction, making this type of simulation less accurate than equivalent atomistic simulations. However, atomistic simulations of  $\mu\text{m}$ -length microtubules would be prohibitively expensive and time consuming to run, making coarse-grained simulations an attractive alternative.

$$F_s = kx$$

**Equation 7. Hooke's Law.**  $F_s$  is spring force;  $k$  is the spring constant;  $x$  is spring expansion or compression.



**Figure 19. Amino acid atom mapping for Martini force field simulations. Adapted from Bradley and Radhakrishnan (2013).** (A) Skeletal structures for the 20 canonical amino acids overlaid with spheres representing differently charged Martini model bead types and the heavy atoms associated with them. (B) Atomistic (left) and Martini coarse-grained (centre) representations of an  $\alpha$ -helix, overlaid (right) to show how polypeptides are mapped for coarse-grained simulations. Backbone beads are shown in grey; Side chain beads are shown in yellow.

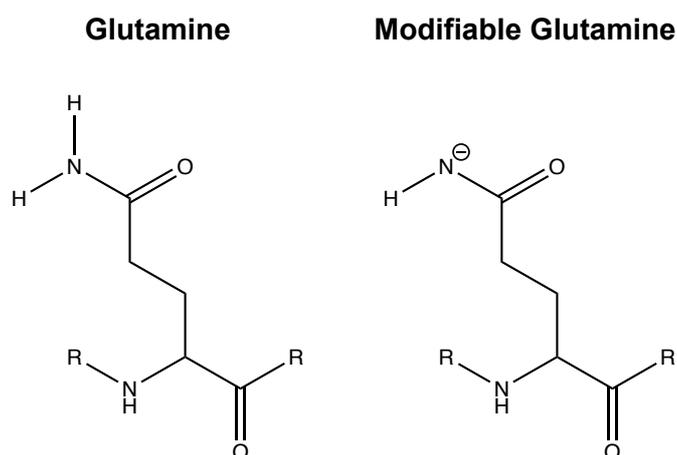
The *Short MT* model could also be used in implicit solvent simulations. These consider the solvent as a structureless medium, as opposed to explicit solvent models in use discrete solvent atoms. Using an implicit solvent model would reduce the number of atoms for *Short MT* simulations to 1/10<sup>th</sup> their explicit counterparts. The *molecular mechanics generalized Born surface area* (MMGBSA) and *molecular mechanics Poisson-Boltzmann surface area* (MMPBSA) models are commonly used implicit solvent models (Miller et al., 2012). Typically, the actual rate of simulation is lower for implicit solvent simulations than for explicit solvent simulations. However, given the same amount of simulation, the solute samples more unique conformations when simulating with implicit solvent than with explicit solvent.

### **Future Experiments**

Whilst it is possible for poly(E) PTMs to be added onto any of the C-terminal Glu residues, *in vitro* studies of purified tubulin, and molecular docking and dynamics simulations of CTTs bound to TTLs have identified 2 major polyglutamylation sites:  $\alpha$ -tubulin Glu 445 and  $\beta$ -tubulin Glu 438 (Eddé et al., 1990; Rüdiger et al., 1992; Natarajan et al., 2017). Future poly(E) simulations should use these sites. These sites are positioned closer to the main body of the corresponding tubulin subunit than  $\alpha$ -tubulin Glu 446 and  $\beta$ -tubulin Glu 445, making it more likely that atoms from PTMs bound to these sites are more likely to overlap and clash with atoms on neighbouring tubulins, causing simulations to crash. To alleviate this issue, we would need run a simulation on a single, unmodified tubulin heterodimer and export a frame in which the tails are positioned so that the major glutamylation sites can be modified without causing clashes with neighbouring heterodimers in the microtubule. Alternatively, we could rebuild the C-terminal tails as  $\beta$ -strands using *Modeller*, making them straight. This would then make positioning the tails and PTMs so that no atoms overlap significantly easier. This initial model would be less accurate to the structure of tubulin CTTs and oligopeptide PTMs, as neither exists as  $\beta$ -strands *in vivo* for any notable period of time, though work published by Wall et al. (2016) using NMR spectroscopy suggests that tubulin CTTs have a slight propensity to form  $\beta$ -sheets. More recent work by Wall et al. (2020) suggests that  $\alpha$ -tubulin CTTs interact with

the tubulin heterodimer surface, rather than extending into the solute, which may have an effect microtubule stability. This could be simulated by rebuilding the CTTs as  $\beta$ -strands and positioning them close to the outer surface of a microtubule. This would also allow us to investigate the effect to adding poly(G) or poly(E) chains on microtubule stability, when positioned along the microtubule surface.

The next step would be to expand the list of PTMs being studied to include things like polyamination, detyrosination,  $\Delta 2$  and  $\Delta 3$ , using the workflow established in this project. Polyamination occurs at Glu 15 of  $\beta$ -tubulin, meaning an *fromod* and *prep* file would have to be generated for an altered glutamine residue, using a similar method to the one used to create 'GLG' and 'GGN' (Fig. 20). Detyrosination,  $\Delta 2$  and  $\Delta 3$  are all common PTMs that can be easily modelled by removing the residues from the PDB file. Lys 40 is a prominent acetylation site for all  $\alpha$ -tubulin isotypes (excluding TUBA8 which has an alanine residue at this position), located in a small, disordered loop positioned at the internal microtubule-solvent interface.

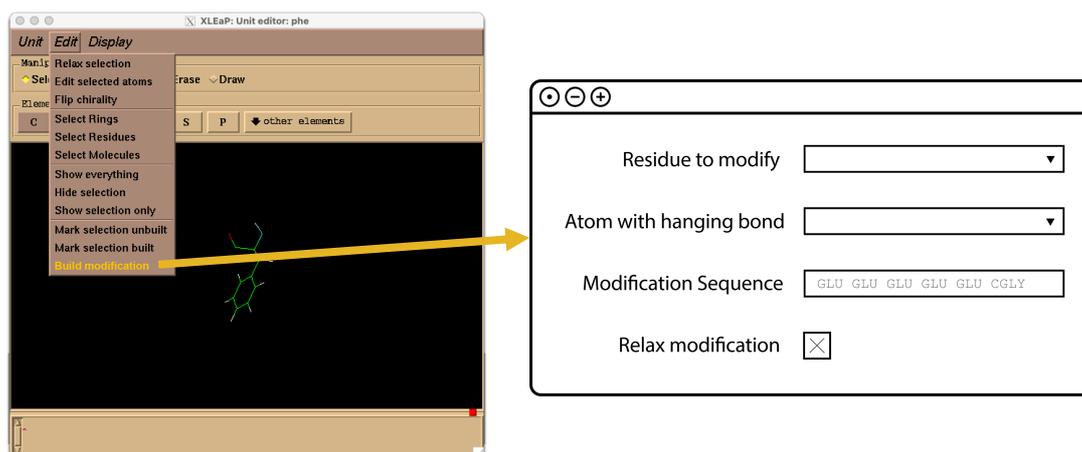


**Figure 20. Alternate glutamine capable of polyamination.** Skeletal representation of Gln (left) and the proposed alternate Gln (right), containing a side amide group with an unsatisfied valence. This results from the removal of a hydrogen from the side chain nitrogen.

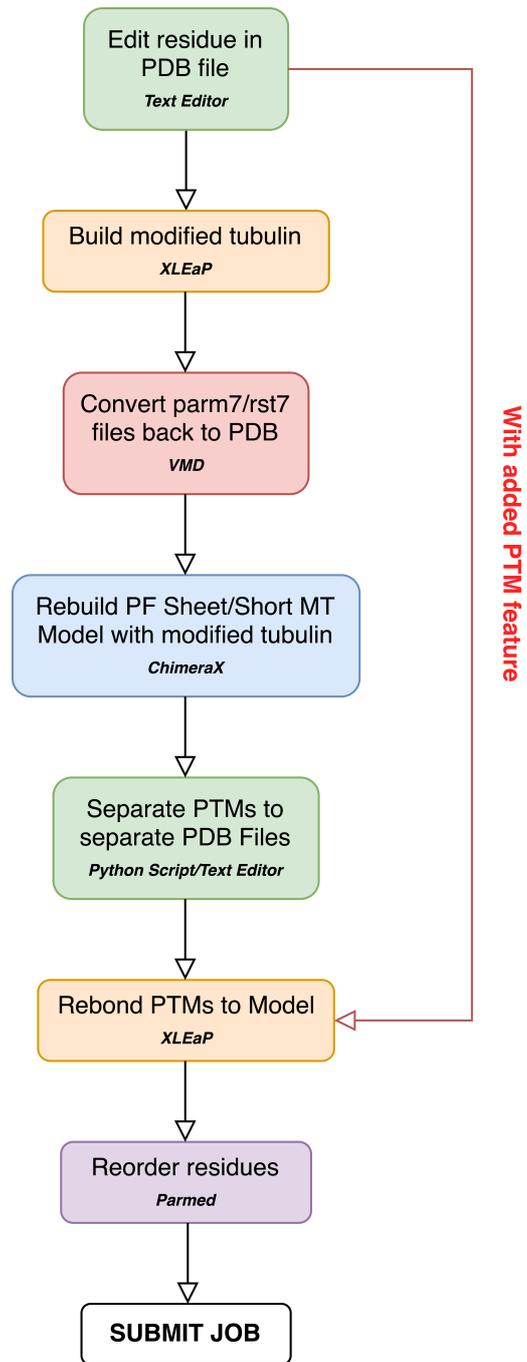
Acetylation at this site has been shown to improve resistance to breakages by mechanical stress (Eshun-Wilson et al., 2019). It would be interesting to study the combinatory effect of acetylating and polyglutamylating a microtubule, to see if this will enhance this resistance, or if they will cancel each other out. LEaP is shipped with sets of modified amino acid parameters, *leaprc.phosaa14SB* and *leaprc.protein.ff14SB\_modAA*, including acetyl-lysine and several phosphorylated amino acids, meaning building and simulating K40 acetylated microtubules can be easily achieved.

### **Improvements for Current Software**

The workflow for modifying multiple tubulins in larger structures presented in this project, whilst functional, is far from user-friendly. As it stands, the workflow requires 5 different pieces of software, as well as some knowledge of PDB file formatting. This is, in part, a result of the outdated XLEaP interface, which makes building and positioning multiple PTMs difficult on large models. A possible fix for this would be a new function and associated graphical interface within XLEaP, capable of building complex oligomer chains and bonding them to a protein at a given residue/atom, using either built-in or user-defined bond length/angle parameters (Fig. 21). This feature would significantly simplify the current modified system workflow by removing the need to convert modified tubulin coordinates and topology files to a PDB file; rebuild the larger model with modified heterodimers; and separate the PTM chains into their own PDB files to import them back into XLEaP as separate chains (Fig. 22). Lowering the barrier to entry for simulating post-translationally modified proteins should increase the likelihood of MD simulations being used as an additional tool for studying these proteins.



**Figure 21.** User interface mock-up for a hypothetical XLEaP PTM creation window. ‘Residue to modify’ is a dropdown menu that lists every residue in the protein using the same syntax as the existing *desc* function; ‘Atom with hanging bond’ is a dropdown menu that lists each atom in the residue selected above. PTMs would be bonded to the atom selected here; ‘Modification Sequence’ is a textbox that allows the user to enter the sequence of their PTM, using the same syntax as the existing *sequence* function; ‘Relax modification’ has the same function as in the edit menu: relaxing strained bonds, angles, and torsions by performing a limited energy minimisation on the defined atoms.



**Figure 22. Software Improvements can simplify the workflow.** Updated flowchart made possible with the hypothetical PTM function added to XLEaP.

## Conclusions

Over the course of this project, I have successfully developed a method for building poly(glutamate) and poly(glycine) chains onto non-terminal glutamate residues in multi-heterodimer models, for use in atomistic molecular dynamics simulations. This method could be adapted for use with more PTMs bound to other tubulin residues with little alteration. Atomistic simulations were run using two representative models: the *Short MT* and *PF Sheet* model. The smaller *PF Sheet* model was modified with 10-residue glycine and glutamate chains. These simulations showed that  $\beta$ -tubulin CTTs were the most mobile regions, the ends of the protofilaments bowed away from the centre of the models, and plus end-bound GDP molecules were more likely to diffuse away from the corresponding nucleotide binding sites. Positioning excess negatively charged residues between protofilaments caused the model to tear, hinting towards a theoretical mechanism for microtubule destabilisation. Inconsistencies between the *PF Sheet* and *Short MT* simulations further confirm previous findings that show smaller representative models (e.g., *PF Sheet*) do not act the same as complete microtubule models (e.g., *Short MT*), meaning they are unlikely to be suitable replacements in atomistic simulations. This project lays the groundwork for improved atomistic simulations of post-translationally modified microtubules in the future, and possibly larger scale coarse-grained simulations of these systems.

## Bibliography

- Adamopoulos, A., Landskron, L., Heidebrecht, T., Tsakou, F., Bleijerveld, O.B., Altelaar, M., Nieuwenhuis, J., Celie, P.H.N., Brummelkamp, T.R. and Perrakis, A. 2019. Crystal structure of the tubulin tyrosine carboxypeptidase complex VASH1–SVBP. *Nature Structural & Molecular Biology*. **26**(7), pp.567–570.
- Aillaud, C., Bosc, C., Peris, L., Bosson, A., Heemeryck, P., Van Dijk, J., Le Fric, J., Boulan, B., Vossier, F., Sanman, L.E., Syed, S., Amara, N., Couté, Y., Lafanechère, L., Denarier, E., Delphin, C., Pelletier, L., Humbert, S., Bogyo, M., Andrieux, A., Rogowski, K. and Moutin, M.-J. 2017. Vasohibins/SVBP are tubulin carboxypeptidases (TCPs) that regulate neuron differentiation. *Science*. **358**(6369), pp.1448–1453.
- Akella, J.S., Wloga, D., Kim, J., Starostina, N.G., Lyons-Abbott, S., Morrisette, N.S., Dougan, S.T., Kipreos, E.T. and Gaertig, J. 2010. MEC-17 is an  $\alpha$ -tubulin acetyltransferase. *Nature*. **467**(7312), pp.218–222.
- Akiyama, T., Kadowaki, T., Nishida, E., Kadooka, T., Ogawara, H., Fukami, Y., Sakai, H., Takaku, F. and Kasuga, M. 1986. Substrate specificities of tyrosine-specific protein kinases toward cytoskeletal proteins in vitro. *Journal of Biological Chemistry*. **261**(31), pp.14797–14803.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. 2002. *Molecular Biology of the Cell* [Online] 4th Editio. New York: Garland Science. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK26862/>.
- Alexander, J.E., Hunt, D.F., Lee, M.K., Shabanowitz, J., Michel, H., Berlin, S.C., MacDonald, T.L., Sundberg, R.J., Rebhun, L.I. and Frankfurter, A. 1991. Characterization of posttranslational modifications in neuron-specific class III beta-tubulin by mass spectrometry. *Proceedings of the National Academy of Sciences*. **88**(11), pp.4685–4689.
- Allnér, O., Nilsson, L. and Villa, A. 2012. Magnesium ion-water coordination and exchange in biomolecular simulations. *Journal of Chemical Theory and Computation*. **8**(4), pp.1493–1502.
- Alushin, G.M., Lander, G.C., Kellogg, E.H., Zhang, R., Baker, D. and Nogales, E. 2014. High-Resolution microtubule structures reveal the structural transitions in  $\alpha$ -tubulin upon GTP hydrolysis. *Cell*. **157**(5), pp.1117–1129.
- Aparna, J.S., Padinhateeri, R. and Das, D. 2020. Regulation of microtubule disassembly by spatially heterogeneous patterns of acetylation. *Soft Matter*. **16**(12), pp.3125–3136.
- Baas, P.W., Deitch, J.S., Black, M.M. and Banker, G.A. 1988. Polarity orientation of microtubules in hippocampal neurons: Uniformity in the axon and nonuniformity in the dendrite. *Proceedings of the National Academy of Sciences of the United States of America*. **85**(21), pp.8335–8339.
- Bekker, H., Berendsen, H., Dijkstra, E., Achterop, S., Van Drunen, R., Van der Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B. and Renardus, M. 1993. Gromacs: A parallel computer for molecular dynamics simulations. *Physics Computing*. **92**(January), pp.252–256.
- Berman, H., Henrick, K. and Nakamura, H. 2003. Announcing the worldwide Protein Data Bank. *Nature Structural Biology*. **10**(12), p.980.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*. **28**(1), pp.235–242.
- Bigman, L.S. and Levy, Y. 2021. Modulating Microtubules: A Molecular Perspective on the Effects of Tail Modifications. *Journal of Molecular Biology*. **433**(13), p.166988.
- Bigman, L.S. and Levy, Y. 2020. Tubulin tails and their modifications regulate protein diffusion on microtubules. *Proceedings of the National Academy of Sciences of the United States of America*. **117**(16), pp.8876–8883.
- Bradley, R. and Radhakrishnan, R. 2013. Coarse-grained models for protein-cell membrane interactions. *Polymers*. **5**(3), pp.890–936.
- Bre, M.H., Redeker, V., Quibell, M., Darmanaden-Delorme, J., Bressac, C., Cosson, J., Huitorel, P., Schmitter, J.M., Rossler, J., Johnson, T., Adoutte, A. and Levilliers, N. 1996. Axonemal tubulin polyglycylation probed with two monoclonal antibodies: widespread evolutionary distribution, appearance during spermatozoan maturation and possible function in motility. *Journal of Cell Science*. **109**(4), pp.727–738.
- Brinkley, B.R. 1985. Microtubule Organizing Centers. *Annual Review of Cell Biology*. **1**(1), pp.145–172.
- Burton, P.R. 1988. Dendrites of mitral cell neurons contain microtubules of opposite polarity. *Brain Research*. **473**(1), pp.107–115.
- Case, D.A., Aktulga, H.M., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cisneros,

- G.A., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., Gilson, M.K., Gohlke, H., Goetz, A.W., Harris, R., Izadi, S., Izmailov, S.A., Jin, C., Kasavajhala, K., Kaymak, M.C., King, E., Kovalenko, A., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Machado, M., Man, V., Manathunga, M., Merz, K.M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., O'Hearn, K.A., Onufriev, A., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Skrynnikov, N.R., Smith, J., Swails, J., Walker, R.C., Wang, J., Wei, H., Wolf, R.M., Wu, X., Xue, Y., York, D.M. and Kollman, P.A. 2021a. Amber 2021 Reference Manual. [Accessed 1 November 2021]. Available from: <https://ambermd.org/doc12/Amber21.pdf>.
- Case, D.A., Aktulga, H.M., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cisneros, G.A., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., Gilson, M.K., Gohlke, H., Goetz, A.W., Harris, R., Izadi, S., Izmailov, S.A., Jin, C., Kasavajhala, K., Kaymak, M.C., King, E., Kovalenko, A., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Machado, M., Man, V., Manathunga, M., Merz, K.M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., O'Hearn, K.A., Onufriev, A., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Skrynnikov, N.R., Smith, J., Swails, J., Walker, R.C., Wang, J., Wei, H., Wolf, R.M., Wu, X., Xue, Y., York, D.M., Zhao, S. and Kollman, P.A. 2020. Amber 20.
- Case, D.A., Aktulga, H.M., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cisneros, G.A., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., Gilson, M.K., Gohlke, H., Goetz, A.W., Harris, R., Izadi, S., Izmailov, S.A., Jin, C., Kasavajhala, K., Kaymak, M.C., King, E., Kovalenko, A., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Machado, M., Man, V., Manathunga, M., Merz, K.M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., O'Hearn, K.A., Onufriev, A., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Skrynnikov, N.R., Smith, J., Swails, J., Walker, R.C., Wang, J., Wei, H., Wolf, R.M., Wu, X., Xue, Y., York, D.M., Zhao, S. and Kollman, P.A. 2021b. AmberTools21.
- Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. 2005. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. **26**(16), pp.1668–1688.
- Cavalier-Smith, T. 1974. Basal body and flagellar development during the vegetative cell cycle and the sexual cycle of *Chlamydomonas reinhardtii*. *Journal of Cell Science*. **16**(3), pp.529–556.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*. **66**(1), pp.12–21.
- Chu, C.-W., Hou, F., Zhang, J., Phu, L., Loktev, A. V., Kirkpatrick, D.S., Jackson, P.K., Zhao, Y. and Zou, H. 2011. A novel acetylation of  $\beta$ -tubulin by San modulates microtubule polymerization via down-regulating tubulin incorporation. E. L. F. Holzbaur, ed. *Molecular Biology of the Cell*. **22**(4), pp.448–456.
- Cowan, N.J., Dobner, P.R., Fuchs, E. V and Cleveland, D.W. 1983. Expression of human alpha-tubulin genes: interspecies conservation of 3' untranslated regions. *Molecular and Cellular Biology*. **3**(10), pp.1738–1745.
- Delandre, C., Amikura, R. and Moore, A.W. 2016. Microtubule nucleation and organization in dendrites. *Cell Cycle*. **15**(13), pp.1685–1692.
- Deloukas, P., Matthews, L.H., Ashurst, J., Burton, J., Gilbert, J.G.R., Jones, M., Stavrides, G., Almeida, J.P., Babbage, A.K., Baggeley, C.L., Bailey, J., Barlow, K.F., Bates, K.N., Beard, L.M., Beare, D.M., Beasley, O.P., Bird, C.P., Blakey, S.E., Bridgeman, A.M., Brown, A.J., Buck, D., Burrill, W., Butler, A.P., Carder, C., Carter, N.P., Chapman, J.C., Clamp, M., Clark, G., Clark, L.N., Clark, S.Y., Clee, C.M., Clegg, S., Cobley, V.E., Collier, R.E., Connor, R., Corby, N.R., Coulson, A., Coville, G.J., Deadman, R., Dhami, P., Dunn, M., Ellington, A.G., Frankland, J.A., Fraser, A., French, L., Garner, P., Graffham, D. V, Griffiths, C., Griffiths, M.N.D., Gwilliam, R., Hall, R.E., Hammond, S., Harley, J.L., Heath, P.D., Ho, S., Holden, J.L., Howden, P.J., Huckle, E., Hunt, A.R., Hunt, S.E., Jekosch, K., Johnson, C.M., Johnson, D., Kay, M.P., Kimberley, A.M., King, A., Knights, A., Laird, G.K., Lawlor, S., Lezvaslaiho, M.H., Leversha, M., Lloyd, C., Lloyd, D.M., Lovell, J.D., Marsh, V.L., Martin, S.L., McConnachie, L.J., McLay, K., McMurray, A.A., Milne, S., Mistry, D., Moore, M.J.F., Mullikin, J.C., Nickerson, T., Oliver, K., Parker, A., Patel, R., Pearce, T.A. V, Peck, A.I., Phillimore, B.J.C.T., Prathalingam, S.R., Plumb, R.W., Ramsay, H., Rice, C.M., Ross, M.T., Scott, C.E., Sehra, H.K., Showkeen, R., Sims, S., Skuce, C.D., Smith, M.L., Soderlund, C., Steward, C.A., Sulston, J.E., Swann, M., Sycamore, N., Taylor, R., Tee, L., Thomas, D.W., Thorpe, A., Tracey, A., Tromans, A.C., Vaudin, M., Wall, M., Wallis, J.M., Whitehead, S.L., Whittaker, P., Willey, D.L., Williams, L., Williams, S.A., Wilming, L., Wray, P.W., Hubbard, T., Durbin, R.M., Bentley, D.R., Beck, S., Rogers, J. and Institute, T.W.T.S. 2001. The DNA sequence and comparative analysis of human chromosome 20. *Nature*. **414**(6866), pp.865–871.
- Desai, A. and Mitchison, T.J. 1997. Microtubule polymerization dynamics. *Annual review of cell and developmental biology*. **13**(1), pp.83–117.

- van Dijk, J., Rogowski, K., Miro, J., Lacroix, B., Eddé, B. and Janke, C. 2007. A Targeted Multienzyme Mechanism for Selective Microtubule Polyglutamylation. *Molecular Cell*. **26**(3), pp.437–448.
- Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. and Shaw, D.E. 2012. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*. **41**(1), pp.429–452.
- Dunn, S., Morisson, E.E., Liverpool, T.B., Molina-París, C., Cross, R.A., Alonso, M.C. and Peckham, M. 2008. Differential trafficking of Kif5c on tyrosinated and detyrosinated microtubules in live cells. *Journal of Cell Science*. **121**(7), pp.1085–1095.
- Ebberink, E., Fernandes, S., Hatzopoulos, G., Agashe, N., Guidotti, N., Reichart, T.M., Reymond, L., Velluz, M.-C., Schneider, F., Pourroy, C., Janke, C., Gönczy, P., Fierz, B. and Aumeier, C. 2022. Tubulin engineering by semisynthesis reveals that polyglutamylation directs detyrosination. *bioRxiv.*, pp.1–14.
- Eddé, B., Rossier, J., Le Caer, J.-P.P., Desbruyères, E., Gros, F. and Denoulet, P. 1990. Posttranslational glutamylation of  $\alpha$ -tubulin. *Science*. **247**(4938), pp.83–85.
- Ersfeld, K., Wehland, J., Plessmann, U., Dodemont, H., Gerke, V. and Weber, K. 1993. Characterization of the tubulin-tyrosine ligase. *Journal of Cell Biology*. **120**(3), pp.725–732.
- Eshun-Wilson, L., Zhang, R., Portran, D., Nachury, M. V., Toso, D.B., Löhr, T., Vendruscolo, M., Bonomi, M., Fraser, J.S. and Nogales, E. 2019. Effects of  $\alpha$ -tubulin acetylation on microtubule structure and stability. *Proceedings of the National Academy of Sciences of the United States of America*. **116**(21), pp.10366–10371.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. 1995. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. **103**(19), pp.8577–8593.
- Feng, W., Liu, R., Xie, X., Diao, L., Gao, N., Cheng, J., Zhang, X., Li, Y. and Bao, L. 2021. SUMOylation of  $\alpha$ -tubulin is a novel modification regulating microtubule dynamics. *Journal of Molecular Cell Biology*. **13**(2), pp.91–103.
- Field, C.R. 2022. Computer Simulations of Post-Translationally Modified Microtubules - Supporting Software/Scripts. *GitHub*. [Online]. [Accessed 19 September 2022]. Available from: <https://github.com/crfield18/AMBER-for-HPC>.
- Fiore, M., Goulas, C. and Pillois, X. 2017. A new mutation in TUBB1 associated with thrombocytopenia confirms that C-terminal part of  $\beta$ 1-tubulin plays a role in microtubule assembly. *Clinical Genetics*. **91**(6), pp.924–926.
- Fourest-Lieuvain, A., Peris, L., Gache, V., Garcia-Saez, I., Juillan-Binard, C., Lantéz, V. and Job, D. 2006. Microtubule Regulation in Mitosis: Tubulin Phosphorylation by the Cyclin-dependent Kinase Cdk1. *Molecular Biology of the Cell*. **17**(3), pp.1041–1050.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H. and Ferrin, T.E. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*. **27**(1), pp.14–25.
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*. **38**(SUPPL. 2), pp.695–699.
- Grafmüller, A. and Voth, G.A. 2011. Intrinsic bending of microtubule protofilaments. *Structure*. **19**(3), pp.409–417.
- Grau, M.B., Curto, G.G., Rocha, C., Magiera, M.M., Sousa, P.M., Giordano, T., Spassky, N. and Janke, C. 2013. Tubulin glycyllases and glutamylases have distinct functions in stabilization and motility of ependymal cilia. *Journal of Cell Biology*. **202**(3), pp.441–451.
- Harris, B.J., Ross, J.L. and Hawkins, T.L. 2018. Microtubule seams are not mechanically weak defects. *Physical Review E*. **97**(6), pp.1–7.
- Huang, K., Diener, D.R. and Rosenbaum, J.L. 2009. The ubiquitin conjugation system is involved in the disassembly of cilia and flagella. *Journal of Cell Biology*. **186**(4), pp.601–613.
- Hubbert, C., Guardiola, A., Shao, R., Kawaguchi, Y., Ito, A., Nixon, A., Yoshida, M., Wang, X.-F. and Yao, T.-P. 2002. HDAC6 is a microtubule-associated deacetylase. *Nature*. **417**(6887), pp.455–458.
- HUGO Gene Nomenclature Committee (HGNC) 2021. Gene Group: Tubulins. HGNC Database. [Accessed 16 September 2022]. Available from: <https://www.genenames.org/data/genegroup/#!/group/778>.
- Humphrey, W., Dalke, A. and Schulten, K. 1996. VMD: visual molecular dynamics. *Journal of molecular graphics*. **14**(1), pp.33–8, 27–8.
- Igaev, M. and Grubmüller, H. 2022. Bending-torsional elasticity and energetics of the plus-end microtubule tip. *Proceedings of the National Academy of Sciences of the United States of America*. **119**(12), pp.1–12.

- Igaev, M. and Grubmüller, H. 2020. Microtubule instability driven by longitudinal and lateral strain propagation. *PLoS Computational Biology*. **16**(9), pp.1–21.
- Ikegami, K., Mukai, M., Tsuchida, J., Heier, R.L., MacGregor, G.R. and Setou, M. 2006. TLL7 Is a Mammalian  $\beta$ -Tubulin Polyglutamylase Required for Growth of MAP2-positive Neurites. *Journal of Biological Chemistry*. **281**(41), pp.30707–30716.
- Ikegami, K. and Setou, M. 2009. TLL10 can perform tubulin glycylation when co-expressed with TLL8. *FEBS Letters*. **583**(12), pp.1957–1963.
- Janke, C. and Magiera, M.M. 2020. The tubulin code and its role in controlling microtubule properties and functions. *Nature Reviews Molecular Cell Biology*. **21**(6), pp.307–326.
- Janke, C., Rogowski, K., Wloga, D., Regnard, C., Kajava, A. V., Strub, J.-M., Temurak, N., van Dijk, J., Boucher, D., van Dorsselaer, A., Suryavanshi, S., Gaertig, J. and Eddé, B. 2005. Tubulin Polyglutamylase Enzymes Are Members of the TTL Domain Protein Family. *Science*. **308**(5729), pp.1758–1762.
- Jiang, N., Bailey, M.E., Burke, J., Ross, J.L. and Dima, R.I. 2017. Modeling the effects of lattice defects on microtubule breaking and healing. *Cytoskeleton*. **74**(1), pp.3–17.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. **79**(2), pp.926–935.
- Katsuki, M., Drummond, D.R. and Cross, R.A. 2014. Ectopic A-lattice seams destabilize microtubules. *Nature Communications*. **5**.
- Kaul, N., Soppina, V. and Verhey, K.J. 2014. Effects of  $\alpha$ -tubulin K40 acetylation and detyrosination on kinesin-1 motility in a purified system. *Biophysical Journal*. **106**(12), pp.2636–2643.
- Kimura, Y., Kurabe, N., Ikegami, K., Tsutsumi, K., Konishi, Y., Kaplan, O.I., Kunitomo, H., Iino, Y., Blacque, O.E. and Setou, M. 2010. Identification of Tubulin Deglutamylase among *Caenorhabditis elegans* and Mammalian Cytosolic Carboxypeptidases (CCPs). *Journal of Biological Chemistry*. **285**(30), pp.22936–22941.
- Koshland, D.E., Mitchison, T.J. and Kirschner, M.W. 1988. Polewards chromosome movement driven by microtubule depolymerization in vitro. *Nature*. **331**(6156), pp.499–504.
- Kunishima, S., Kobayashi, R., Itoh, T.J., Hamaguchi, M. and Saito, H. 2009. Mutation of the  $\beta$ 1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. *Blood*. **113**(2), pp.458–461.
- Kunishima, S., Nishimura, S., Suzuki, H., Imaizumi, M. and Saito, H. 2014. TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia. *European Journal of Haematology*. **92**(4), pp.276–282.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**(21), pp.2947–2948.
- Lasser, M., Tiber, J. and Lowery, L.A. 2018. The role of the microtubule cytoskeleton in neurodevelopmental disorders. *Frontiers in Cellular Neuroscience*. **12**(June), pp.1–18.
- Lawson, C.L., Patwardhan, A., Baker, M.L., Hryc, C., Garcia, E.S., Hudson, B.P., Lagerstedt, I., Ludtke, S.J., Pintilie, G., Sala, R., Westbrook, J.D., Berman, H.M., Kleywegt, G.J. and Chiu, W. 2016. EMDataBank unified data resource for 3DEM. *Nucleic Acids Research*. **44**(D1), pp.D396–D403.
- LeDizet, M. and Piperno, G. 1987. Identification of an acetylation site of *Chlamydomonas* alpha-tubulin. *Proceedings of the National Academy of Sciences*. **84**(16), pp.5720–5724.
- Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S. and Gascuel, O. 2019. NGPhylogeny.fr: New generation phylogenetic services for non-specialists. *Nucleic Acids Research*. **47**(W1), pp.W260–W265.
- Li, F., Hu, Y., Qi, S., Luo, X. and Yu, H. 2019. Structural basis of tubulin detyrosination by vasohibins. *Nature Structural & Molecular Biology*. **26**(7), pp.583–591.
- Li, S., Fernandez, J.J., Marshall, W.F. and Agard, D.A. 2012. Three-dimensional structure of basal body triplet revealed by electron cryo-tomography. *EMBO Journal*. **31**(3), pp.552–562.
- Liao, S., Rajendraprasad, G., Wang, N., Eibes, S., Gao, J., Yu, H., Wu, G., Tu, X., Huang, H., Barisic, M. and Xu, C. 2019. Molecular basis of vasohibins-mediated detyrosination and its impact on spindle function and mitosis. *Cell Research*. **29**(7), pp.533–547.

- Löwe, J., Li, H., Downing, K.H. and Nogales, E. 2001. Refined structure of  $\alpha\beta$ -tubulin at 3.5 Å resolution. *Journal of Molecular Biology*. **313**(5), pp.1045–1057.
- Ludueña, R.F., Zimmermann, H.-P. and Little, M. 1988. Identification of the phosphorylated  $\beta$ -tubulin isotype in differentiated neuroblastoma cells. *FEBS Letters*. **230**(1–2), pp.142–146.
- Maahs, L., Sanchez, B.E., Gupta, N., Van Harn, M., Barrack, E.R., Reddy, P.V. and Hwang, C. 2019. Class III  $\beta$ -tubulin expression as a predictor of docetaxel-resistance in metastatic castration-resistant prostate cancer. *PLoS ONE*. **14**(10), pp.1–10.
- MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D. and Karplus, M. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*. **102**(18), pp.3586–3616.
- Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*. **11**(8), pp.3696–3713.
- Mandelkow, E.M., Mandelkow, E. and Milligan, R.A. 1991. Microtubule dynamics and microtubule caps: a time-resolved cryo-electron microscopy study. *Journal of Cell Biology*. **114**(5), pp.977–991.
- Mani, N. and Subramanian, R. 2021. The tail wags the tubulin. *Developmental Cell*. **56**(14), pp.2007–2009.
- Margolin, G., Gregoret, I. V., Cickovski, T.M., Li, C., Shi, W., Alber, M.S. and Goodson, H. V. 2012. The mechanisms of microtubule catastrophe and rescue: Implications from analysis of a dimer-scale computational model. *Molecular Biology of the Cell*. **23**(4), pp.642–656.
- Margolis, R.L. and Wilson, L. 1981. Microtubule treadmills—possible molecular machinery. *Nature*. **293**(5835), pp.705–711.
- Marrink, S.J., De Vries, A.H. and Mark, A.E. 2004. Coarse Grained Model for Semiquantitative Lipid Simulations. *Journal of Physical Chemistry B*. **108**(2), pp.750–760.
- Mary, J., Redeker, V., Le Caer, J.P., Rossier, J. and Schmitter, J.M. 1997. Posttranslational modifications of axonemal tubulin. *Journal of Protein Chemistry*. **16**(5), pp.403–407.
- Matten, W.T., Aubry, M., West, J. and Maness, P.F. 1990. Tubulin is phosphorylated at tyrosine by pp60c-src in nerve growth cone membranes. *Journal of Cell Biology*. **111**(5), pp.1959–1970.
- McKenney, R.J., Huynh, W., Vale, R.D. and Sirajuddin, M. 2016. Tyrosination of  $\alpha$ -tubulin controls the initiation of processive dynein–dynactin motility. *The EMBO Journal*. **35**(11), pp.1175–1185.
- Meagher, K.L., Redman, L.T. and Carlson, H.A. 2003. Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry*. **24**(9), pp.1016–1025.
- Miller, B.R., McGee, T.D., Swails, J.M., Homeyer, N., Gohlke, H. and Roitberg, A.E. 2012. MMPBSA . py : An Efficient Program for End-State Free Energy Calculations.
- Mitchison, T. and Kirschner, M. 1984a. Dynamic instability of microtubule growth. *Nature*. **312**(5991), pp.237–242.
- Mitchison, T. and Kirschner, M. 1984b. Microtubule assembly nucleated by isolated centrosomes and dynamic instability of microtubule growth. *Nature*. **312**(V), pp.232–242.
- Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P. and Marrink, S.J. 2008. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*. **4**(5), pp.819–834.
- Mungall, A.J., Palmer, S.A., Sims, S.K., Edwards, C.A., Ashurst, J.L., Wilming, L., Jones, M.C., Horton, R., Hunt, S.E., Scott, C.E., Gilbert, J.G.R., Clamp, M.E., Bethel, G., Milne, S., Ainscough, R., Almeida, J.P., Ambrose, K.D., Andrews, T.D., Ashwell, R.I.S., Babbage, A.K., Bagguley, C.L., Bailey, J., Banerjee, R., Barker, D.J., Barlow, K.F., Bates, K., Beare, D.M., Beasley, H., Beasley, O., Bird, C.P., Blakey, S., Bray-Allen, S., Brook, J., Brown, A.J., Brown, J.Y., Burford, D.C., Burrill, W., Burton, J., Carder, C., Carter, N.P., Chapman, J.C., Clark, S.Y., Clark, G., Clee, C.M., Clegg, S., Cobley, V., Collier, R.E., Collins, J.E., Colman, L.K., Corby, N.R., Coville, G.J., Culley, K.M., Dhami, P., Davies, J., Dunn, M., Earthrowl, M.E., Ellington, A.E., Evans, K.A., Faulkner, L., Francis, M.D., Frankish, A., Frankland, J., French, L., Garner, P., Garnett, J., Ghori, M.J.R., Gilby, L.M., Gillson, C.J., Glithero, R.J., Grafham, D. V., Grant, M., Gribble, S., Griffiths, C., Griffiths, M., Hall, R., Halls, K.S., Hammond, S., Harley, J.L., Hart, E.A., Heath, P.D., Heathcott, R., Holmes, S.J., Howden, P.J., Howe, K.L., Howell, G.R., Huckle, E., Humphray, S.J., Humphries, M.D., Hunt, A.R., Johnson,

- C.M., Joy, A.A., Kay, M., Keenan, S.J., Kimberley, A.M., King, A., Laird, G.K., Langford, C., Lawlor, S., Leongamornlert, D.A., Leversha, M., Lloyd, C.R., Lloyd, D.M., Loveland, J.E., Lovell, J., Martin, S., Mashreghi-Mohammadi, M., Maslen, G.L., Matthews, L., McCann, O.T., McLaren, S.J., McLay, K., McMurray, A., Moore, M.J.F., Mullikin, J.C., Niblett, D., Nickerson, T., Novik, K.L., Oliver, K., Overton-Larty, E.K., Parker, A., Patel, R., Pearce, A. V, Peck, A.I., Phillimore, B., Phillips, S., Plumb, R.W., Porter, K.M., Ramsey, Y., Ranby, S.A., Rice, C.M., Ross, M.T., Searle, S.M., Sehra, H.K., Sheridan, E., Skuce, C.D., Smith, S., Smith, M., Spraggon, L., Squares, S.L., Steward, C.A., Sycamore, N., Tamlyn-Hall, G., Tester, J., Theaker, A.J., Thomas, D.W., Thorpe, A., Tracey, A., Tromans, A., Tubby, B., Wall, M., Wallis, J.M., West, A.P., White, S.S., Whitehead, S.L., Whittaker, H., Wild, A., Willey, D.J., Wilmer, T.E., Wood, J.M., Wray, P.W., Wyatt, J.C., Young, L., Younger, R.M., Bentley, D.R., Coulson, A., Durbin, R., Hubbard, T., Sulston, J.E., Dunham, I., Rogers, J. and Beck, S. 2003. The DNA sequence and analysis of human chromosome 6. *Nature*. **425**(6960), pp.805–811.
- Nasedkin, A., Ermilova, I. and Swenson, J. 2021. Atomistic molecular dynamics simulations of tubulin heterodimers explain the motion of a microtubule. *European Biophysics Journal*. **50**(7), pp.927–940.
- Natarajan, K., Gadadhar, S., Souphron, J., Magiera, M.M. and Janke, C. 2017. Molecular interactions between tubulin tails and glutamylases reveal determinants of glutamylation patterns. *EMBO reports*. **18**(6), pp.1013–1026.
- Nieuwenhuis, J., Adamopoulos, A., Bleijerveld, O.B., Mazouzi, A., Stickel, E., Celie, P., Altaalar, M., Knipscheer, P., Perrakis, A., Blomen, V.A. and Brummelkamp, T.R. 2017. Vasohibins encode tubulin detyrosinating activity. *Science*. **358**(6369), pp.1453–1456.
- NIH National Library of Medicine 2023. TUBB3 tubulin beta 3 class III [ Homo sapiens (human) ] Gene ID: 10381. [Accessed 6 February 2023]. Available from: <https://www.ncbi.nlm.nih.gov/gene/10381#gene-expression>.
- Nogales, E., Whittaker, M., Milligan, R.A. and Downing, K.H. 1999. High-resolution model of the microtubule. *Cell*. **96**(1), pp.79–88.
- North, B.J., Marshall, B.L., Borra, M.T., Denu, J.M. and Verdin, E. 2003. The Human Sir2 Ortholog, SIRT2, Is an NAD<sup>+</sup>-Dependent Tubulin Deacetylase. *Molecular Cell*. **11**(2), pp.437–444.
- Ori-McKenney, K.M., McKenney, R.J., Huang, H.H., Li, T., Meltzer, S., Jan, L.Y., Vale, R.D., Wiita, A.P. and Jan, Y.N. 2016. Phosphorylation of  $\beta$ -Tubulin by the Down Syndrome Kinase, Minibrain/DYRK1a, Regulates Microtubule Dynamics and Dendrite Morphogenesis. *Neuron*. **90**(3), pp.551–563.
- Ozols, J. and Caron, J.M. 1997. Posttranslational modification of tubulin by palmitoylation: II. Identification of sites of palmitoylation. *Molecular Biology of the Cell*. **8**(4), pp.637–645.
- Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.-C., Davis, I.J., Fahey, C.C., Hacker, K.E., Verhey, K.J., Bedford, M.T., Jonasch, E., Rathmell, W.K. and Walker, C.L. 2016. Dual Chromatin and Cytoskeletal Remodeling by SETD2. *Cell*. **166**(4), pp.950–962.
- Paturle-Lafanechere, L., Edde, B., Denoulet, P., Van Dorselaer, A., Mazarguil, H., Le Caer, J.P., Wehland, J. and Job, D. 1991. Characterization of a major brain tubulin variant which cannot be tyrosinated. *Biochemistry*. **30**(43), pp.10523–10528.
- Peris, L., Wagenbach, M., Lafanechère, L., Brocard, J., Moore, A.T., Kozielski, F., Job, D., Wordeman, L. and Andrieux, A. 2009. Motor-dependent microtubule disassembly driven by tubulin tyrosination. *Journal of Cell Biology*. **185**(7), pp.1159–1166.
- Peters, J.D., Furlong, M.T., Asai, D.J., Harrison, M.L. and Geahlen, R.L. 1996. Syk, Activated by Cross-linking the B-cell Antigen Receptor, Localizes to the Cytosol Where It Interacts with and Phosphorylates  $\alpha$ -Tubulin on Tyrosine. *Journal of Biological Chemistry*. **271**(9), pp.4755–4762.
- Piedra, F.-A.A., Kim, T., Garza, E.S., Geyer, E.A., Burns, A., Ye, X. and Rice, L.M. 2016. GDP-To-GTP exchange on the microtubule end can contribute to the frequency of catastrophe S. Reck-Peterson, ed. *Molecular Biology of the Cell*. **27**(22), pp.3515–3525.
- Ploussard, G., Terry, S., Maillé, P., Allory, Y., Sirab, N., Kheuang, L., Soyeux, P., Nicolaiw, N., Coppolani, E., Paule, B., Salomon, L., Culine, S., Buttyan, R., Vacherot, F. and De La Taille, A. 2010. Class III  $\beta$ -tubulin expression predicts prostate tumor aggressiveness and patient response to docetaxel-based chemotherapy. *Cancer Research*. **70**(22), pp.9253–9264.
- Prota, A.E., Bargsten, K., Zurwerra, D., Field, J.J., Díaz, J.F., Altmann, K.-H. and Steinmetz, M.O. 2013. Molecular Mechanism of Action of Microtubule-Stabilizing Anticancer Agents. *Science*. **339**(6119), pp.587–590.
- Prota, A.E., Magiera, M.M., Kuijpers, M., Bargsten, K., Frey, D., Wieser, M., Jaussi, R., Hoogenraad, C.C., Kammerer, R.A., Janke, C. and Steinmetz, M.O. 2013. Structural basis of tubulin tyrosination by tubulin tyrosine ligase. *Journal of Cell Biology*. **200**(3), pp.259–270.

- PubChem 2022. PubChem Compound Summary for CID 33032, Glutamic Acid. *National Center for Biotechnology Information*. [Online]. [Accessed 29 September 2022]. Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/Glutamic-acid>.
- Ranganathan, S., Dexter, D.W., Benetatos, C.A. and Hudes, G.R. 1998. Cloning and sequencing of human  $\beta$ (III)-tubulin cDNA: Induction of  $\beta$ (III) isotype in human prostate carcinoma cells by acute exposure to antimicrotubule agents. *Biochimica et Biophysica Acta - Gene Structure and Expression*. **1395**(2), pp.237–245.
- Redeker, V., Levilliers, N., Schmitter, J.-M.M., Le Caer, J.-P.P., Rossier, J., Adoutte, A. and Bré, M.-H.H. 1994. Polyglycylation of Tubulin: a Posttranslational Modification in Axonal Microtubules. *Science*. **266**(5191), pp.1688–1691.
- Ren, Y., Zhao, J. and Feng, J. 2003. Parkin Binds to  $\alpha/\beta$  Tubulin and Increases their Ubiquitination and Degradation. *The Journal of Neuroscience*. **23**(8), pp.3316–3324.
- Rocha, C., Papon, L., Cacheux, W., Marques Sousa, P., Lascano, V., Tort, O., Giordano, T., Vacher, S., Lemmers, B., Mariani, P., Meseure, D., Medema, J.P., Bièche, I., Hahne, M. and Janke, C. 2014. Tubulin glycyllases are required for primary cilia, control of cell proliferation and tumor development in colon. *The EMBO Journal*. **33**(22), pp.2735–2735.
- Rogowski, K., van Dijk, J., Magiera, M.M., Bosc, C., Deloulme, J.-C., Bosson, A., Peris, L., Gold, N.D., Lacroix, B., Grau, M.B., Bec, N., Larroque, C., Desagher, S., Holzer, M., Andrieux, A., Moutin, M.-J. and Janke, C. 2010. A Family of Protein-Deglutamylating Enzymes Associated with Neurodegeneration. *Cell*. **143**(4), pp.564–578.
- Rogowski, K., Juge, F., van Dijk, J., Wloga, D., Strub, J.-M., Levilliers, N., Thomas, D., Bré, M.-H., Van Dorsseleer, A., Gaertig, J. and Janke, C. 2009. Evolutionary Divergence of Enzymatic Mechanisms for Posttranslational Polyglycylation. *Cell*. **137**(6), pp.1076–1087.
- Rosas-Acosta, G., Russell, W.K., Deyrieux, A., Russell, D.H. and Wilson, V.G. 2005. A Universal Strategy for Proteomic Studies of SUMO and Other Ubiquitin-like Modifiers. *Molecular & Cellular Proteomics*. **4**(1), pp.56–72.
- van Rossum, G. and L. Drake, F. 2009. Python 3.9 Reference Manual. [Accessed 1 February 2022]. Available from: <https://docs.python.org/3.9/reference/>.
- Rüdiger, M., Plessman, U., Klöppel, K.-D.D., Wehland, J. and Weber, K. 1992. Class II tubulin, the major brain  $\beta$  tubulin isotype is polyglutamylated on glutamic acid residue 435. *FEBS Letters*. **308**(1), pp.101–105.
- Šali, A. and Blundell, T.L. 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*. **234**(3), pp.779–815.
- Schneider, A., Plessmann, U., Felleisen, R. and Weber, K. 1998. Posttranslational modifications of trichomonad tubulins; identification of multiple glutamylation sites. *FEBS Letters*. **429**(3), pp.399–402.
- Shida, T., Cueva, J.G., Xu, Z., Goodman, M.B. and Nachury, M. V. 2010. The major  $\alpha$ -tubulin K40 acetyltransferase  $\alpha$ TAT1 promotes rapid ciliogenesis and efficient mechanosensation. *Proceedings of the National Academy of Sciences*. **107**(50), pp.21517–21522.
- Sirajuddin, M., Rice, L.M. and Vale, R.D. 2014. Regulation of microtubule motors by tubulin isotypes and post-translational modifications. *Nature Cell Biology*. **16**(4), pp.335–344.
- Song, Y., Kirkpatrick, L.L., Schilling, A.B., Helseth, D.L., Chabot, N., Keillor, J.W., Johnson, G.V.W. and Brady, S.T. 2013. Transglutaminase and Polyamination of Tubulin: Posttranslational Modification for Stabilizing Axonal Microtubules. *Neuron*. **78**(1), pp.109–123.
- Stearns, M.E., Connolly, J.A. and Brown, D.L. 1976. Cytoplasmic Microtubule Organizing Centers Isolated from *Polytomella agilis*. *Science*. **191**(4223), pp.188–191.
- Stengel, C., Newman, S.P., Leese, M.P., Potter, B.V.L., Reed, M.J. and Purohit, A. 2010. Class III B-tubulin expression and in vitro resistance to microtubule targeting agents. *British Journal of Cancer*. **102**(2), pp.316–324.
- Sui, H. and Downing, K.H. 2010. Structural basis of interprotofilament interaction and lateral deformation of microtubules. *Structure*. **18**(8), pp.1022–1031.
- Ti, S.-C.C., Alushin, G.M. and Kapoor, T.M. 2018. Human  $\beta$ -Tubulin Isoforms Can Regulate Microtubule Protofilament Number and Stability. *Developmental Cell*. **47**(2), pp.175–190.e5.
- Ti, S.C., Pamula, M.C., Howes, S.C., Duellberg, C., Cade, N.I., Kleiner, R.E., Forth, S., Surrey, T., Nogales, E. and

- Kapoor, T.M. 2016. Mutations in Human Tubulin Proximal to the Kinesin-Binding Site Alter Dynamic Instability at Microtubule Plus- and Minus-Ends. *Developmental Cell*. **37**(1), pp.72–84.
- Tort, O., Tanco, S., Rocha, C., Bièche, I., Seixas, C., Bosc, C., Andrieux, A., Moutin, M.-J., Avilés, F.X., Lorenzo, J. and Janke, C. 2014. The cytosolic carboxypeptidases CCP2 and CCP3 catalyze posttranslational removal of acidic amino acids S. Doxsey, ed. *Molecular Biology of the Cell*. **25**(19), pp.3017–3027.
- Tweedie, S., Braschi, B., Gray, K., Jones, T.E.M., Seal, R.L., Yates, B. and Bruford, E.A. 2021. Genenames.org: The HGNC and VGNC resources in 2021. *Nucleic Acids Research*. **49**(D1), pp.D939–D946.
- UniProt 1987. Human Tubulin alpha-1B chain (P68363). [Accessed 28 February 2021]. Available from: <https://www.uniprot.org/uniprot/P68363>.
- UniProt 2001. Human Tubulin beta-1 chain (Q9H4B7). [Accessed 21 September 2022]. Available from: <https://www.uniprot.org/uniprotkb/Q9H4B7/entry#sequences>.
- UniProt 2002. Human Tubulin beta-3 chain (Q13509). [Accessed 28 February 2021]. Available from: <https://www.uniprot.org/uniprot/Q13509>.
- UniProt 2004. Human Tubulin beta chain (P07437). [Accessed 21 September 2022]. Available from: <https://www.uniprot.org/uniprot/Q6ZT98>.
- Valenstein, M.L. and Roll-Mecak, A. 2016. Graded Control of Microtubule Severing by Tubulin Glutamylation. *Cell*. **164**(5), pp.911–921.
- Verhey, K.J. and Gaertig, J. 2007. The tubulin code. *Cell Cycle*. **6**(17), pp.2152–2160.
- Walker, R.C. 2008. AMBER Workshop - Tutorial A1 - Section 2 'Simulating a Solvated Protein that Contains Non-Standard Residues (Simple Version)'. *AMBERMD Advanced Tutorials*. [Online]. [Accessed 20 September 2022]. Available from: [http://ambermd.org/tutorials/advanced/tutorial1\\_orig/section2.htm](http://ambermd.org/tutorials/advanced/tutorial1_orig/section2.htm).
- Wall, K.P., Hart, H., Lee, T., Page, C., Hawkins, T.L. and Hough, L.E. 2020. C-Terminal Tail Polyglycylation and Polyglutamylation Alter Microtubule Mechanical Properties. *Biophysical Journal*. **119**(11), pp.2219–2230.
- Wall, K.P., Pagratis, M., Armstrong, G., Balsbaugh, J.L., Verbeke, E., Pearson, C.G. and Hough, L.E. 2016. Molecular Determinants of Tubulin's C-Terminal Tail Conformational Ensemble. *ACS Chemical Biology*. **11**(11), pp.2981–2990.
- Wang, N., Bosc, C., Ryul Choi, S., Boulan, B., Peris, L., Olieric, N., Bao, H., Krichen, F., Chen, L., Andrieux, A., Olieric, V., Moutin, M.-J., Steinmetz, M.O. and Huang, H. 2019. Structural basis of tubulin detyrosination by the vasohibin–SVBP enzyme complex. *Nature Structural & Molecular Biology*. **26**(7), pp.571–582.
- Wehenkel, A. and Janke, C. 2014. Towards elucidating the tubulin code. *Nature Cell Biology*. **16**(4), pp.303–305.
- Wloga, D., Webster, D.M., Rogowski, K., Bré, M.-H., Levilliers, N., Jerka-Dziadosz, M., Janke, C., Dougan, S.T. and Gaertig, J. 2009. TTLL3 Is a Tubulin Glycine Ligase that Regulates the Assembly of Cilia. *Developmental Cell*. **16**(6), pp.867–876.
- Wolff, A., Houdayer, M., Chillet, D., de Néchaud, B. and Denoulet, P. 1994. Structure of the polyglutamyl chain of tubulin: Occurrence of alpha and gamma linkages between glutamyl units revealed by monoreactive polyclonal antibodies. *Biology of the Cell*. **81**(1), pp.11–16.
- Word, J.M., Lovell, S.C., Richardson, J.S. and Richardson, D.C. 1999. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of Molecular Biology*. **285**(4), pp.1735–1747.
- wwPDB 2022. PDB File Format Documentation. *wwPDB Foundation*. [Online]. [Accessed 23 August 2022]. Available from: <https://www.wwpdb.org/documentation/file-format>.
- Yu, W., Centonze, V.E., Ahmad, F.J. and Baas, P.W. 1993. Microtubule nucleation and release from the neuronal centrosome. *Journal of Cell Biology*. **122**(2), pp.349–359.
- Zha, J., Zhang, Y., Xia, K., Gräter, F. and Xia, F. 2021. Coarse-Grained Simulation of Mechanical Properties of Single Microtubules With Micrometer Length. *Frontiers in Molecular Biosciences*. **7**(February), pp.1–11.
- Zhou, C., Yan, L., Zhang, W. and Liu, Z. 2019. Structural basis of tubulin detyrosination by VASH2/SVBP heterodimer. *Nature Communications*. **10**(1), p.3212.
- Zhuang, Z., Cummings, S.W., Roll-Mecak, A. and Tanner, M.E. 2022. Phosphinic acid-based inhibitors of tubulin polyglycylation. *Chemical Communications*. **58**(45), pp.6530–6533.

## Appendices

### Appendix A: Optimising AMBER input files for efficient simulations

The amount of time a molecular dynamics simulation takes to finish depends heavily on the size of the system (i.e., the number of atoms) and the parameters for each stage of the simulation. Improving average simulation rates allows them to finish in a shorter time frame, meaning longer simulations can be run, and the time between starting and analysing the data from a simulation is greatly reduced. This also has the added benefit of reduced total power consumption, as the HPC node running the simulation is running for less time. As this project was based on simulating systems containing over 1 million atoms, it was vital that optimal parameters are set for the simulation for it to run as quickly as possible.

A series of simulations containing a single tubulin heterodimer, without bound GTP/GDP, was run to test different MD parameters. Each system was solvated with 150 mM KCl in water, using a minimum distance from the periodic box bounds of 10 Å. System contained a total of 42,983 atoms. A series of 10 energy minimising steps (min1, min2, md1-8) was run for each simulation. 4 different production MD steps were run in parallel, each using a different set of parameters (Table 3). Water molecule trajectories are typically discarded, as far more focus is placed on analysing the protein in question, so we are not discarding any important data by not writing these trajectories.

The largest increase in simulation rate was caused by reducing *cut* from 12 Å to 8 Å. Doubling *ntwx* from 500 to 1000 caused the resulting output trajectories file to halve in size, from 121.04 GB to 60.52 GB. In combination with using *barostat* rather than *ntb*, these changes further increased the simulation rate by 25%. Setting *ntwprt* to a non-zero value had effectively no effect on the simulation rate. The decrease in rate of 0.73 ns/day can reasonably be explained by run-to-run variance. The trajectories file for Run 3 was just 13.73% of that from Run 2 (8.31

GB compared to 60.52 GB), greatly decreasing the volume of wasted storage being used to hold trajectories for atoms that are going to be stripped from the visualisation, regardless.

Parameter \ Input file	Original	Run 1	Run 2	Run 3
cut	12.0	8.0	8.0	8.0
ntwx	500	500	1000	1000
ntwr	500	500	1000	1000
ntb	2	2	-	-
barostat	-	-	2	2
ntwprt	0	0	0	13858
<b>Mean Rate (ns/day)</b>	115.74	201.92	242.47	241.74
<b>Rate relative to original</b>	100%	174%	209%	209%

**Table 3. Changes made to AMBER input files and resulting simulation performance.** *cut* is the non-bonded cut-off distance in Angstroms, at which long range electrostatics are calculated using the Particle Mesh Ewald (PME) method (Essmann et al., 1995; Case et al., 2021a); *ntwx* and *ntwr* determine the interval between which trajectories are written to the coordinates and restart files, respectively; *ntb* and *barostat* control the pressure exerted on the system (*ntb* = 2 applies constant pressure to the system) (*barostat* = 2 sets this to use the Monte Carlo barostat, which is better optimised for GPU simulations) (Case et al., 2021a); *ntwprt* determines for which atoms the trajectories are written to the output file (0 if unspecified).

The unmodified *PF Sheet* model contains over 10 times as many protein atoms as the single tubulin heterodimer used in these tests (195202 and 13943 atoms, respectively). A typical simulation using this model produces approximately 120 GB of trajectories files when using the optimisations detailed above, at a rate of between 10-15 ns/day depending on periodic box size and number of water residues. Without these optimisations, the trajectories would theoretically total around 1.75 TB, and simulations would run at less than half the rate of the optimised simulations. The optimisations made have, therefore, made simulating large, multi-dimer systems like the *PF Sheet* and *Short MT* models possible. This was achieved by speeding up the rate of simulation so that 100 ns can be simulated within two weeks, and by minimising storage space requirements by not retaining unwanted water molecule trajectories.

## Appendix B: Scripts/Input Files

Up to date versions of python scripts can be found at:

<https://github.com/crfield18/AMBER-for-HPC>

### (I) Building a modified tubulin heterodimer (LEaP)

This is an input file for LEaP that speeds up the process of building a tubulin heterodimer with a single PTM. *solvatebox* prints the periodic box volume to the terminal, which can be used to calculate the number of salt ions needed to reach 150 mM. Run the script without the *addionsrand* and *saveamberparm* steps, calculate the number of salt ions, and input the remaining commands manually. This input file only work in XLEaP (not TLEaP), as the PTM chain needs to be moved manually using the edit window. This step could be replaced if a command existed for translating atom coordinates. The reconstructed 5ij0 PDB file contains 901 protein residues. GTP, Mg<sup>2+</sup> and GDP are residues 902-904, respectively. PTMs typically begin at residue 905 but can sometimes occur before the ligands. For the larger *PF Sheet* and *Short MT* models, the second tubulin heterodimer will either begin at residue 905, if the previous tubulin is unmodified, or 905 + n where n is the length of the PTM chain (e.g., for a heterodimer modified with a 10 residue PTM, the N-terminus of the next heterodimer will be residue 912).

'example-leap-dimer-prep.in'

```
# Load the ff14SB force field for protein simulations
source leaprc.protein.ff14SB

# Load force field parameters for GTP, GDP, Mg2+, GLG and GGN
loadamberprep GDP.prep
loadamberprep GTP.prep
loadamberprep magnesium.prep
loadamberprep GLG.prep
loadamberprep GGN.prep

loadamberparams GTP.frcmod
loadamberparams GDP.frcmod
loadamberparams magnesium.frcmod
loadamberparams GLG.frcmod
loadamberparams GGN.frcmod

# Load the modifiable tubulin PDB
pdb = loadpdb example.pdb

# Create a unit for the ptm chain (e.g., 5 residue poly(E))
ptm = sequence { GLU GLU GLU GLU CGLU }

# Combine the pdb and ptm into a single unit
sum = combine {pdb ptm}

# Form a covalent bond between GLG and the N-terminus of the PTM
bond sum.889.CD sum.905.N

# Open the edit window
# Select the ptm and position as best you can with the mouse
# Then select the ptm and the surrounding residues and choose 'relax selection'
edit sum

# Load the water model and add water molecules
source leaprc.water.tip3p
solvatebox sum TIP3PBOX 10.0

# Add salt ions to the system. Concentration is calculated separately (replace n)
addionsrand sum K+ n
addionsrand sum Cl- n

# Export coordinates and topology for the solvated system
saveamberparm sum example.parm7 example.rst7
```

## (II) Modifying tubulin in a multi-dimer model (ChimeraX)

This is a batch script that builds a modified microtubule/protofilament sheet model using an unmodified scaffold model and a modified tubulin PDB file. This example uses a 4-dimer model, but it can easily be adapted to the desired number of heterodimers by increasing the number of modified tubulin PDB files loaded, and by incrementing the matchmaker selections (e.g., matchmaker #6 to #1.5, matchmaker #7 to #1.6, etc.).

### 'example.cxc'

```
# Load the unmodified PF Sheet/Short MT model
Open scaffold.pdb

# Load n copies of the modified tubulin pdb where n is the number of tubulins in the scaffold
open modified-tubulin.pdb
open modified-tubulin.pdb
open modified-tubulin.pdb
open modified-tubulin.pdb

# Position each modified tubulin into a different unmodified tubulin
matchmaker #2 to #1.1
matchmaker #3 to #1.2
matchmaker #4 to #1.3
matchmaker #5 to #1.4

# Save a ChimeraX project file
save "/path/to/directory/example.cxc"

# Select all the modified tubulins
sel #2-*

# Save a pdb file containing only modified tubulins, positioned as in the scaffold pdb.
save "/path/to/directory/example.pdb" selectedOnly true relModel #1
```

### (III) Split PTM chains into separate PDB files automatically

'*ptm-split.py*' automates the process of separating the PTM chains and protein body into separate PDB files (Field, 2022). It takes a single PDB file with multiple modified heterodimers as an input using *-i* as a command line argument.

#### '*ptm-split.py*' (Field, 2022)

```
import argparse

# This script uses -i as a command line argument to pick the input pdb file
parser = argparse.ArgumentParser()
parser.add_argument('-i', '-input', type=str, required=True)
args = parser.parse_args()

input = args.input

check_list = ('ATOM', 'HETATM')
tail_res = ()

# Check whether gtp/gdp have residue number before or after the ptm
# Currently set to check for 5-residue PTMs
# Add/remove residues from each 'tail_res' tuple as needed
with open(input, r) as file:
    for line in file:
        if not any(i in line for i in check_list): # Ignore any lines without 'ATOM' or 'HETATM'
            pass
        else:
            if 'gdp X 909' in line:
                print('PTM before ligands')
                tail_res = ('X 902', 'X 903', 'X 904', 'X 905', 'X 906')
                break
            else:
                print('PTM after ligands')
                tail_res = ('X 905', 'X 906', 'X 907', 'X 908', 'X 909')
                break

# Creates pdb files for the body and all tails, using the same filename as the pdb file
body = open(input.split('.')[0] + '-body.pdb', 'w')
tail_all = open(input.split('.')[0] + '-tails.pdb', 'w')

# Write body and tails out to separate pdb files
with open(input) as file:
    for line in file:
        if any(r in line for r in tail_res):
            if 'OXT' not in line or 'GGN' in line:
                tail_all.writelines(line)
            else:
                tail_all.writelines(line + 'TER\n')
        else:
            body.writelines(line)
    tail_all.writelines('END')
file.close()

# Split tails into separate pdb files
with open(input.split('.')[0] + '-tails.pdb', 'r') as tail_all:
    t = 1
    while t <= 14: # Currently set for 14 heterodimers. Change as needed.
        tail = open(input.split('.')[0] + '-tail' + str(t) + '.pdb', 'w')
        for line in tail_all:
            if 'TER' not in line:
                tail.writelines(line)
            else:
                tail.writelines(line + 'END\n')
        tail.close()
        t += 1
    break
```

**(IV) Reconstructing modified microtubules (LEaP)**

This script takes the separated body and PTM PDB files and rebuilds them in LEaP, making the modified microtubule/PF sheet compatible with . The sequence of the PTM must be set before importing the PTM PDB files to ensure they are interpreted correctly when importing. The N-terminus is not defined as an N-terminal residue (i.e., GLU nor NGLU), leaving an unsatisfied valence that will be used to draw the covalent bond between it and the tubulin residue being modified. If left unspecified, LEaP will make this an N-terminal residue when exporting coordinate or topology files. The C-terminus is set to CGLU, as this residue is not being modified further. Since all the PTM chains are loaded in separately from the main protein, once summed, their residue numbers will start after all the residues associated with the main protein and ligands. This influences which numbers need to be set for the bond commands.

'example-leap-md-prep.in'

```
# Load the ffl4SB force field for protein simulations
source leaprc.protein.ffl4SB

# Load force field parameters for GTP, GDP, Mg2+, GLG (and/or GGN)
loadamberprep GDP.prep
loadamberprep GTP.prep
loadamberprep magnesium.prep
loadamberprep GLG.prep
loadamberparams GTP.frcmod
loadamberparams GDP.frcmod
loadamberparams magnesium.frcmod
loadamberparams GLG.frcmod

# Load the pdb file containing only the main body of the model
body = loadpdb body.pdb

# Set the sequence for the PTM. This example is a 5-residue poly(E) chain
ptmseq = { GLU GLU GLU GLU CGLU }

# Load the separate tail pdb files using the PTM sequence defined previously
tail1 = loadpdbusingseq tail1.pdb ptmseq
tail2 = loadpdbusingseq tail2.pdb ptmseq

# Combine all the proteins into one summed unit
sum = combine {body tail1 tail2}

# Form covalent bonds between the body and PTMs. This example has the PTMs at beta tubulin E438
bond sum.889.CD sum.12657.N
bond sum.1793.CD sum.12667.N

# Load the water model and add water molecules
source leaprc.water.tip3p
solvatebox sum TIP3PBOX 10.0

# Add salt ions to the system. Concentration is calculated separately (replace n)
addionsrand sum K+ n
addionsrand sum Cl- n

# Export coordinates and topology for the solvated system
saveamberparm sum example.parm7 example.rst7
```

## (V) Running MD with PMEMD

These scripts are used as input files for PMEMD to set the parameters for each simulation. The first is one of multiple energy minimisation scripts which are used to reduce the potential energy in the simulation, by simulating the solvent with the protein being restrained. During this project I used 10 minimisation steps (min1, min2, md1-8), totalling 590000 cycles. The second is the production MD script, used for running a 100 ns protein simulation. For larger systems (i.e., systems containing many tubulin heterodimers), multiple copies of this script were used in series to reach 100 ns.

### 'min1.in' (energy minimisation)

```
# Change n in 'RES 1 n' to final C-terminal protein residue

&cntrl
  imin   = 1,
  maxcyc = 10000,
  ncyc   = 5000,
  ntb    = 1,
  ntr    = 1,
  cut    = 12.0
/
Hold the DNA fixed
500.0
RES 1 n
END
END
```

### 'md9.in' (production MD)

```
# Set ntwprt equal to the final atom associated with your protein/ligand (replace n)
# i.e., not water or salt ions

&cntrl
  irest=1, ntx=5,
  ntf=2, barostat=2, cut=8.0,
  nstlim=50000000, dt=0.002,
  temp0=300.0, ntt=1,
  ntp=1,
  ntc=2,
  ntwprt=n,
  ntwx=1000, ntwr=1000,
/
```

## (VI) ARC4 Submission Scripts

These input files contain the commands needed to run PMEMD jobs on ARC4. They worked on ARC4 from October 2021 until September 2022. Energy minimisation jobs were run on a full CPU node (40 cores). The large number of atoms comprising the *Short MTIPF Sheet* model requires the entire 192 GB of RAM of the node. Production MD jobs were run on 2 GPUs. *filename* should be set to the name as the *parm7* and *rst7* files used for the job. *previous-md* should be the same as *current-md* for the previous job. *current-md* should use the name of the current md input file. A typical series of jobs used in this project began with *min1.in* (l=rst7; f=min1) and continue through min2 (l=min1; f=min2), then md1 (l=min2; f=md1) through to md8. Production MD started at md9.

### CPU jobs:

```
#!/bin/sh

#$ -cwd -v
#$ -l nodes=1
#$ -l h_rt=48:00:00
#$ -N jobname

module purge
module add user
module add amber/20

n=filename
l=previous-md
f=current-md

mpirun pmemd.MPI -O -i $f.in -o $f.out -inf $f.inf -c $n.$l -ref $n.$l -r $n.$f -p $n.parm7 -x $n.$f.x
```

### GPU jobs:

```
#!/bin/sh

#$ -cwd -v
#$ -l coproc_v100=2
#$ -l h_rt=48:00:00
#$ -N jobname

module purge
module add user
module add cuda
module add amber/20gpu

n=filename
l=previous-md
f=current-md

mpirun -np 2 pmemd.cuda_SPPF.MPI -O -i $f.in -o $f.out -inf $f.inf -c $n.$l -ref $n.$l -r $n.$f -p $n.parm7 -x $n.$f.x
```

## Appendix C: PDB file formatting

The different pieces of software used for the modified tubulin/microtubule workflow (Fig. 7) output PDB files with small differences in formatting. Slight inconsistencies between how each program interprets and exports PDB files can prevent them from working in LEaP, causing the program to crash without providing an error message that explains why. Information on the PDB file format can be found on the wwPDB Foundation website (Berman et al., 2003; wwPDB, 2022).

When mutating a Glu residue to 'GLG', first remove hydrogen atoms from the protein using the *-trim* flag for the *Reduce*. Then change the residue name in columns 18-20 from *GLU* to *GLG* in the trimmed PDB file (Fig. 23). Delete the line associated with atom OE2. This is the side chain carboxylic acid oxygen atoms that is not present in 'GLG'. Using non-standard residues to build PTMs can introduce issues when exporting files from ChimeraX. ChimeraX exports PDB files containing non-standard residues, such as 'GLG', using the HETATM record type. By default, residues marked as HETATM are not bonded to adjacent amino acids. Therefore, any non-standard residues with the HETATM record type should be changed to ATOM. The PDB file format is dependent on the position of the data in each column. ATOM needs to be followed by 2 spaces so that correct column spacing is maintained for the rest of the line.

TER indicated the end of a molecule or protein chain. A TER line should be after the final residue in each protein chain, and at the end of each ligand molecule. This includes at the end of each tubulin subunit; between GTP, GDP and  $Mg^{2+}$ ; and either side of any PTM chains. When using the GTP and GDP parameters from Meagher et al., (2003), any reference to these molecules needs to be lower case (i.e. gtp instead of GTP). GTP and GDP atoms use ` or \* to indicate ribose atoms. When using these parameter files, use ` to indicate GTP ribose atoms, and \* to indicate GDP ribose atoms.

**Unmodified**

ATOM	7020	N	GLU	B 896	471.161	344.172	360.014	1.00221.58	N
ATOM	7021	CA	GLU	B 896	470.533	344.586	361.230	1.00221.58	C
ATOM	7022	CB	GLU	B 896	469.613	345.807	361.064	1.00221.58	C
ATOM	7023	CG	GLU	B 896	469.225	346.452	362.395	1.00221.58	C
ATOM	7024	CD	GLU	B 896	468.496	347.754	362.097	1.00221.58	C
ATOM	7025	OE1	GLU	B 896	468.159	347.985	360.905	1.00221.58	O
ATOM	7026	OE2	GLU	B 896	468.272	348.537	363.058	1.00221.58	O
ATOM	7027	C	GLU	B 896	469.721	343.438	361.720	1.00221.58	C
ATOM	7028	O	GLU	B 896	469.047	342.762	360.945	1.00221.58	O

**Modified**

ATOM	7020	N	GLG	B 896	471.161	344.172	360.014	1.00221.58	N
ATOM	7021	CA	GLG	B 896	470.533	344.586	361.230	1.00221.58	C
ATOM	7022	CB	GLG	B 896	469.613	345.807	361.064	1.00221.58	C
ATOM	7023	CG	GLG	B 896	469.225	346.452	362.395	1.00221.58	C
ATOM	7024	CD	GLG	B 896	468.496	347.754	362.097	1.00221.58	C
ATOM	7025	OE1	GLG	B 896	468.159	347.985	360.905	1.00221.58	O
ATOM	7027	C	GLG	B 896	469.721	343.438	361.720	1.00221.58	C
ATOM	7028	O	GLG	B 896	469.047	342.762	360.945	1.00221.58	O

**Figure 23. Mutating Glu to ‘GLG’ in a PDB file.** Changes made (green) to mutate a CTT Glu residue to ‘GLG’. Deletions are in cyan.

For large PDB files exported from VMD, such as the *Short MT* and *PF Sheet* models, atom and residue numbering can cause issues when loading these same files into other software like ChimeraX and LEaP. Past an atom number (serial) of 99999 and residue number (resSeq) of 9999, these values will begin to be written to the output file as hexadecimal values, rather than decimal. This is likely to maintain the strict column width rules imposed on PDB files to make them as inter-compatible with as many software packages as possible. The atom number can only occupy columns 7-11 and the residue number can only occupy columns 23-26 (wwPDB, 2022). ChimeraX and LEaP are not set up to interpret hexadecimal values in these columns. ChimeraX interprets atoms currently up to 99999, then ignores the remaining atoms, leaving large holes in the model. Loading the same PDB file into LEaP will cause it to crash. ‘*vmd-numifx.py*’ is design to take a PDB file exported from VMD and convert hexadecimal values in the atom and residue number columns (serial and resSeq, respectively) and convert them to base 10, fixing compatibility with ChimeraX and LEaP (Field, 2022).

'vmd-numfix.py' (Field, 2022)

```

import argparse

# # Convert base 16 values to base 10, then conform to column width restrictions of PDB files
def hex_to_dec(hex:str, clmn:int):
    dec = int(hex, 16)
    if len(str(dec)) > clmn:
        dec = str(dec)
        dec = dec[-int(clmn):]
    return dec

# # Check to see if atom/residue number is in hexadecimal
def check(b:int, e:int, field:bool, clmn:int):
    # If column values are not hexadecimal
    if field == False:
        # Check if columns contain any letters (i.e., a hexadecimal value)
        try:
            n = ''.join(line[b:e]).replace(' ', '')
            n = int(n)
            # Convert any hexadecimal values to decimal
        except ValueError:
            n = hex_to_dec(n, clmn)
            # All future values in these columns will be treated as hexadecimal
            field = True
        else:
            n = ''.join(line[b:e])
    # If column values are hexadecimal
    else:
        n = ''.join(line[b:e]).replace(' ', '')
        n = hex_to_dec(n, clmn)

    # Return the column values (converted or not converted)
    return n

if __name__ == "__main__":
    # # Command line argument for the input PDB file
    parser = argparse.ArgumentParser()
    parser.add_argument('-i', '--input', type=str, required=True)
    args = parser.parse_args()
    input_pdb = args.input

    # Are atom/residue numbers from the current line all hexadecimal or not?
    serial_hex, resSeq_hex = False, False
    # Previous atom/residue number to compare against current
    serial_prev, resSeq_prev = 0, 0
    # Output PDB file with have -numfix at the end of the file name
    output = open(input_pdb.split('.')[0] + '-numfix.pdb', 'w')

    with open(input_pdb, 'r') as file:
        for line in file:
            # Ignore any lines without 'ATOM' in the record columns (columns 1-6)
            if 'ATOM' not in line:
                pass
            else:
                # Convert the line into a list of characters and spaces
                line = list(line)

                # Set new atom number (serial) column values (columns 7 - 11)
                serial = check(6,11,serial_hex,5)
                line[6:11] = list(str(serial))

                # Set new residue num (resSeq) column values (columns 23 - 26)
                resSeq = check(22,26,resSeq_hex,4)
                if resSeq_prev == '9999' and resSeq != '9999':
                    resSeq_hex = True
                    resSeq = check(22,26,resSeq_hex,4)
                line[22:26] = list(str(resSeq))

                # Set current serial and resSeq values to previous
                serial_prev = serial
                resSeq_prev = resSeq

            # Write to output file ('*-numfix.pdb')
            output.writelines(''.join(line))

```