



POLITECNICO DI TORINO  
MASTER'S DEGREE IN BIOMEDICAL ENGINEERING  
Academic Year 2023/2024

MICROTUBULE TARGETING AGENTS  
A bridge between Paclitaxel and Microtubules lateral contacts

SUPERVISORS

PROF. J. A. TUSZYNSKI

PHD M. AMINPOUR

PHD P. VOTTERO

Candidate

Martina Centroni

## **ABSTRACT**

Microtubules (MT) are small cellular components that play a fundamental role in development, proliferation, spreading, and signaling of cells. The basic lattice of MTs is composed by the alternation of  $\alpha$  and  $\beta$  protein subunits called tubulins. Over the decades, they have become the main target of drugs involved in the treatment against tumors. Chemotherapy drugs can be divided into two main categories: stabilizing and destabilizing agents, which, according to their properties, can stabilize or destabilize the lattice of MTs, leading to cell apoptosis. This is the case of Paclitaxel (PTX), a prominent chemotherapy drug used in breast cancer treatment. It aims to stabilize MTs through the firming of longitudinal and lateral contacts. Even though PTX is one of the most powerful and useful drugs used nowadays, often there are cases of resistance towards this compound. Several reasons have been suggested to understand how this mechanism developed. Surely, mutations in the amino acid sequences in  $\beta$  tubulin isotypes can affect the way PTX binds to its luminal site, leading to an incorrect stabilization of MTs. In the following study, different mutations of  $\beta$  tubulin isotypes will be analyzed and how the latter could affect the lateral contacts between adjacent  $\alpha$  and  $\beta$  subunits. Through computational techniques, including molecular dynamics simulations and docking studies, it will be delineated how mutations in  $\beta$  tubulin isotypes can influence PTX's binding affinity to its target site, consequently affecting its ability to stabilize MTs. In conclusion, the aim of the project will be to better understand the mechanisms of ligand-receptor interaction, even in the presence of mutations, to make chemotherapy treatments increasingly precise and reliable.

## **Acknowledgement**

I would like to thank Professor Tuszynski for giving me the opportunity to carry out my thesis abroad, at the University of Alberta, allowing me to learn new concepts and improve my skills.

Thanks also to Maral Aminpour for welcoming me into her team and guiding me through the project.

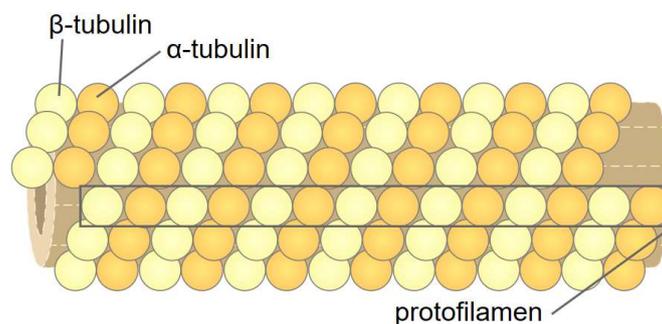
Finally, I would like to thank Paola Vottero for providing significant technical support for this project.

# TABLE OF CONTENTS

<b>1. BIOLOGICAL BACKGROUND</b> .....	5
1.1 <i>Microtubules and cancer</i> .....	7
1.2 <i><math>\beta</math>-isotypes</i> .....	7
1.3 <i>Tubulin binding sites</i> .....	8
1.4 <i>Paclitaxel binding site</i> .....	8
1.5 <i>Drug resistance</i> .....	9
<b>2. COMPUTATIONAL METHODS</b> .....	11
2.1 <i>Template and <math>\beta</math> isotypes choice</i> .....	13
2.2 <i>Mutations</i> .....	14
2.3 <i>Homology Model – MOE and Validation</i> .....	15
2.4 <i>Molecular Dynamics – AMBER</i> .....	17
2.4.1 <i>Simulating the environment</i> .....	18
2.4.2 <i>Energy minimization</i> .....	19
2.4.3 <i>Molecular dynamics</i> .....	20
2.4.4 <i>Molecular dynamic steps</i> .....	21
2.5 <i>Docking - VINA</i> .....	25
2.5.1 <i>How drugs work?</i> .....	25
2.5.2 <i>Docking programs</i> .....	27
<b>3. RESULTS</b> .....	29
3.1 <i>Homology Model &amp; Validation</i> .....	29
3.2 <i>MD &amp; clustering</i> .....	33
3.2.1 <i>MD files</i> .....	33
3.3 <i>Vina results</i> .....	36
3.3.1 <i>Vina files</i> .....	36
3.3.2 <i>Vina scores</i> .....	37
3.4 <i>Second MD &amp; new RMSD</i> .....	43
3.5 <i>Mmgsba</i> .....	44
3.5.1 <i>MMGBSA results</i> .....	45
<b>4. DISCUSSION</b> .....	49
4.1 <i>Homology Model &amp; Validation</i> .....	49
4.2 <i>MD &amp; Clustering</i> .....	49
4.3 <i>Vina scores</i> .....	50
4.4 <i>Second MD &amp; New RMSD</i> .....	51
4.5 <i>MMGBSA</i> .....	51
<b>5. CONCLUSIONS</b> .....	53
<b>6. ACKNOWLEDGMENTS</b> .....	54
<b>7. REFERENCES</b> .....	55

# 1. BIOLOGICAL BACKGROUND

Microtubules (MTs) are a dynamic component of cells, playing multiple roles within organisms, supporting critical structures such as centrosomes, axonemes, and cytoplasmic arrays. They are comprised of 13 protofilaments (PFs), constituted by  $\alpha$  and  $\beta$  tubulin, which rearrange to shape cylindrical proteins with an outer diameter of 25 nm as depicted in *figure 1.1*. Among their main functions, we mention the cell shape maintenance, cellular transport, and the formation of the mitotic spindle [1][2]. Tubulin  $\alpha$  and  $\beta$  are approximately 40% identical at the sequence level [3] and they are arranged in a specific head-to-tail sequence, where the  $\alpha$ -subunit of one dimer interacts with the  $\beta$ -subunit of the next, forming the MT lattice. The  $\beta$ -subunit is designated as positive, while the one housing  $\alpha$ -tubulin is termed negative [4]. There are two main types of MT lattices: a B-type, where the  $\alpha$  subunit forms lateral contacts with another  $\alpha$  subunit, while the  $\beta$  subunit contacts another  $\beta$ ; and an A-type, where the PFs are 'shifted', resulting in lateral contact between  $\alpha$  and  $\beta$  subunits. In mammalian cells, the B-type is prevalent. [5].

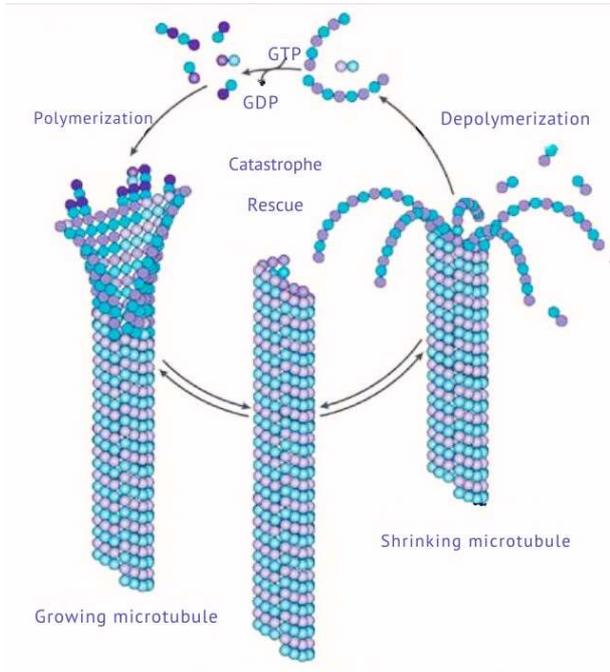


*Figure 1.1 MT structure, protofilaments  $\alpha$  and  $\beta$ .*

[https://commons.wikimedia.org/w/index.php?title=File:Microtubule\\_id.svg&oldid=796577661](https://commons.wikimedia.org/w/index.php?title=File:Microtubule_id.svg&oldid=796577661)

MT growth occurs through the formation of a reversible covalent bond between  $\alpha$  and  $\beta$  tubulin dimers. The  $\alpha$  subunit displays lower dynamicity in contrast to the  $\beta$  subunit, which experiences quicker growth and shrinkage [4]. Both  $\alpha$  and  $\beta$  ends contain a site for guanosine triphosphate (GTP) however, unlike GTP in the  $\alpha$ -subunit, which remains unaltered, the GTP in the  $\beta$ -subunit undergoes hydrolysis into guanosine diphosphate (GDP) [6]. The presence of GTP in the bond renders MTs more "stable," facilitating uninterrupted growth, whereas when the  $\beta$ -subunit is bound with GDP, they become more susceptible to depolymerization. MTs demonstrate two

forms of dynamism: treadmilling and dynamic instability. The latter entails a series of stochastic occurrences involving growth, degradation, and attenuation. This alternate mechanism involves the addition of tubulin dimers to the positive terminal ends and their release at the negative pole [7] *figure 2.1*. These distinct growth and degradation phases can be leveraged to impede tumor growth. Certain chemotherapy drugs, like Paclitaxel (PTX), operate by altering the properties of MTs.



*Figure 2.1: MTs dynamic process. [8]*

## *1.1 Microtubules and cancer*

In recent years, cancer has emerged as one of the leading causes of death globally. There are various strategies for cancer treatment, among which are those targeting MTs, as they are indispensable for mitosis and structural maintenance [9]. Agents acting on MTs are known as microtubule targeting agents (MTAs) or tubulin binding agents (TBAs), which can induce two distinct responses: stabilization or destabilization of MTs. Primarily of natural origin, MTAs bind to specific sites on tubulin. They can be categorized into two main groups: Microtubule stabilizing agents (MSAs) which stabilize lateral interactions between heterodimers, promoting polarization and stabilization of MTs. Microtubule destabilizing agents (MDAs) which, instead, stabilize longitudinal interactions between heterodimers [10]. Subsequent sections will elucidate the role of MTAs, particularly MSAs like Paclitaxel (PTX), the primary drug in this category, and how it can be optimally utilized in cancer treatment. Additionally, the role of tubulin isotypes will be explained, as they play a crucial role in drug resistance.

## *1.2 $\beta$ -isotypes*

In mammals there are various  $\beta$  tubulin isotypes that have homologous amino acids sequence and all of them have a crucial role at the cellular level. Every isotype affects MTs dynamics by influencing dimer-dimer interaction. Varying the expression/level of specific isotypes can thus lead to altered MTs dynamics [3]. Different studies have shown that not all the isotypes are equally distributed in the human body, but there are a few predominant in specific organs, (like  $\beta$ I and  $\beta$ IVB), while others are specific for different tissue (like  $\beta$ II,  $\beta$ II,  $\beta$ IVA) [11]. Numerous details about  $\beta$  tubulins can be found on this website: <https://www.proteinatlas.org/search/TUBB>. It provides an overview of the different isoforms, the genes they are encoded by, and the tissues where they can be more easily found.

### *1.3 Tubulin binding sites*

Tubulin possesses six binding sites, four located on  $\beta$ -tubulin (taxane, laulimalide/peloruside, vinca, maytansine). The remaining two, colchicine and pyronetin, are situated near the surface between  $\alpha$  and  $\beta$  dimers, and on the  $\alpha$  surface respectively [10]. The taxane site, initially identified in 1966, resides near the lateral surface of MTs and tends to stabilize them. Key drugs targeting this site include PTX, Docetaxel, and the Epothilones family [2]. PTX typically polarizes MTs by stabilizing  $\beta$ -tubulin, leading to the formation of 12 PFs compared to the standard 13 [12]. Unfortunately, it's not free from side effects such as high neurotoxicity, low water solubility, and frequent drug resistance. Docetaxel, a semi-synthetic agent akin to PTX, is employed in treating solid tumors such as breast and ovarian cancers. Multiple studies have underscored concerns with these agents; tubulin may undergo mutations induced by certain malignancies, potentially exacerbating drug resistance [13]. This study will solely concentrate on PTX since it is the principal drug binding to this site, predominantly employed across a wide array of cancer types.

### *1.4 Paclitaxel binding site*

For many decades scientists have been trying to understand how PTX binds to his site, because is not so clear, without any doubt, how this happens. Obviously, there are different theories about this topic and now, in the following section, I'll go to explain the most likely hypotheses that have been evaluated over the last years. PTX has a very fast kinetics of binding, so it's complicated making hypothesis [14]. The first hypothesis describes how PTX reaches his site through the hollow space within the MT, with an opening and closing mechanism. In this scenario, the proximity of the drug to the MT induces a conformational change with the opening of an access point which allows the drug to access the luminal site. However, this hypothesis has been discarded by Diaz et al [15], who used a fluorescent marker to look at how two derivatives of PTX, Flutax-1 and Flutax-2, reach their destination. In their work, if this hypothesis were true, they would have to see, in fluorescence, points either completely white or completely black, however it did not happen. The second hypothesis, which is the most probable, shows the possibility that PTX reaches its destination in two steps, passing first through an intermediate site near the H6/H7 loop and the serine 275 residue. This residue, in  $\beta$ III and  $\beta$ IV, is replaced by alanine. It is not perfectly clear how PTX binds to his site, but the substitution in this residue, in case of mutations, could somehow affect PTX kinetics and the

effect on MTs [14]. When PTX tries to stabilize MTs, it acts through the strengthening of the entire shape (longitudinally speaking) acting on the lateral interactions. When PTX bind to the luminal site the M-loop is forced to get outside of his normal position, resulting in its repositioning relative to the interprotofilament contact [14]. Probably, in presence of mutations, this mechanism is compromised and so the entire effect of the drug, resulting in a power reduction or the failure of the cure. Freedman et al [14] have shown that PTX may use the H6 and H7 loop as a hinge to reach its final destination in the luminal site, so mutations near these structures can lead resistance to the drug. Another hypothesis suggests that in MTs, there are other structures called nanopores through which it is possible for PTX to pass and reach the binding site. There are two types of pores: an interdimer nanopore (type 1) and an intradimer nanopore (type 2). Both are located in the type B and type A lattice, but it appears that in type B lattice interdimer nanopores are larger than intradimer nanopores. Type 1, however, appear to be closest to the PTX binding site [14]. This hypothesis was also described and simulated by Magnani et al [16], who focused on the interactions between PTX and H6/H7 loops and studied what type of nanopores is used by PTX to go through the MT wall. The problem with this hypothesis is that to have this very fast kinetics the drug should go through larger pores, which are not found in MTs. The explanation could be the rearrangement in the position of the H6 and H7 loops, so, thanks to that there is the formation of a suitable binding pocket (and temporarily) for PTX.

### *1.5 Drug resistance*

As already discussed, drug resistance is one of the main side effects which can manifest during the treatment. There are lots of possible causes for the rise of drug resistance, but nowadays the discussion is centered around the ones that involve the mutations, punctual and not, of the  $\beta$  isotypes [17]. In this work, only mutations for the  $\beta$  subunit will be treated, since the aim of the project is to look how these mutations could affect PTX behavior and, consequentially, the lateral interactions between dimers. The isotypes of tubulin have specific characteristics for each tissue and are distinguished by different sequences at the end of the carboxi-terminals. Tumors often display altered tubulin isotypes compared to healthy tissues. These mutated isotypes are often the cause of early rise for drug resistance with a related poorer chance of survival [18].

There are residues, in the amino acid sequence, which are most critical since they are near particular structures on MTs, like H6,H7 and M-loops. Therefore, mutations affecting these

residues could have a crucial role in the PTX dynamics, affecting the way in which it binds to site, resulting in a non-complete stabilization of MT.

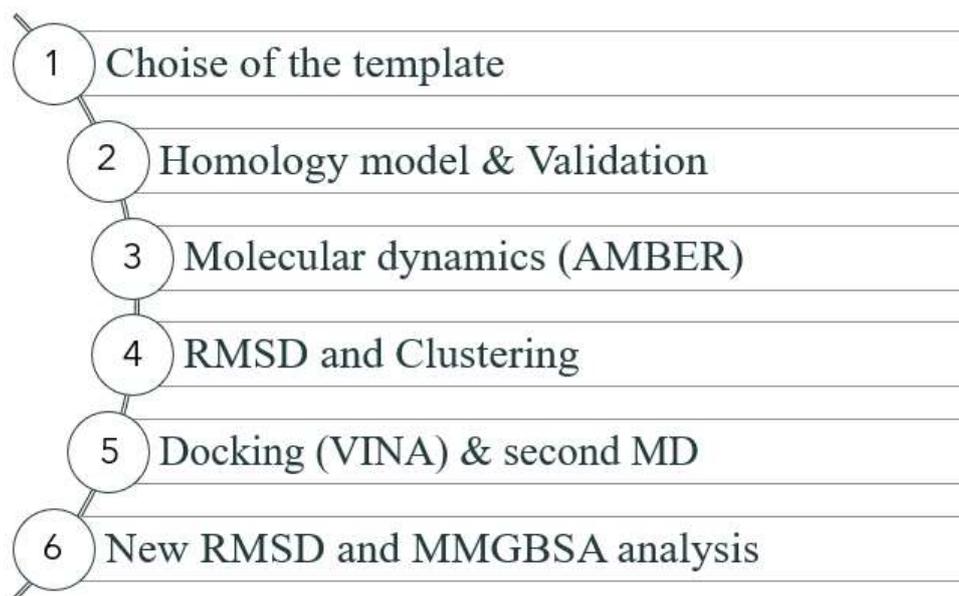
## 2.COMPUTATIONAL METHODS

In this and the following section will be explained the computational methods used for the analysis and evaluation of PTX at its binding site. Through an explanation of the software used for the preparation of the various models, the choices made for the optimization of the results, we will ultimately arrive at the evaluation of the binding energy of the drug with the MT, as well as an overall evaluation of the RMSD, that will allow us to understand the stability of the entire structure and the ligands contained therein.

Below I'll present a short list of the steps followed during all the phases of the project:

- 1) Choice of reference template, evaluation of the different options in the Protein data bank.
- 2) Choice of  $\beta$  tubulin isotypes (download of '.fasta' amino acid sequences from the Uniprot database).
- 3) Selection of mutations, single and not, to compare, in the end, the difference between mutated models and not mutated.
- 4) Construction of the model by homology model, using MOE software.
- 5) Validation of previous models, including mutated models.
- 6) Molecular dynamics analysis, using Amber software.
- 7) Docking of the previous structures, with Autodock-Vina software.
- 8) Second molecular dynamics, always with Amber.
- 9) Data analysis and discussion.

*Figure 3.2* illustrates the project steps in a schematic manner. All configuration parameters will be thoroughly discussed in the following chapters.



*Figure 3.2 Project steps.*

## 2.1 Template and $\beta$ isotypes choice

For the choice of the template, I evaluated the various possibilities in the Protein Data Bank (<https://www.rcsb.org/>). The selection criteria, which led me to the choice of the final model, were different:

- Since the project is based on lateral interactions, so how PTX can affect them, the first skimming was to find models with, possibly, at least two adjacent dimers.
- The next criterion concerns the characteristics of MT. As mentioned above, the most common model of MT in mammals is type B, therefore with tubulin coupled in this way:  $\beta$ - $\beta$  and  $\alpha$ - $\alpha$ .

A good resolution is around 2.5 Å or less. The resolution depends on the method by which the models were created, generally electron microscopy is quite accurate, but tends to have a worse resolution than X-rays or electron crystallography.

- A final selection criterion was the model's publication year. Obviously the more the model has a recent publication date the better it is, especially because in the last few years there have been many innovations in this field, so having the most up-to-date data possible allows us to have more accurate results.

Considering the criteria listed above, the choice for the model fell on the 6WVL model. The structure has two adjacent dimers of type B with 13 PTs with a slight curvature that tries to simulate the natural fold that MT has inside the cell [19]. The model has also docked four ligands that are respectively: TA1 (Taxol), GDP, GTP and  $Mg^{+2}$ , has a good resolution of 3.20 Å and a recent publication year, 2020. The origin of the dimers is from "BOS TAURUS" species. Since in the Protein Data Bank there are no tubulins of human origin, to have a good reliability of the results, it's convenient to use organisms that share a good degree of similarity with human amino acid chains. Generally, bulls, pigs or mice have a good degree of similarity; the differences arise mainly in the terminal zones of the chains, often eliminated precisely because they do not add relevant information.

Regarding  $\beta$  tubulin, for the creation of the models, the choice was based mainly on two factors. The first concerns the paper of Wang et al [20], where isotypes of  $\beta$  tubulin I, IIA, III, IVB are analyzed, with the related mutations. Only "these" isotypes have been taken as they are the most present in a large amount of tumor pathologies. Two drugs are used in the paper, PTX and docetaxel (DTX), which is a compound similar to the previous, newer, smaller in size and tends to have lower toxicity [21].

The selection of these isotypes is partly influenced by a previous factor. Among them,  $\beta$ I and  $\beta$ III stand out as particularly prevalent in various types of tumors. This aligns with the primary goal not only of this study but also of many others: to understand the mechanisms behind different cancers and develop effective treatments. Consequently, extensive research has focused on these isotypes due to their widespread presence, especially in cancers such as breast and ovarian cancer.  $\beta$ I, the most abundant isotype in the human body, experiences numerous mutations, many of which occur near the PTX binding site and crucial loops like H6, H7, and M, which play a pivotal role in microtubule stabilization.

## 2.2 Mutations

The selection of mutations was informed by a thorough review of pertinent literature and supplemented using the selector tool on MOE. This tool enabled an examination of residues located within or near the PTX binding site. As previously discussed, the primary criterion for selection revolved around identifying residues crucial for PTX binding, particularly those near loops H6, H7, and M. Mutations occurring in these regions have the potential to disrupt PTX binding, thereby compromising MT stabilization and diminishing the drug's efficacy, potentially leading to treatment failure. The focus was exclusively on  $\beta$ -tubulin since the PTX binding site resides within this isotype. Notably, certain isotypes exhibit a higher frequency of mutations compared to others; for instance,  $\beta$ I displays a propensity for numerous mutations, particularly at key binding site residues, consistent with its prevalence in various human tissues. Conversely,  $\beta$ III presents a distinct profile characterized by specific mutations, including the notable arginine substitution at position 275 in the amino acid chain [14]. Additionally,  $\beta$ III is associated with aggressive cancer phenotypes and recurrent drug resistance, attributed to mutations both within and external to the binding site. The mutations analyzed are shown in *table 1.2*.

<b>TUBB1</b>	<b>TUBB2A</b>	<b>TUBB3</b>	<b>TUBB4B</b>
<b>V60F<sup>2</sup></b>	<b>N48S<sup>3</sup></b>	<b>A231T<sup>1,2</sup></b>	<b>A124C<sup>3</sup></b>
<b>A185T<sup>4</sup></b>		<b>F270V<sup>1,2</sup></b>	<b>S126N<sup>3</sup></b>
<b>L215R<sup>1,2</sup></b>		<b>T274I<sup>1,2</sup></b>	<b>I155V<sup>3</sup></b>
<b>L217R<sup>1,2</sup></b>		<b>R282N<sup>1,2</sup></b>	<b>V189I<sup>3</sup></b>
<b>L228F<sup>1,2</sup></b>		<b>Q292E<sup>2</sup></b>	<b>T218A<sup>3</sup></b>
<b>S234G<sup>1,2</sup></b>		<b>S364A<sup>1,2</sup></b>	<b>C239S<sup>3</sup></b>
<b>A248V<sup>4</sup></b>			
<b>L273P<sup>1</sup></b>			
<b>R306C<sup>4</sup></b>			
<b>A364T<sup>1,2</sup></b>			

*Table 1.2:  $\beta$  tubulin isotype with related mutations. 1) Information gathered from the Tubulin Mutation Database the oldest one and the new created by Abbaali et al [22]. The other mutations are all from different papers 2) [23], 3) [20], 4) [24].*

### 2.3 Homology Model – MOE and Validation

Once the starting model, isotypes, and mutations were selected, the next step was to create models through the Homology model process. As we have seen before, in the protein data bank it is not possible to find secondary structures of proteins of human origin, but only of animals. This method allows us to recreate the secondary structure of a human protein using the human amino acid sequence, that is the primary structure known to us, and an animal protein that will act as a template. The only aspect to consider is the similarity between structures; a good similarity is around 40% and up, better values from 70% if you work with human proteins. In this way it will be possible to work with fairly accurate human models, despite the starting structure being of animal origin. The amino acid chain sequences, crucial for structural and functional protein analysis, were obtained by downloading data from the Uniprot database (<https://www.uniprot.org/>), which provides a wide range of biological and biochemical information on proteins. To build the various models the MOE software has been used. MOE (Molecular Operating Environment) is a molecular visualization software developed by Chemical Computing Group Inc (<https://www.chemcomp.com/Products.htm>), that can be used on a wide variety of platforms [25]. Thanks to a very intuitive interface, has allowed me to view the amino acid sequences of my structures, through the SEQ button, and to create my models using the homology model panel. The parameters set in the homology model panel are as follow:

- Gradient limit: 200, since a MD will be performed later, so it is not necessary to it set up now.
- Temperature at 300 like an in vitro test.
- Sidechain samples: 5.
- Initial protonation to fix some charges.
- Addition of the N and C terminal. They were considered because they might be relevant to see side interactions.

Once the models were generated, validation became imperative to ensure the reliability of the structures. Validation can be conducted through various methods. One such method is the Ramachandran plot, which visualizes whether amino acid residues reside within "allowed" regions. Since each atom experiences specific steric hindrance, it can only adopt certain conformations with precise angles. Occasionally, following homology modeling, some residues may fall into disallowed regions. While this is not inherently problematic, provided the number of such residues remains low, a truly dependable model should ideally exhibit a minimal percentage of residues in disallowed areas. Additionally, there are online servers like Q-Means or SAVEsv.06, which offer comprehensive model evaluations based on multiple criteria. As a first analysis, the Ramachandran plot on MOE has been visualized in the geometry section, for a first evaluation. Subsequently, to have a more "effective" validation, the SAVE SV.06 server was used. Before loading the file as *'pdb'*, an initial energetic minimization was undertaken. This process aimed to identify atoms with potentially incorrect positions, as visualized beforehand. Consequently, there exists the possibility of overlap between atoms, which could prove problematic during MD simulations. Given that each atom possesses its own steric hindrance, such overlap may lead to errors or even simulation failure. Moving forward with geometric analysis, where various graphs are displayed, any atom clashes are highlighted. These clashes are then addressed by selecting the respective atoms and reversing their positions within the amino acid chain (using the SEQ button and the invert function). Subsequently, under the constraint tab in MOE, the *fix* option is selected. With the structure now stabilized, energy minimization can proceed by adjusting the gradient to 0.01 and ensuring atoms are checked. Following this optimization, the model can be validated using the server.

## 2.4 Molecular Dynamics – AMBER

Once the validation is completed, we move to the study of MD. Before delving into the steps followed during the project, it is necessary to briefly introduce what MD means and the underlying physics. MD is a branch of molecular mechanics (MM), which can be used to study small molecules or large biological assemblies. MM, or the "forcefield method," is based on Newtonian mechanics, which, through mass-spring systems (almost never dampened), studies the kinematics and dynamics for modeling molecular systems [26].

The particles, components of these systems, are point masses described by the following parameters:

- The geometry of the system.
- Atomic type.
- A set of atoms can be considered as a single particle.

The particles within the considered space interact with each other through primarily bonding interactions but also non-bonding interactions. What we can calculate is the potential energy associated with the various interactions present in the system and can be defined as follows, as highlighted in *formula 1.2*:

$$V = V_{bond} + V_{non-bond} \quad (1.2)$$

V is a function of 3N variables, considering the x, y, and z directions and both bonding and non-bonding interactions [27].

Bonding interactions [27]:

1. **Bonds:** Interactions between two atoms bonded by a covalent bond. In MM, it is a harmonic interaction, dependent on the bond length.
2. **Angles:** These terms consider the relative motion between atoms describing an angle on a plane.
3. **Dihedrals:** Identifies the rotation of one plane relative to another involving 4 bonded atoms. In the "AMBER" force field, dihedrals are represented by complex equations.

4. **Improper Dihedrals:** While proper dihedrals allow rotations up to  $360^\circ$ , improper dihedrals add a term for potential energy calculation. Specific to atoms bonded in a cycle.

Non-bonding interactions [27]:

1. **Van der Waals:** Forces that arise between different charges, even with particles without a net charge, are short-range forces that 'deplete' after 1nm or just beyond. They are further divided into very short-range and long-range interactions (up to 1.5 nm) or London dispersion forces. Van der Waals forces are well described by Lennard-Jones potential.
2. **Coulomb:** Part of electrostatic interactions and computationally more complex to represent.
3. **Hydrogen Bonds:** Described by the Lennard-Jones potential; for AMBER, they use a specific function.

All the characteristics listed above are typical of force fields. Depending on the simulations to be performed, there will be force fields specific to the biological system under consideration. Examples include AMBER, GAFF (Generalized Amber Force Field), and GROMOS, where the main differences lie in the parameters that constitute the basic equations.

#### 2.4.1 Simulating the environment

When discussing simulations, reference is always made to more complex systems of particle-particle interactions, with the goal of representing more intricate biological systems such as proteins, lipids, carbohydrates, etc., present in aqueous environments. Water is crucial for simulating molecules, but at the same time, it poses certain challenges [27]. Certainly, an aqueous environment can cause screening phenomena but can also promote electrostatic interactions. Among the main problems associated with water molecules, there are certainly characteristics that make their description difficult through classical mechanics. Additionally, simulating certain environments requires many molecules, increasing computational costs. There are two modeling approaches:

1. **Explicit Solvent:** Water molecules are explicitly represented. This is undoubtedly a more accurate method but computationally expensive. Among these models are TIP3P (Jorgensen) or SPC (Single Point Charge, Berendsen). It is an excellent representation

model but comes with some limitations such as the rigidity of molecules and high computational costs.

2. **Implicit Solvent:** In these models, the solvent can be represented as a medium with a different dielectric constant. Some models consider the so-called solvation energy on the protein surface. Implicit solvent models are still widely discussed in the literature, as they are not as accurate. However, simulations are often based on models using explicit solvent. Among these models are those based on Solvent Accessible Surface Area (SASA) and those based on continuous-level electrostatic interactions.

#### 2.4.2 Energy minimization

The goal of MM during a simulation is to map the Potential Energy Surface (PES). PES, or landscape, consists of various points representing local and global minima, which in the biological context identify points of metabolic activity. By calculating all local minima of the PES, access is gained to all microstates of the system, providing insight into the behavior of the biological system [28]. For the calculation of local minima, energy minimization algorithms are used, which can be:

1. **Derivative method:** Algorithms based on Steepest Descent and Conjugate Gradient, first order derivative methods. Through successive minimizations, the algorithm tends towards progressively lower energies. Second-order derivative methods like Newton-Raphson and L-BFGS involve the inversion of the Hessian function.
2. **Non derivative method:** Among these, the SIMPLEX method is based on constructing a geometric figure with  $N + 1$  vertices connected to each other, where  $N$  is the dimensionality of the considered potential energy function. The potential energy is specific to a particular set of coordinates. A disadvantage of this method is that it cannot provide information about the slope.

For both algorithms, identifying the local minimum of the potential energy function is challenging, especially when using first order derivative and non-derivative methods. Typically, the output of simulations provides the value of the nearest local minimum to the starting point. Energy minimization does not study the macrostates composing a system but is a preliminary step to investigate its dynamic behavior [26].

Since the primarily used method is explicit solvent, the choice of energy minimization algorithms tends to lean towards first-order methods, particularly the steepest descent. In any

case, what an energy minimization algorithm accomplishes is merely the search for a local minimum without delving into the characteristics of that minimum. Energy minimization algorithms, therefore, cannot macroscopically describe the system, a limitation that can be bypassed with the study of MD [26].

### 2.4.3 Molecular dynamics

The dynamics of a particle system is a computational approach that calculates the average properties of a system by sampling its microstates over time in a specific statistical ensemble. It can be studied using different methods, such as the Lagrangian and Hamiltonian methods, which involve the use of coordinates ( $q_1, q_2, \dots, q_N$ ) and generalized momenta ( $p_1, p_2, \dots, p_N$ ). MD is a deterministic method, as the state of the system in a future configuration is entirely determined by its present state. As an output result, there will be a trajectory, which is very useful for system analysis. The number of possible trajectories is high, but with a sufficiently long simulation, all these trajectories will overlap [27]. The goal of MD is the resolution of Newton's equations of motion in a system of atoms interacting through a known potential energy function.

Starting from the second Newton equation

$$F_i = m_i \cdot a_i \quad (2.2)$$

The force can be expressed as the potential energy gradient

$$F_i = -\nabla_i V \quad (3.2)$$

Combining the previous equations

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (4.2)$$

If the acceleration is constant, the equation will be developed, obtaining.

$$r = at^2 + v_0 + r_0 \quad (5.2)$$

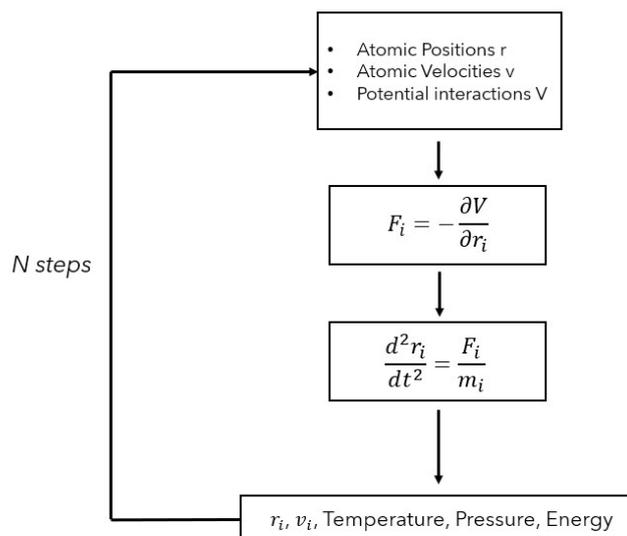
$$a = -\frac{1}{m} \frac{dV}{dr} \quad (6.2)$$

What can be obtained from this process are:

- The potential energy, from which force and acceleration can be derived.
- The position.

#### 2.4.4 Molecular dynamic steps

MD steps can be listed as show in *figure 4.2*:



*Figure 4.2 Molecular Dynamic steps.*

Here are the steps [28] with a brief description:

1. **Initial Configuration:** As mentioned in the previous paragraphs, when dealing with proteins, the files considered are PDB (Protein Data Bank) type, collected in a corresponding database. Initial parameters are extracted from X-ray experiments or crystallography. In these files, information about atom positions and spatial localization relative to an arbitrary reference point can be found. The file is in ASCII format, describing atoms and their positions. PDB format files have a header containing bibliographic information and list cartesian coordinates of atoms in angstroms (Å) along with connectivity data. Below is an excerpt from a *‘.pdb’* file. An example is reported in *figure 5.2*.

```

HETATM 1 C 1 -1.576 -0.433 -0.004
HETATM 2 C 2 -1.301 -0.777 -0.643
HETATM 3 C 3 -0.072 -1.413 -0.444
HETATM 4 C 4 -0.898 -0.848 -0.394
.
.
CONNECT 1 2 6 13
CONNECT 2 1 3 14
.
.
END

```

*Figure 5.2 PDB file excerpt.*

The "HETATM" label signifies that these atoms do not belong to the protein (but, for example, to a ligand). The second column shows the atom numbering, the third column identifies the atoms (in this case, they are carbon atoms). The fourth column provides a numbering that indicates the atom's membership in a monomer, and the last three columns report the spatial coordinates.

The "CONNECT" label indicates which atoms are in contact. For example, the first line indicates that the atom is bonded to atoms 2, 6, and 13. This part of the PDB file describes the topology of the structure.

2. **Energy Minimization:** Given the initial coordinates and the force field parameters (information not traceable in the PDB file), it is possible to proceed with the calculation of the potential energy function.
3. **Initial velocity:** After completing the minimization step, a new PDB file will be obtained containing the updated, more homogeneous, and representative coordinates of the system. The next step will be to update the initial velocity values, which can be set to zero or assigned according to a Maxwell-Boltzmann distribution at the desired temperature.
4. **Heating:** The next step will be the heating phase, only if the velocities are initially set to zero. For proper sampling of all microstates in the system, it is necessary to simulate the presence of a thermal bath: by combining Newton's equation with an equation for modifying velocities, the velocities will be rescaled based on the temperature of the external thermal bath. Since temperature is closely related to kinetics, it's necessary to control velocity values to achieve a system based on a certain temperature, being cautious not to excessively impose certain values that could introduce biases.

These initial phases of minimization and heating are delicate because they involve working with atom velocities and positions. For example, a minimal displacement could cause instability and high forces among atoms, leading to simulation failure. To address this issue, restraints are introduced, potential energy constraints that anchor atoms as springs, adding a penalty to their movement. As the simulation progresses, the system tends to stabilize, allowing these constraints to be gradually removed.

5. **Balance conditions:** The simulation will continue to progress until it is completely dissociated from the initial conditions. Once completed, various analyses of the obtained results can be performed.
6. **Production phase:** In this phase, the thermodynamic properties of the system will be calculated.
7. **Data analysis:** Once the simulation is completed, various properties of the system can be calculated. Below are listed some of the calculable parameters:

1. Average energy

$$\langle E \rangle = \frac{1}{N} \sum_{i=1}^N E^i \quad (7.2)$$

2. RMS

$$RMS = \langle (r_i^\alpha - r_i^\beta)^2 \rangle^{\frac{1}{2}} = \sqrt{\left( \frac{1}{N_i} \sum_i (r_i^\alpha - r_i^\beta)^2 \right)} \quad (8.2)$$

The value of RMS, or more precisely RMSD (Root Mean Squared Deviation), calculates the fluctuations of atoms at the end of the simulation. Generally, the RMSD value rises rapidly in the initial nanoseconds of the simulation, then stabilizes, reaching a plateau indicative of convergence. A good RMSD value hovers around 4-5 Å, indicating small oscillations, while values in the order of nanometers indicate high oscillations. The equilibrium observed with this parameter is a structural equilibrium, meaning that at equilibrium, the structures will not be identical but will all hover around the same equilibrium conformation, with minimal fluctuations.

3. RMS of fluctuation

$$RMS_i^{fluct} = \sqrt{\left( \frac{1}{N_f} \sum_f (r_i^f - r_i^{ave})^2 \right)} \quad (9.2)$$

This type of RMS always calculates fluctuations not on individual atoms but on frames, which represent the discretization of time in a simulation. It measures the deviations from a hypothetical average structure over the course of the simulation.

#### 4. Gyration radius

$$RadiusGyration = \sqrt{\frac{1}{N_i} \sum_i (r_i - r_{cm})^2} \quad (10.2)$$

The gyration radius assesses, over time, the variation in the position of the atoms within the structure relative to its center of mass.

In conclusion, MD simulations provide a good overall representation of the structure. However, it is important to acknowledge their limitations, including high computational costs and the shape of the energy landscape. There are methods that aim to overcome these limitations, based on Enhanced Sampling techniques, which allow to overcome energetic barriers. These models have not been covered in the present work.

## 2.5 Docking - VINA

Once the MD part is completed and repeated twice, the next step will be the docking. In this case, it is necessary to briefly introduce what is meant by ‘docking’ and why it is necessary to perform it. Briefly recapping the steps of the project completed so far: an initial preparation of the models was carried out using MOE, followed by subsequent MD using AMBER. From the simulation output files will be obtained, which will be partially analyzed by assessing the RMSD (all results will be presented in *Chapter 3*) and subjected to a clustering process to obtain frames, which will be our new starting models. The PTX comes into play, which, through the docking process, will be docked to our various frames using the Autodock-Vina program on the Linux terminal. Before proceeding with the actual docking, it is necessary to investigate the PTX binding site, as Vina will use a box to explore the area where the drug binds. To proceed with this step, the analysis of the spatial coordinates of the binding site will be performed using the Pymol software, which, through loading the models and selecting the residues involved in the PTX binding site, will provide the coordinates of the site itself for the specific model considered. Once the coordinates are known, the necessary files will be created to perform docking on Vina. Below, a brief explanation of what docking will be provided, while the description of the selected parameters will be discussed in detail in the next chapter.

### 2.5.1 How drugs work?

When attempting to dock a drug to its target, certain aspects need to be considered:

- Binding site location.
- Binding mode.
- Binding energy: indicating stronger or weaker binding energies.

The process of how a drug binds to its target can be described as the lock-and-key mechanism, meaning there is a specific position and timing for the process to occur. Regarding ligand-protein interactions, bonds can be either covalent or non-covalent [26]. Docking is a virtual screening approach used to predict and rank the ways in which a ligand binds to its receptor. Generally, the receptor is a protein, and docking is employed to classify the most probable binding mode of the ligand. Docking programs evaluate compound activity by analyzing ligand-receptor interactions and estimating binding affinity. The main challenge is determining the most probable binding configuration to form a stable protein-ligand complex. The docking process is divided into two phases: exploring the conformational space of the ligand and

selecting the optimal target-ligand alignment. Since docking is probabilistic, the probability of events occurring in nature is calculated, considering protein flexibility, multiple binding sites and factors, like pH [26].

Molecular docking is highly useful in the following areas:

1. Prediction of intermolecular biological interactions: Molecular docking is valuable for predicting the formation and dissociation of specific interactions between various molecules, playing a crucial role in the functions of biological systems.
2. Determination of molecular interactions: While methods like X-ray crystallography and nuclear magnetic resonance (NMR) are expensive and unsuitable for large-scale studies, molecular docking provides a rapid, cost-effective computational method suitable for large-scale studies (virtual screening) to predict protein-molecule interactions.

The docking process has two main goals: predicting the geometry of the interaction (binding poses) using a search algorithm and estimating binding energy. The definition of docking states that it seeks to find the energetically most feasible 3D arrangement of two molecules in close contact to predict their binding energy.

What is required as input to initiate the docking process includes:

- 3D structure for the protein and ligand
- Information about the binding site

What will be obtained as output includes:

- Ligand poses
- Binding affinity

Docking can be described as solving a 3D "puzzle," which involves generating all possible combinations of ligands and proteins. The characteristics involved in the final evaluation of docking algorithms include the quality of fit, such as shape complementarity, Coulomb potential for charge complementarity, and Van der Waals potential for hydrophobic surface complementarity.

## 2.5.2 Docking programs

Different molecular docking programs vary based on the following parameters:

1. **Protein Representation:** Programs differ in how they represent the protein, whether it's explicit (with 3D coordinates of all atoms) or based on a grid or molecular surface.
2. **Search Method:** The methods for searching vary, including rotations, translations, and internal degrees of freedom.
3. **Scoring Function:** Differences exist in the functions used to evaluate the quality of the alignment between the protein and ligand.
4. **Protein Flexibility:** Considerations for protein flexibility, recognizing that proteins are not rigid. Different programs handle protein flexibility differently, using approaches like MD models, soft docking, and inducing conformational adaptations during docking. The mentioned programs, including AUTODOCK, GOLD, DOCK and FlexX, each have specific approaches to representing and managing the flexibility of the molecules involved in docking [26].

When using these programs, it's essential to pay attention to several aspects:

1. **Metal Ions:** Metals like zinc or iron can form complexes with specific amino acids and water molecules. The coordination often involves amino acids like cysteine and histidine, forming structures known as "zinc fingers".
2. **Protein Flexibility:** Protein flexibility is managed in a complex way, with flexible regions called "hinge regions" allowing relative movements between protein domains. Some docking programs incorporate modest flexibility in the target during docking simulations.
3. **Ligand Flexibility:** While ligands are often treated as flexible in docking programs, there are sampling techniques falling into three main categories: systematic search routines, stochastic exploration, and simulation techniques.
4. **Protonation State:** The pKa is used to estimate whether a group is protonated or deprotonated. The protein environment can significantly influence the protonation state of ionizable groups in the ligand.
5. **Molecular Interactions:** Interactions such as hydrogen bonding, ionic interactions, hydrophobic interactions, and the importance of water in protein-ligand interactions play a crucial role in representing the environment.

Hydrophobic interactions, for instance, result from the release of water molecules from the hydrophobic environment during complex formation.

6. **Enthalpic and Entropic Contribution:** Enthalpic and entropic optimization aims to enhance the hydrophobic surface of a ligand buried during binding, considering the physical interactions (enthalpy) and changes in degrees of freedom (entropy).

Understanding these aspects is crucial for effectively utilizing docking programs in predicting and analyzing protein-ligand interactions.

### 3. RESULTS

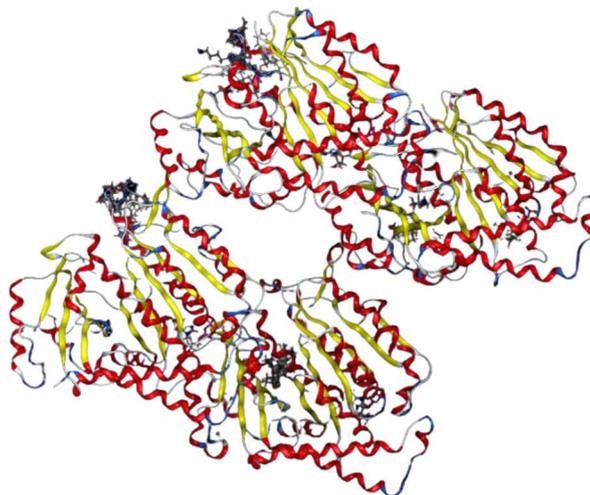
In the following section, the results obtained at each step will be analyzed in more detail, also examining specifically the parameters used in the various files to obtain them.

#### *3.1 Homology Model & Validation*

The first step involved searching for the template to be used as a reference. To proceed with the homology modeling, some modifications were made to the template to optimize it, using the structures preparation tools on MOE, following the suggestion of *Paola Vottero's* thesis.

- Alternates: The positions of amino acids with alternative variations were adjusted, selecting the ones with the highest occupancy.
- Termini: Missing atoms in the backbone of the protein chain at the C- or N-termini were removed, and the ends were capped.
- Breaks: Gaps found within the protein chain were filled by constructing a loop segment.
- Library: Discrepancies between the amino acid names and their structures, or missing atoms, were corrected using sequence information relative to the three-dimensional structure.

*Figure 6.3* shows the template used for this project, after all the corrections.



*Figure 6.3* Template created from 6WVL model. The model shows the  $\alpha$  and  $\beta$  tubulin subunits. B type MT.

What is shown in figure above, the template is for the 2-dimers (2ds) model. To compare the final results more accurately, the project was also conducted with only one of the two dimers (1d). The model downloaded from the Protein Data Bank already contained two PTXs within the crystallography; the chosen dimer to be retained was the one with the internally docked PTX.

Homology modeling is a powerful technique that allows us to predict the three-dimensional structure of a protein based on its similarity to structurally known proteins. Protein models were created using this technique, striving to capture the structure and function of the target proteins as accurately as possible. Once the models were obtained, it was crucial to assess their reliability. The results of homology modeling are presented in *table 2.3* summarizing key metrics, such as the percentage of identity between the target protein and its closest homologs. It is possible to say that homology values above 90% are indicative of structures very similar to those found in crystallography. Values between 75% and 90% can be considered excellent models, while values below 75% can still be regarded as good models, but only with some refinement operations [28].

Table 2 Homology model panel.

<b>ISOTYPE-MUTATIONS</b>	<b>HOMOLOGY ALFA (%)</b>	<b>HOMOLOGY BETA (%)</b>
<b>TUBB1</b>	96.90	76.30
V60F	96.90	76.10
A185T	96.90	76.10
L215R	96.90	76.10
L217R	96.90	76.10
L273P	96.90	76.10
L228F	96.90	76.10
R306C	96.90	76.10
S234G	96.90	76.10
A248V	96.90	76.10
A364T	96.90	76.30
<b>TUBB2A</b>	96.90	95.70
N48S	96.90	95.30
<b>TUBB3</b>	96.90	88.40
A231T	96.90	88.20
F270V	96.90	88.20
Q292E	96.90	88.20
R282N	96.90	88.20
S364A	96.90	88.20
T274I	96.90	88.20
<b>TUBB4B</b>	96.90	93.50
A124C	96.90	93.50
C239S	96.90	93.30
I155V	96.90	93.30
S126N	96.90	93.30
T218A	96.90	93.30
V189I	96.90	93.30

To ensure the robustness and accuracy of the models, I subjected them to validation using the SAVESv6.0 server. SAVESv6.0 is a cutting-edge tool that meticulously examines the models, assessing their structural integrity and identifying any issues or discrepancies. The results of

this validation were compiled into *table 3.3* reporting the percentage of ERRAT [29] and RAMACHANDRAN plot outcomes, crucial metrics for evaluating the quality of the models.

*Table 3 Errat e Ramachandran results.*

<b>ISOTYPE-MUTATIONS</b>	<b>ERRAT (%)</b>	<b>RAMACHANDRAN (%)</b>
<b>TUBB1</b>	93.95	91.00
V60F	93.07	90.70
A185T	93.26	90.90
L215R	94.23	90.09
L217R	93.18	91.20
L273P	93.38	91.20
L228F	93.01	90.07
R306C	93.45	91.20
S234G	93.28	91.30
A248V	93.47	90.60
A364T	93.01	90.50
<b>TUBB2A</b>	93.88	91.30
N48S	93.26	91.90
<b>TUBB3</b>	92.80	91.90
A231T	91.67	91.20
F270V	92.91	91.60
Q292E	92.48	91.30
R282N	91.88	91.00
S364A	92.96	90.50
T274I	92.96	91.80
<b>TUBB4B</b>	94.00	91.00
A124C	93.92	91.70
C239S	93.73	91.70
I155V	93.56	91.50
S126N	93.02	92.60
T218A	93.43	92.30
V189I	93.51	91.70

As you can see from the table above, the values of both parameters are good, meaning the reliability of the models. Ramachandran plot displays all possible values of  $\phi$  and  $\psi$  for each amino acid residue in a protein. The permitted areas in the plot are defined by constraints closely associated with the stability of the protein structure, such as steric hindrances and non-covalent interactions. Thus, having values above 90% allows for asserting a good reliability of the models.

### 3.2 MD & clustering

The next step will focus on the application of MD using Amber software for all models created with the homology model. During this phase, MD simulations will be conducted to explore the dynamic behavior of protein structures in a simulated environment. Understanding how proteins interact and move over time will be crucial, providing valuable insights into their dynamic properties and structural stability. During the presentation of the MD results, I will provide a detailed description of the files used to conduct the simulations. The results obtained will be useful to evaluate the RMSD (Root Mean Square Deviation) to assess whether the simulations reach equilibrium. This parameter permits us to evaluate how much the protein structures fluctuate over time and whether the simulations are reliable.

#### 3.2.1 MD files

In the following subsection, some of the parameters contained in the files used for MD will be briefly described. Firstly, the *‘.pdb’* files obtained with MOE had to be properly prepared for AMBER. These *‘.pdb’* files were loaded after removing the ligands from the models, as the creation of the *‘.lib’*, *‘.mol2’*, and *‘.frcmod’* parameters in the various steps was simplified. Subsequently, the *“pdb4amber”* command was executed with the addition of the *“--nohyd”* option, which removes the hydrogen atoms from the original *‘.pdb’* file. This conversion allows for a more manageable file for tLeap [30]. The conversion with the above command allows AMBER to properly rename some of the residues that would otherwise not be readable in the pdb format [31]. Among these are those listed in *table 2.3* below.

Table 4 Residues accepted by Amber software.

Residue	PDB	AMBER
Histidine	HIS	HIE
Cysteine	CYS	CYX
Aspartate	ASP	ASH
Glutamate	GLU	GLH
Lysine	LYS	LYN

After the conversion, the necessary files were created with tLeap to represent the ligands and, finally, the files for the topology and trajectory, '*prmtop*' and '*inpcrd*'. The procedure included placing the compounds in a cubic box of 10.0 Å and then solvating with water. This step is essential to establish simulation conditions in a realistic environment, where the presence of water more accurately reproduces the behavior of biomolecules in solution. The next step involved creating files for the dynamics, starting from energy minimization of the solute only, followed by another minimization of the entire structure. All parameters describing the system to be minimized were defined in these two files, paying particular attention to restraints, the number of cycles, and the type of minimization. Subsequently, the system was heated providing the starting temperature. Next, relaxation in the NPT ensemble was performed, followed by the actual production of the dynamics. The duration of the simulations was set to 100 ns, with a *nstlim* of 50000000, corresponding to the number of MD steps to be evaluated. The *dt* was set to 0.002 ps, which is the time interval between individual steps of the dynamics. Once the simulations were completed, the RMSD calculation was carried out using *cpptraj* [32] and a *MATLAB* code. This step is crucial as it allows us to assess whether the simulations reach a sort of equilibrium by the end of the 100 ns. The typical trend of RMSD involves high initial oscillations followed by stabilization from approximately halfway through the simulation onwards [33]. RMSD was calculated for both the binding site and the receptor to identify any significant differences. Two examples of RMSD are shown in *figure 9.3*.

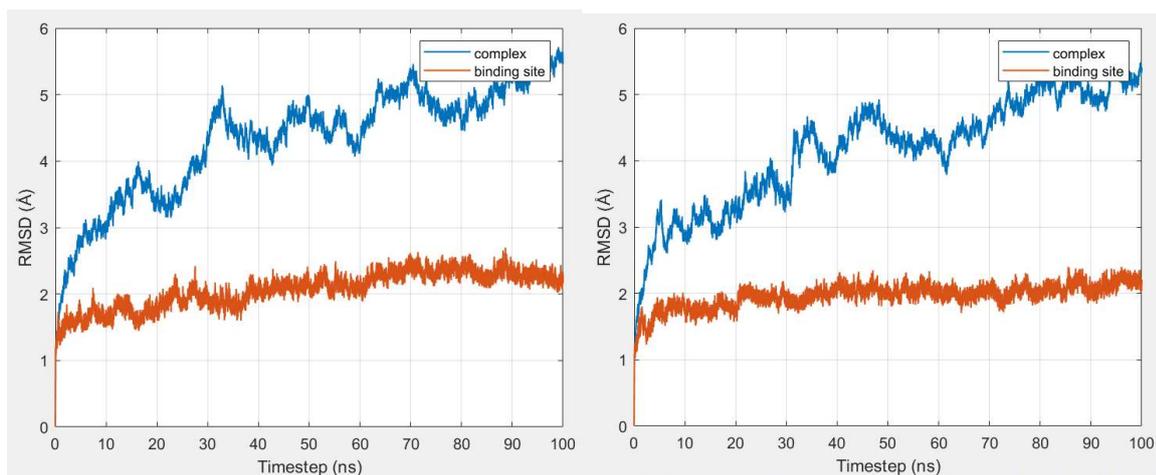


Figure 7.3 TUBB1 and TUBB4B RMSD, 100ns simulation.

Alongside the various MD simulations, dynamics were also performed on four randomly chosen models where PTX was not removed from the starting *.pdb* files. This additional step was executed because there was uncertainty about the stability of the models after MD, as PTX could potentially introduce greater instability. Figure 10.3 shows only two of the four models chosen for this test, TUBB1 and TUBB4B. It is evident that the presence of PTX significantly increases receptor instability, and the binding site exhibits greater oscillations. Therefore, removing the drug before conducting the dynamics was the better solution.

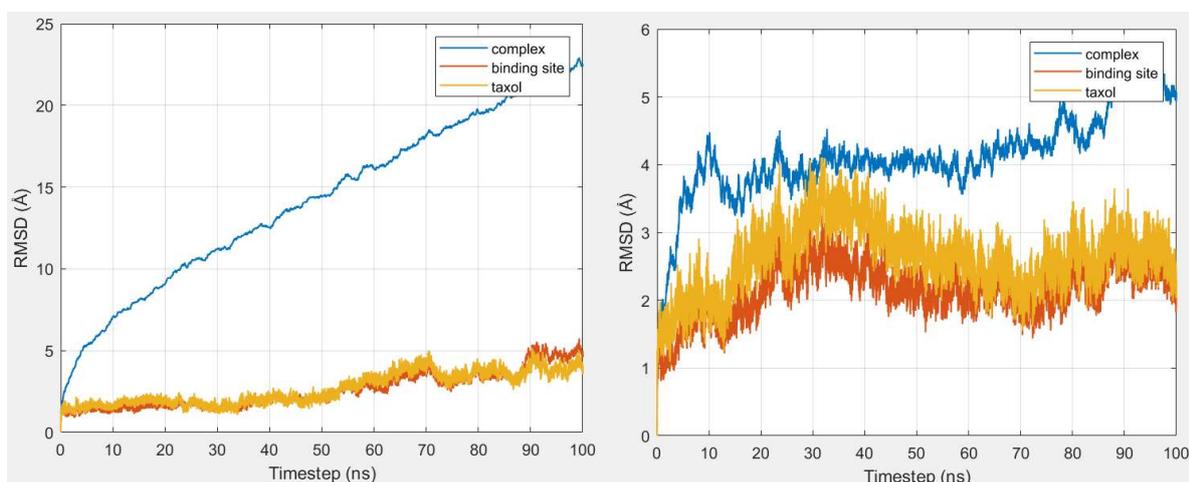


Figure 8.3 TUBB1 and TUBB4B RMSD with PTX 100ns simulation.

The next steps will involve proceeding with a clustering process of the entire structure obtained from the MD simulations. The clustering process resulted in a number of frames, each

representing specific configurations, with the initial ones deemed the most representative of the models.

### 3.3 *Vina results*

#### 3.3.1 Vina files

Once the clustering process was completed, it was necessary to select the top 3 frames, representing the most frequent structures within the dataset. This selection is crucial as it allows us to focus on the most representative and significant conformations for further analysis. Initially, we proceeded with the conversion of the *‘.pdb’* files into *‘.pdbqt’* format, as this is the format accepted by the Vina software used for molecular docking [34]. To ensure proper file conversion, codes provided by the Autodock package were used: *receptor.py* and *ligand.py*. This operation was carried out to ensure that our models were compatible with the docking software and could be correctly evaluated. Subsequently, the definition of a docking box was required, within which the Vina software would examine the best binding positions between the ligand (PTX) and the protein receptor [34][35]. To obtain the specific coordinates of this box, the *Pymol* software was used. By visualizing the amino acid chain and selecting the residues constituting the PTX binding site, we were able to determine the exact coordinates for each model. The coordinates obtained were then integrated into the input file for the Vina software. Additionally, we selected appropriate parameters for docking, including exhaustiveness (which indicates the probability of finding the global maximum), the maximum number of binding modes to generate, and the maximum energy difference between the best and worst binding modes. All information previously described can be found at this link: <https://vina.scripps.edu/manual/>. The results produced by Vina provided a database of docking configurations, representing the best interactions between the drug and the receptor. These outputs were subsequently converted into *‘.mol2’* format using OpenBabel, enabling detailed visualization of molecular interactions within the MOE software. Finally, the first docking configuration was selected as the most significant and representative. These models will undergo further MD simulations to assess the stability of molecular interactions in a dynamic context.

### 3.3.2 Vina scores

The docking results have been represented in two ways:

1. The first one is an average of the most significant docking scores obtained for three different clusters (c0, c1, c2), compared to the reference value of the wild type (also averaged across the three clusters). I'll call these models 'averaged models' (AM).
2. In the second case, the 3 clusters weren't averaged between them but the most negative values, over the 3 clusters, have been chosen. I'll call these models 'not averaged model' (NAM).

Each isoform will be analyzed to identify any mutations of interest. The results were conducted for both single dimer (1d) models and for those with two dimers (2ds). The mutations discussed in this report are mutations of residues located in the PTX binding site or near it, therefore they should decrease the drug binding affinity.

For example:

- **L215R**, and **S234G** for TUBB1
- **R282N**, **F270V** and **T274I** for TUBB3
- **T218A** and **C239S** for TUBB4B

Those mutations are very close to PTX binding site. Many of the treated mutations were sourced from various papers, as indicated in *table 1.2*, which claimed that these mutations lead to PTX resistance. Forthcoming lines will undergo analysis for isotype mutations, both for 1d and 2ds models.

## 1 DIMER MODELS

**TUBB1:** Not all mutations in **TUBB1** show lower affinity towards PTX. The tendency between the models is similar, some mutations have the same behavior in the two cases, as **A364T**, **L217R**, **L215R** and **A185T**, that show lower values compared to the wild type, how should it be. Other mutations like **L273P**, **A248V**, **S234G** and **V60F** show significantly higher values compared to the **TUBB1**. The only differences are in mutations **R306C** and **L228F**, that have higher affinity for the drug, in the NAM. Results are shown in *table 5.3*.

Table 5.3 TUBB1 and related mutations, docking results.

<b>TUBB1 1d</b>	<b>Averaged (kcal/mol)</b>	<b>Not averaged (kcal/mol)</b>
<b>Wild type</b>	<b>-7.47</b>	<b>-7.94</b>
A364T	-6.89	-7.61
R306C	-7.44	-8.14
L273P	-7.78	-9.00
A248V	-7.97	-8.90
S234G	-8.63	-8.89
L228F	-7.47	-8.16
L217R	-7.28	-7.87
L215R	-7.23	-7.38
V60F	-8.79	-9.68
A185T	-7.16	-7.74

**TUBB2A:** In this case, the only mutation discussed, **N48S**, shows higher affinity in both cases compared to the wild type, as if the mutation improves the binding affinity. Results are shown in *table 6.3*.

Table 6.3 TUBB2A and related mutation, docking results.

<b>TUBB2A 1d</b>	<b>Averaged (kcal/mol)</b>	<b>Not averaged (kcal/mol)</b>
<b>Wild type</b>	<b>-8.20</b>	<b>-8.34</b>
N48S	-8.32	-9.05

**TUBB3:** Like **TUBB1**, mutations in **TUBB3** do not always exhibit lower affinity towards PTX, as theoretically expected. However, this discrepancy occurs only in the NAM, where only mutation **F270V** shows an increase in binding affinity, while all other mutations have lower values compared to the wild type. Results are shown in *table 7.3*.

*Table 7.3 TUBB3 and related mutations, docking results.*

<b>TUBB3 1d</b>	<b>Averaged (kcal/mol)</b>	<b>Not averaged (kcal/mol)</b>
<b>Wild type</b>	<b>-8.53</b>	<b>-9.38</b>
S364A	-7.26	-7.70
Q292E	-7.91	-9.32
R282N	-7.64	-9.07
T274I	-7.72	-7.89
F270V	-8.33	-10.00
A231T	-8.36	-9.05

**TUBB4B:** In AM only mutations **T218A** and **S126N** have docking score values higher than the reference. Instead, in NAM, mutations with higher values than the wild type are: **T218A**, **V189I**, **S126N**. Only mutation **V189I** shows an increase of the docking score, for the NAM. Results are shown in *table 8.3*.

Table 8.3 TUBB4B and related mutations, docking results.

TUBB4B 1d	Averaged (kcal/mol)	Not averaged (kcal/mol)
Wild type	-8.01	-8.32
C239S	-7.54	-7.77
T218A	-8.27	-10.22
V189I	-7.93	-8.77
I155V	-7.83	-8.05
S126N	-8.27	-8.59
A124C	-7.52	-7.80

## 2 DIMERS MODELS

**TUBB1:** The tendency is similar for both models, the only differences are in few mutations: **L273P, S234G** which have higher docking scores in the NAM. All the other mutations have a similar behavior, mutations that show higher values in the AM have kept them in the NAM. The same thing happens for mutations with lower docking score values. Results are shown in *table 9.3*.

Table 9.3 TUBB1 and related mutations, docking results.

<b>TUBB1 2ds</b>	<b>Averaged (kcal/mol)</b>	<b>Not averaged (kcal/mol)</b>
<b>Wild type</b>	<b>-7.81</b>	<b>-8.20</b>
A364T	-7.95	-8.31
R306C	-6.69	-7.47
L273P	-7.50	-8.66
A248V	-7.75	-7.99
S234G	-7.53	-8.25
L228F	-6.70	-7.20
L217R	-7.25	-8.01
L215R	-8.19	-8.84
V60F	-7.03	-7.32
A185T	-8.64	-8.85

**TUBB2A:** As for the 1d models, both mutated models show higher affinity for PTX. Results are shown in *table 10.3*.

Table 10.3 TUBB2A and related mutation, docking results.

<b>TUBB2A 2ds</b>	<b>Averaged (kcal/mol)</b>	<b>Not averaged (kcal/mol)</b>
<b>Wild type</b>	<b>-7.66</b>	<b>-8.35</b>
N48S	-8.03	-8.48

**TUBB3:** The tendency is similar, the only difference is in a single mutation: **Q292E** which has higher affinity in the NAM, instead in the AM the behavior is opposite. All the other mutations have a similar trend, those that show higher values in the AM have kept them in the NAM. Results are shown in *table 11.3*.

*Table 11.3 TUBB3 and related mutations, docking results.*

<b>TUBB3 2ds</b>	<b>Averaged (kcal/mol)</b>	<b>Not averaged (kcal/mol)</b>
<b>Wild type</b>	<b>-7.76</b>	<b>-8.09</b>
S364A	-8.16	-9.21
Q292E	-7.71	-8.41
R282N	-8.17	-8.32
T274I	-8.02	-8.49
F270V	-7.95	-8.82
A231T	-7.65	-7.98

**TUBB4B:** The tendency is the same for both cases. All mutations show better affinity compared to the wild type, controversial results since these mutations are known to lead PTX resistance. Results are shown in *table 10.3*.

Table 12.3 TUBB4B and related mutations, docking results.

TUBB4B 2ds	Averaged (kcal/mol)	Not averaged (kcal/mol)
Wild type	-7.81	-7.89
C239S	-8.79	-9.15
T218A	-8.50	-9.01
V189I	-7.86	-8.31
I155V	-8.76	-9.40
S126N	-7.94	-8.05
A124C	-8.27	-8.92

The upcoming phase of the project involves initiating a new MD simulation, followed by evaluating the RMSD and computing MMGBSA energy. To advance, it was imperative to select the most negative docking scores among the three calculated for each model.

### 3.4 Second MD & new RMSD

After obtaining the models with Vina, we conducted a second series of MD simulations. This time, in addition to the previously included ligands, we also incorporated PTX, the main compound of interest. Consequently, we prepared the necessary files for PTX as well, ensuring a comprehensive representation of molecular interactions within the system. Unlike the first MD, the simulations were shortened to 20 ns each but were executed a total of 5 times for each model. This approach allowed us to gather a greater amount of useful data for subsequent analysis. The reduction of the total simulation duration to 20 ns led to a decrease in the MD steps, from 50 million to 10 million, while maintaining the same time interval between individual steps (dt). All other parameters remained unchanged to ensure consistency and comparability of results between the two simulation phases. Following the acquisition of the new MD results, we re-evaluated the RMSD, this time also including the behavior of the ligands: GDP, GTP, and PTX. *Figure 11.3* shows two of many RMSD graphs (TUBB1 and TUBB4B, 2ds model) obtained after the second MD.

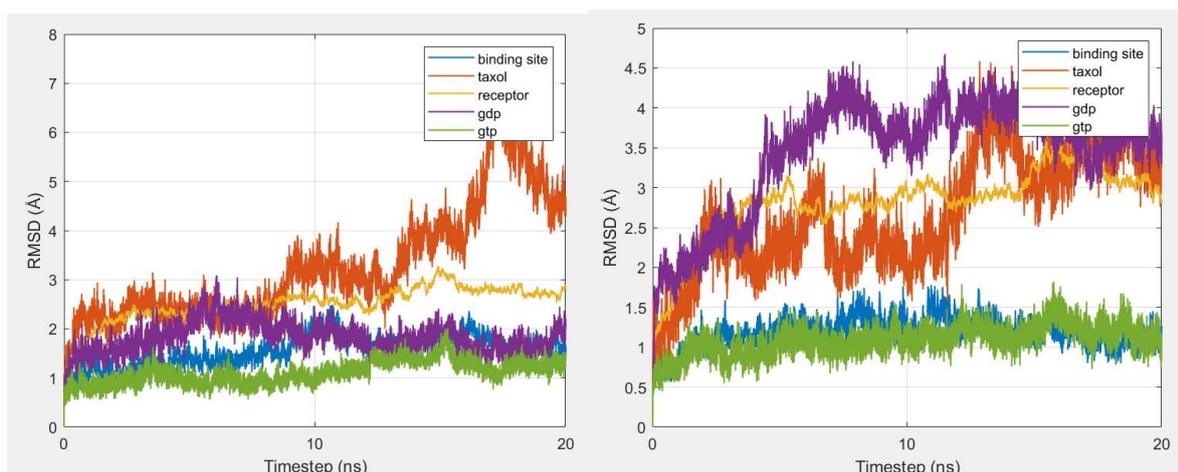


Figure 9.3 TUBB1 and TUBB4B first of 5 simulations performed, 20 ns simulation.

Oscillations for the TUBB1 and TUBB4B 2ds models are relatively contained, even with shorter duration simulations. However, it's noteworthy to observe higher oscillations in the ligands. This phenomenon may suggest greater dynamics and flexibility of the ligands compared to the proteins themselves. It's important to emphasize that, despite the reduced duration of the simulations, there is an increase in ligand oscillations towards the midpoint of the simulation. This could be attributed to various factors, including the complexity of ligand-receptor interactions and the surrounding environment, which could influence the dynamic behavior of the ligands over time. These results highlight the importance of considering not only the dynamics of the target proteins but also those of the ligands during MD simulations.

### 3.5 *Mmgsba*

Once MD simulations are performed to obtain the conformations of protein-ligand complexes, the MMGBSA (Molecular Mechanics Generalized Born Surface Area) method was applied to calculate the binding free energy. This estimated value provides an indication of the relative affinity of the ligand for the target protein. Overall, MMGBSA is a computational method that combines MM with an implicit solvent model to estimate the binding affinities between ligand molecules and target proteins. This approach is useful for drug design and understanding molecular interactions that play a critical role in biological processes [36]. To proceed with the energy calculation, the files generated from the second MD simulation, along with the associated trajectory, were used. In the input files used in this step, several key pieces of information were specified. Firstly, the salt concentration was set to zero to examine molecular interactions in the absence of salt ions. Additionally, the number of trajectory frames to be used

was defined based on the RMSD of each model, indicating the start and end points of the MD simulation to be considered. These parameters allowed focus on a specific portion of the simulation where molecular interactions are most relevant for the study. Following the energy calculation, an output file containing various important information was obtained. Specifically, binding affinities of the ligand, receptor, and entire molecular complex were analyzed. These data provide a detailed overview of molecular interactions and their strengths, enabling assessment of interaction effectiveness and better understanding of the nature of the studied molecular complex. The numerical values chosen to represent the results were taken from the '.dat' file, resulting from the calculation of Mmgbsa: Differences (Complex - Receptor - Ligand), Delta total. In conclusion, analysis of the output files obtained from this process provides valuable information on energy and molecular interactions within the studied system. These data are essential for deepening understanding of molecular structure and dynamics and can provide important insights for the development of new drugs and treatments. The MMGBSA calculation was performed on the 1d and 2ds models. The final results for each model are an average of the five different results obtained in the energy calculation, and they will be presented in the following subsection.

### 3.5.1 MMGBSA results

**TUBB1:** As shown in *table 11.3*, not all mutations exhibit lower MMGBSA energy compared to the wild-type model. Almost all mutations in the 1d models show significantly higher values, such as: **V60F**, **S234G**, **R306C**, **L273P**, **L217R** and **A364T**, as if they have a better binding affinity. In the 2ds models mutations with the same trend are: **S234G**, **L273P**, **L215R**, **A364T**, and **A185T**.

Table 13.3 TUBB1 and related mutations, MMGBSA results.

<b>TUBB1</b>	<b>MMgbsa 1d (kcal/mol)</b>	<b>MMgbsa 2ds (kcal/mol)</b>
<b>Wild type</b>	<b>-36.7092</b>	<b>-29.7222</b>
A364T	-39.8426	-41.9793
R306C	-48.3122	-21.4393
L273P	-38.9059	-36.5821
A248V	-32.1825	-21.7169
S234G	-42.6708	-41.9569
L228F	-36.3547	-20.0549
L217R	-38.3010	-23.9187
L215R	-21.5064	-31.6757
V60F	-56.1779	-28.2175
A185T	-30.1965	-54.1109

**TUBB2A:** The only mutation considered shows a different behavior in the two cases, as shown in *table 12.3*. In the 1d model the mutation **N48S** has a higher value of binding energy, while the 2ds model shows a significantly lower energy compared to the wild type.

Table 14.3 TUBB2A and related mutation, MMGBSA results.

<b>TUBB2A</b>	<b>MMgbsa 1d (kcal/mol)</b>	<b>MMgbsa 2ds (kcal/mol)</b>
<b>Wild type</b>	<b>-23.9073</b>	<b>-44.7004</b>
N48S	-42.1740	-41.1737

**TUBB3:** Also in this case, as shown in table 13.3, the behavior between the two models presents few differences. While in the 1d models all mutations have lower binding energy compared to the not mutated model, a different scenario is shown in the 2ds models results in which mutations like: **T274I**, **F270V**, and **A231T**, have higher binding energy, almost as if they have a better binding affinity.

*Table 15.3 TUBB3 and related mutation, MMGBSA results.*

<b>TUBB3</b>	<b>MMgsa 1d (kcal/mol)</b>	<b>MMgsa 2ds (kcal/mol)</b>
<b>Wild type</b>	<b>-45.5240</b>	<b>-38.2490</b>
S364A	-42.1645	-30.1397
Q292E	-30.0652	-28.5373
R282N	-36.6339	-24.4691
T274I	-23.2257	-38.6300
F270V	-44.2962	-42.2019
A231T	-36.0630	-44.7884

**TUBB4B:** In the last isotype it's possible to say that, for 1d models, only mutations **T218A** and **S126N** show higher values of MMGBSA energy. For the 2ds models the same behavior is presented by mutations **S126N**, **C239S** and **A124C**, almost as if they have a better binding affinity.

*Table 16.3 TUBB4B and related mutation, MMGBSA results.*

<b>TUBB4B</b>	<b>MMgbsa 1d (kcal/mol)</b>	<b>MMgbsa 2ds (kcal/mol)</b>
<b>Wild type</b>	<b>-42.4325</b>	<b>-28.7648</b>
C239S	-31.0816	-31.4651
T218A	-48.7129	-26.1912
V189I	-34.1629	-22.4256
I155V	-29.4149	-25.8606
S126N	-41.8914	-31.6701
A124C	-25.9534	-41.5492

In conclusion what can also be observed, as for the docking scores, is that not all mutations exhibit lower binding energies compared to their respective non-mutated models, almost as if the mutations improve the binding affinity. Moreover, the results vary for nearly all mutations between the 1d and 2ds models.

## 4.DISCUSSION

The previous chapter provides all the results obtained during the project. In this chapter, each step will be revisited, offering an in-depth analysis and discussion of the outcomes.

### *4.1 Homology Model & Validation*

The homology modeling process, essential for predicting the three-dimensional structure of proteins, involved meticulous adjustments to optimize the template structure. Modifications included addressing alternate amino acid positions, resolving missing backbone atoms, and filling gaps within the protein chain. These adjustments, validated through MOE tools and referencing to MOE manual, ensured the structural integrity of the models. Further validation through the SAVESv6.0 server provided critical insights into model reliability, evidenced by favorable ERRAT and RAMACHANDRAN plot outcomes. As discussed earlier, ERRAT and Ramachandran plots are two crucial metrics used to assess the quality of protein structures. For both metrics, values above 90% indicate good structural quality and a low presence of errors or distortions in the protein conformation [29], [30]. From the data presented in *table 3.3 on pages 33-34*, it emerges that the models used in the research obtained ERRAT and Ramachandran values above 90%. This suggests that the generated protein structures are reliable and accurately representative of the three-dimensional conformation of the target proteins. Consequently, we can conclude that the analyzed models have good reliability and can be confidently used for research purposes.

### *4.2 MD & Clustering*

Following homology modeling, MD simulations using Amber software provided dynamic insights into protein behavior. These simulations, essential for understanding protein interactions over time, utilize meticulously prepared MD files. Parameters such as system preparation, energy minimization, and dynamics initiation were carefully defined to ensure accurate representation and reliable results. RMSD analysis revealed equilibrium attainment and provided valuable information on structural fluctuations. All obtained RMSD plots exhibited consistent behavior, aligning well with theoretical expectations. Overall, the trend across all models adhered to the theoretical trajectory: there was an initial rapid increase followed by stabilization of fluctuations around the midway point or shortly thereafter in the simulations. The initial plots solely depict the RMSD of the entire receptor and the binding site, excluding the GDP and GTP ligands. This was done to initially assess receptor stability, to

ascertain whether, following an initial MD run, the models could attain sufficient stability to proceed with subsequent steps.

### 4.3 *Vina scores*

The subsequent application of Vina software for molecular docking facilitated the exploration of protein-ligand interactions. By converting and preparing input files, defining docking parameters, and analyzing docking configurations, Vina provided valuable insights into ligand binding affinities. The selection of top frames and conversion of output files enabled further analysis and visualization of molecular interactions, laying the groundwork for subsequent investigations. The results obtained from these analyses provide docking energy values, expressed in kcal/mol, which represent the energy associated with the interactions between the molecules involved. It is important to contextualize these docking energy values to fully understand their biological relevance. To do so, we can connect these values to the concept of thermal energy, which represents thermal agitation of molecules at room temperature. The thermal energy at room temperature is approximately 0.59 kcal/mol, <https://gehrcke.de/2014/05/thermal-energy-in-calories/#:~:text=First%20of%20all%2C%20the%20thermal,values%20on%20a%20regular%20basis>. This value provides us with a reference point to evaluate the importance of interactions predicted by docking. If the docking values are significantly higher than the thermal energy, this may indicate that the detected interactions are much stronger than the thermal agitation of molecules at room temperature. This is important because it suggests that such interactions could be biologically relevant and significantly influence the behavior of molecules in a biological context. If the docking values are similar to or higher than thermal energy, it may indicate that the predicted interactions are strong enough to overcome thermal agitation and therefore may be significant in a biological context. Therefore, by connecting the docking values obtained with Vina to the concept of molecular interaction energy and thermal energy, we can gain a better understanding of the importance and biological relevance of predicted molecular interactions. One last aspect to pay attention is to highlight that the docking calculations were conducted using only one docking program. There are several docking programs available, each associated with an accuracy percentage. For Vina, this percentage is very high, around 80%, as highlighted by Astalakshmi et al. [37] To conduct a more reliable analysis, it would have been advisable to resort to other docking programs, even just MOE or Autodock, in order to obtain more results to compare with each other. The discriminating factor

could have been the RMSD, by setting a threshold, many results could have been retained or discarded.

#### *4.4 Second MD & New RMSD*

The second phase of MD simulations, incorporating PTX and shortened durations, offered additional insights into protein dynamics and ligand interactions. By assessing RMSD and comparing results with the initial simulations, we observed consistent trends in protein stability and ligand dynamics. Notably, ligands exhibited higher oscillations, underscoring their dynamic nature and potential implications for biological activity. As with the first RMSD, it can be asserted that simulations tend toward stability after approximately halfway through, despite the total time being reduced from 100 ns to 20 ns. The only observable difference lies in the oscillations, which are slightly higher compared to before, likely due to the shortened duration of the simulations. Furthermore, RMSD was calculated for all ligands: GDP, GTP, and PTX, in order to observe the final state of the models in a comprehensive manner.

#### *4.5 MMGBSA*

MMGBSA analysis provided critical insights into protein-ligand binding energies, essential for understanding molecular interactions and potential drug design applications. By evaluating binding free energies and comparing them across mutated and non-mutated models, we gained valuable insights into the impact of mutations on protein-ligand interactions. The obtained results certainly deviate from what would have been expected theoretically. Anomalous behaviors are evident for many of the mutations, with some showing significantly higher binding affinities compared to their unmutated counterparts, almost as if the mutations were enhancing the drug-target binding. The MMGBSA energy values are averaged from five values calculated for each model, aiming to obtain a value that better approximates reality. Despite this adjustment, the results do not fully align with what is expected from theory and require further investigation and a deeper understanding of the models. There could be various interpretations of the results obtained at different steps. Generally, it is expected that docking and MMGBSA scores would be lower for mutated models compared to their wild type counterparts. Some mutations indeed follow this trend, while others exhibit opposite behavior. Moreover, mutations showing this peculiarity are not always the same, but vary between single dimer and double dimer models, indicating inconsistency in the results. Certainly, one reason for this behavior lies in the preparation of the models themselves. Mutations were selected by cross-referencing

data from articles cited in *table 1.2* of mutations with those from the tubulin mutation database. Models have a single mutation in their amino acid sequence, except for the N48S mutation of TUBB2A, taken from the article by Wang et al. [20]. This could influence the final results; it is possible that in nature, when these mutations occur, there are other associated mutations that have not been added to the models. Another factor to consider is the template. The starting model includes two adjacent dimers, providing a basis for discussing lateral contacts and potential effects that PTX brings once bound to its binding site. The treated template is not fully representative but is still an approximation of the MT wall; therefore, having a larger model would certainly be more faithful to reality, albeit it would entail higher computational costs. The primary goal of this thesis was to evaluate whether  $\beta$ -tubulin mutations could influence PTX behavior, especially in lateral contacts, as the aim of this drug is to stabilize the entire length of MTs. Further analysis could assess hydrogen bonds within the binding site to see how formation and breaking of such bonds vary between mutated and non-mutated models, assessing if this contribution is significant. More analyses are needed to better understand how PTX regulates MT dynamics, and certainly a more accurate representation of the models can be a starting point.

In conclusion, the comprehensive analysis presented in this chapter offers valuable insights into protein structure, dynamics, and interactions. Through homology modeling, MD simulations, molecular docking, and MMGBSA analysis, we gained a deeper understanding of protein-ligand interactions and potential implications for biological activity. Further investigations based on these findings could contribute to drug design efforts and advance our understanding of molecular mechanisms underlying protein function.

## 5.CONCLUSIONS

MTs are exceptional cellular components that, owing to their properties, enable cells to carry out various activities, including cell division. When tumor forms develop, cells undergo a process of ecosystem alteration leading to uncontrolled growth. For several years now, especially in recent decades, new therapies have targeted microtubules to slow the spread of these malignant cells. Through agents that modify the dynamics of these small cellular components, scientific research has made significant progress, improving expectations and the quality of life for countless individuals. Unfortunately, despite the significant strides made, there are still some aspects partly unknown or to be further investigated to make chemotherapy more effective. When a tumor disease develops, affecting the first cell, cellular mutations often lead to drug resistance. Tubulins, especially  $\beta$ -tubulin, are often subject to mutations that result in various side effects, including treatment failure. Research becomes crucial to understand how to overcome the challenges posed by this possibility and to create increasingly specific and accessible treatments for all. This was the goal of this project: to provide a deeper understanding of one of the most famous chemotherapeutic drugs, PTX. Through the creation of protein models, a validation process, and a study of MD, I was able to delve into aspects related to various  $\beta$ -tubulin mutations and how they interact in the presence of this potent drug. Although the data obtained did not provide exhaustive information and, in some cases, differed from theoretical expectations, they can still serve as a starting point for future analyses. By further exploring aspects related to mutations and utilizing more advanced computational equipment, it will be possible to obtain more significant data.

## **6.ACKNOWLEDGMENTS**

We have finally reached the conclusion of this journey. A huge chapter of my life has just ended, characterized by mood swings, various difficulties, and many satisfactions. It's hard to thank everyone who has actively participated in keeping my hope alive during this crazy journey. The first thanks go to my family, a huge and noisy gang that has supported me throughout these years, encouraged me to believe in myself, and never stop. A continuous support, without judgment, just a constant presence that has shaped the person I am today. To my father, who with jokes and pearls of wisdom (of any kind) is guiding me towards my future. To my mother, who with her empathy and courage has taught me what is truly important in life. To my noisy sister Miriam, simply called Mi or "capretta", who with her constant and "exhausting" affection has always supported me, lifting my spirits even on rainy days. To my grandparents and my uncles, who have always had blind trust in me, even in the darkest moments. To my university mates and all my friends, a wild squad with whom I had the honor and pleasure of sharing the years of my university career and my entire life, thanks for your constant presence and your relentless support. To Cinzia, a soul sister, a constant source of inspiration, who taught me to never give up. Finally, I would like to thank my boyfriend, my partner, Stefano. Thank you for encouraging and supporting me all this time. I am not an easy person, but with your constant presence, you managed to bring me calm even in the most complicated moments. I promise to all of you to continue making you proud and as I always say, "To infinity and beyond," towards a new adventure.

## 7. REFERENCES

- [1] J. Zhou and P. Giannakakou, "Targeting Microtubules for Cancer Chemotherapy," 2005.
- [2] E. Mukhtar, V. M. Adhami, and H. Mukhtar, "Targeting microtubules by natural agents for cancer therapy," *Molecular Cancer Therapeutics*, vol. 13, no. 2. pp. 275–284, Feb. 2014. doi: 10.1158/1535-7163.MCT-13-0791.
- [3] J. T. Huzil, R. F. Ludueña, and J. Tuszynski, "Comparative modelling of human  $\beta$  tubulin isotypes and implications for drug binding," *Nanotechnology*, vol. 17, no. 4, Feb. 2006, doi: 10.1088/0957-4484/17/4/014.
- [4] E. A. Perez, "Microtubule inhibitors: Differentiating tubulin-inhibiting agents based on mechanisms of action, clinical activity, and resistance," *Molecular Cancer Therapeutics*, vol. 8, no. 8. pp. 2086–2095, Aug. 01, 2009. doi: 10.1158/1535-7163.MCT-09-0366.
- [5] M. Katsuki, D. R. Drummond, and R. A. Cross, "Ectopic A-lattice seams destabilize microtubules," *Nat Commun*, vol. 5, Jan. 2014, doi: 10.1038/ncomms4094.
- [6] G. M. Alushin, G. C. Lander, E. H. Kellogg, R. Zhang, D. Baker, and E. Nogales, "High-Resolution microtubule structures reveal the structural transitions in  $\alpha\beta$ -tubulin upon GTP hydrolysis," *Cell*, vol. 157, no. 5, pp. 1117–1129, May 2014, doi: 10.1016/j.cell.2014.03.053.
- [7] P. M. Checchi, J. H. Nettles, J. Zhou, J. P. Snyder, and H. C. Joshi, "Microtubule-interacting drugs for cancer treatment," *Trends in Pharmacological Sciences*, vol. 24, no. 7. Elsevier Ltd, pp. 361–365, Jul. 01, 2003. doi: 10.1016/S0165-6147(03)00161-5.
- [8] M. O. Steinmetz and A. E. Prota, "Microtubule-Targeting Agents: Strategies To Hijack the Cytoskeleton," *Trends in Cell Biology*, vol. 28, no. 10. Elsevier Ltd, pp. 776–792, Oct. 01, 2018. doi: 10.1016/j.tcb.2018.05.001.
- [9] D. A. Fletcher and R. D. Mullins, "Cell mechanics and cytoskeleton," *Nature*, vol. 463, no. 7280, pp. 485–492, 2010, doi: 10.1038/nature08908.Cell.
- [10] F. Borys, E. Joachimiak, H. Krawczyk, and H. Fabczak, "Intrinsic and Extrinsic Factors Affecting Microtubule Dynamics in Normal and Cancer Cells," *Molecules*, vol. 25, no. 16. MDPI AG, Aug. 01, 2020. doi: 10.3390/molecules25163705.
- [11] L. J. Leandro-García *et al.*, "Tumoral and tissue-specific expression of the major human  $\beta$ -tubulin isotypes," *Cytoskeleton*, vol. 67, no. 4, pp. 214–223, 2010, doi: 10.1002/cm.20436.
- [12] A. L. Risinger, F. J. Giles, and S. L. Mooberry, "Microtubule dynamics as a target in oncology," *Cancer Treatment Reviews*, vol. 35, no. 3. pp. 255–261, May 2009. doi: 10.1016/j.ctrv.2008.11.001.
- [13] M. Mariani *et al.*, "Clas- III  $\beta$ -tubulin in normal and cancer tissues," *Gene*, vol. 563, no. 2. Elsevier B.V., pp. 109–114, Jun. 01, 2015. doi: 10.1016/j.gene.2015.03.061.
- [14] H. Freedman, J. T. Huzil, T. Luchko, R. F. Ludueña, and J. A. Tuszynski, "Identification and characterization of an intermediate taxol binding site within microtubule nanopores and a

- mechanism for tubulin isotype binding selectivity," *J Chem Inf Model*, vol. 49, no. 2, pp. 424–436, Feb. 2009, doi: 10.1021/ci8003336.
- [15] J. F. Díaz, I. Barasoain, and J. M. Andreu, "Fast kinetics of Taxol binding to microtubules: Effects of solution variables and microtubule-associated proteins," *Journal of Biological Chemistry*, vol. 278, no. 10, pp. 8407–8419, Mar. 2003, doi: 10.1074/jbc.M211163200.
- [16] M. Magnani, G. MacCari, J. M. Andreu, J. F. Díaz, and M. Botta, "Possible binding site for paclitaxel at microtubule pores," *FEBS Journal*, vol. 276, no. 10, pp. 2701–2712, May 2009, doi: 10.1111/j.1742-4658.2009.06994.x.
- [17] H. K. Berrieman, M. J. Lind, and L. Cawkwell, "Do  $\beta$ -tubulin mutations have a role in resistance to chemotherapy?," *Lancet Oncology*, vol. 5, no. 3, pp. 158–164, Mar. 01, 2004. doi: 10.1016/S1470-2045(04)01411-1.
- [18] A. L. Parker, W. S. Teo, J. A. McCarroll, and M. Kavallaris, "An emerging role for tubulin isotypes in modulating cancer biology and chemotherapy resistance," *International Journal of Molecular Sciences*, vol. 18, no. 7, MDPI AG, Jul. 04, 2017. doi: 10.3390/ijms18071434.
- [19] G. E. Debs, M. Cha, X. Liu, A. R. Huehn, and C. V Sindelar, "Dynamic and asymmetric fluctuations in the microtubule wall captured by high-resolution cryoelectron microscopy", doi: 10.1073/pnas.2001546117/-/DCSupplemental.
- [20] W. Wang *et al.*, "Novel mutations involving  $\beta$ I-,  $\beta$ IIA-, or  $\beta$ IVB-tubulin isotypes with functional resemblance to  $\beta$ III-tubulin in breast cancer," *Protoplasma*, vol. 254, no. 3, pp. 1163–1173, May 2017, doi: 10.1007/s00709-016-1060-1.
- [21] W. Krause, "Resistance to anti-tubulin agents: From vinca alkaloids to epothilones," *Cancer Drug Resistance*, vol. 2, no. 1. OAE Publishing Inc., pp. 82–106, 2019. doi: 10.20517/cdr.2019.06.
- [22] I. Abbaali *et al.*, "The tubulin database: Linking mutations, modifications, ligands and local interactions," *PLoS One*, vol. 18, no. 12, p. e0295279, Dec. 2023, doi: 10.1371/journal.pone.0295279.
- [23] J. T. Huzil, K. Chen, L. Kurgan, and J. A. Tuszynski, "The Roles of  $\beta$ -Tubulin Mutations and Isotype Expression in Acquired Drug Resistance," 2007.
- [24] S. Yin, R. Bhattacharya, and F. Cabral, "Human mutations that confer paclitaxel resistance," *Mol Cancer Ther*, vol. 9, no. 2, pp. 327–335, 2010, doi: 10.1158/1535-7163.MCT-09-0674.
- [25] U. Roy and L. A. Luck, "Articles Molecular Modeling of Estrogen Receptor Using Molecular Operating Environment", doi: 10.1002/bambd.65.
- [26] Andrew R. Leach, "Molecular modelling," 2001.
- [27] Daan Frenkel and Berend Smit, "Understanding molecular simulation \_ from algorithms to applications," 2002.
- [28] D. C. Young, *Computational chemistry : a practical guide for applying techniques to real world problems*. Wiley, 2001.
- [29] Chris Colovos and Todd O. Yeates, "Verification of protein structures," 1992.

- [30] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *J Appl Crystallogr*, vol. 26, no. 2, pp. 283–291, Apr. 1993, doi: 10.1107/s0021889892009944.
- [31] D. A. Case Ross C Walker and A. M. Roitberg Kenneth Merz Pengfei Li, "Amber 2020 Reference Manual Principal contributors to the current codes." [Online]. Available: <http://ambermd.org/contributors.html>
- [32] D. R. Roe and T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data," *J Chem Theory Comput*, vol. 9, no. 7, pp. 3084–3095, Jul. 2013, doi: 10.1021/ct400341p.
- [33] D. A. Case Ross C Walker and A. M. Roitberg Kenneth Merz Pengfei Li, "Amber 2020 Reference Manual Principal contributors to the current codes." [Online]. Available: <http://ambermd.org/contributors.html>
- [34] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J Comput Chem*, vol. 31, no. 2, pp. 455–461, Jan. 2010, doi: 10.1002/jcc.21334.
- [35] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, "AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings," *J Chem Inf Model*, vol. 61, no. 8, pp. 3891–3898, Aug. 2021, doi: 10.1021/acs.jcim.1c00203.
- [36] S. Genheden and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities," *Expert Opinion on Drug Discovery*, vol. 10, no. 5. Informa Healthcare, pp. 449–461, May 01, 2015. doi: 10.1517/17460441.2015.1032936.
- [37] Dr. N. Astalakshmi *et al.*, "Over View on Molecular Docking: A Powerful Approach for Structure Based Drug Discovery," *Int. J Pharm Sci Rev Res*, vol. 77, no. 2, Dec. 2022, doi: 10.47583/ijpsrr.2022.v77i02.029.