POLITECNICO DI TORINO

Master Degree in Biomedical Engineering

Master Thesis

Characterization of novel colchicine derivatives to selectively inhibit tubulin polymerization in humans

Thesis Supervisors

Prof. Marco Agostino Deriu

Co-supervisors

Prof. Umberto Morbiducci

Prof. Jacek Adam Tuszynski

Candidate

Antonio Rocca

ACADEMIC YEAR 2018-2019

Summary

ABSTRACT

Microtubules are protein assemblies made of tubulin dimers shaped as hollow cylinders. They are key players of many cellular functions, such as cell mechanical behaviour, signal transportation and mitosis. Because of this central role in cell mitosis, they are an interesting target of pharmacological treatment, as it is possible to block the mitotic process by altering the dynamic equilibrium between microtubule's polymerization and depolymerization. From this perspective, it is interesting to design and develop antitumoral drugs able to selectively bind the tubulin in the tumoral cells driving them to apoptosis. There are several tubulin isotypes. Tumoral cells are characterized by an overexpression of αβIII tubulin, so a drug able to specifically target the above-mentioned tubulin is assumed to be effective in targeting tumoral cells.

For example, colchicine is an interesting compound because it has a strong antimitotic-activity, but it is quite toxic for the organism, so the purpose is to develop derivatives that on one hand are capable to enhance the antitumoral activity and on the other hand to reduce its poisonous side effects. However, bringing a drug to market is not easy and implies enormous costs.

In this context, computational modelling can be a valuable tool to develop and test several compounds. The present study focuses on 61 colchicine derivative compounds targeting the tubulin dimer. More in depth, Drug Discovery and Virtual Screening methods supported by Molecular Dynamics simulations have been employed to characterize the above-mentioned ligands' ability to bind different tubulin isotypes and their local effects on the target.

ESTRATTO

I microtubuli sono degli aggregati proteici dalla forma di cilindri cavi. Sono dei componenti chiave per mole funzioni cellulari, come il comportamento meccanico della cellula, il trasporto del segnale e la mitosi. A causa del loro ruolo centrale nella mitosi cellulare, sono degli interessanti target per il trattamento farmacologico, essendo possibile bloccare il processo mitotico alterando l'equilibrio dinamico tra la polimerizzazione e la depolimerizzazione del microtubulo. Da questa prospettiva, è interessante progettare e sviluppare farmaci antitumorali in grado di legare selettivamente la tubulina nelle cellule tumorali, portandole all'apoptosi. Ci sono diversi isotipi di tubulina. Le cellule tumorali sono caratterizzate da una over espressione della tubulina αβIII, perciò un farmaco in grado di legarsi in maniera specifica a questo isotipo di tubulina si può considerare efficace per mirare selettivamente alle cellule tumorali.

Ad esempio, la colchicina è un composto interessante, poiché ha una forte attività antimitotica, ma è tossica per l'organismo; pertanto l'obiettivo è quello di sintetizzare dei suoi derivati che da un lato siano in grado di aumentarne l'attività antitumorale e dall'altro siano in grado di ridurne gli effetti collaterali. Però, portare un farmaco sul mercato non è semplice ed ha costi elevati.

Pertanto, il modelling computazionale può aiutare a sviluppare e testare diversi farmaci. Questo studio si focalizza su 61 derivati della colchicina per colpire i dimeri di tubulina. Il Drug Discovery e il Virtual Screening, supportati dalla Dinamica Molecolare, sono stati utilizzati per caratterizzare l'abilità dei ligandi sopra menzionati di legare differenti isotipi di tubulina e i loro effetti locali sul bersaglio.

1. Introduction

The present chapter deals with a general introduction of this master thesis work, summarizing the biological background, the aims of the research and the organization of the dissertation.

Microtubules are cytoskeletal filaments involved in several cell functions, including cell shape maintenance and cell division. They are composed by α and β heterodimers of tubulin, aligned in head-to-tail fashion, forming linear protofilaments, which bound laterally to form a hollow cylindrical polymer. Microtubule functionality is linked to a complex process of polymerization and depolymerization, called *dynamic instability*, that is characterized by switching between phases of growing and disassembling. This polymerization process is fundamental in many cellular functions. By altering this dynamics equilibrium, it is possible to arrest the mitotic process, leading the cell to the apoptosis. In this context, many antimitotic drugs have been developed in the past years for anti-cancerous therapies. Colchicine is one of the most used, but it has the side effect of being toxic and not very effective and specific on the tumoral cells.

There are many tubulin isotypes, which are found to have a role in the development of drug resistance cell lines, as different isotypes show dissimilar binding energies to various drugs. In particular, tumoral cells show an overexpression of the β-tubulin isotype βIII, which can lead the tumoral cells to become drug-resistant. Because of this overexpression, this isotype can be used to target selectively the tumoral cells, by designing drugs able to selectively bind it over the other isotypes.

Developing and bringing a drug to the market implies enormous cost and it is not a very simple process, as it needs adequate assays or animal models. Therefore, computational methods can be used to design and test a huge number of drugs with significant lower costs and resource.

The aim of the present work is to characterize and evaluate several colchicine derivatives on different human tubulin isotypes, focusing on their local effects, and to individuate among them some drugs able to overcome colchicine's limitations.

The thesis is divided in five Chapters as described in the following.

Chapter 1 is the present introduction.

Chapter 2 is an overview of the methods used in this work. Molecular mechanics and molecular dynamics are firstly described in general, focusing on physical and theoretical aspects, followed

by the description of Molecular Docking and Virtual Screening, with focus on the search algorithms, the binding free energy calculation and the scoring functions. The software used are also presented and described.

Chapter 3 is dedicated to a biological background of microtubule, describing the molecular structure, the polymerization dynamics, the role in anti-cancerous therapies and the isotype expression in humans. The microtubule targeting agents are also presented and described.

Chapter 4 is devoted to the characterization and the evaluation of several colchicine derivative drugs bonded to the human tubulin isotypes αβIIA, αβIII and αβIVA. Molecular dynamics simulations and molecular docking are performed on the tubulin dimers. The results are analyzed with classical analysis and with binding free energy analysis, focusing on the differences between the drugs inter the isotypes and intra the same isotype.

Chapter 5 is dedicated to the comparison of the effects of the colchicine and a novel compound in the isotype αβIII. Molecular dynamics simulations on free tubulin dimer are analyzed with classical analysis and with the binding free energy analysis, in order to understand if the novel compound can be better than the colchicine against the tumoral cells.

2. Materials and Methods

The present chapter provides the theoretical background for the present master thesis work, with the aim of explaining the physical basis behind the computational approach. In detail, after a quick introduction to the molecular modelling (section 2.1), sections 2.2 and 2.3 are devoted to the Molecular Mechanics and Molecular Dynamics, describing the theoretical method together with algorithm implementation details. Finally, the Drug Discovery approaches Molecular Docking and Virtual Screening are described in detail in section 2.4.

2.1 Computational Modelling of Biomolecular Systems

Molecular Modelling includes a set of theoretical and computational techniques used to describe complex chemical systems (e.g., proteins, polymers, molecules, nucleic acids), in terms of a realistic atomistic description. The aim is to understand and predict macroscopic properties of these systems by solving the equations of quantum and classical physics. Molecular systems generally consist of a large number of molecules and for this reason is difficult to estimate thermodynamic or kinetic properties of these systems. Molecular Dynamics (MD) is a powerful tool for understanding molecular processes like protein folding/unfolding, transport of drugs, docking of molecules and much more, because it represents a connection point between laboratory experiments and theory. The most accurate description of a hierarchical system is at the Quantum Mechanical (QM) level, by solving the Schrodinger equation[1]. However, Schrodinger equation can be solved only for very simple systems. So, despite their high accuracy, QM simulations are computational demanding and only hundreds of atoms can be simulated in a reasonable time-frame, leading to limited knowledge of the biological phenomena. At the molecular level, Molecular Mechanics (MM) methods enable to consider systems containing thousands of atoms. The result of MD is a trajectory that specifies the position and the velocity of each particle at any time[1]. MM and MD will be discussed in detail in the next sections.

2.2 Molecular Mechanics

Molecular Mechanics methods use Newtonian mechanics to model molecular systems. Due to the Born-Oppenheimer approximation[2], the atoms are treated as spherical particles with a radius and an overall charge obtained from both theory and experimental measurements and the bonds are treated as springs, whose stiffness depends on the elements bonded together. This allows to describe the bond stretching, bending and twisting. The aim of molecular mechanics is to

predict the energy associated with a given molecular conformation. A molecular force field (FF) is a set of equations that allows to estimate the potential energy of the system as a function of the atoms' positions.

2.2.1 Potential Energy Function

The energy of a molecule in the ground electronic state can be considered as a function of the nuclear coordinates only. Any changes in the system causes the variation of a multidimensional "surface" called energy surface. The potential energy surface for a molecular system of N atoms in a given conformation is the sum of two contributions:

$$V = V_{Bonded} + V_{non-Bonded}$$

where the components of the bonded and non-bonded terms are given by the following equations:

$$V_{Bonded} = V_{Bonds} + V_{Angles} + V_{Dihedrals}$$

$$V_{non-Bonded} = V_{Van\ der\ Waals} + V_{Electrostatics}$$

Each of the terms mentioned above can be modelled in a different way, depending on the particular simulation settings being used.

2.2.2 Treatment of Bond and Non-Bond interactions

The central core of the MM method is the definition of a potential energy function. A simple molecular mechanics energy equation contains four components[3]:

$$V(r_N) = \sum_{Bonds} \frac{1}{2} k_l [l - l_0]^2 + \sum_{Angles} \frac{1}{2} k_\theta [\theta - \theta_0]^2 + \sum_{Dihedrals} k_\varphi [1 + \cos(n\varphi + \delta)] + \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}^{12}}{r_{i,j}} \right) - \left(\frac{\sigma_{i,j}^{6}}{r_{i,j}} \right) \right] + \frac{Q_i Q_j}{4\pi\varepsilon_0 \varepsilon_r r_{i,j}}$$

In the previous equation, the first term describes the interaction between bonded atoms, modelled as a harmonic potential inducing potential energy to increase when the bond length l departs from the reference value l_0. The k_l represents the bond stiffness. Both l_0 and k_l are assigned for each couple of atoms and their values are different depending on the atoms' types. The second term of the equation describes the angle among three atoms covalently bonded: k_θ is the stiffness of the bond angle, θ_0 is the equilibrium angle and θ is the bond angle. The third term represents the dihedral bond interactions. That usually includes a torsional potential describing bond rotates: k_φ is the energetic barrier related to the angle deformation, δ is the phase that determines the minimum position for the torsional angle and n is the multiplicity. The last term represents the non-bond interactions. They are usually modelled as a function of an inverse power of the distance and are built up from two components: Van der Waals forces and electrostatic interactions.

The Van der Waals potential represents the sum of the attractive and the repulsive forces between two neutral entities within a molecule. These forces are caused by London dispersion and Pauli exclusion principle phenomena of nearby particles. Compared to the covalent bonds, the Van der Waals interactions are relatively weak, but play a fundamental role in defining many properties of organic compounds, such as their solubility in polar and non-polar media, contribute to the steric effect and to the three-dimensional molecular conformation. This potential is often represented with the Lennard-Jones equation[4], which consists in an approximate model for the isotropic part of a total (repulsion plus attraction) Van der Waals force as a function of distance r:

$$V_{Lennard-Jones} = 4\varepsilon_{i,j}\left[\left(\frac{\sigma_{i,j}^{12}}{r_{i,j}}\right) - \left(\frac{\sigma_{i,j}^{6}}{r_{i,j}}\right)\right]$$

The first term represents the repulsive component, while the second one represents the attractive component. The equation contains two parameters: the collision diameter $\sigma_{i,j}$, which is the distance after that the Van der Waals interactions produce a zero contribution, and the depth of the potential minimum $\varepsilon_{i,j}$.

The electrostatic interactions can be described by using the Coulomb's law:

$$V_E = \frac{Q_i Q_j}{4\pi\varepsilon_0\varepsilon_r r_{i,j}}$$

The denominator factor is called electric conversion factor, where ε_0 is the free space permittivity and ε_r is the relative permittivity. This interaction is defined as a long-range interaction, because their contribution goes down with $\frac{1}{r}$, slower than the other non-bond interactions.

The number of the non-bond interactions increases as the square of the number of atoms in the system, so the calculation of non-bond forces is extremely expensive in terms of computational effort. To properly reduce their computational effort, non-bond interactions can be computed by applying various methods, like the *distance cutoff*, in which every interaction between two atoms is computed only if their distance is smaller than the cutoff, or the *Ewald summation*, in which the interaction energies summation in real space is replaced by an equivalent summation in the Fourier space which is employed in the long-range interactions' calculation.

2.2.3 Periodic Boundary Conditions

The treatment of the boundaries effects is crucial in simulations, since they strongly influence the properties of the whole system. The *Periodic Boundary Conditions* (PBC) allow to minimize the edge effects in a finite system. All the atoms are put in a box, usually filled with implicit or explicit models of water, surrounded by translated copies of itself, in order to remove the boundaries of the system. The presence of the PBC causes imprecisions, but they are still less severe than the errors resulting from an artificial boundary with vacuum. The particles in the box interact with any other particles in adjacent box, which are repeated infinitely if the PBC are settled on. To avoid this problem, many approaches are possible. Applying a Lennard-Jones cutoff distance, which decreases very rapidly, is not a problem for short-range interactions, but the long-range interactions require a more accurate method, with the aim of avoiding discontinuities in the potential energy calculation. Some methods proposed to overcome these kinds of problems are Particle Mesh Ewalds[5], Multipole Cells[6] and Reaction Fields[7].

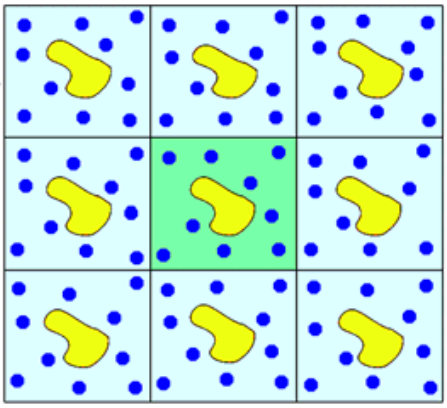


Figure 1: Periodic boundary conditions. The central box, filled with water, is replicated in copies of itself.

2.2.4 Potential energy minimization

The potential energy function is a very complex multidimensional function of molecular system coordinates, a N-atoms system being defined by 3N cartesian and 3N-6 internal coordinates (bonds, angles and torsional angles). A minimum point of the Potential Energy Surface (PES) corresponds to a stable state of the system: any movement away from this configuration is characterized by higher energy. There are a lot of *local minima* in the PES and the minimum with the lowest energy is known as *global minimum*. The energy minimization is able to reduce the potential energy of the system, starting from the MM description of the molecular system in term of force field. To identify the minimum point of the PES there are two different kinds of approaches: *derivative methods* and *non-derivative methods*. In the *first order derivative methods* (e.g., Steepest Descent[8], Conjugate Gradient[9]) the direction of the gradient (the first derivative of the energy) indicates where the minimum lies. In the *second order derivative methods* (e.g., Newton-Raphson[10], L-BFGS[11]) the second derivatives provide information about the curvature of the function, showing where the energy function change direction; the inverse Hessian matrix of second derivatives is calculated. These methods can be computationally demanding and may also require a significant amount of storage.

Especially for simulations of complex systems (e.g., macromolecules), energy minimization is widely present before a molecular dynamics simulation.

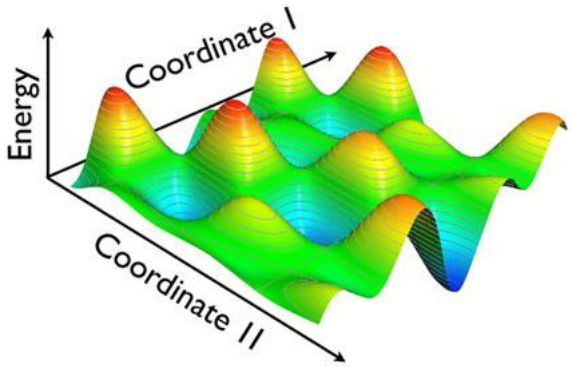


Figure 2: Potential Energy Surface (PES). It is a multi-dimensional surface characterized by local and global minima.

2.3 Molecular Dynamics

Molecular Dynamics (MD) is a computational approach able to investigate the dynamics of large molecular systems, such as proteins in solution, by providing trajectories, in terms of positions, velocities and forces. The method allows to calculate average properties of a system by solving the Newton's equations of motion, that can be expressed as follows for a N particles system:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i$$

Where m_i represents the mass and r_i the position of the i^{th} particle of the system respectively. F_i are the forces, which can be written in terms of the Potential Energy:

$$F_i = -\frac{\partial U(r_i, \ldots, r_N)}{\partial r_i}$$

2.3.1 Statistical Ensemble

There are two categories of macroscopic properties in a chemical system: static equilibrium properties (e.g., radial distribution function, averaged potential energy) and non-equilibrium properties (e.g., diffusion processes, dynamics of phase changes). According to statistical mechanics, it is necessary to average the microscopic states of a system in a time independent fashion, distributed according with certain statistical ensemble, in order to compute macroscopic properties of a system. The space in which all possible physical states of a system are represented is known as *phase space,* which often consists of all possible values of position *r* and momentum *p* variables. It means that in a system containing N atoms, 6N coordinates (three coordinates of position and three coordinate of momentum) are necessary to identify a point in the phase space. A single point in the space represents a particular state of the system,

and all the plotted points represent all the states the system can be. A collection of points of the phase state with different microscopic states but satisfying some particular thermodynamic state it is called *statistical ensemble*. There are different kinds of ensembles:

- The Micro-Canonical Ensemble (NVE): it is characterized by a fixed number of atoms (N), a fixed volume (V) and a fixed energy (E). It corresponds to an isolated system.
- The Canonical Ensemble (NVT): it is characterized by a fixed number of atoms (N), a fixed volume (V) and a fixed temperature (T). It corresponds to a closed system.
- The Grand Canonical Ensemble (μVT): it is characterized by a fixed chemical potential (μ), a fixed volume (V) and a fixed temperature (T). It corresponds to an open system.
- The Isobaric-Isothermal Ensemble (NPT): it is characterized by a fixed number of atoms (N), a fixed pressure (P) and a fixed temperature (T).

The MD method samples the phase space in a way that is representative of the equilibrium state and is able to calculate the ensemble average of the macroscopic property. The ensemble average of property A, which is function of the position and the momenta, is calculated as follow:

$$\langle A \rangle_{Ensemble} = \iint dp^N dr^N A(p^N, r^N) \rho(p^N, r^N)$$

where r is the atomic position, p is the momenta and $\rho(p^N, r^N)$ is the probability density function. Under canonical statistical ensemble (NVT) the probability density is the Boltzmann distribution:

$$\rho(p^N, r^N) = \frac{1}{Q} exp[-H(p^N, r^N)/k_b T]$$

where k_b is the Boltzmann factor, H is the Hamiltonian and T is the temperature. Q is the partition function, which is usually written in more generally terms as function of the Hamiltonian H:

$$Q = \iint dp^N dr^N exp[-H(p^N, r^N)/k_b T]$$

The partition function is a sum of Boltzmann factor over all microstates of the system and it relates microscopic thermodynamics variables to microscopic properties. It is extremely important, and it should be calculated over all possible states corresponding to the particular thermodynamics constrains. However, the analytical solution is impossible to calculate because

it is extended to all possible states of the system. To overcome this problem, the *ergodic hypothesis* can be used: over long periods of time, the time average of a certain property is equal to the ensemble average of the same property. The time average can be computed by:

$$\langle A \rangle_{Time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \cong \frac{1}{M} \sum_{t=1}^{M} A(p^N, r^N)$$

where t is the simulation time, M is the number of the time steps in the simulation and $A(p^N, r^N)$ is the instantaneous value of the property A.

2.3.2 Molecular Dynamics implementation scheme

The basic idea of the MD method is to solve Newton's equations of motion for a system of N interacting particles. It is a deterministic method, so the state of the system at any time can be predicted from its previous state. According to the Newton's second law, the acceleration for each atom is:

$$a = -\frac{1}{m} \frac{dV}{dr}$$

The potential energy is a function of the atomic position and because to the complex nature of this function there is no analytical solution of the previous equation. The solution can be obtained using some numerical integration scheme. There are various integration methods for MD: the Verlet algorithm[12], the Leap-Frog algorithm[13], the Velocity Verlet[14] and many more. A very important choice is the integration time-step (usually fs), in order to sample correctly the phase space and to avoid instability. The flowchart in Figure 3 describes the MD:

- The initial structure is provided as input data. The initial atomic positions are known (e.g., from Protein Data Bank), while the initial atoms velocities are chosen randomly from a Maxwell-Boltzmann distribution at a given temperature. Starting from this input, the potential energy is calculated.
- The algorithm continues with the calculation of the forces F_i by deriving the potential energy function.
- A new set of atoms coordinates r_i and velocities v_i is generated step by step and the cycles goes on until the equilibrium is reached.
- Lastly, the output trajectory is defined.

Using the output trajectory of the MD, the macroscopic thermodynamics properties (e.g., energy, temperature, pressure) can be calculated as time averages.

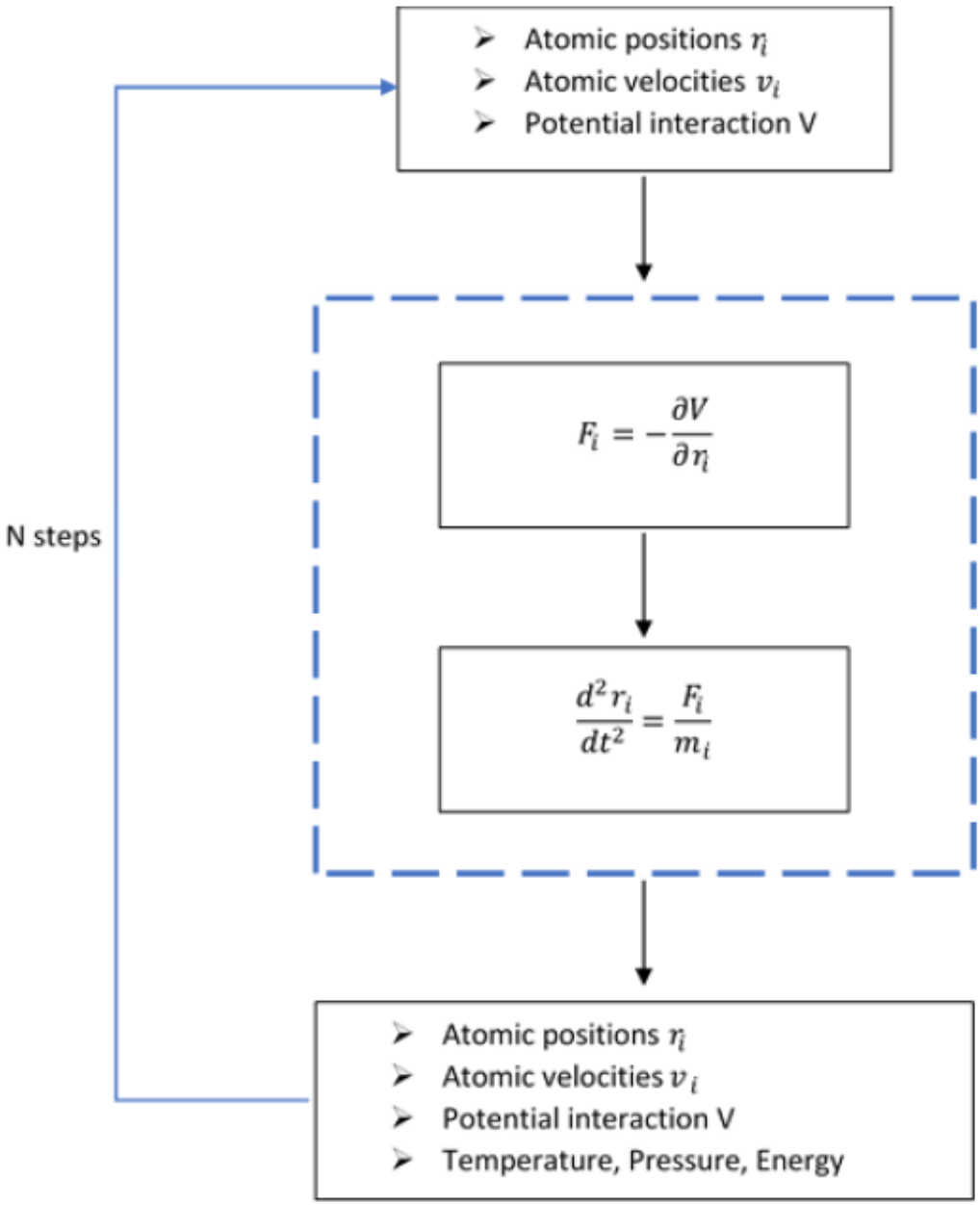


Figure 3: MD flow-chart: the initial positions and velocities are provided as input. Stating from the input, the potential interaction is calculated. The algorithm goes on with the calculation of forces and, through integration methods, new positions and velocities are generated; the cycle continues for a certain number of steps, until the equilibrium is reached.

2.3.3 Software package

There is a wide variety of MD codes for biomolecular simulation: AMBER[15], CHARMM[16], GROMACS[17], etc. In this work, both GROMACS and AMBER software have been used. GROMACS (GROningen MAchine for Chemical Simulations) is a molecular dynamics simulation package originally developed in the University of Groningen, now maintained and extended at different places and university. GROMACS is written for Unix-like operating systems and the entire package is available under the GNU General Public License. GROMACS is a versatile tool to perform molecular dynamics simulations and energy minimization. It is primarily designed for biochemical molecules like proteins, lipids and nucleic acids that have a lot of bonded interactions, but it is very fast in the calculation of non-bonded interactions, and for this reason many research groups are also using it for non-biological systems, e.g. polymers. AMBER (Assisted Model Building and Energy Refinement) is another software package developed to perform MD simulations on macromolecules and consists of several programmes such as:

- Leap: is a software that allows to build the coordinates, parameters and topology of the system to correctly start the simulation. It provides an editor for molecular manipulation.
- Antechamber: automates the process of parametrization compatible with the Amber Force Field.
- Sander: It is the central simulation program used for MD simulations.
- Pbsa and mmpbsa: software used to estimate the free energy between two states.

2.4 Molecular Docking and Virtual Screening

In order to discover and design new therapeutic agents, many techniques can be used, such as X-ray crystallography and NMR. This led to a large number of available molecules[18]. However, these methods could be expensive and not always applicable. Discovering or designing a drug takes up to 15 years to get the point of starting for a clinical trial, even for a pharmaceutical company, that usually have all the needed resources. These difficulties are consequences of several reasons: not adequate assay or animal models, challenges of organic synthesis and expensive costs. Due to the rise in computer power and storage capacity, the use of Computational Molecular Docking and Virtual screening (VS) tools have become crucial to make the development process more efficient and economic.

Molecular Docking can be used to model ligand-receptor interaction by means of prediction of the ligand conformation in the active site (pose prediction) and predict the binding affinity. Docking outputs are the ligand-receptor complex structure at an atomistic level and the binding energy between protein and ligand. Virtual Screening has the aim to rank active molecules above inactive ones. These tools automatically sample many different 3D conformations of the ligand in the active site of the receptor and compute a score for each of them. The aim of this score is to have comparable results with the experimental biochemical assay, in which an inhibition constant K_i is evaluated[19].

2.4.1 Ligand and receptor

In order to perform a structural based molecular docking, both ligand and receptor structures are needed. Receptors are 3D-structures (usually proteins), with atomic detail. Structural methods such as NMR studies or X-ray crystallography are used to determine the starting target structure. These structures can be downloaded by free from the Protein Data Bank website[18]. Even small deviations in structures can largely alter the results of a docking experiment, so the choice of the protein is a crucial step for the docking process. In general, a good rule is to opt for high resolution structures (below 2.5 Å[20]). However, not all the structures available on databases are from the human organism and not all the protein structures that the human genome can encode are disclosed. Nevertheless, a powerful bioinformatic tool, the *homology modelling*[21], permits to build a new atomic scale model by proteins that are phylogenetically similar to the target, called *templates*. Before performing docking, the protein 3D structure has to be prepared: a good practice is to perform a MD simulation to relax the system and explore the conformational space to have more precise information about the structure. The most critical step is the identification of the *active site*, which is usually a pocket or a protuberance that has hydrogen bonds acceptors or donors, hydrophobic properties, etc. The active site is often known by experimental works. However, if it is unknown, several approaches can be used to find it, based either on geometrical or energetic criteria. In order to represent the receptor, different methods are applied: atomic, surface, or *grid representation*[22], which is more efficient and computationally less demanding, as it creates only a potential grid field per each atom/interaction types, and it is then read during ligand scoring.

Several millions of 2D and 3D virtual models of ligands are available on different databases, so the compounds of interest can be found in one of these libraries, such as PubChem[23]. The extracted ligands have to be processed before docking: if the information about the ligand is

stored in 2D format, the 3D structure has to be created; atom and bonds have to be checked, protonation state and charges have to be assigned. If the structure is not present in databases, it is possible to directly design the molecules as they are usually simple and with a small number of atoms, using a design software such as *Avogadro*[24]. The 3D structure is then optimized by using the tools cited above.

2.4.2 Search Algorithms

Searching the correct pose of ligands requires to explore many different conformations. There are many degrees of freedom, so an *automatic algorithm* is needed. The aim of the search is to identify the pose with the lowest energy. Different algorithms can perform a local search, where the method finds the closest minimum on the energy landscape, or a global search, where the landscape is explored more widely. Different docking programs implements different searching methods, which could be *Matching algorithms*, *Systematic Search* or *Stochastic methods*.

Matching algorithms match ligand shape and chemical complementarity with the binding site. An example is the pharmacophore-based docking. It requires known binders (active compounds) and the representation of molecular feature is necessary for ligand binding. It matches the pharmacophore map of the binding site to the ligand's one. The native binding pose can be determined using an energy-based scoring or calculating the lowest Root Mean Square Deviation (RMSD) between centres of features.

Systematic search methods sample the conformational space at a predefined interval. They are deterministic methods. The aim of the search is to explore all the possible degrees of freedom in a molecule. They can be divided in two classes: exhaustive search and fragmentation algorithms. Exhaustive search ideally explores all the possible conformations by rotating all the possible bonds at a given interval. However, due to the combinatory explosion:

$$N_{conform} = \prod_{i=1}^{N} \prod_{j=1}^{n=inc} \left(\frac{360}{\theta_{ij}} \right)$$

that for large ligands diverges, a series of geometrical constrains and a post filtering step are used in order to reduce the huge amount of different conformations. This is especially useful during virtual screening experiments, where millions of small molecules are tested. In the equation above, N is the number of rotatable bonds and θ_{ij} is the size of the incremental rotational angle j for bond i. In the fragmentation method, the ligand is divided into fragments

and the conformational space is sampled by incrementally adding one fragment per time in the active site by linking them covalently, in order to minimize the energy of conformation.

Stochastic methods reach quickly a near-optimal solution by randomly changing the degrees of freedom of the ligand. They are more suitable for large molecules, where the degrees of freedom are too high for an efficient systematic search. Examples are genetic algorithms (GA)[25], simulated annealing (SA)[26] and Monte Carlo[27] (MC). Genetic Algorithms try to optimize a problem by mimic the process of evolution. The generated configurations for the ligand represent the individuals and a function, called fitness function, decides which configuration (individual) survives and shall produce the next generation of configurations. The survived generation is used for further iteration-optimization steps by means of combining it with other selected conformations. To match this theory to the docking searching method, some assumptions have to be done:

- Each individual (configuration) has a unique chromosome with its associated genes that equals the state variables that describe a representation of a configuration, usually in a linearized for such a bit string. For example, a bit in the string can encode the torsion angles conformations and another bit the mapping of the hydrogen bond parameters in the protein and the ligand.
- A fitness function should evaluate the best configurations.
- Genetic operators such as a mutations or crossover are allowed in order to randomize the new generated offspring.
- The search can terminate either when convergence, a maximum number of iterations or a particular value of the fitness function is reached.

In the Monte Carlo search, several random changes in the ligand's configuration are done. The method keeps track of the already samples states, in order to avoid a random movement that samples them again. It can also be coupled with simulated annealing methods: random changes are made as well, but a consecutive step is accepted or rejected according to the *Metropolis Criterion*. Starting from a configuration with energy E1, the Metropolis criterion accepts the change if the new energy is lower or with a probability $P = \exp\{-[(E2 - E1)/k_b T]\}$ where T is the temperature and k_b is the Boltzmann constant. Moreover, the temperature is reduced during the simulation in order to lower the provability to accept high energy conformations. This approach permits to overcome the problem of the stuck in a local minimum, as a conformation with higher (worse) binding energy could be accepted, but with a lower probability.

2.4.3 Binding Energy

A crucial step in docking and VS is the evaluation of binding energy and the ranking of generated conformations. The standard binding energy in a non-covalent ligand-protein complex is associated to the association constant K_i by this equation:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT \ln K_i$$

Where H is the enthalpy, S is the entropy, T is the temperature, R is the gas constant and the superscript "0" means that the binding free energy is evaluated ad standard conditions. Therefore, to correctly evaluate the binding energy, both enthalpic and entropic effects have to be accounted. The enthalpy accounts for the electrostatic, Van der Waals, hydrogen, polar, aromatic and more others[28], while the entropy effect accounts somehow for the flexibility of the ligand. Despite the theoretical method being well known, how to implement it in biomolecular systems is not easy and is computationally demanding. For this reason, approximated methods such as Molecular Mechanics Generalized Born/(Poisson Boltzmann) Surface Area model (MMGB/(PB)SA)[29] can be used to define the binding energy reaching a good compromise between computational time and accuracy. This method calculates the difference between two different states that are the bound and the unbound state of two solvated molecules (the drug and the target) and not the absolute free energy. It can be used only as a post-processing method of simulation searching algorithms such as MD. In fact, representative frames from an ensemble of conformations are used to calculate the free energy between the two states. The solvent-solvent interactions, due to the higher order of magnitude fluctuations, can hide the binding energy. Therefore, the calculation is divided in a thermodynamic cycle and the free energy of binding can be written as follows:

$$\Delta G^0_{bind\,,solv} = \Delta G^0_{bind\,,vacuum} + \Delta G^0_{solv\,,complex} - \Delta G^0_{solv\,,ligand} - \Delta G^0_{solv\,,receptor}$$

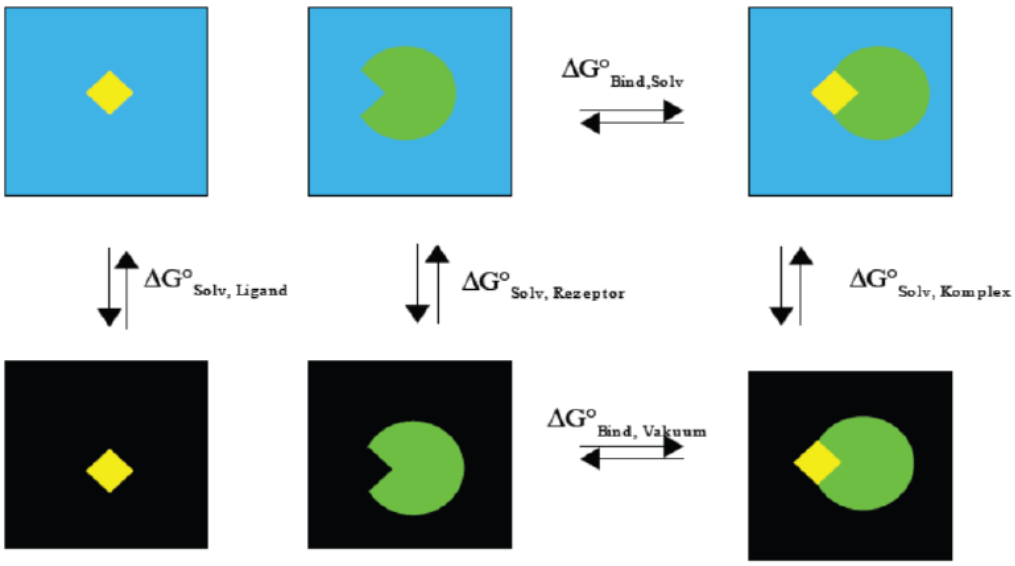


Figure 4: Thermodynamic cycle used in MMGB/PBSA method.

In this way the final binding free energy is divided in solvation contributes and a gas phase contribute. The free energy in vacuum can be approximated to:

$$\Delta G_{vacuum} = \langle E_{MM} \rangle - T\Delta S$$

where the first term is the average molecular mechanics, which consists of the sum of Van der Waals, electrostatics and bond, angle, torsion energy terms, while the second term is the internal entropy, which can be evaluated by either performing normal mode analysis[30] or a quasi-harmonic analysis[31]. However, it is computationally demanding, thus the entropy contribution is usually calculated by using low quality empirical evaluations, or in particular cases, such as a docking approach with a common target (where the only interest is the rank order) it can be neglected[32].

On the other hand, the solvation free energy contribute is subdivided into a sum of two components, one due to the electrostatic/polar interactions and the second due to the non-polar interactions. The polar contribute represents the energy stored in the continuum dielectric in response to the presence of the solute's charge distribution and is typically obtained by solution of the Poisson-Boltzmann (PB) equation or (GB) equation. For example, in the generalized Born equation the electrostatic contribution[33] is given by:

$$\Delta G_{polar} = -\frac{1}{2}\sum_{i,j}\frac{q_i q_j}{f(r_{ij}a_{ij})}\left(1-\frac{1}{\varepsilon}\right)$$

where $f(r_{ij}a_{ij})$ depends upon the inter-particle distances and the Born radii and ε is the solvent dielectric constant[34]. The non-polar contribution is proportional to the solvent accessible surface area (SASA) of the solute:

$$\Delta G_{non-polar} = \gamma SASA(x)$$

where SASA(x) is the area accessible to the solvent for the conformation x and γ is the microscopic surface free energy for formation of a cavity in water (surface tension).

2.4.4 Scoring Function

Free energy calculations such as MMPBSA or MMGBSA are enough reliable in prediction of binding affinity, but are impractical for high-throughput virtual screening, where a high number of protein-ligand poses have to be evaluated. In this case, the binding free energy is calculated by means of simplified models, called *scoring functions*, in order to reach a balance between speed and accuracy. There are three classes of scoring functions: the force-field (FF) based, the empirical and the knowledge-based.

The force-field based functions are based on the potential energy; thus, the entropy and the solvation energy are usually ignored. In basic FF scoring functions the goodness of coupling between ligand and target are described by the electrostatic and the Van der Waals terms. An example of this kind of energy functions is:

$$E_{nonbond} = \sum_{i}^{lig} \sum_{j}^{prot} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{D r_{ij}} \right]$$

where A_{ij} and B_{ij} are respectively the Van Der Waals repulsion and attraction terms of the Lennard-Jones potential; r_{ij} is the distance between atoms i and j, q is the point charge and D the dielectric constant. These functions are well understood, and the scoring is very fast when a grid is precomputed, but larger molecules are used to have a better score because of the larger number of interactions (the entropy term should account for that).

Empirical scoring functions aim to reproduce the experimental binding energy. The scoring function is given as a weighted sum of different energy terms such as ionic interaction, hydrogen bond, binding entropy and hydrophobic effect. The weights are obtained from linear regression analysis based on a test set of complexes with experimental known binding affinity. A general empirical scoring function is given by:

$$\Delta G_{empiric} = \Delta G_0 + \Delta G_{HB} \sum_{HB} f(\Delta R, \Delta\alpha) +$$
$$\Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{metal} \sum_{metal} f(\Delta R, \Delta\alpha) +$$
$$\Delta G_{lipo} \sum_{lipo} f(\Delta R, \Delta\alpha) + \Delta G_{aromatic} \sum_{aromatic} f(\Delta R, \Delta\alpha) + \Delta G_{rot} N_{rot}$$

Where the HB, ionic, metal, lipophilic, and aromatic are the interaction terms, f is a penalty function reflecting deviations from the ideal geometry in terms of distance ΔR and angles Δα, ΔG_{rot} is the energy loss per degrees of freedom and N_{rot} are the ligand rotatable bonds. The limitation of this approach is that the result strongly relies on the training set used, but these scoring functions are usually fast and useful in large virtual screening.

Knowledge-based scoring functions are based on the deviation of statistical preferences of ligand-protein contacts gained from experimentally solved structures: if a ligand-protein atom is found more frequently at a given distance in experimental structures, than this kind of interactions is a "plus" for the scoring. Therefore, the aim of these scoring functions is to reproduce experimental structures rather than binding free energy. The basic idea leading to these functions is based on the inverse Boltzmann law. This leads to the master equation of knowledge-based scoring functions:

$$w(r) = -kT \ln \left[\frac{\rho(r)}{\rho^{ref}(r)} \right]$$

Where k is the Boltzmann constant and T is the absolute temperature measured in Kelvin degrees, ρ is the probability function of the protein-ligand atom pair at distance (r) in the training set, and ρ^{ref} is the reference state density function where the inter-atomic interactions are considered zero. In other cases, the reference state is the one of mean interaction.

2.4.5 Software package

Nowadays, a huge number of docking software is available. The most widely used are AutoDock[35], AutoDock Vina[36], DOCK[37], LigandScout[38]. In this work AutoDock Vina has been used. AutoDock Vina achieves an approximately two orders of magnitude speed-up compared to AutoDock. It also significantly improves the accuracy of the binding mode predictions. Moreover, AutoDock Vina automatically calculates the grid maps and clusters the results in a way transparent to the user.

3. Biological Background

The present chapter provides the biological background for the present master thesis work. In detail, section 3.1 describes the microtubules, whit focus on the dynamic instability, on the tubulin structures available and on the different human tubulin isotypes, while section 3.2 describes the Microtubule targeting agents, with particular focus on the colchicine and its derivatives, as they are the ones used in this work.

3.1 Microtubule

Microtubules are cylindrical protein filaments. Along with actin filaments and intermediate filaments, they form the cytoskeleton, which is a complex network with the function to maintain the cell shape and give mechanical resistance. Microtubules are formed by polymerization of a dimer composed by two globular proteins, the α and the β tubulins. They are hollow tubes with internal diameter of about 12 nm and external one about 25 nm and can achieve 50 μm of length. A microtubule is usually composed by 13 linear protofilaments assembled around a hollow core. Each protofilament is composed by tubulin dimers linked head-to-tail[39]. The process of growth and destruction of microtubules is dynamic: at the same time both polymerization and depolymerization occur in the process called *dynamic instability*[40]. This process is fundamental for many basic cellular functions, such as mitosis: in fact, the microtubules form the mitotic spindle, which is able to separate eukaryotic chromosomes. Because of this central role in cell life, microtubules are an interesting target of pharmacological treatments: it is possible to block the mitotic process or to lead the cell to apoptosis by enhancing microtubules' depolymerization or polymerization.

3.1.1 Microtubule polymerization

As mentioned above, tubulin dimers are composed by two globular protein, α and β tubulins, which are very similar (indistinguishable at a resolution of 6 Å). Each tubulin monomer has a weight of about 50 kDa and can be divided into three distinct domains: the amino terminal domain, or Rossmann fold, (composed of residues 1–205), an intermediate domain (residues 206– 381) and a carboxy terminal domain, which contains the flexible C-terminal tail (residues 382–444)[41]. The microtubules are polar filaments, because they have two different ends: the minus end, where the α-tubulin subunit is exposed, and the plus end, where the β-tubulin subunit is exposed.

α and β monomers start the polymerization binding together on the plus end of the microtubule, which is characterized by a fast-growing process, while on the minus end the growing process is quite slow: usually, while the plus ends are free to explore the cytoplasmic space by stochastically switching between periods of growing and shortening (dynamic instability), the minus ends are typically anchored to the centrosome. It is important to study the seam region, where the MT wall is closed in the cylindrical structure, because it is the only region were α-α and β-β bonds are found and it seems to be a high instable region[42].

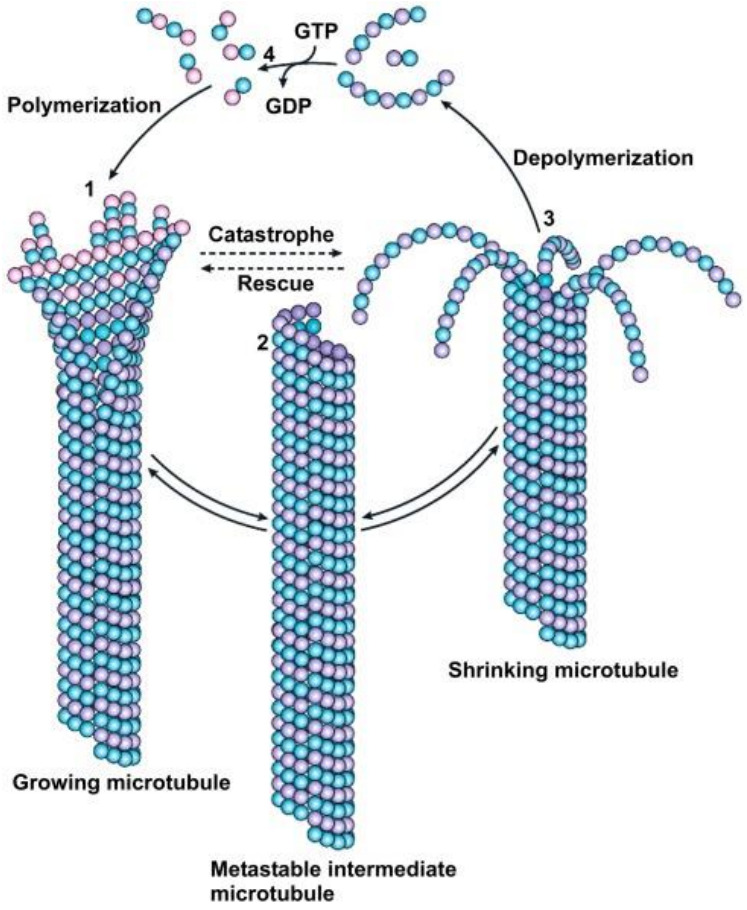


Figure 5: Microtubule dynamic instability: the closure of the terminal sheet structure generates a metastable intermediate microtubule, which can go under polymerization or depolymerization phase.

Before the integration in the microtubule wall, the tubulin dimers bind a GTP molecule, that is hydrolysed into GDP after the polymerization. Only the tubulin dimers at the end of the microtubules are bonded to the GTP, which forms the so called GTP-cap. The GTP or GDP state influences the stability of the microtubules, as the dimers bounded to GTP tend to assemble into microtubules, while the GDP-bounded tubulin tends to fall apart. For this reason, the loss of the GTP-cap leads to the opening of one end of the microtubule and to its depolymerization[43].

Microtubule polymerization is associated with a structural modification of the dimers from a curved conformation to a straight conformation. It is still not very clear what determines this switch. Nowadays, two models are proposed:

- Lattice model[44]: both in GDP and GTP state the free tubulin dimers adopt curved conformations. After the integration in the microtubule wall, they are somehow forced into assuming the straight conformation. The binding to GTP makes stronger the dimer-dimer bonds in the lattice but does not cause a conformational transition.
- Allosteric model: the GTP binding to free tubulin before assembly causes a curved-to-straight conformational switch, making the tubulin dimer sterically compatible for microtubule integration.

There are evidences for both the two models, thus the debate between them is still open.

3.1.2 Tubulin Structures

In the last twenty years several crystallographic structures of tubulin have been released and are now available from the RSCB Protein Data Bank (PDB). The first tubulin structure, 1TUB, was crystallized as a flat Zn^{2+} induced sheet of antiparallel protofilament-like end-to-end α/β dimer repeats, using docetaxel as a stabilizing agent[41]. A new structure, 1FFX[45], was derived from the 1TUB using a stathmin-like domain in 2000. This structure contains misalignments due to the difficulties in fitting electron density. Therefore, it was superseded by the 1JFF[46], in which paclitaxel was used as a stabilizing agent. Then, the 1TVK[47] was produced using the epothilone A, that is bounded to the taxane binding site, stabilizing the microtubule. In 2004, the structure 1SA0[48] was developed, using a stahmin-like domain. Then, a structure where both the colchicine and the vinblastine binding site are observed was developed, the 1Z2B[49]. In the end, two high resolution structures were built, the 3J6E[50] and the 4O2B[39]. The first one contains nine α/β dimers arranged to form a portion of the microtubule wall, while the second one contains two dimers, a Stathmin-4 and a tubulin-tyrosine ligase. The following table summarizes and confronts the information about these models.

Table 1: Tubulin models available at Protein Data Bank (PDB) website.

PDB code	Author	Resolution	Characteristic	Missing Residues
1TUB	Nogales et al. 1998	3.7 Å	Docetaxel as stabilizing agent	/
1FFX	Gigant et al. 2000	3.95 Å	stathmin-like domain	A 48-64, 441-451 B 34-50, 438-455
1JFF	Lowe et al. 2001	3.5 A	Paclitaxel as stabilizing agent	A 1, 35-60, 440-451 B 1, 438-455
1TVK	Nettles et al. 2004	2.89 Å	Epothilone A as stabilizing agent	A 1, 35-60, 440 B 1
1SA0	Ravelli et al. 2004	3.58 Å	Colchicine as destabilizing agent	A 1, 38-46, 438-451 B 1, 278-285, 439-455
1Z2B	Gigant et al. 2005	4.1 Å	Colchicine and vinblastine both binding	A 1, 38-46, 438-448 B 1, 278-285, 440-455
3J6E	Alushin et al. 2014	4.7 Å	Microtubule wall	A 1, 39-48 B 1
4O2B	Prota et al. 2014	2.3 Å	Two dimers, colchicine bonded	A 440-451 B 276-281, 439-455

3.1.3 Human Tubulin isotypes

In every organism there are several genomic copies of α or β tubulin, that differ for restricted or punctual variations in terms of aminoacidic sequence. In particular, if they are common through different species, tubulin characterized by the same mutations form an isoform, while, if they are common only in the same organism, tubulin characterized by the same mutations form an isotype.

These variations highly affect microtubule dynamics; thus, the organism produces specific isotype of tubulin to correctly maintain the microtubule assembly/disassembly equilibrium, according to external stimuli.

In the last years, several isotypes of α and β tubulin in humans were identified in the last years. However, there is more available data in literature about β tubulin isotypes, because this subunit is a target for the drugs more than the α subunit. Ten different beta tubulin isotypes have been

characterized: βI, βIIA, βIIB, βIII, βIVA, βIVB, βV, βVI, βVII and βVIII[51]. Different cells express different tubulin isotypes, as it can be seen in the following figures[52].

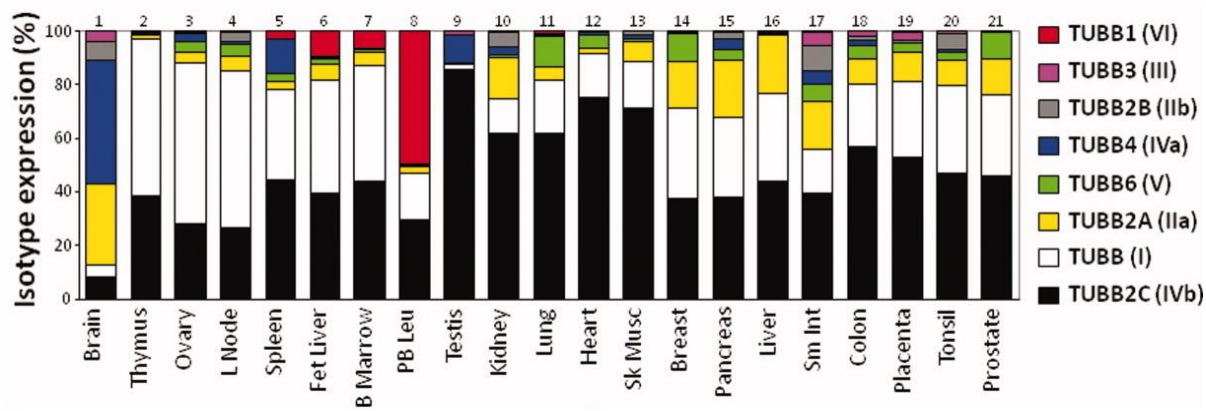


Figure 6: Different tubulin isotypes expression in normal human cells. The most ubiquitous ones are the I and the IV.

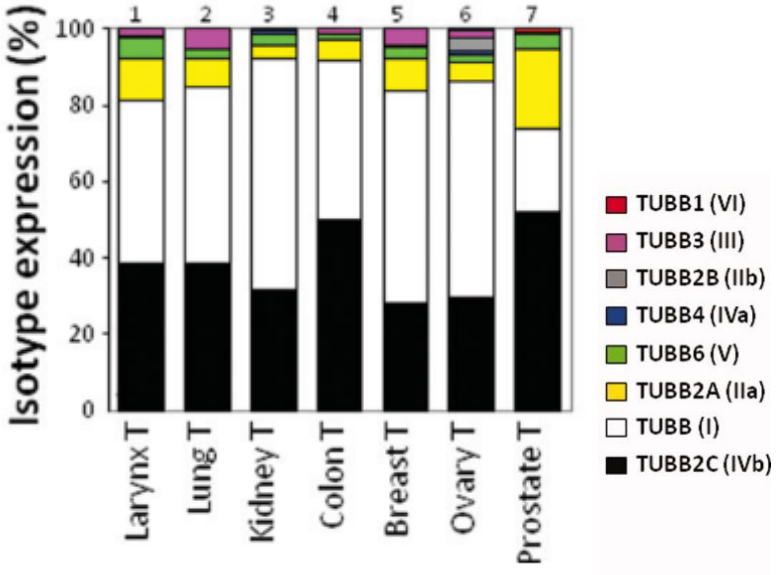


Figure 7: Different tubulin distribution in cancerous human cells. Isotype III is slightly more expressed than in normal cells.

As it can be seen in the pictures above, in normal cells, βI, βIVB, and βV are ubiquitously expressed, βVI is hematopoietic cell-specific, and βIIA, βIIb, βIII, and βIVA have high expression in brain. In tumoral cells, isotype βIII is usually over expressed, in comparison to nontumoral cells.

3.2 Microtubule Targeting Agents

Several classes of drugs can be used in order to alter the dynamic instability of the microtubule. The drugs can either move the equilibrium towards an enhanced polymerization or an enhanced depolymerization. The drugs interacting with the tubulin can be divided in three major classes[53],

based on their binding sites: taxol binding site drugs, vinca domain binding site drugs and colchicine binding site drugs. The drugs in the first class act as microtubule stabilizers, while the drugs in the other two classes act as microtubule destabilizers.

3.2.1 Taxol

Taxol was one of the first small molecules discovered to show the potential to be an important antitumor agent. It was isolated by Wall and Wani from the bark of the tree *Taxus Brevifolia*. They also determined the correct structure of the Taxol[54].

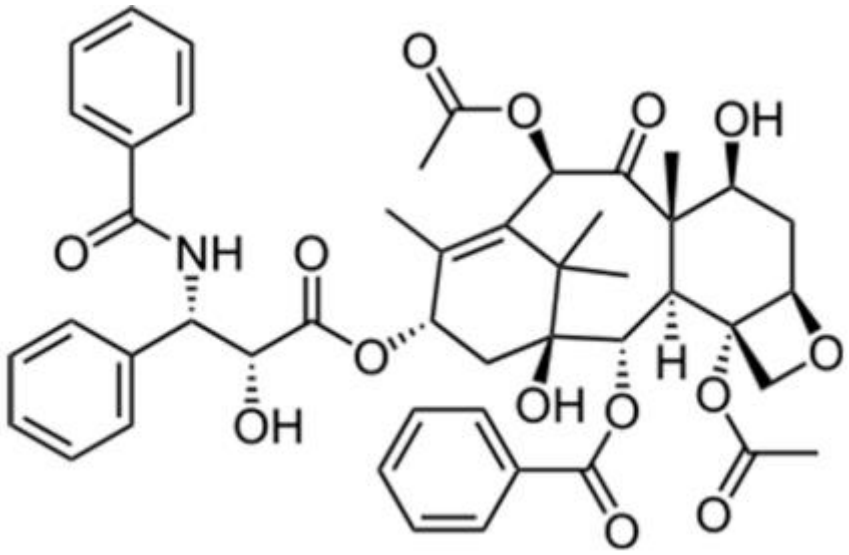


Figure 8: Structure of Taxol.

Taxol promotes microtubule assembly, reducing the critical concentration of tubulin required to form the microtubule and blocking the mitotic process in the cells[55].

Taxol binding site is primarily located in the β-tubulin. It contains its amino acid residues 217-231[56], the M-loop[57] (residues 273-285) and the N-terminal 31 amino acids[58].

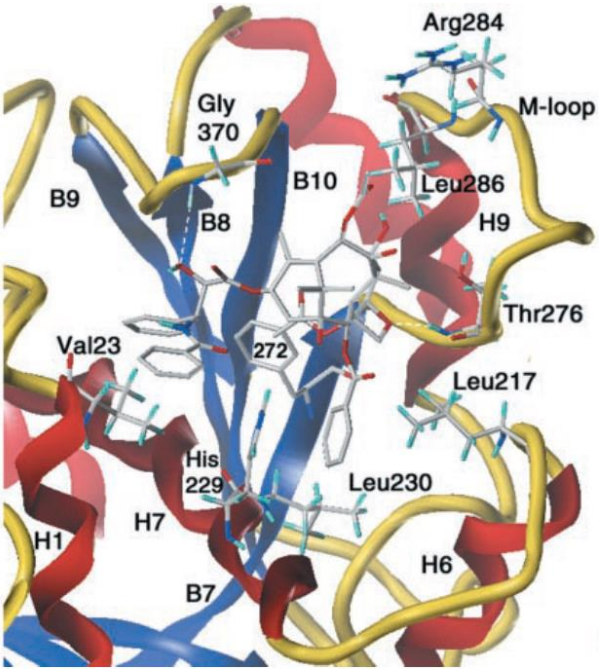


Figure 9: Taxol binding site[59]

Taxol has a complex structure and its extreme hydrophobicity makes it difficult to use for humans. To enhance the solubility, Taxol is usually mixed with ethanol in clinical use. This can lead to toxic effects like neutropenia and peripheral neuropathies, so a strategy to overcome this problem is to design Taxol derivatives drugs whit enhanced solubility.

3.2.2 Vinca alkaloids

Vinca alkaloids have shown to be potential antitumoral agents against numerous cell types. Three major vinca alkaloids, vincristine, vinblastine and vinorelbine have been approved for clinical use[60]. They are microtubule destabilizing agents. However, they all show some side toxic effects: vincristine is highly neurotoxic, vinblastine causes various disturbs primarily in the digestive system, while vinorelbine causes diseases in the circulatory system[61].

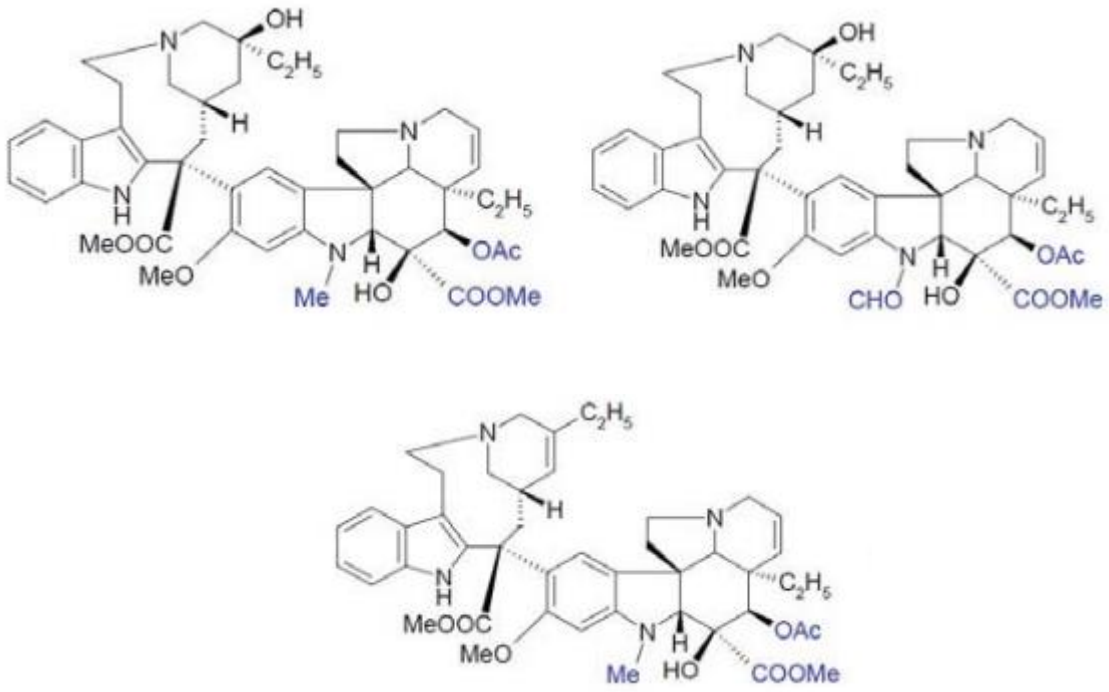


Figure 10: Major Vinca alkaloids structures: vinblastine, vincristine and vinorelbine

Vinca alkaloids bind to the microtubule and their binding site seems to be across two tubulin dimers. In particular they seem to contact helix H10, strand S9 and loop T7 in the α2-subunit and loops T5 and H6–H7 in the β1-subunit[62], where α2 and β1 are referred to tubulin subunits belonging to two different dimers.

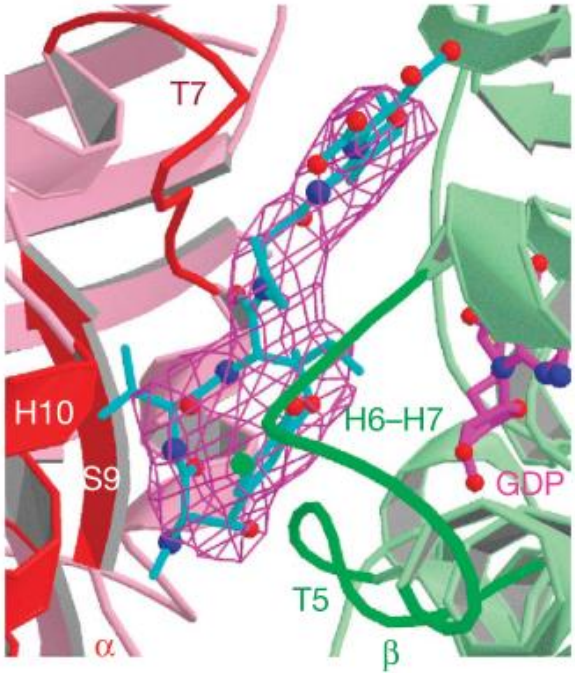


Figure 11: Vinca Alkaloids domain[62]: the α2 subunit is represented in pink, while the β1 subunit is in green.

3.2.3 Colchicine and colchicine derivatives

Colchicine is a compound with a relatively high antiproliferative activity. It is a major alkaloid isolated from Colchicum autumnale's and Gloriosa superba's seeds and corms. In recent years it has been studied for its antimitotic activity[63]. This activity is related to the formation of a colchicine–tubulin complex, which prevents microtubule polymerization due to a conformational inflexibility making tubulin dimers incompetent for microtubule assembly. Therefore, the cells exposed to it tend to undergo mitotic arrest during the cell cycle, followed by apoptosis.

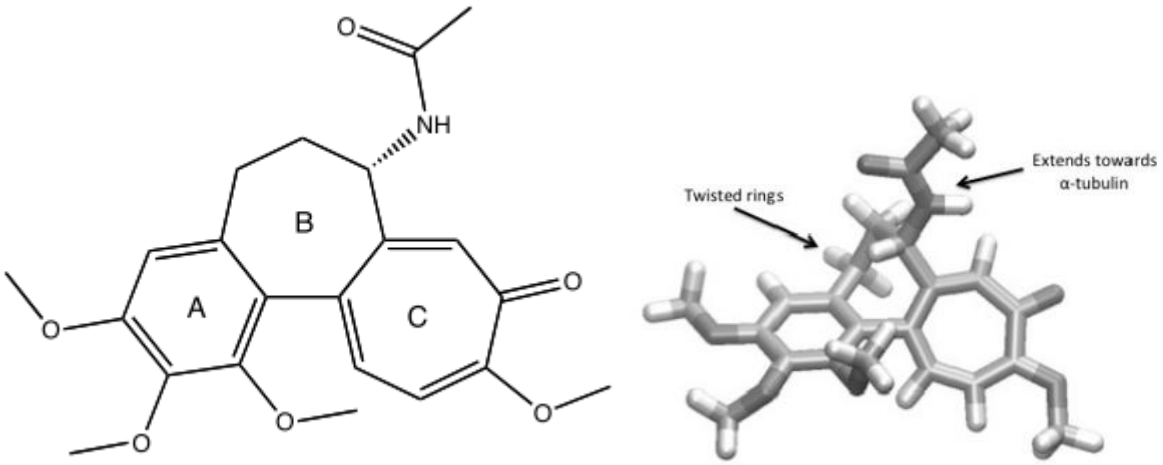


Figure 12: Colchicine 2D and 3D structures.

Cancer cells are more susceptible to colchicine poisoning than normal cells, as they undergo mitosis at higher rate. However, the colchicine is very toxic for the organism (especially for the gastrointestinal tract)[51]. In high doses, it can also cause hepatocellular failure, marrow failure, disseminated intravascular coagulation and late central-nervous-system dysfunction, among other effects. To overcome this problem, many colchicine derivatives having higher power and lower toxicity have been designed and synthetized over the years.

The most relevant structures belonging to the colchicine binding site are S9 (349-354) and S8 (310-318) sheets, H7 (223-241) and H8 (250–258) helices, and T7 loop (242-249) of the β-tubulin, and αT5 loop (173-181) of the α-tubulin.[64]

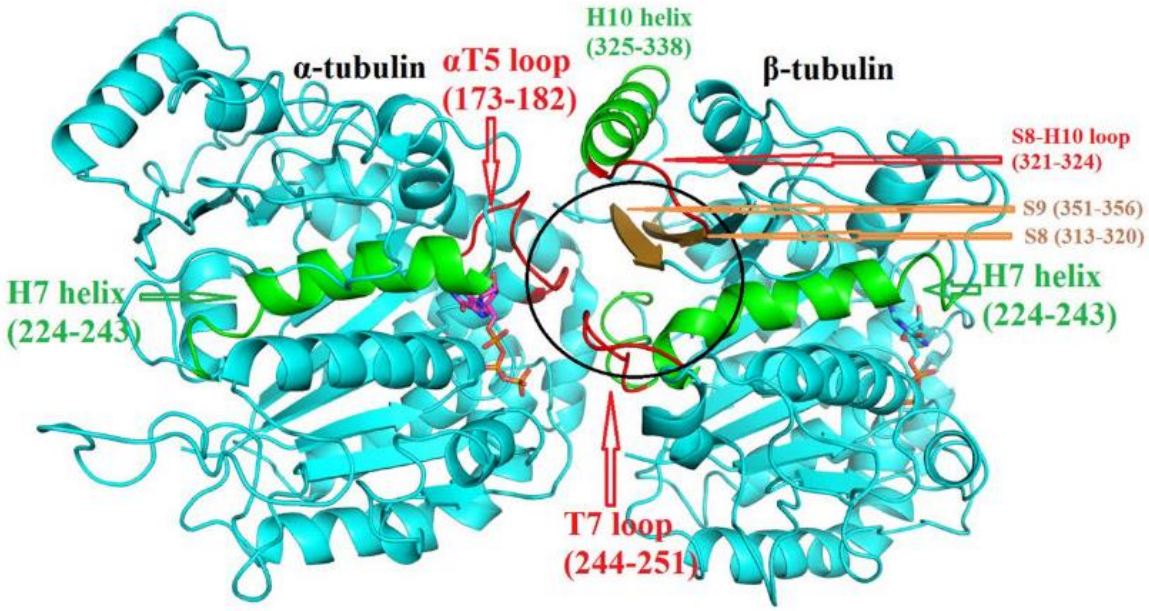


Figure 13: Colchicine binding site[64]*. The important secondary structures near CBS are named with respective residue numbers.*

In this study, colchicine site has been chosen among the others for several reasons. First, colchicine is a drug with a long clinical history[65] and its chemistry is simpler that of other tubulin-binding-drugs, so it is easier to design novel derivatives[66]. Furthermore, colchicine has strong antimitotic-activity which slightly differs among different tubulin isotypes[51]. In particular, it would be interesting to synthetize a colchicine derivative that interacts specifically on tubulin isotype βIII, as it is overexpressed in tumoral cells and seems to be linked to tumour resistance to various drugs[67].

4. Comparative study over novel colchicine derivatives on human tubulin isotypes

In this chapter a comparative study based on molecular dynamics and molecular docking over novel colchicine derivatives on human tubulin dimers αβIIA, αβIII and αβIVA is presented. The results provide an investigation of drugs binding energies, their positions in the cleft and their specificity.

Abstract

Microtubules are key players in cell mitosis. For this reason, they are an interesting target of pharmacological treatment, as it is possible to block the mitotic process by altering the dynamic equilibrium between microtubule's polymerization and depolymerization. From this perspective, it is interesting to design and develop antitumoral drugs able to selectively bind the tubulin in the tumoral cells, in order to lead them to apoptosis. There are several tubulin isotypes: in particular, tumoral cells are characterized by an overexpression of αβIII tubulin. The affinity between tubulin and drugs can vary sensibly in relation to the particular isotypes expressed. This becomes more evident if the mutations are located in the drug's binding site. Therefore, a drug showing a higher affinity with αβIII isotype, compared to the others, can be considered able to specifically target this dimer and can be assumed to be an efficient drug to target tumoral cells. In this study, Molecular Dynamics (MD) simulations and Ensemble Docking were carried out for characterizing and comparing the effect of 61 colchicine derivative compounds on different human tubulin isotypes. The outcome of this work shows the differences among the compounds on the different human tubulins and shows that one of these drugs has the potential to be an effective anticancer drug, as it shows specificity for the isotype αβIII.

4.1 Introduction

Nowadays, finding a cure for cancer is an open ambitious challenge. Tumoral cells are characterized by a higher rate of mitotic process compared to nontumoral cells, so the use of drugs able to arrest the mitosis during the cell cycle is an established strategy to fight the cancer. In order to do that, the drugs have to be able to alter microtubule dynamic instability by enhancing the depolymerization or the polymerization process. Tubulin isotypes are found to contribute in the development of drug resistance cell lines, as they show differential binding energy to various anticancer drugs[68]. Therefore, an overexpression of a particular isotype can

lead the tumoral cells to become drug-resistant. In particular, the overexpression of isotype βIII seems to be linked to the tumour resistance[67]. Thus, the aim is to find a drug that is effective on isotype αβIII, as targeting it selectively can be a good strategy to fight the tumor. In this study, 61 colchicine derivatives have been tested over different human tubulin isotypes. Colchicine is microtubule depolymerizing agent, with the side effect of being toxic for the organism; thus, the purpose is to design a derivative which is a good microtubule destabilizer, and which is less poisonous for the organism. Basing on the mutations in the colchicine binding site, tubulin isotypes can be grouped in three clusters[69]; thus, in this work only isotypes αβIIA, αβIVA, αβIII have been used. These three isotypes show important differences in their sequences and, in particular, in the colchicine binding site: isotype αβIII has three mutations, Cys239 to Ser, Ala315 to Thr and Thr351 to Val, while isotype αβIIA contains only one modification, such as Val316 to Ile. Isotype αβIVA has no modifications in the residue composition of the colchicine binding site. Computational approaches can be a powerful tool able to yield an insight into molecular level and provide detailed studies of molecular mechanisms, like discovering the position of the drug in the binding site, its affinity to the protein or its destabilizing power. In this study, the effect of the colchicine and its derivatives mentioned above have been confronted over the different human tubulin isotypes, using computational tools as Molecular Dynamics (MD) simulation, Ensemble Docking and binding free energy calculation.

4.2 Materials and Methods

4.2.1 Atomic models of investigated compounds

61 colchicine derivative compounds were considered in this work. All compounds have proved in vitro antiproliferative effect on normal and cancer cells: in particular, they were tested on human lung adenocarcinoma (A549), which is a cell line displaying various levels of drug resistance, human breast adenocarcinoma (MCF-7), human colon adenocarcinoma cell line (LoVo) and doxorubicin-resistant subline (LoVo/DX). For better evaluation of the cytotoxicity, their effect was also studied on normal murine embryonic fibroblast cell line (BALB/3T3). These compounds can be divided in five classes: 4-Br-Amides (10 compounds), 4-Cl-Amides (10 compounds), DT-and-4I-Amides (19 compounds), 4-Cl-Carbamates (11 compounds) and 4-I-Carbamates (11 compounds). In addition to these compounds, the colchicine was used in every phase of this work as a control. Starting from now, the colchicine will be called C01, while the other compounds will be called as follows:

Table 2: List of the compounds used in this study

4-Br-Amides	C02-C11
4-Cl-Amides	C12-C21
DT-and-4I-Amides	C22-C40
4-Cl-Carbamates	C41-C51
4-I-Carbamates	C52-C62

The 2D structures of the above-mentioned colchicine-derivative compounds employed in this work are reported in Supporting Information. In order to perform molecular docking, their 3D structure was designed using the software AVOGADRO[24]. Ligand topologies were defined using Antechamber program[70], with general amber force field and bcc charge method, and ACPYPE[71].

4.2.2 Homology modelling of tubulin isotypes

The atomic structures of human tubulin isotypes are not present in RCSB Protein Data Bank-PDB, so they have been created by using homology modelling. To create the template structure, the Protein Data Bank entry 4O2B[39] was downloaded. This structure was chosen among the others because it has the colchicine already bounded in the cleft and because it is more recent with a higher resolution (2.3 Å) compared to other structures with the bounded colchicine, such as 1SA0[48]. The structure downloaded from Protein Data Bank presents more than one dimer, so only the dimer composed by protein chains A (α tubulin) and B (β tubulin), along with Mg^{2+} and GTP linked to chain A and GDP and colchicine linked to chain B has been considered, as they present fewer missing atoms and residues compared to the other chains in the PDB file. Missing atoms were filled using the Swiss PDB Viewer software, while missing residues were filled using the software MODELLER[72] 9.20. The best model was selected on the basis of DOPE (Discrete optimized protein energy) score[73] and was used as the template structure to build the human isotypes. The template structure was validated using various software as PROCHECK, ERRAT and VERIFY3D.

Human isotypes FASTA sequences have been downloaded from the UNIPROT database[74] and have been used to build the human tubulin on the template by using MODELLER 9.20. In particular, for the chain α it was chosen the sequence Q71U36, for the βIIa the Q13885, for the βIII the Q13509 and for the βIVa the P04350. Only β tubulin isotypes have been considered

because the study is focused on the interaction between tubulin and the colchicine derivatives. α-tubulin residues from 440 and β-tubulin residues from 428 were cut, as they represent the C-terminus, which is a region characterized by high fluctuations. For each isotype, the best model was chosen on the basis of the DOPE score. Human tubulin dimers were then evaluated by using the same methods used for template validation. GDP, GTP and Mg^{2+} and the colchicine were added from the starting template (4O2B).

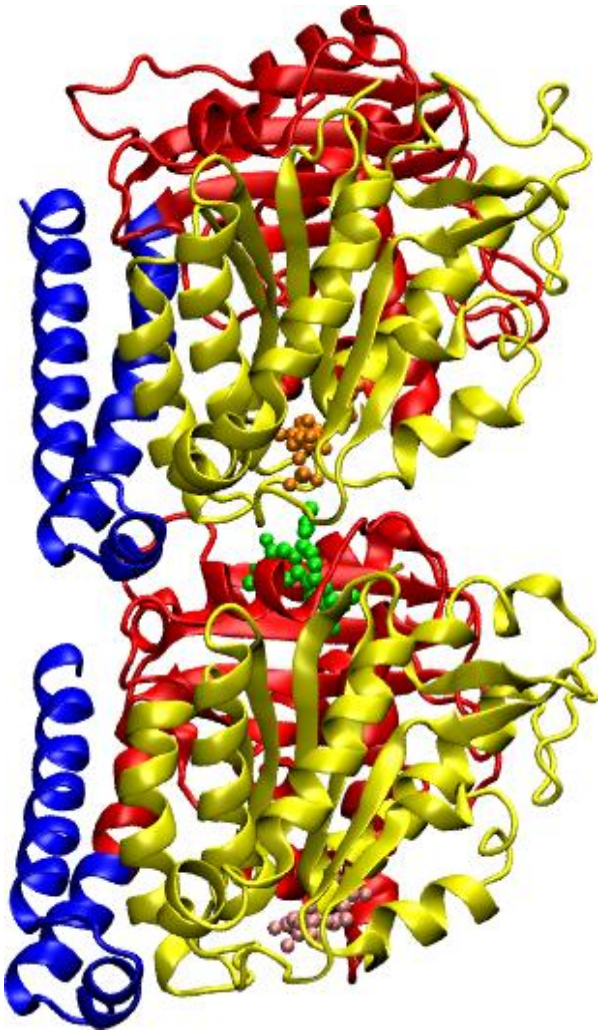


Figure 14: Representation of a simulated structure: tubulin dimer bound to GTP (orange), GDP (pink) and colchicine (green). Tubulin domain are highlighted in different colors: in yellow the Rossman fold (1-205), in red the intermediate domain (206-381) and in blue the C-terminal domain (382-440/427)

4.2.3 Molecular Dynamics of tubulin isotypes bound to colchicine

Conformational dynamics of the three isotype models with a bound colchicine molecule has been investigated by MD simulations following the procedure described below.

The AMBER-ff99SB-ILDN force-field was used to define protein topology, while Mg^{2+}, GDP, GTP and colchicine topologies were defined using Antechamber program of the AMBER16 package, with general amber force field and bcc charge method, and ACPYPE. A triclinic box with 0.6 nm of minimum distance was used and all systems were solvated using TIP3P explicit water model, with the addition of chlorine and sodium ions to neutralize the charge, maintaining an ionic strength at 150 mM. Periodic boundary conditions were applied along the xyz coordinates. A 1000-step energy minimization by steepest descent was then performed, with a maximum force of 100 $kJmol^{-1}nm^{-1}$. After that, restraining potentials were applied on water molecules and sodium and chlorine ions and system equilibration in NVT and NPT ensembles were performed: NVT equilibrium phase was performed for 100 ps using velocity-rescale thermostat with the tau constant at 0.1 and the reference temperature at 300 K, while the NPT one for 300 ps using Berendsen barostat to maintain the pressure at 1 atm. Electrostatic interactions were treated by means of Particle Mesh Ewald (PME) approach, with 1.0 nm cutoff, Fourier spacing of 0.2 nm and interpolation order of 4. Lennard-Jones interactions were cut-off at a distance of 1 nm. Finally, each system was simulated for at least 200 ns without any restrain with a time step of 2 fs and coordinates saved at every 2 ps.

All the simulations were performed by using the GROMACS 5.1.4 package. The Visual Molecular Dynamics (VMD) package was employed for the visual inspection of the simulated systems. Dedicated GROMACS tools were used for a quantitative analysis in terms of Root-Mean-Square Deviation (RMSD), Radius of Gyration (RG) and clustering, while analysis of secondary structure was performed by applying the STRuctural IDEntification (STRIDE)[75].

4.2.4 Ensemble Docking of investigated compounds on tubulin models

Ensemble docking was performed using AUTODOCK VINA 1.1 on the three human tubulin isotypes. The centre of the search space has been fixed by calculating the centre of mass of the colchicine bonded to the original dimer downloaded from the Protein Data Bank database; then the colchicine has been removed and the docking has been performed using a grid space 20*20*20 nm around the centre. An exhaustiveness = 64 has been imposed. Each compound was docked to ten different configurations of each isotype: these configurations have been extracted from the last 50 ns of the MD simulation, one every 5 ns. For each docking, the best ligand pose was extracted. Mean and standard deviation were calculated over the best poses of each ligand on all the ten tubulin configurations, for each isotype. Furthermore, it has been analysed how much the drugs positions in the cleft differ to the colchicine's one. Before doing

that, a MD simulation of 1 ns was performed on each structure, after solvation, neutralization and equilibration steps in the same way described in the previous section. The complexes used in this phase were the best poses of each ligand in every isotype.

4.2.5 Binding free energy calculation

The binding free energies between proteins and ligands were calculated by using the molecular mechanics Generalized Born surface area (MMGBSA) theory. The free energy was thus calculated with the software AMBER16. This calculation was performed using the trajectory frames from the 1 ns MD simulations mentioned above and igb=8. The parameters were set according to the literature[76–78]. The free energy was decomposed on every residue, in order to analyse the energy on each residue in the colchicine binding site. This analysis has been performed after clustering the drugs with the dedicate GROMACS tool.

4.3 Results

4.3.1 Validation of Homology Models

The DOPE scores for the template and the three isotypes are very similar and around -11000.

The PROCHECK analysis shows that more than the 95.4% of the residues are in most favored regions in the Ramachandran plot for all isotypes and no residues are in the disallowed regions, except for the template structure that exhibit 0.1% of the residues in those regions. A good quality model is expected to have at least the 90% of the residues in the most favoured regions[79]. Therefore, this analysis proves that all the built models are good ones.

The Overall Quality Factor obtained whit the ERRAT tool for the template model is 86.68 for chain α and 91.17 for chain β. For the isotype αβIIA these values are respectively 79.12 and 88.31, for the isotype αβIII 80.29 and 84.73 and for the isotype αβIVA 77.26 and 84.96. The generally accepted range is >50 for a high quality model[80], so all the models passed the test.

In the end, the VERIFY3D test has been performed. This test is considered passed if a high percentage of the residues has an averaged 3D-1D score higher than 0.2[80]. All models pass the VERIFY3D.

More details can be found in the Supporting Information.

4.3.2 Conformational dynamics of tubulin isotypes

The analysis on molecular dynamics simulations was performed to check the stability of the structures. First, the RMSD of the backbone atoms was calculated for the template structure and for each isotype during the 200 ns of simulation. As it can be seen in the following picture, all the simulated structures reach the convergence, whit values under 0.4 nm.

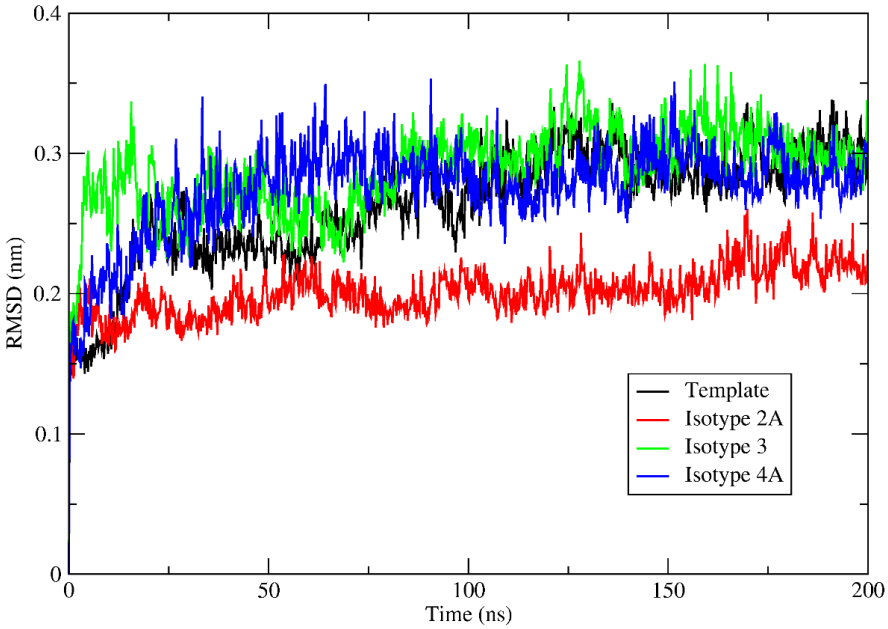


Figure 15: RMSD values for each simulated structure: all the structures show to have reached the convergence. The lowest value is for isotype αβIIA (around 0.25 nm), while the others are higher, but always under 0.4 nm.

Then, the secondary structure modifications in the colchicine binding site have been investigated. The most relevant structures belonging to the colchicine binding site, are S9 (349-354) and S8 (310-318) sheets, H7 (223-241) and H8 (250–258) helices, and T7 loop (242-249) of the β-tubulin, and αT5 loop (173-181) of the α-tubulin; these structures where analyzed with STRIDE using the last 50 ns from simulation trajectories for the three isotypes.

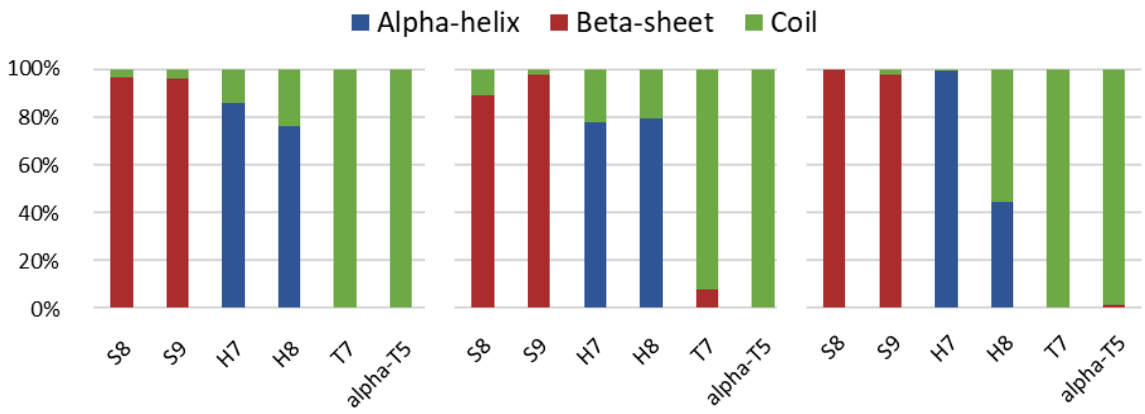


Figure 16: Secondary structure probability in the colchicine binding site during the last 50 ns of simulation for isotypes a) αβIIA, b) αβIII and c) αβIVA with colchicine. The structures seem stable during the simulations.

The structures seem to be all stable, as they don't change very much during simulations.

After that, the Radius of Gyration has been analyzed. It is a measure of protein structure compactness, so a low and stable value of radius of gyration indicates a stable protein, which does not unfold during the simulation.

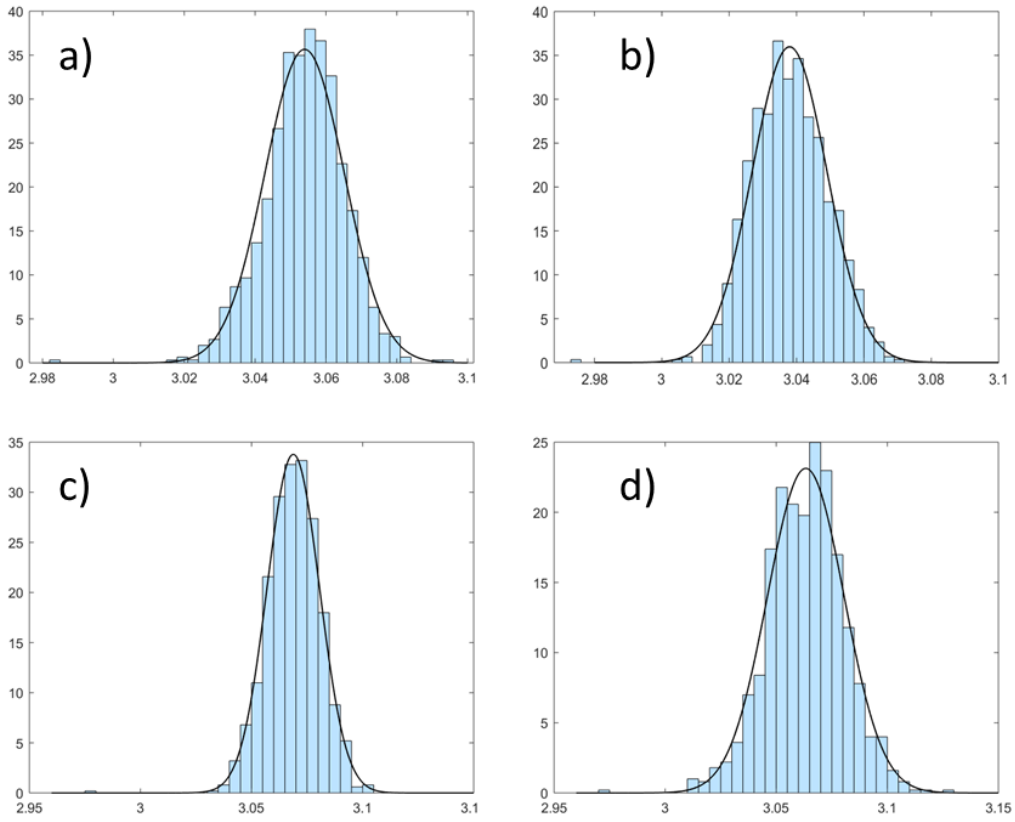


Figure 17: Radius of Gyration distribution for the template (a), the isotype αβIIA (b), the isotype αβIII (c) and the isotype αβIVA (d). They all have a gaussian distribution around 3 nm, indicating their stability.

The values are around 3 nm with a gaussian distribution in every simulated structure, indicating that they are all stable.

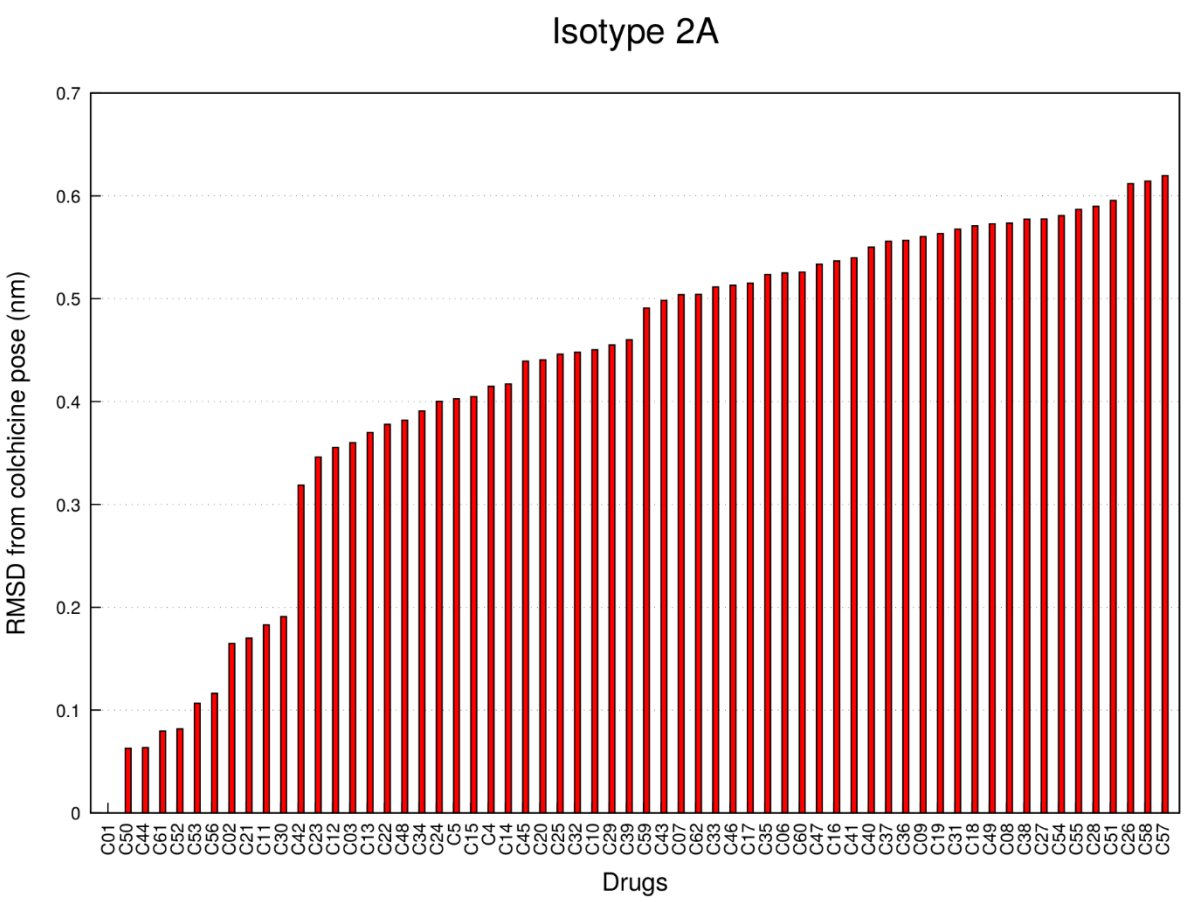
Last, the cluster GROMACS tool has been used on the last 50 ns of simulations, with a 0.15 nm cutoff. All the frames resulted grouped in only one cluster, indicating the stability of the structures.

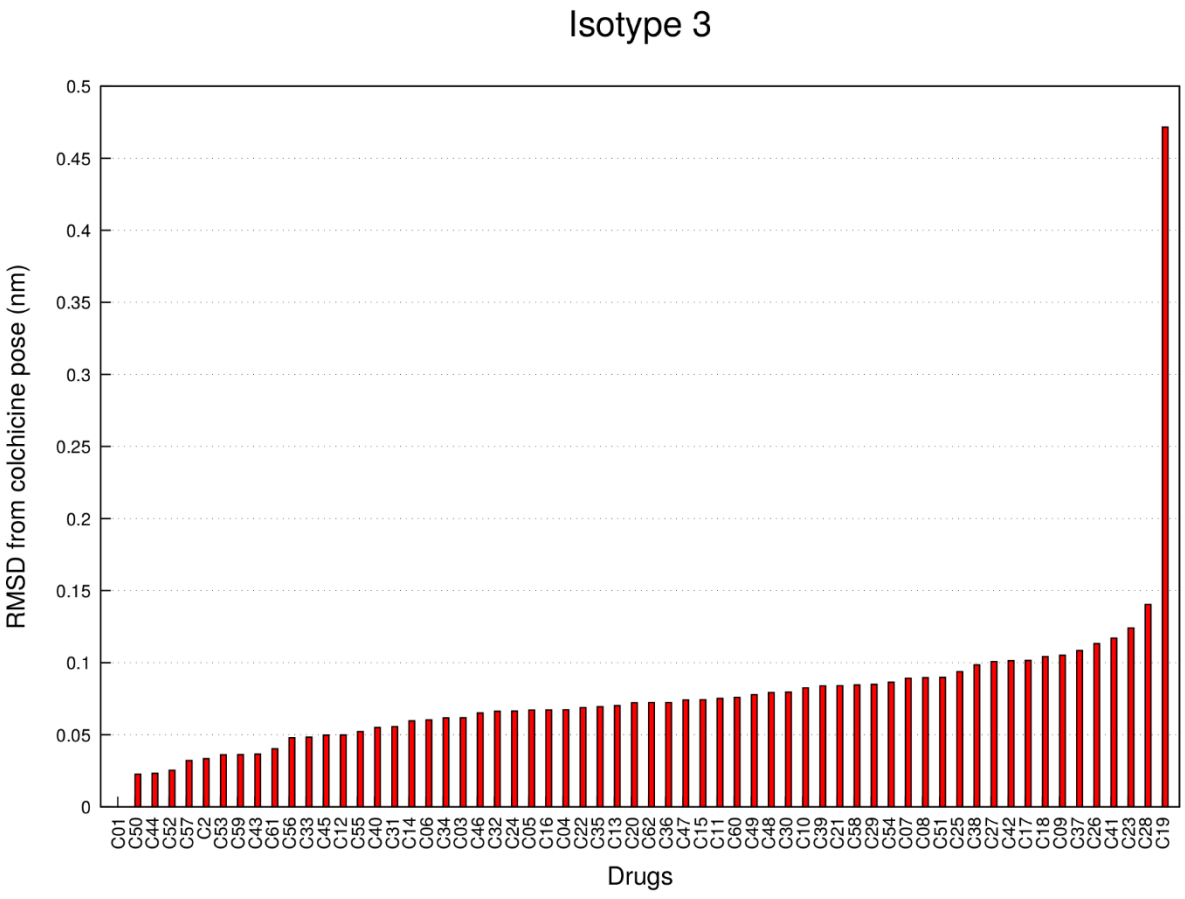
4.3.3 Ensemble Docking

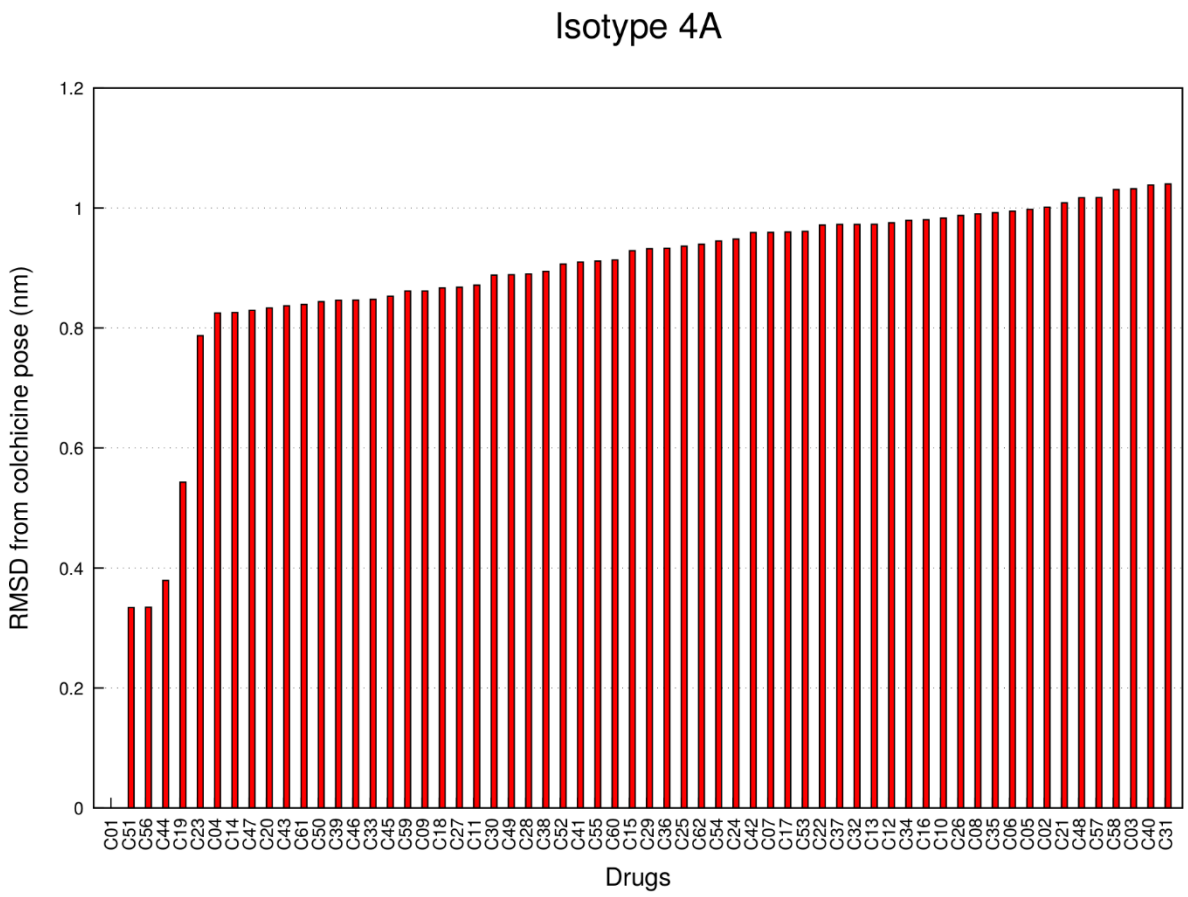
The affinities calculated by the ensemble docking show that all the analyzed compounds are binders. There are no significant differences in these affinity values, either confronting the compounds intra the same isotype and confronting them inter the different isotypes. The values are between -7.5 and -9.5 kcal*mol^{-1}. Details can be found in Supporting Information.

To perform a more accurate analysis, the positions of the ligands in the cleft have been analyzed, using the structures with the best pose of each ligand. The positions of the ligands were confronted to the colchicine's one, in order to investigate the presence of some differences. The compounds are all colchicine derivatives, so they have a zone in the structure, which is in common with the colchicine. This zone is the one composed by the three structural-rings, which will be referred as the "head" of the drug, while the variable regions will be called the drug's "tails".

Each of the protein structures containing the drugs was aligned to the one bonded to the colchicine for every isotype (reference structure), using the GROMACS tool *confrms*; only the protein structures were aligned, in order to analyze if the ligands are in the same positions or not. The trajectories were aligned to the reference structure using the GROMACS tool *trjconv.* To analyze the difference between the drugs positions, the GROMACS tool *rms* has been used: only the "heads" of the drugs have been considered, since they are the common regions. A low RMSD value means that the drug is approximatively in the same region and the same position of the colchicine, while a high value indicates that it is in a different position confronted to the colchicine and a very high value indicates that it is in a different region of the cleft. The results of this analysis can be seen in the following figures.


Isotype 2A
RMSD from colchicine pose (nm)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
C01 C50 C44 C61 C52 C53 C56 C02 C21 C11 C30 C42 C23 C12 C03 C13 C22 C48 C34 C24 C5 C15 C4 C14 C45 C20 C25 C32 C10 C29 C39 C59 C43 C07 C62 C33 C46 C17 C35 C06 C60 C47 C16 C41 C40 C37 C36 C09 C19 C31 C18 C49 C08 C38 C27 C54 C55 C28 C51 C26 C58 C57
Drugs


Isotype 3
RMSD from colchicine pose (nm)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C01 C50 C44 C52 C57 C2 C53 C59 C43 C61 C56 C33 C45 C12 C55 C40 C31 C14 C06 C34 C03 C46 C32 C24 C05 C16 C04 C22 C35 C13 C20 C62 C36 C47 C15 C11 C60 C49 C48 C30 C10 C39 C21 C58 C29 C54 C07 C08 C51 C25 C38 C27 C42 C17 C18 C09 C37 C26 C41 C23 C28 C19
Drugs



Figures 18-19-20: Ranking of the RMSD distance from colchicine pose for isotypes 2A, 3 and 4A. Overall, the highest values are for isotype 4A, followed by isotype 2A and isotype 3, which has the lowest values.

Looking at the figures, it is evident that isotype 3 shows the lowest values of RMSD (under 0.2 nm) of RMSD, indicating that every drug (except for the C19) is inside the cleft in a very similar position to the colchicine's one. In isotype 2A, many drugs show intermediate values of RMSD (between 0.2 nm and 0.7 nm), while in isotype 4A the majority of drugs has a RMSD near 1 nm. To see how different the poses are, on the basis of these RMSD values, VMD package was used to visualize the drugs in the clefts. The differences can be seen in the following figures.

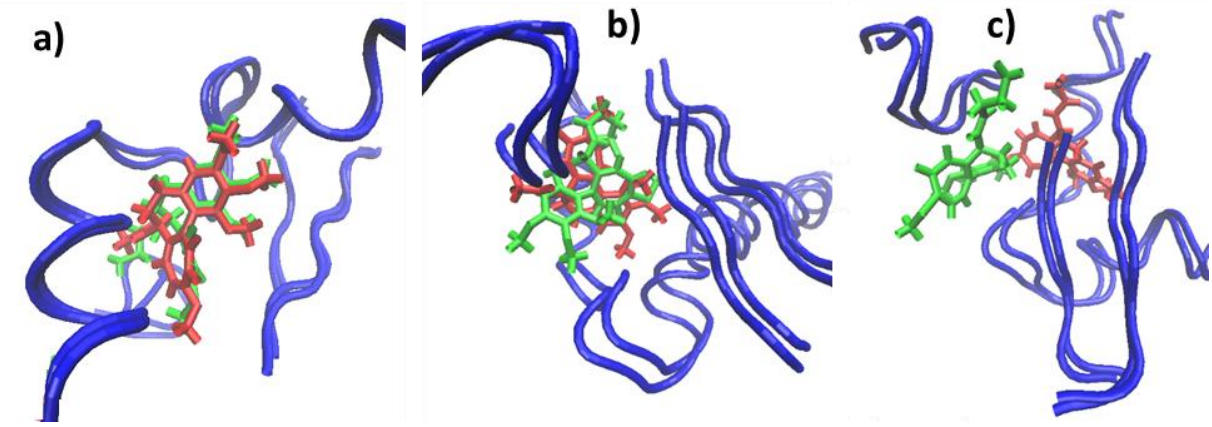


Figure 21: Different ligands poses (red) compared to the colchicine pose (green).

In all the pictures, the colchicine is represented in green, while the other drug is colored in red. Figure 21-A) is referred to a drug that have a RMSD value under 0.2 nm confronted to the colchicine position. It is clear that it means that the drug is almost in the same position of the colchicine. This is what happens for some drugs in isotype 2A and for all the drugs, except the C19, in the isotype 3. Figure 21-B) is referred to a drug that have a RMSD value over 0.4 nm, but under than 0.8 nm. Compared to the colchicine, these drugs are in the same spatial region in the cleft, but are not in the same position, as they are rotated. This behavior is observed in many drugs in isotype 2A, in C19 in isotype 3 and in a few drugs in isotype 4A. Figure 21-C) is referred to a drug that have a RMSD value near 1 nm. These drugs are in a different region of the cleft. This is how the majority of the drugs behaves in isotype 4A.

4.3.4 Binding free energy calculation

The binding free energy calculation results show values between -225 kcal/mol and -40 kcal/mol. Isotype αβIII shows, on average, the highest values, while isotype αβIVA shows the lowest ones, on average. Details about the binding energy of each compound for all the isotypes can be found in the Supporting Information.

In order to investigate the specificity of the drugs, the binding free energies of each compound were averaged over the three isotypes. Standard deviations were calculated as well. Drugs with a high mean binding energy and a low standard deviation can be considered good to target all the isotypes and so they are non-specific, while drugs with a high mean binding energy and a high standard deviation are probably specific for one or two isotypes.

In the following graph, these results are summarized. The colour patterns are chosen in relation to the colchicine's energy value, which is set equal to 100%.

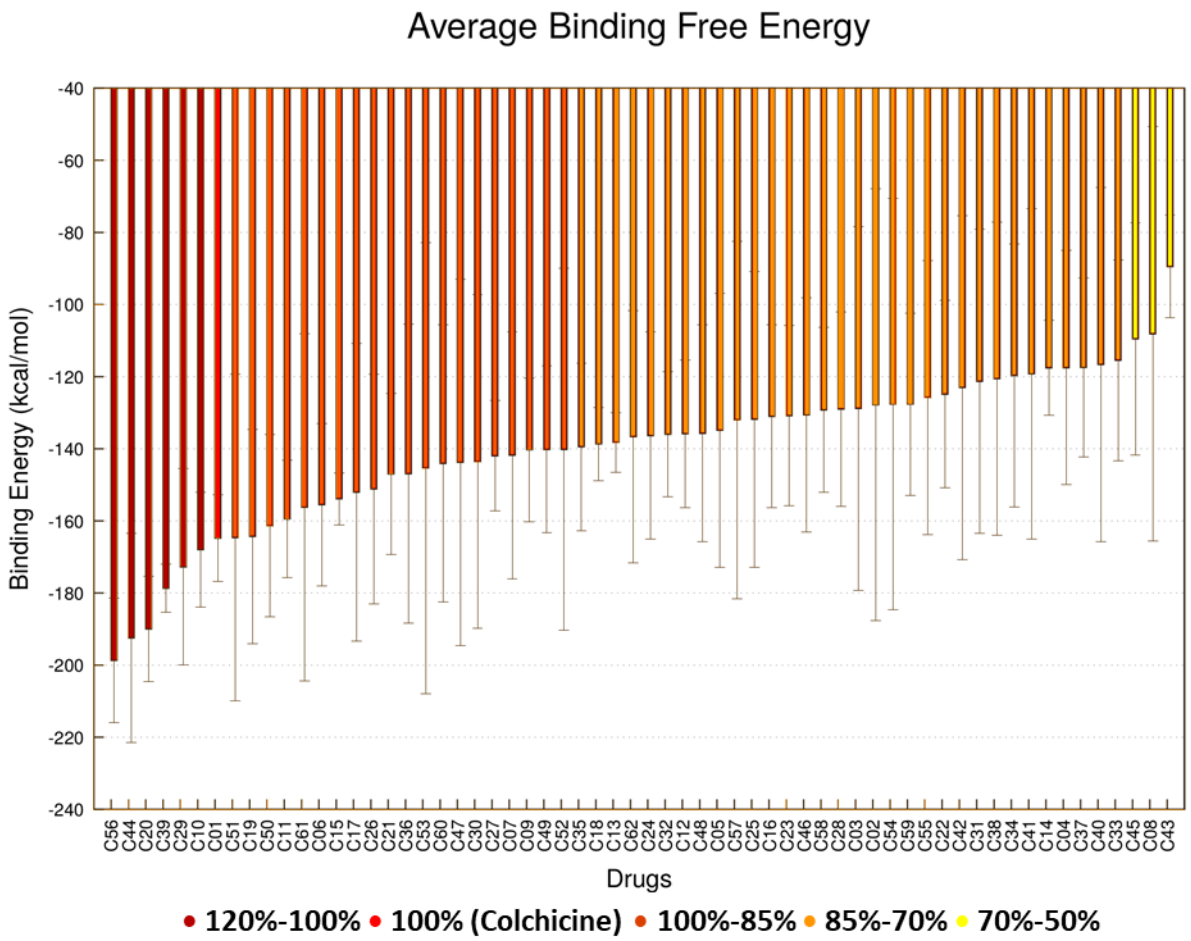


Figure 22: Average binding free energy. The colour patterns show how much the average energy values of the drugs differ from the colchicine's one. Only few drugs have a higher energy and show low values of standard deviation, indicating that they are non-specific.

Colchicine is represented in red and shows a higher average binding energy compared to the majority of the other compounds. It has a very low standard deviation, showing that it is a non-specific drug. The drugs C56, C44, C20, C39, C29 and C10 show higher binding energies than the colchicine and quite low standard deviation values, so they can be considered non-specific drugs, as they seem able to bind the three isotypes with the same affinities.

To better understand and investigate the specificity of the other drugs, it is more useful to plot only their standard deviations, as they are the most significant indicator of the specificity. The values are normalized on the colchicine's value, which is set equal to 1.

The other values are calculated as follows:

$$Std_{drug}^{normal} = \frac{Std_{colchicine}}{Std_{drug}}$$

The results can be seen in the following figure.

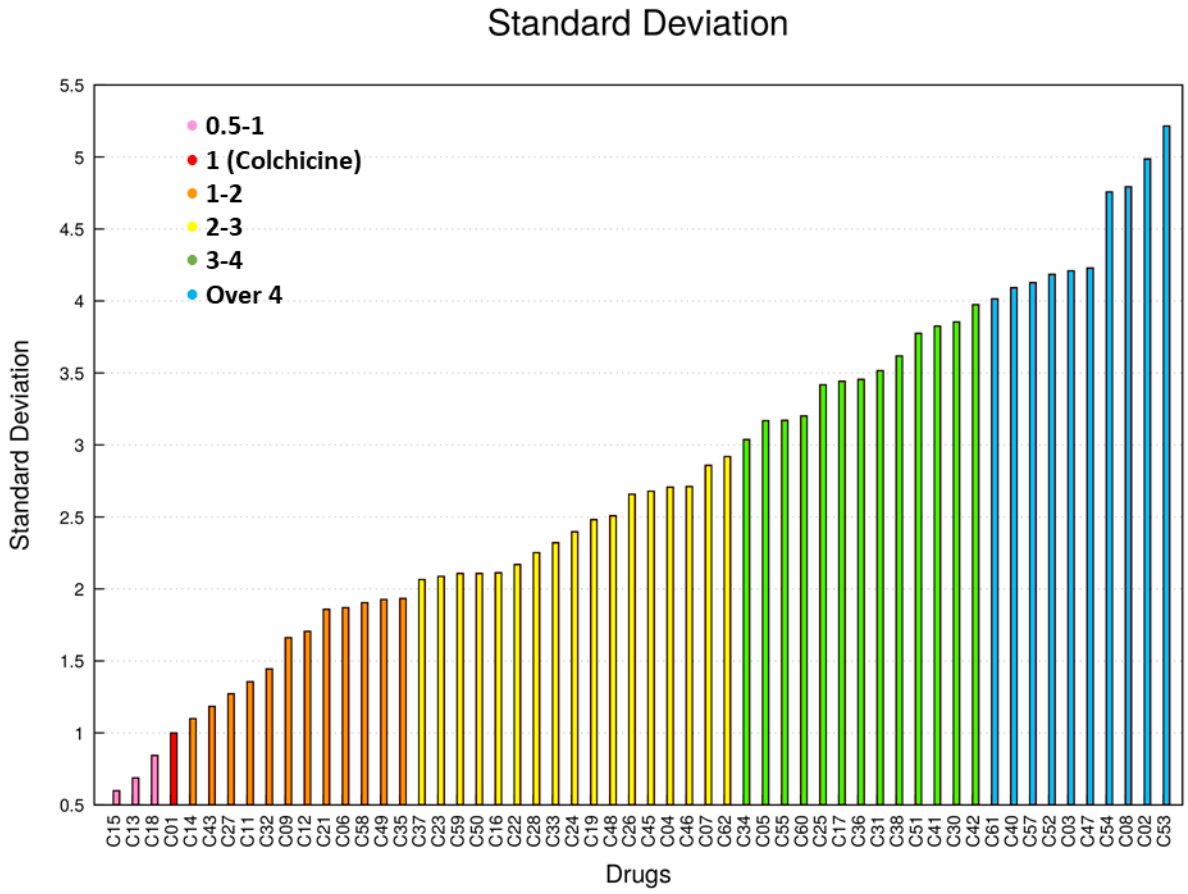


Figure 23: Standard deviations of the drugs binding energies over the three isotypes. The values are normalized on the colchicine's one, which is set equal to 1. The colour pattern shows how much the standard deviations differ from the colchicine's one. Only few drugs have a lower value, while the majority of them shows values from 2 to 5 time higher.

Looking at the figure, it is clear that C15, C13 and C18 are non-specific, as they show a lower standard deviation than the colchicine, which is a non-specific drug. Drugs showing values between 1 and 2 can also be considered non-specific, as the colchicine's standard deviation value is very low, so the double can be considered low, too. Thus, the compounds coloured in orange, C14, C43, C27, C11, C32, C09, C12, C21, C06, C58, C49 and C35 can be considered non-specific. The other compounds are probably specific for one or two isotypes. To better understand the specificity, it useful to plot the binding energies of the remaining compounds over the isotypes. The values are normalized on the binding energy of the isotype αβIIA, which is set equal to 1. The results can be seen in the following figure.

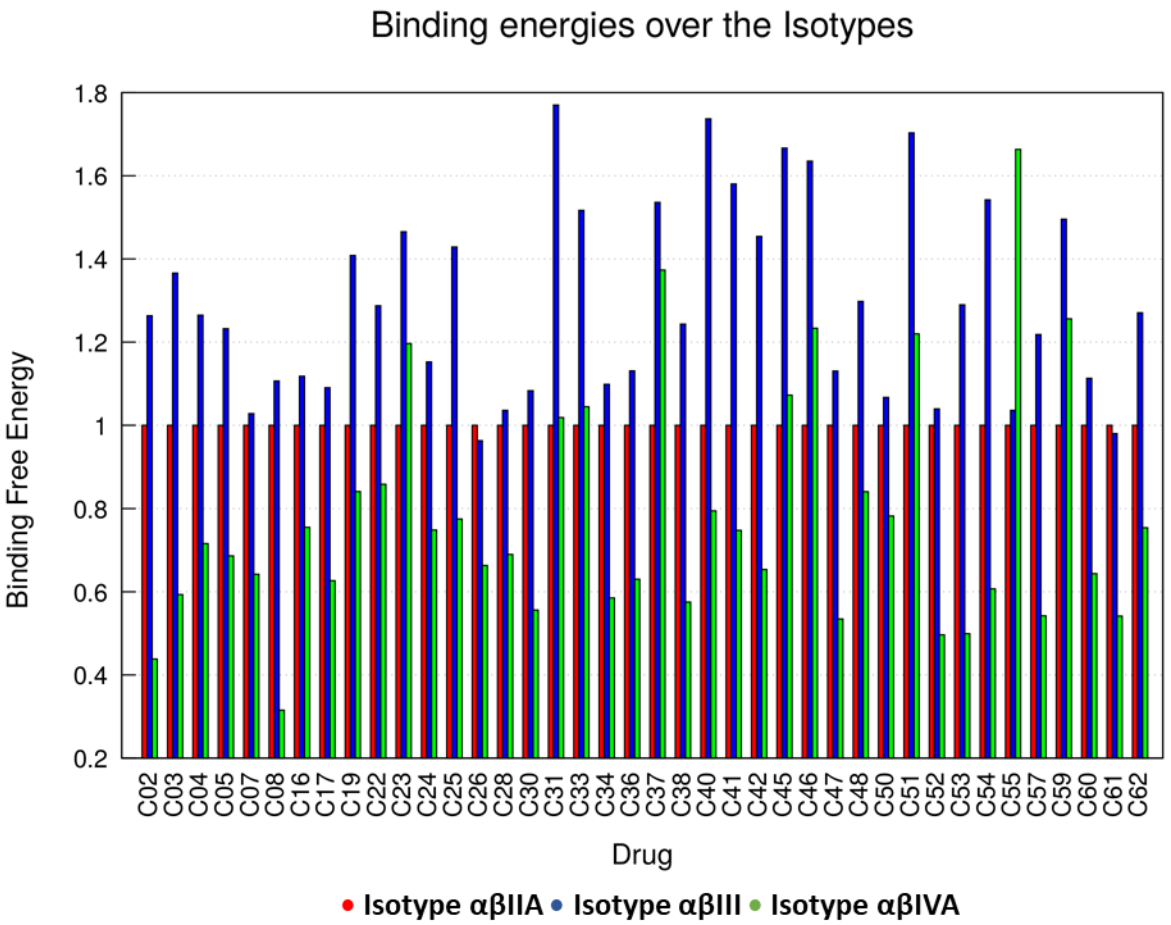


Figure 24: Binding free energies over the isotypes for the identified specific drugs. The values are normalized on the energy value of the isotype αβIIA for each drug. The colours indicate the different isotypes.

Looking at the figure, it is possible to identify the specific compounds for each isotype. In particular:

- Compounds C02, C03, C04, C05, C19, C22, C23, C25, C31, C33, C40, C41, C45, C46, C48, C51, C53, C54, C57, C59 and C62 seem to be specific for isotype αβIII.
- Compound C55 seems to be specific for isotype αβIVA.
- Compound C37 seems to be specific for isotypes αβIII and αβIVA.
- Compounds C07, C08, C16, C17, C24, C26, C28, C30, C34, C36, C38, C47, C50, C52, C60 and C61 seem to be specific for isotypes αβIII and αβIIA.
- There are no compounds specific only for isotype αβIIA.

It is interesting to see if the drugs with the highest binding energies for each isotype are specific or not. The best five drugs, in terms of binding energy, for each isotype can be found in the following table:

Table 3: Drugs with the highest binding energies for each tubulin isotype.

Best Drugs (Binding energy)					
Isotype 2A		Isotype 3		Isotype 4A	
Drug	*Binding energy*	*Drug*	*Binding energy*	*Drug*	*Binding energy*
C44	-225 kJ/mol	C51	-214 kJ/mol	C20	-197 kJ/mol
C56	-218 kJ/mol	C53	-202 kJ/mol	C56	-191 kJ/mol
C61	-186 kJ/mol	C20	-200 kJ/mol	C39	-183 kJ/mol
C29	-178 kJ/mol	C29	-197 kJ/mol	C44	-170 kJ/mol
C10	-175 kJ/mol	C19	-195 kJ/mol	C55	-169 kJ/mol

For what concerns isotype αβIIA, there are no specific drugs. Among the best five, C44, C56, C29 and C10 are non-specific at all, while C61 is specific for both αβIIA and αβIII. Among the best drugs of isotype αβIII, C51, C53 and C19 are specific, while C20 and C29 are non-specific. In the end, for what concerns isotype αβIVA, C20, C56, C39 and C44 are non-specific, while C55 is specific for it.

After this analysis, all the drugs were clustered on the basis of the RMSD using the dedicated GROMACS tool. The RMSD cutoff has been set to 0.035 nm. In all the isotypes, the drugs are clustered in five clusters which are quite similar among the different structures. The colchicine is always included in cluster 1. The following tables summarize the principal information about each cluster: the number of drugs contained and the centroid structure. Details about the drugs included in each cluster can be found in the Supporting Information.

Table 4: Number of drugs and centroid of each cluster.

Clusters								
Isotype 2A			Isotype 3			Isotype 4A		
Cluster	*Drugs number*	*Centroid*	*Cluster*	*Drugs number*	*Centroid*	*Cluster*	*Drugs number*	*Centroid*
1	41	C22	1	42	C16	1	38	C5
2	13	C27	2	8	C19	2	15	C18
3	1	C41	3	3	C51	3	3	C42
4	3	C42	4	4	C45	4	3	C51
5	4	C45	5	5	C60	5	3	C55

At this point, the results of the binding free energy calculation and the investigation about the poses of the drugs in the cleft have been merged in radar graphs, in order to understand if there is some correlation between them. Each drug is represented by a point, coloured on the basis of the cluster of membership. The distance from the centre represents the RMSD from the colchicine's pose, while the angular coordinate represents the binding energy value.

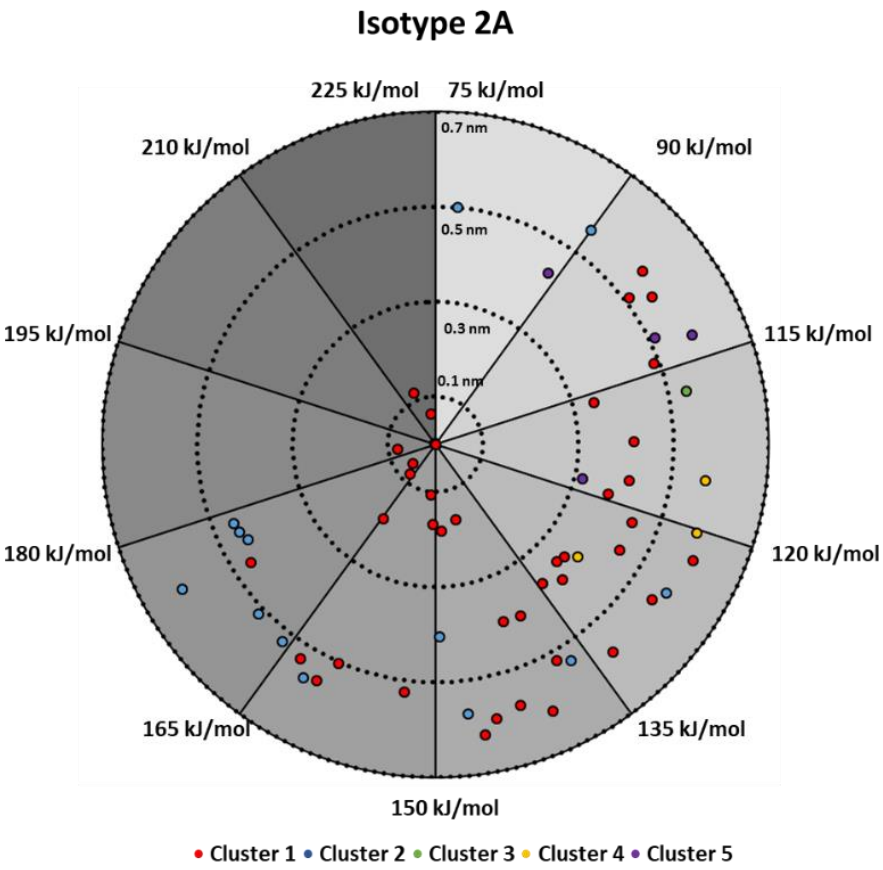


Figure 25: Radar graph for isotype 2A. The drugs are represented by the points and coloured based on the cluster of membership. The drugs with the highest binding energies are the ones with the lowest RMSDs from colchicine's pose.

From the graph, it is clear that the drugs with the highest binding energies have also a low RMSD from the colchicine's pose. They are all included in the cluster 1, which includes the colchicine too, so the most effective drugs are the ones most similar to the colchicine. Drugs in clusters 3, 4 and 5 have low binding energies and high RMSD. The compounds included in cluster 2 have slightly higher binding energies compared to the the other three just mentioned, but high RMSD.

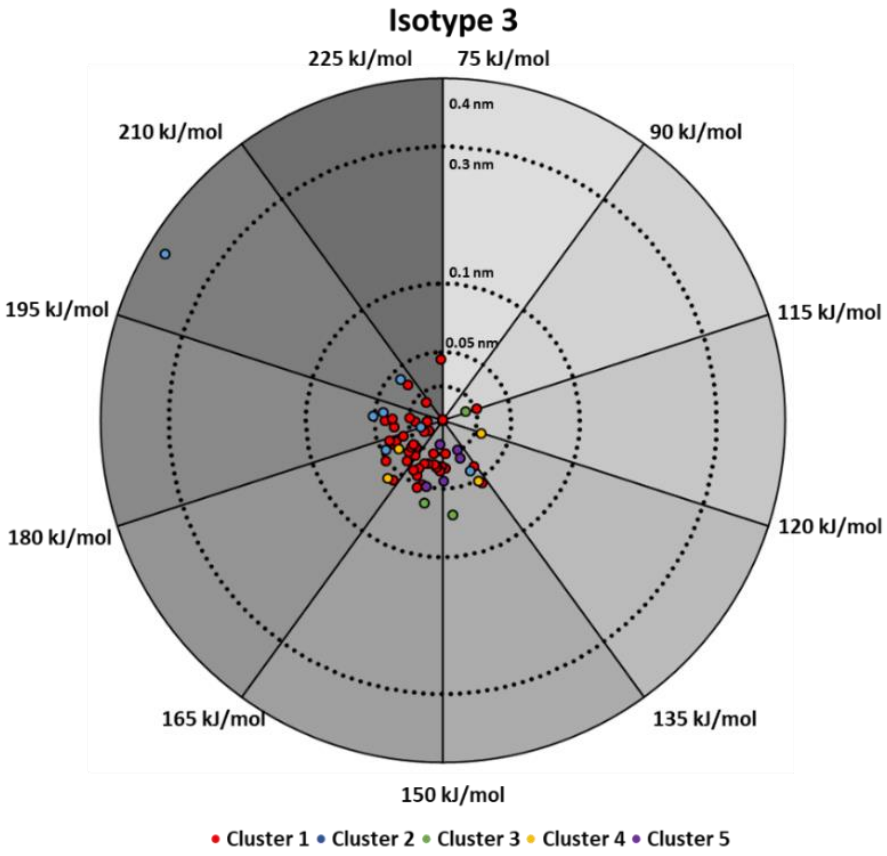


Figure 26: Radar graph for isotype 3. The drugs have all very low RMSDs from colchicine's pose. The only exception is the compound C19, showing also a high binding energy in addition to the high RMSD:

For what concerns the isotype 3 graph, the only exception in terms of RMSD is represented by C19 (the centroid of cluster 2) which is also one of the drugs with the highest binding energy. The other drugs have all very low RMSDs, and can be noticed that drugs in cluster 3,4 and 5 have lower binding energies compared to the ones in clusters 1 and 2.

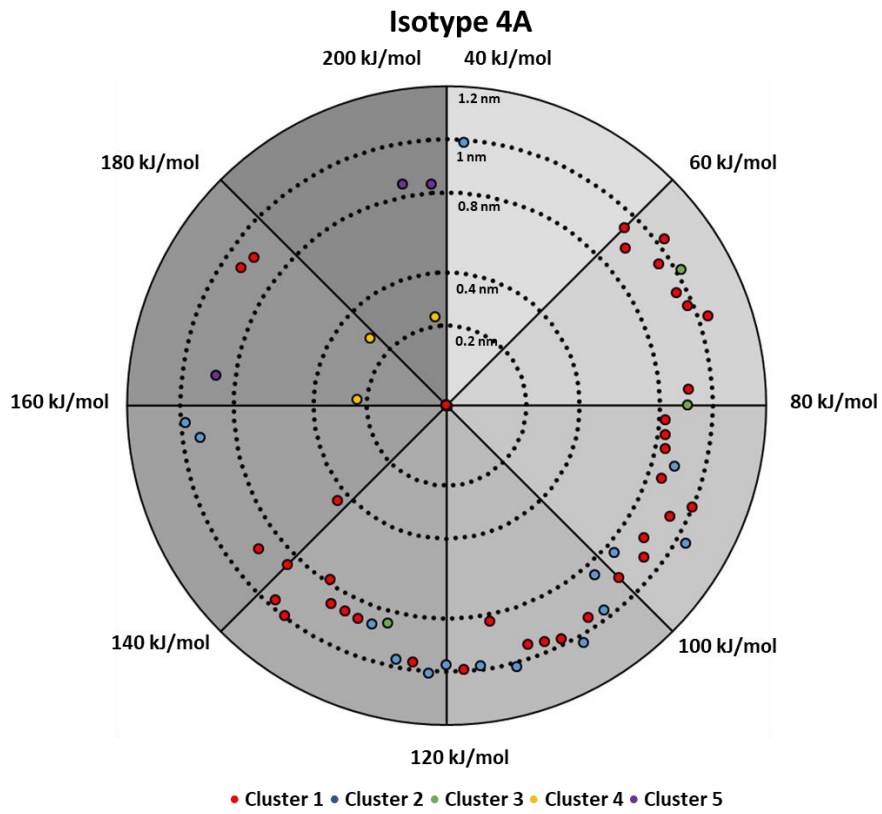


Figure 27: Radar graph for isotype 4A. The highest binding energies belong to the drugs in cluster 4 (that have low RMSDs from colchicine's pose) and in cluster 5 (that show high RMSD values).

In the end, isotype 4A graph shows that the drugs in cluster 4 and 5 are the ones with the highest binding energies, but while the drugs in cluster 4 are the only ones having low RMSDs, the ones in cluster 5 have very high RMSDs compared to the colchicine. The drugs included in clusters 1 and 2 (which represent the majority of the drugs) have lower binding energies compared to them, while the drugs included in cluster 3 seem to have the lowest energy values. The values of the binding energies overall are lower compared to the other two isotypes.

To go deep in the investigation, the binding free energy on each of the residues composing the colchicine binding site has been evaluated. In order to make the graph more intelligible, the energy has been mediated on the drugs of each cluster, so only the mean and the standard deviation for each cluster are shown. Furthermore, only the residues that have high energy and significant differences among the clusters are shown. In particular, among all the three isotypes, residues 178, 179, 239, 247, 256 and 351 are polar, residue 249 is acid (charge -) and residue 252 is basic (charge +), while the others are not polar. For isotype III, the mutations cause differences in residue 351, which is not polar in this isotype.

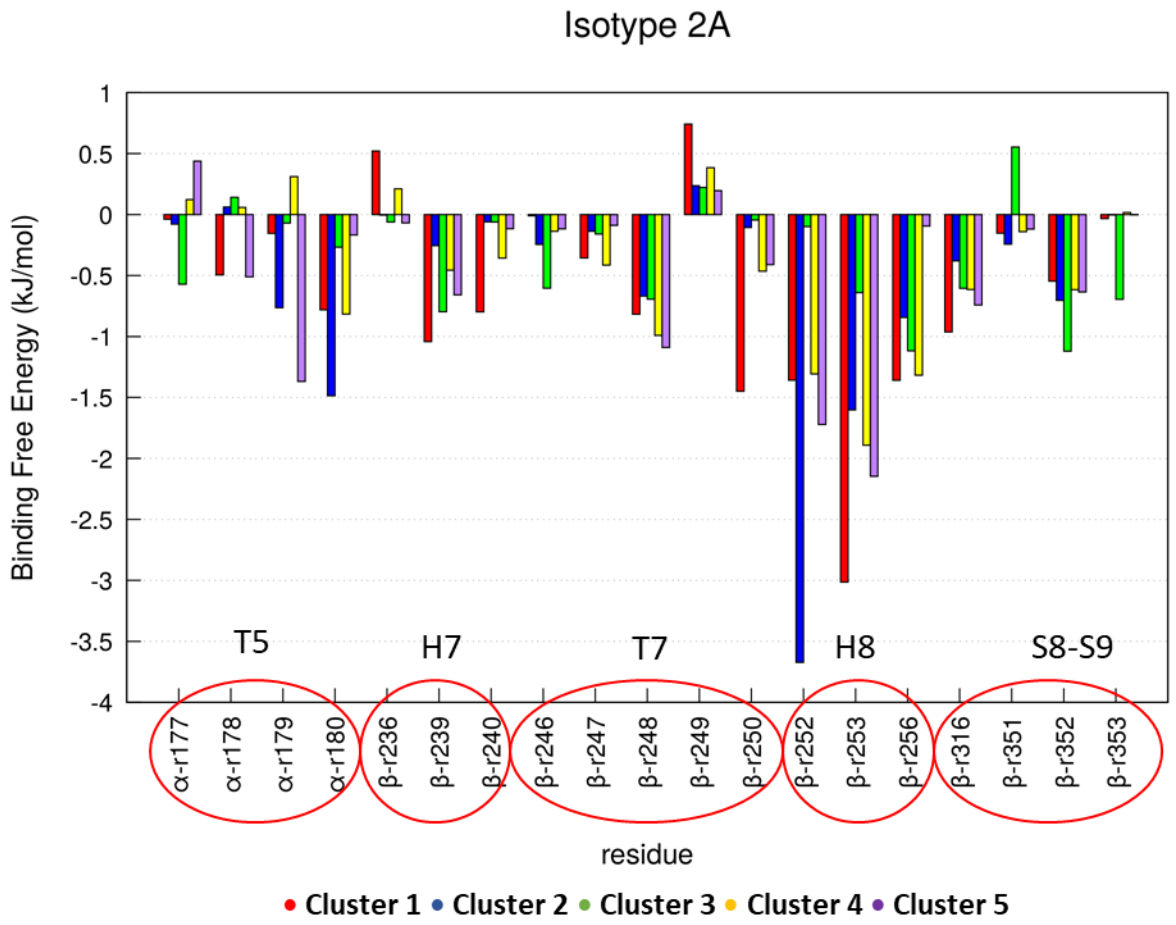


Figure 28: Binding free energy decomposed on the residue for each cluster in Isotype 2A. The energy is mediated on the drugs in each cluster. For all the clusters except cluster 3 the highest binding energies are on the H8 residues, while cluster 3 does not show a particular region with higher energy.

In isotype 2A, except for the cluster 3, the major energy values are on residues β-252 and β-253 (basic residue) for the other clusters. These reisudes are located in the H8 region. The drugs in cluster 1 have the highest binding energies overall, followed by cluster 2. Residue β-249 presents positive values for all the clusters. This is probably due to the fact that it is an acid residue, with negative charge, which repulse the ligands.

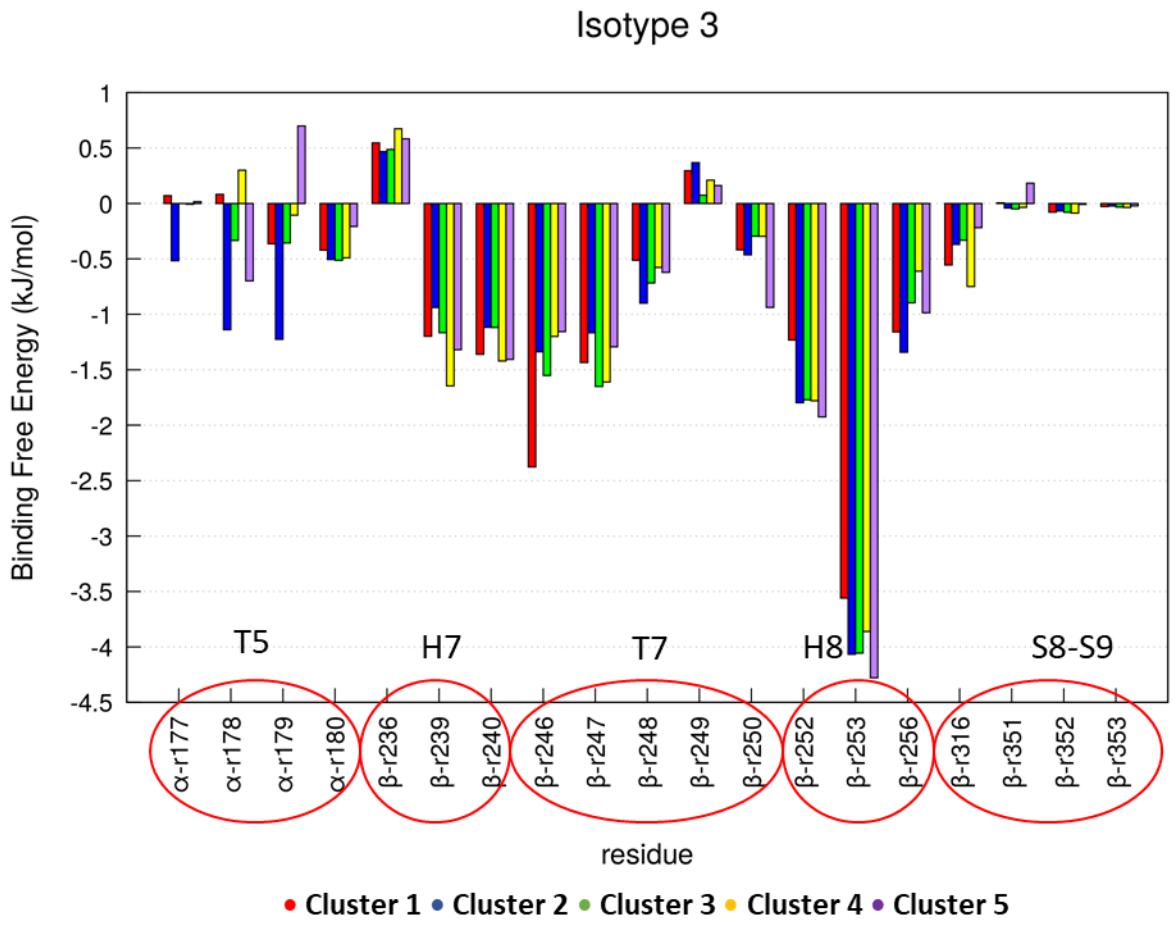


Figure 29: Binding free energy decomposed in the residues for each cluster in isotype 3. The highest binding energies are on H8 for each cluster, followed by H7 and T7 regions. On T5 region, cluster 2 shows higher binding energies compared to the others.

For what concerns isotype 3, the energy values vary very little among the clusters for almost every residue. Every cluster show higher binding energies on H8, followed by H7 and T7 regions. The S8-S9 regions have the lowest binding energies, near zero. The only evident difference is on the residues of the α-T5-Loop, where cluster 2 shows higher values of binding energies. This time too residue β-249 has positive values of energy, while residue β-253 has the highest ones.

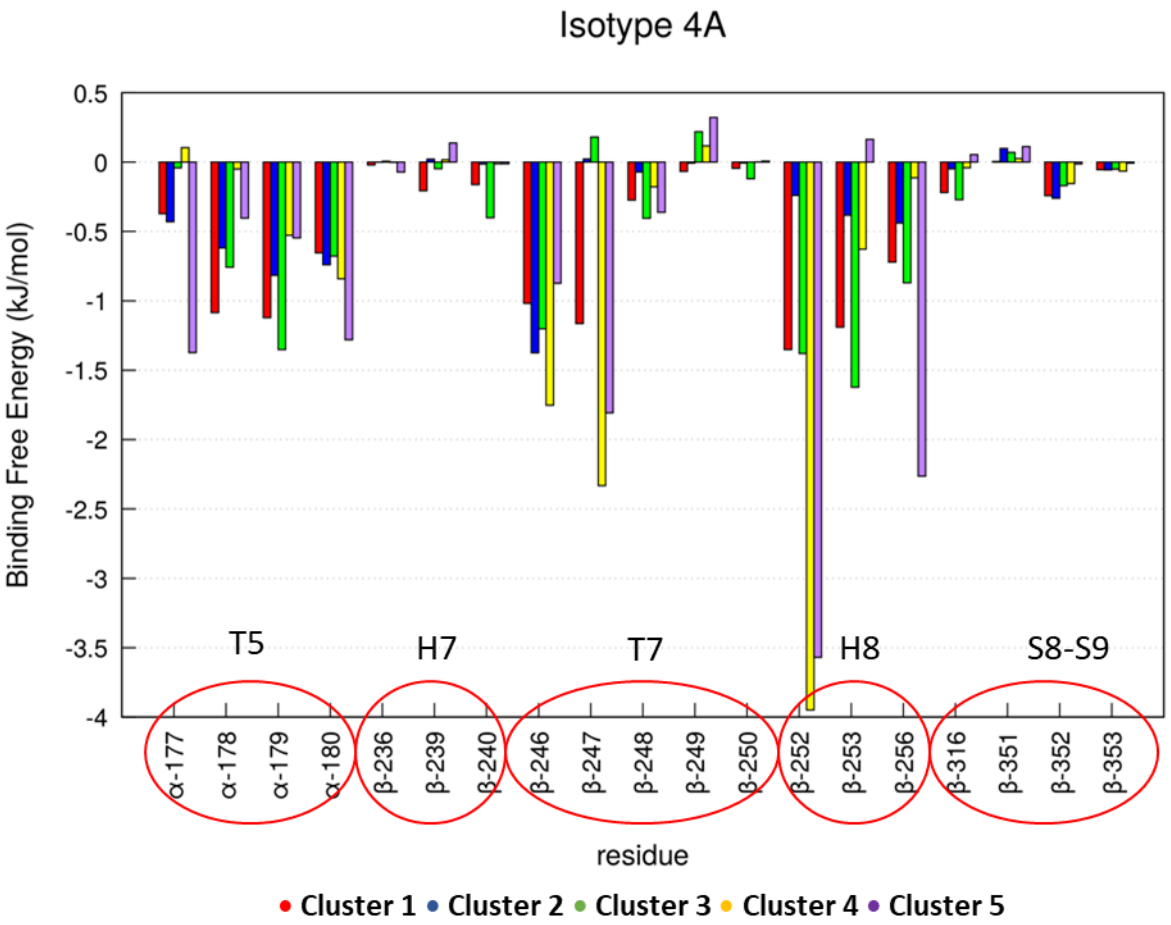


Figure 30: Binding free energy decomposed on the residues for each cluster in isotype 4A. Cluster 4 and cluster 5 show the highest binding energies. Both have high energies on H8, while on T5 cluster 4 shows the lowest values and cluster 5 shows high energy values.

In the end, in isotype 4A, cluster 4 shows the highest values of binding energies in the H8 region, alongside cluster 5. Cluster 4 has the lowest values on the residues of the T5 region. The acid residue β-249 has repulsive energy values. T7 and T5 regions show energy values on average for every cluster, while H7 and S8-S9 regions show energy values near zero for all the clusters.

4.4 Discussion

Microtubules are an interesting target for many drugs, as by altering their dynamic instability it is possible to arrest the mitotic process, leading the cell to apoptosis. In the cell, microtubules present several kinds of tubulin isotypes, that differ from cell to cell and from normal and tumoral cells. The most interesting isotypes are the ones in the β-chain, as they are more involved in the binding of the drugs. Targeting isotype αβI is not a very good choice, as it is widely expressed over all the cells, so tubulin isotype αβIII is the most interesting for inhibit

the cancer cells multiplication[81]. This isotype is also reported as one of the reasons that lead the tumoral cell to drug-resistance[82]. For what concerns the drugs, colchicine is one of the simplest and widely used, but has the side effect of being very cytotoxic[51]. In this study, the effect of several colchicine derivatives has been confronted over the isotypes αβIIA, αβIII and αβIVA, in order to find a drug that can bind the αβIII isotype in a more efficient way than the colchicine and that it is less cytotoxic for the normal cells.

Figures 18, 19 and 20 show the differences between the colchicine's pose and the other drugs poses in terms of RMSD. A low value indicates that they are in a similar position as the colchicine, while a high value indicates that they are in a different one. For what concerns isotype αβIII, only one of the drugs, the compound C19, shows a quite high value of RMSD, indicating that it is in a different position compared to the colchicine. This is an interesting result, as this isotype is the most resistant to the drugs (and in particular to the colchicine), thus a different pose can make this drug more efficient than the colchicine. Probably, the drugs are all in a similar position because the cleft in this isotype is quite narrow. In isotype αβIVA, many of the drugs have a RMSD near 1 nm, indicating that they can assume very different poses. This is probably due to the particular conformation of the cleft, that makes the drugs able to bind it in very different positions. In isotype αβIIA the values are mostly intermediate, indicating that the drugs are in the same overlapped and overturned. These differences can be observed in figure 21.

Then, the binding free energies of the drugs over the three isotypes have been investigated in order to classify them on the basis of their specificity. Figures 22, 23 and 24 summarize these results. From figure 22, it can be acknowledged that six drugs are non-specific and show a very high binding energy for all the isotypes, so they can be useful to target many cells without interest in the specificity. There are other non-specific drugs, as it is shown in figure 23, but they show lower average binding energies values compared to the others. Finally, figure 24 helps in figuring out which drugs are specific for a particular isotype. 21 drugs are specific for isotype αβIII, so they can be used to specifically target cancerous cells, that overexpress this isotype. Many others are specific for both the isotypes αβIII and αβIIA, so the cleft conformation in isotype αβIVA is probably the cause of the lower binding energies values for this isotype. Only one drug, C55, is specific for this isotype. There are no specific drugs for isotype αβIIA, as they are all shared with isotype αβIII. Moreover, most of the drugs with the highest binding energies of this isotype are not specific at all. For what concerns the isotype

αβIII, the compounds C51, C53 and C19 are specific and their binding energies are among the highest for the isotype. Therefore, they can be efficient drugs to selectively target the tumoral cells.

Figures 25, 26 and 27 merge the results from the previous graphs in radar graphs. The drugs are coloured on the basis of the cluster of belonging. The details of the energy on each significant residue for each cluster can be seen in figures 28, 29 and 30. In isotype αβIIA radar graph, it can be noticed that the drugs with the highest binding energies have all a very low RMSD and are all in the cluster 1, which is the colchicine's one. This indicates that the most effective drugs are the ones more similar to the colchicine. The drugs in clusters 3, 4 and 5 have higher RMSD and lower binding energies, compared to the ones in cluster 1. These results are supported by figure 28, where it can be seen that clusters 1 and 2 have higher binding energies compared to the others. Overall, the binding energies for cluster 1 are just slightly higher than the ones in cluster 2. The most involved region seems to be the H8.

The radar graph for isotype αβIII (figure 26) is very interesting, as it clear that the behaviour of one drug differs significantly from the others. This drug is the already mentioned C19, that is the only drug with a high RMSD and has one of the top binding energies. Overall, al the drugs in its cluster have high binding energies, so it can be due to their positions in the cleft. Looking at the decomposition over the residues (figure 29), the most evident difference is that the drugs in cluster 2 have higher binding energies on the residues of the α-T5-loop, supporting the thesis that they are is their different position in the cleft, that lead to the higher values of binding energy overall and in this particular region. For what concerns the other residues, there are not evident differences: the highest binding energies are on H8, followed by H7 and T7 regions, while the S8-S9 regions have binding energies near zero. This result could have been expected, as the drugs are all in a similar position. Cluster 2 higher energy on α-T5-loop can be probably due to the lateral groups of these drugs and to the different pose of the compound C19.

Last, the radar graph for isotype αβIVA (figure 27) shows that the drugs in both cluster 4 and 5 have the higher binding energies. Cluster 4 drugs have low RMSD, so they are in a position near the colchicine's one, while the drugs in cluster 5 are in a different position, as they have very high RMSD. The majority of the drugs (cluster 1 and 2), have lower binding energies. It could be expected, as in this isotype the binding energies overall have lower values compared to the others. Looking at the decomposition graph (figure 30), this difference becomes more evident. The region with the highest binding energy is H8, followed by T7 and H7. The beta-

sheets and the H7 have energy values near zero. For what concerns the α-T5-loop, the values are on average; the only exception in cluster 4, where these values are near zero.

Overall, the helix H8 is the highest binding energy region in every isotype, while the beta-sheets have the lowest values in both isotypes 3 and 4A. The other regions have different values depending both to the drugs and to the particular isotype.

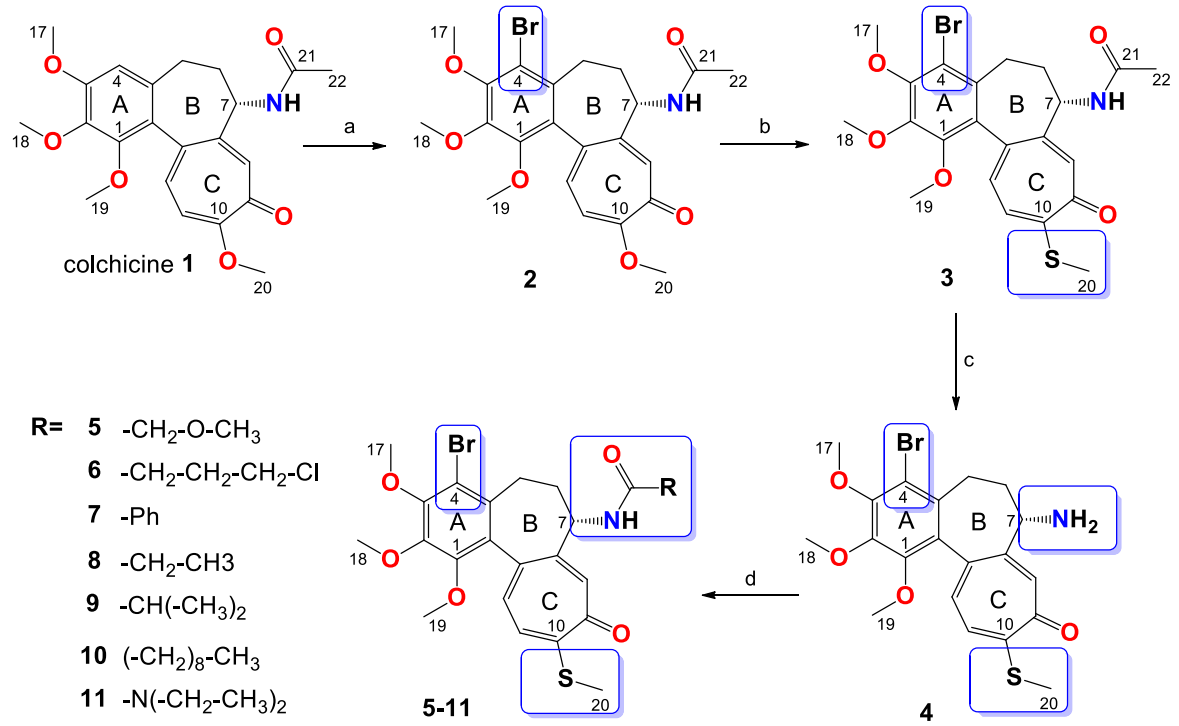
As the aim was to find a drug able to bind specifically the isotype αβIII and able to overcome the colchicine in the performance, the drug C19 seems to be an interesting compound: it has one of the highest binding energies on isotype αβIII and has a different pose compared to the colchicine's one. Moreover, it is a specific drug for the isotype αβIII, thus it is interesting to perform more analysis of this compound.

4.5 Conclusions

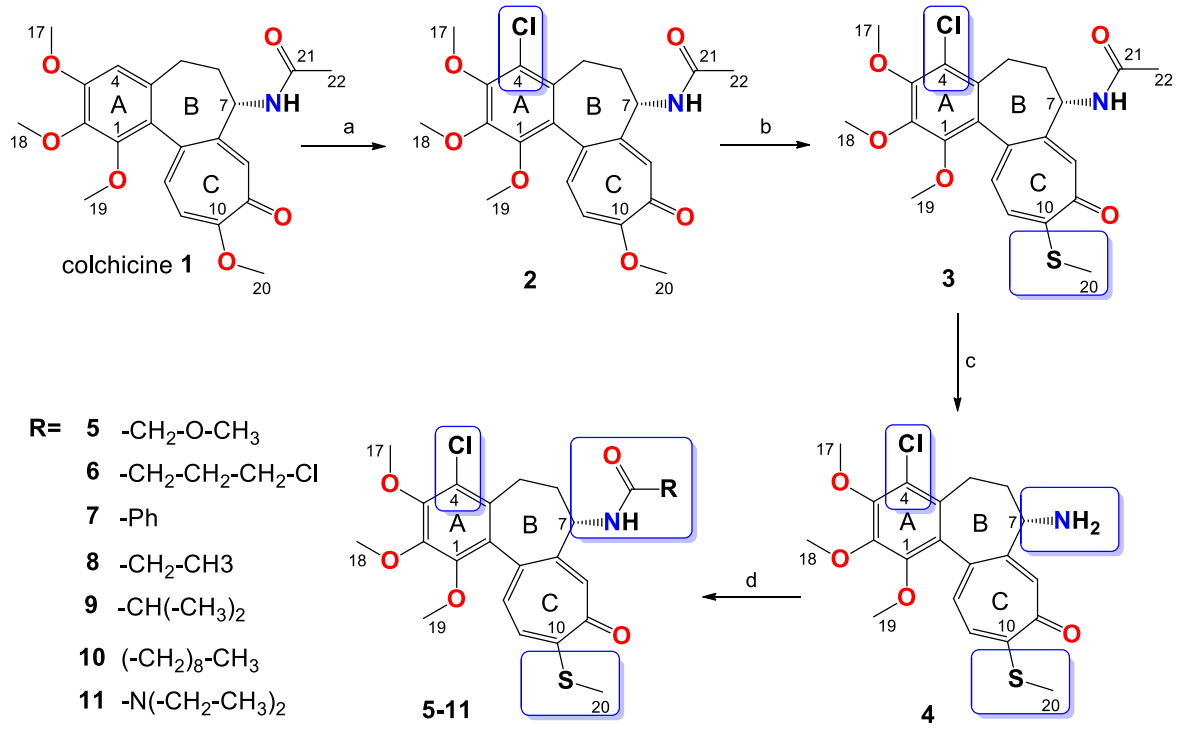
In this chapter, several colchicine derivatives compounds have been analysed and compared over three models of different human tubulin isotypes: αβIIA, αβIII and αβIVA. The aim was to find a drug potentially able to act selectively on isotype αβIII, the most overexpressed in cancer cells. The results show that one compound, C19, could be a good drug to test. In the next chapter, a more detailed study of this compound on isotype αβIII will be performed.

4.6 Supporting information

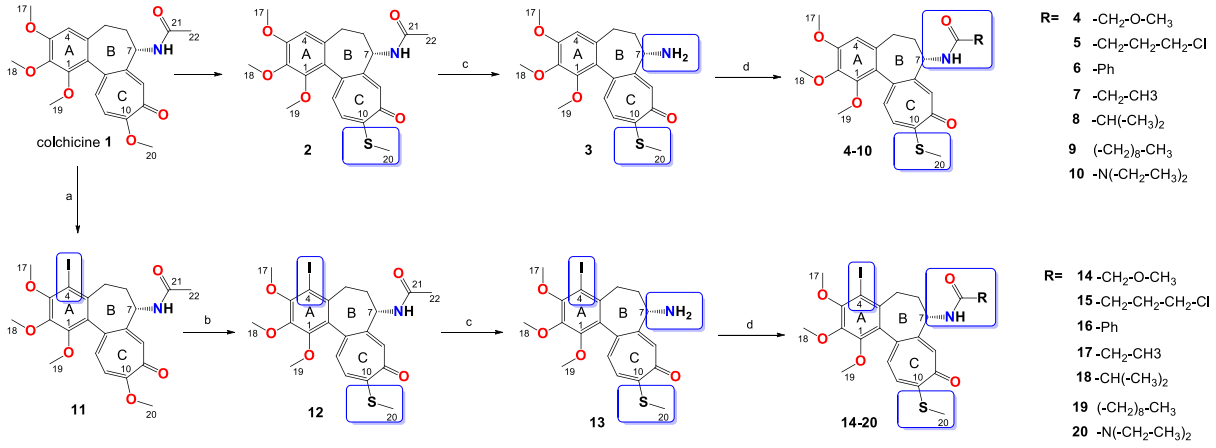
4.6.1 Drugs chemical structures



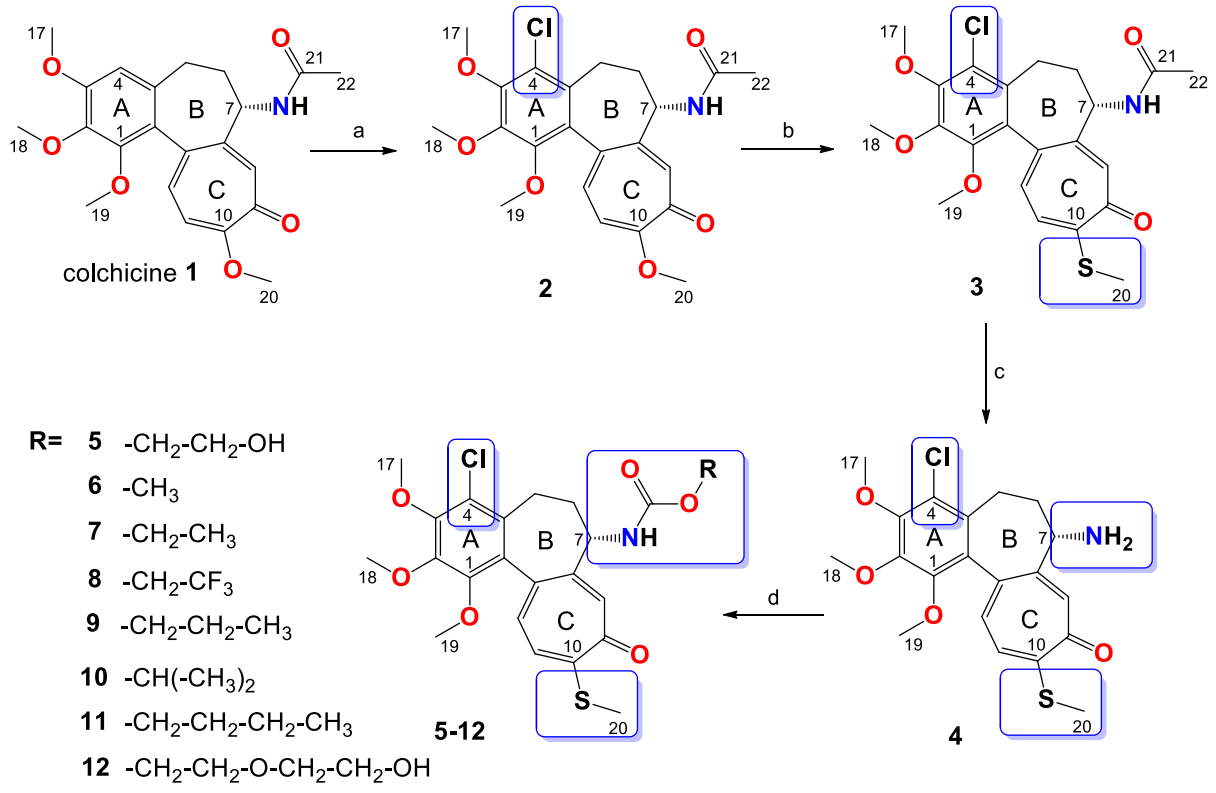
Supporting Figure 1: Chemical structures of 4-Br-Amides



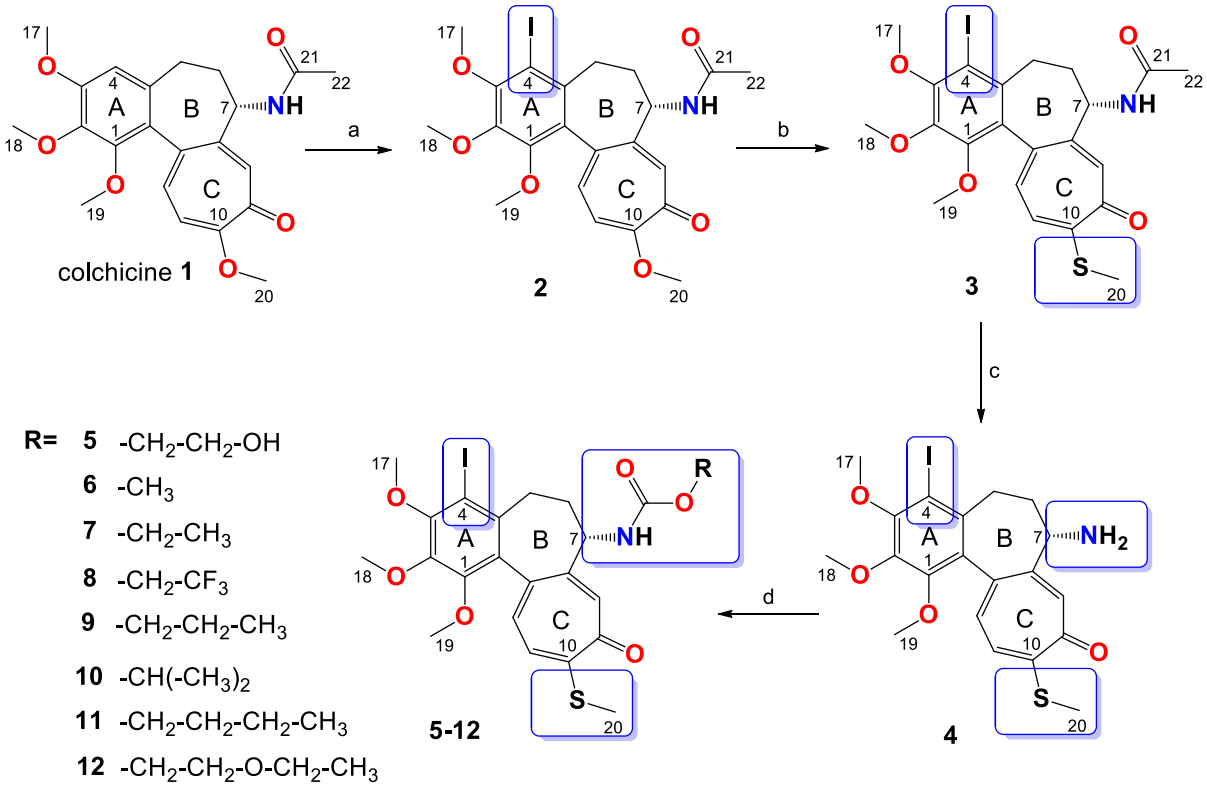
Supporting Figure 2: Chemical structures of 4-Cl-Amides



Supporting Figure 3: Chemical structures of DT-and-4-I-Amides



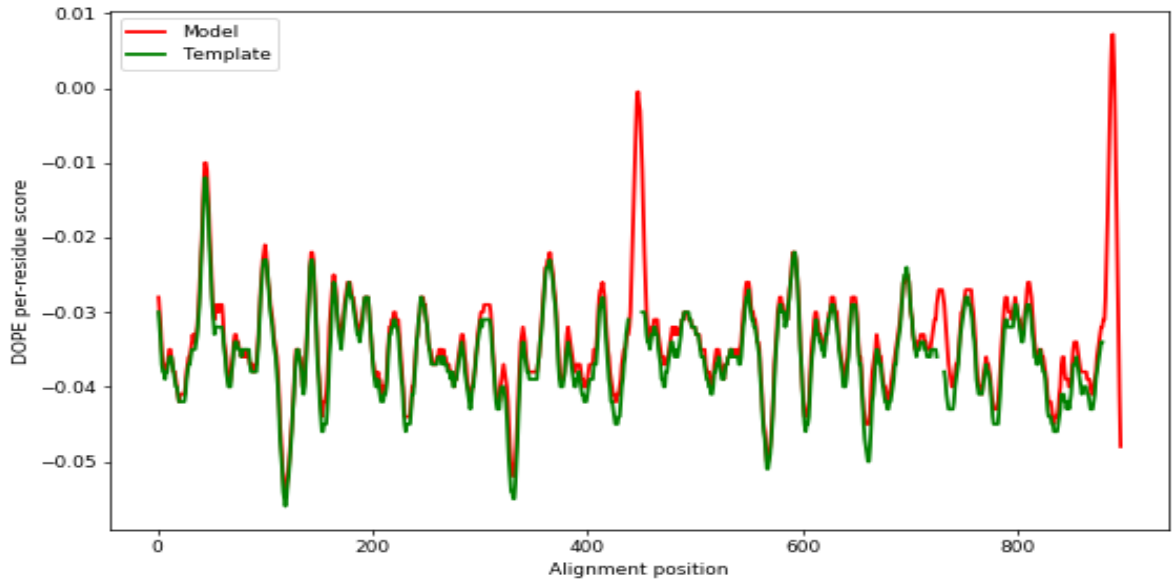
Supporting Figure 4: Chemical structures of 4-Cl-Carbamates

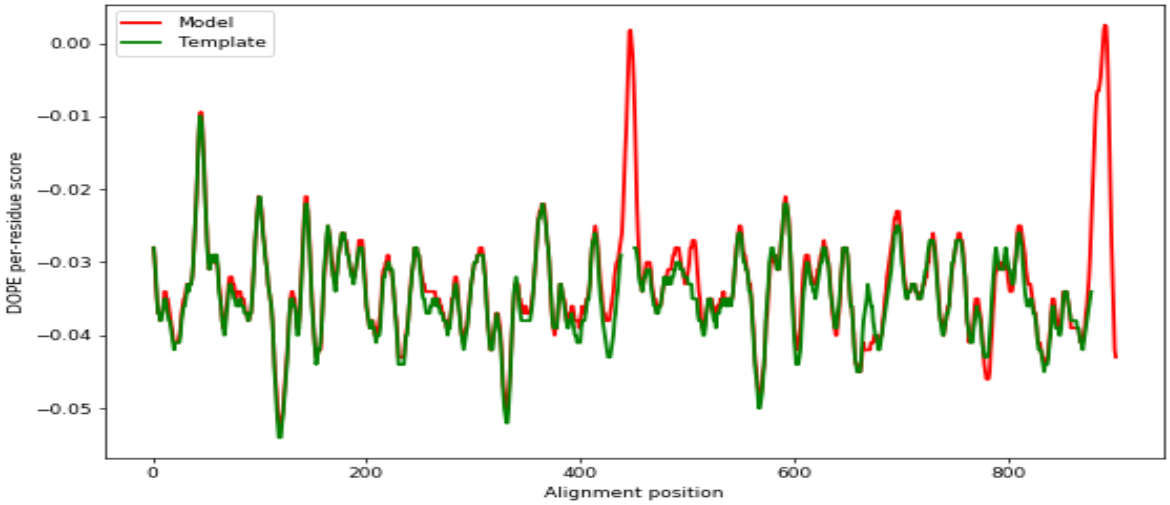


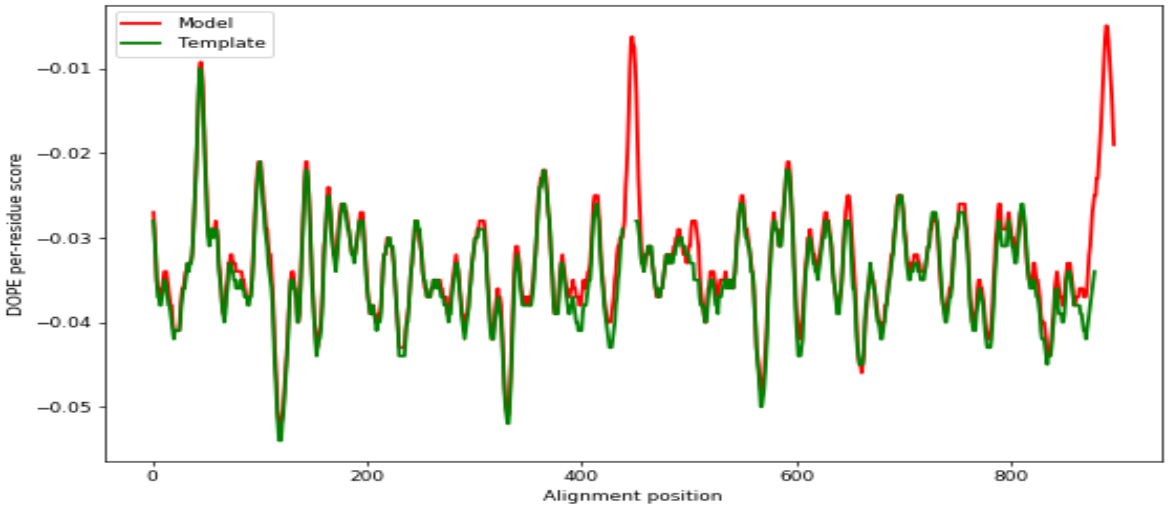
Supporting Figure 5: Chemical structures of 4-I-Carbamates

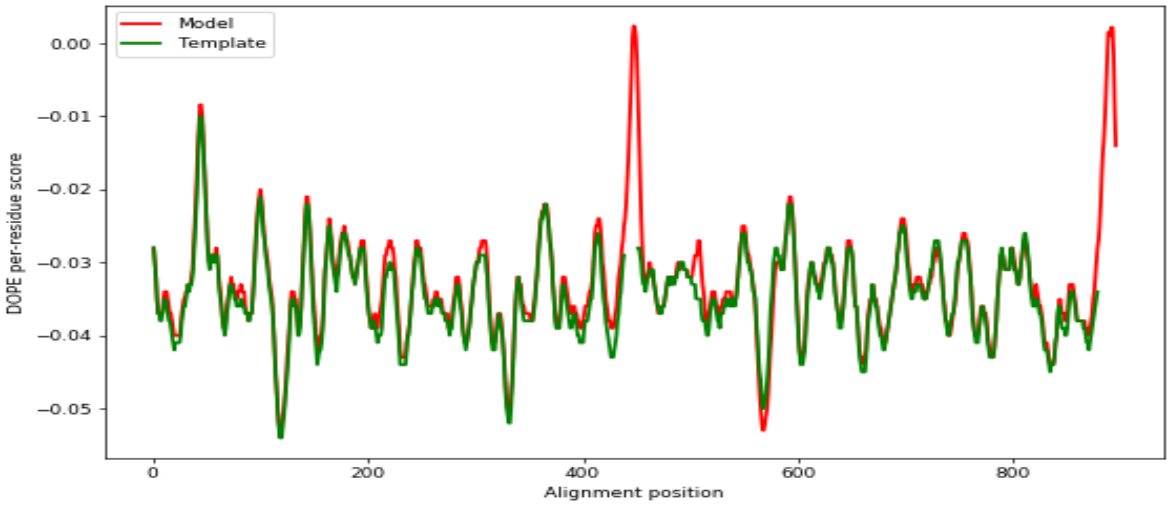
4.6.2 Homology Modelling

The best model built with the homology modelling was chosen based on the lowest values of the DOPE score for each structure. The comparison between the dope profiles of the template and the built model are represented in following figures.




Model
Template
DOPE per-residue score
0.00
−0.01
−0.02
−0.03
−0.04
−0.05
0
200
400
600
800
Alignment position


Model
Template
DOPE per-residue score
−0.01
−0.02
−0.03
−0.04
−0.05
0
200
400
600
800
Alignment position


Model
Template
DOPE per-residue score
0.00
−0.01
−0.02
−0.03
−0.04
−0.05
0
200
400
600
800
Alignment position

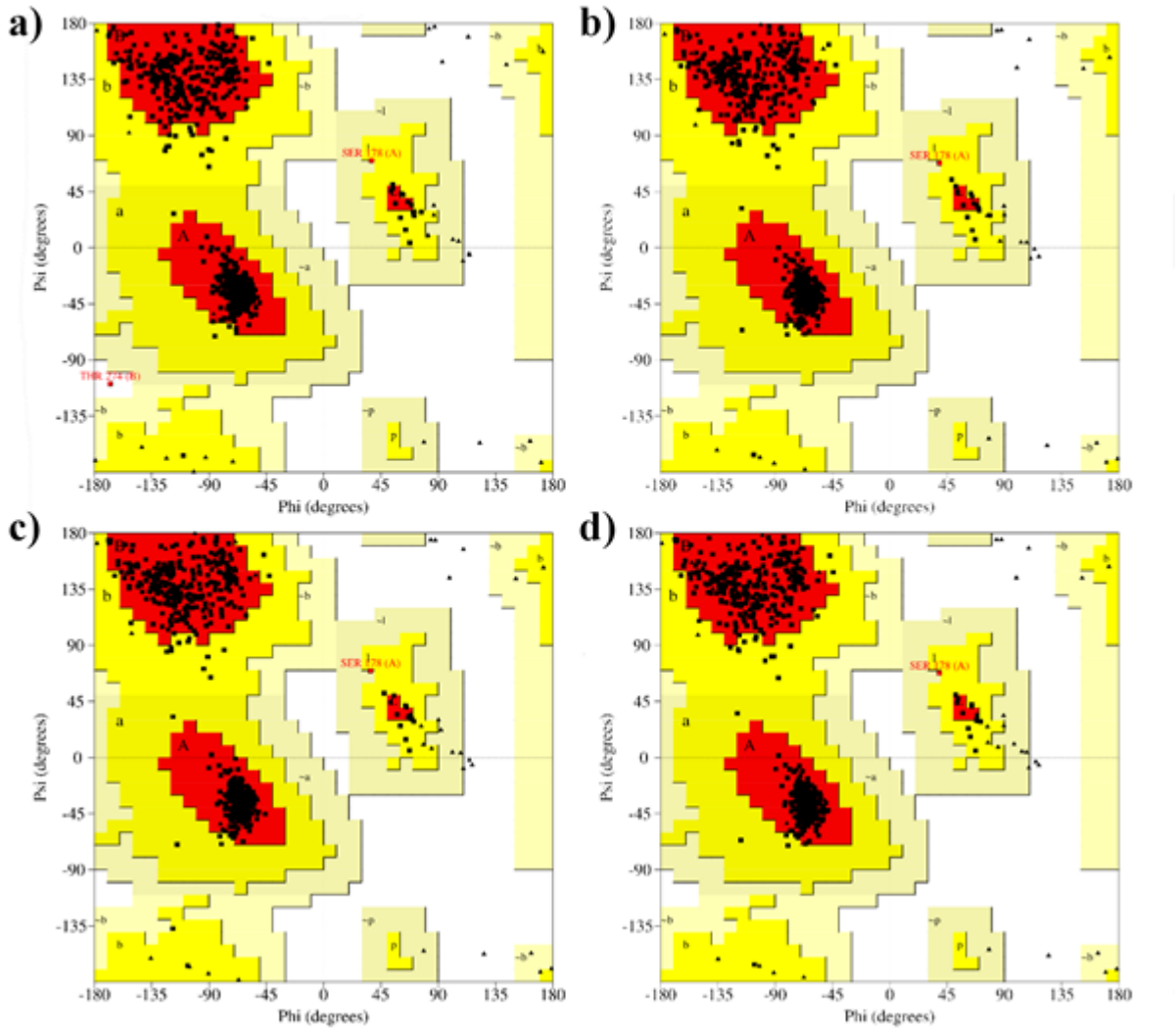
Supporting Figures 6-7-8: DOPE per-residue score for the bovine template structure and the three isotypes αβIIA, αβIII and αβIVA respectively: the template structure is represented in green, while the models are represented in red. More the curves are superimposed, more the model is good.

The quality of the general models was then evaluated using the PROCHECK, VERIFY3D and ERRAT tools.

PROCHECK analysis shows the Ramachandran plot for each generated model: Phi and Psi angles for each residue are plotted, defining the most favoured, the allowed and the not allowed regions.

The template model has 95.9% of the residues in the most favoured regions, 3.8% in the additional allowed region, 0.1% in the generously allowed regions and 0.1% in the disallowed regions. For what concerns the three isotypes, the percentages of the residues in each region are respectively:

- αβIIA: 95.8% in the most favoured regions, 4.1% in the additional allowed regions, 0,1% in the generously allowed regions and no residues in the disallowed regions.
- αβIII: 95.6% in the most favoured regions, 4.2% in the additional allowed regions, 0.1% in the generously allowed regions and no residues in the disallowed regions.
- αβIVA: 95.4% in the most favoured regions, 4.5% in the additional allowed regions, 0.1% in the generously allowed regions and no residues in the disallowed regions.



Supporting Figure 9: Ramachandran Plots for the template structure (a), the isotype αβIIA (b), αβIII (c) and αβIVA (d). Over the 90% of the residues is in the most favoured regions in every model.

As more than 90% of the residues are in the most favoured regions, the models can be assumed as good ones.

The Overall Quality Factor obtained whit the ERRAT tool for the template model is 86.68 for chain α and 91.17 for chain β. For the isotype αβIIA these values are respectively 79.12 and 88,31, for the isotype αβIII 80.29 and 84.73 and for the isotype αβIVA 77.26 and 84.96.

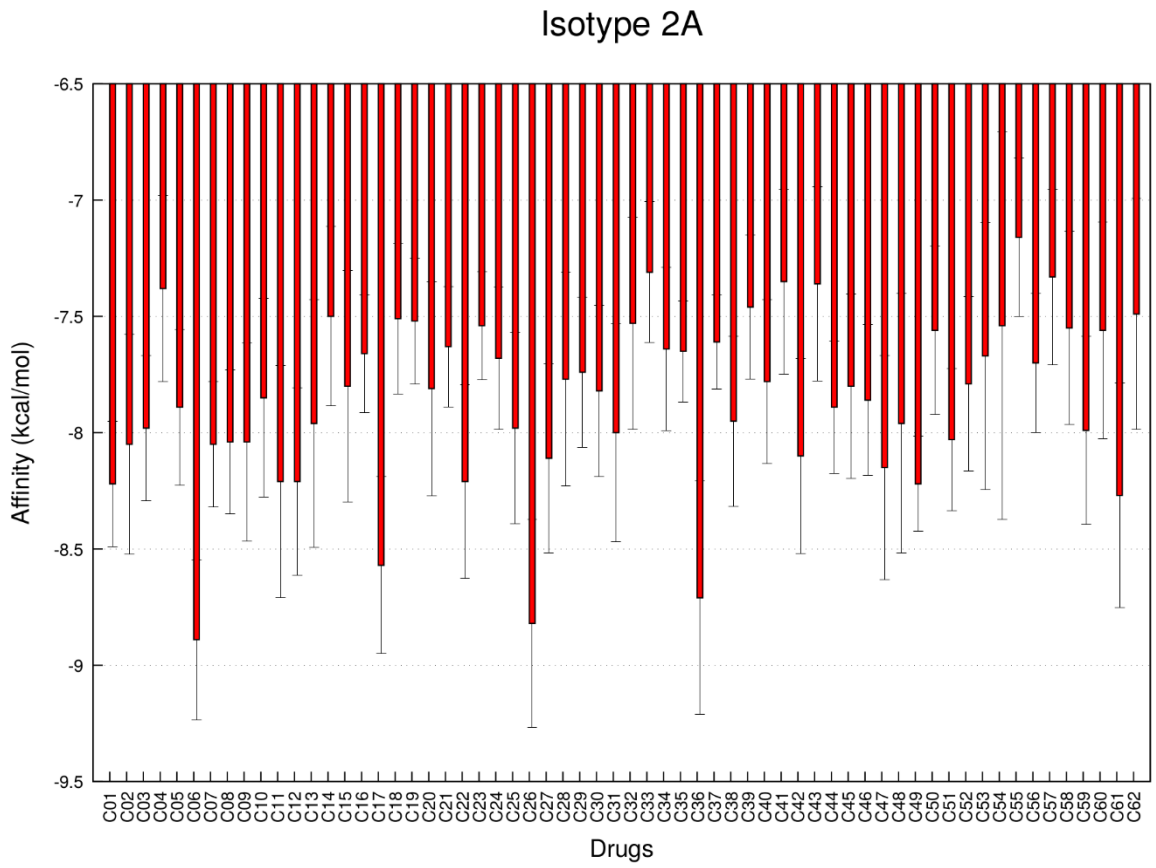
The generally accepted range is >50 for a high quality model, so all the models passed the test.

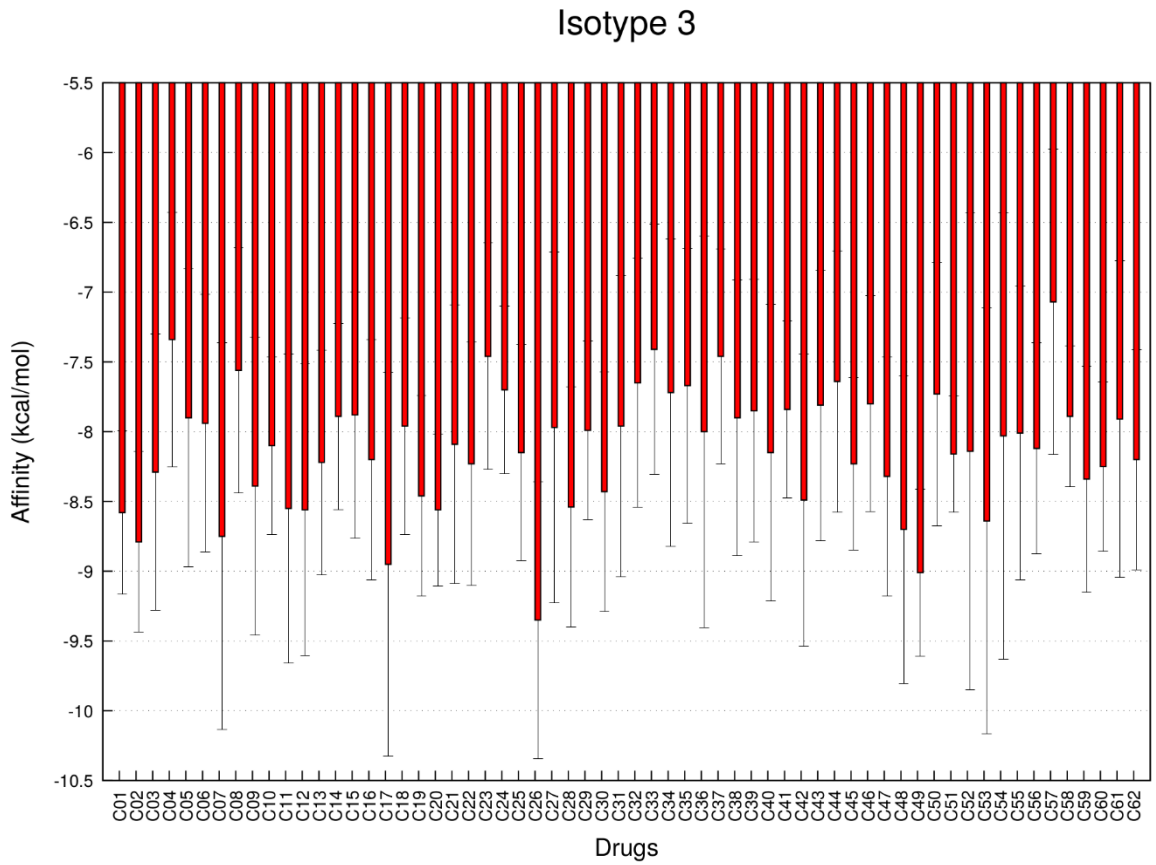
The VERIFY3D is considered passed if a high percentage of the residues has an averaged 3D-1D score higher than 0.2. All models pass the VERIFY3D showing percentages of 96.77% for

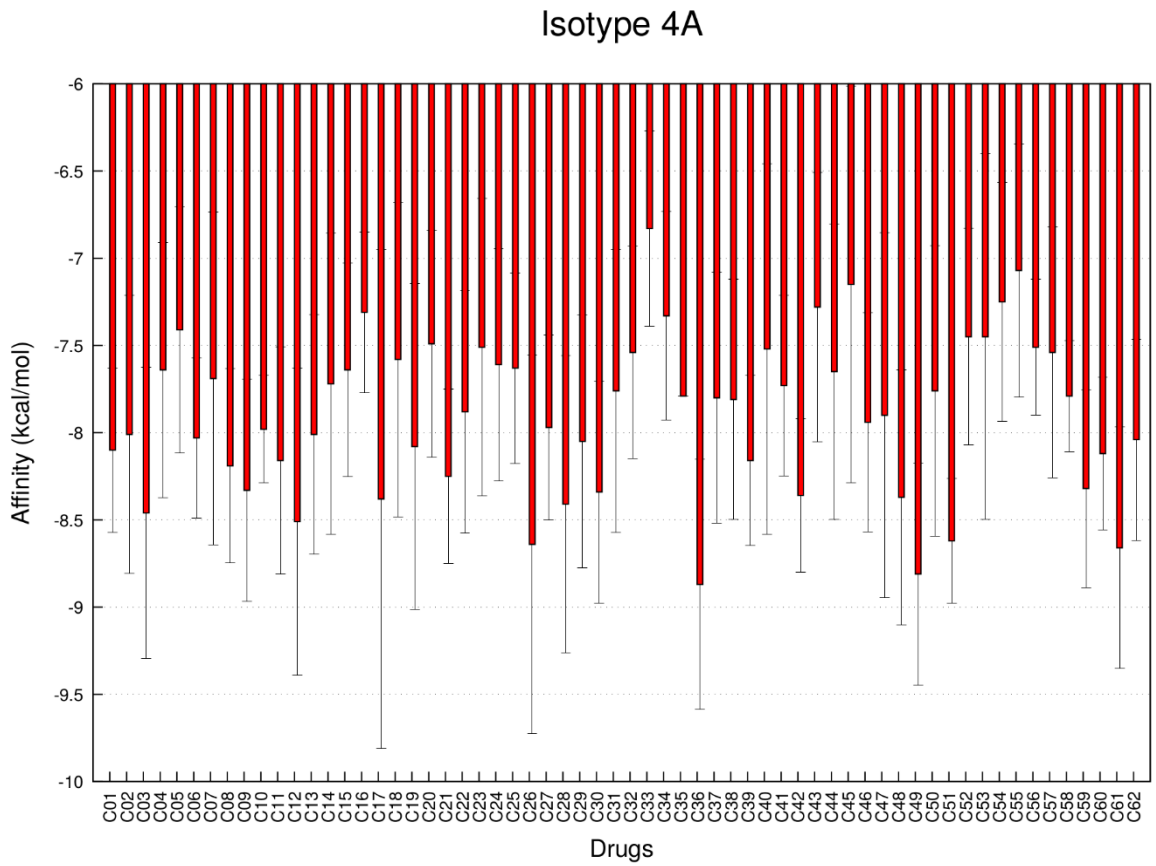
the template model, 98.85% for isotype αβIIA, 98.15% for isotype αβIII and 97.69% for isotype αβIVA.

All the results show that the models are acceptable.

4.6.3 Ensemble Docking results

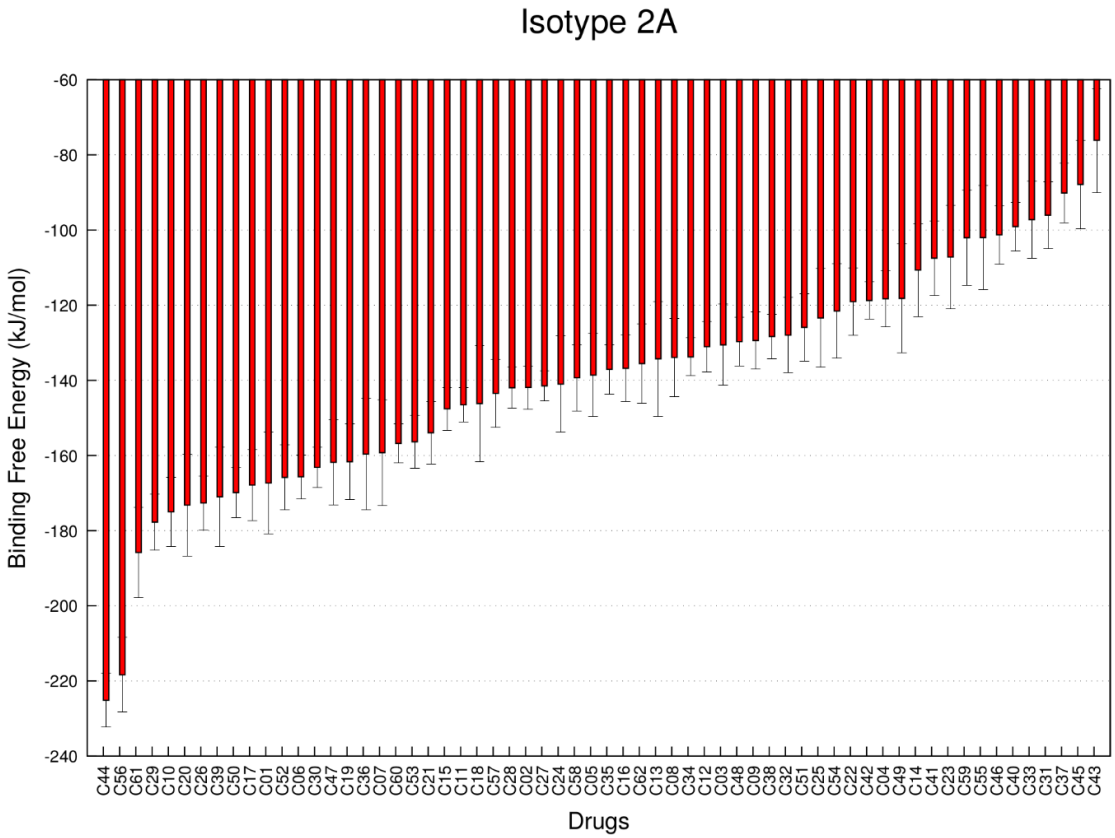


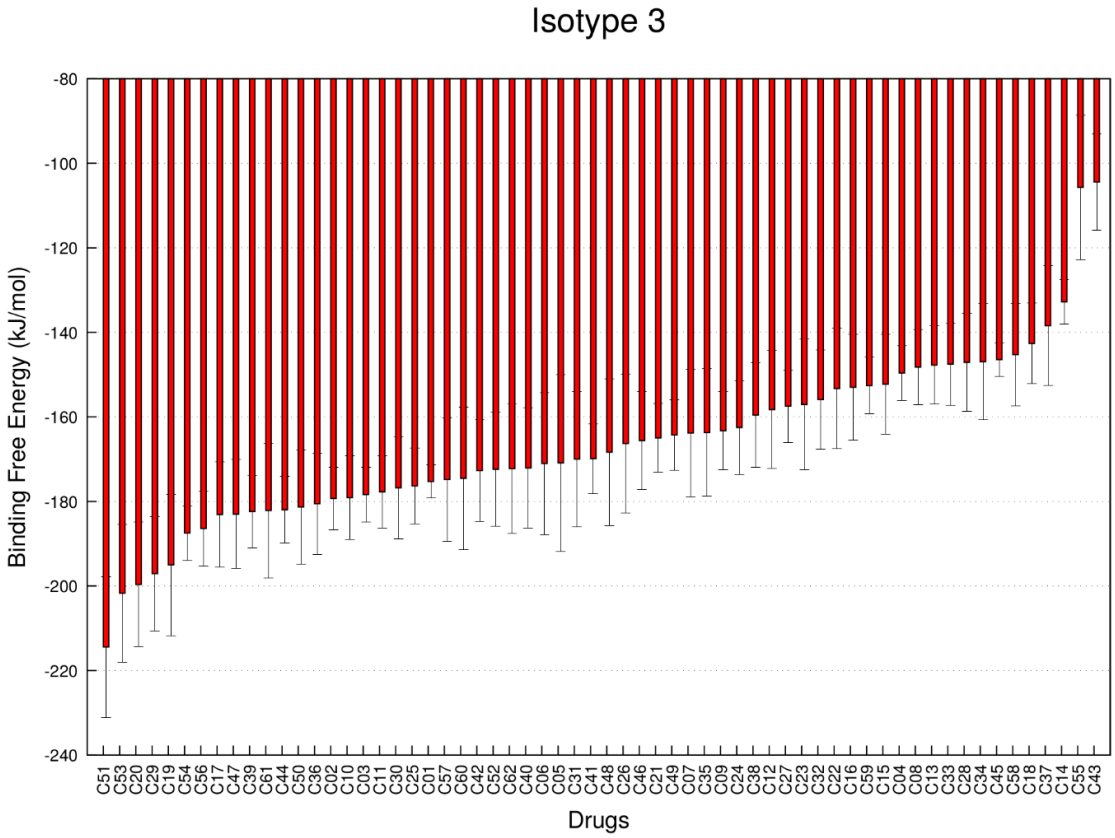


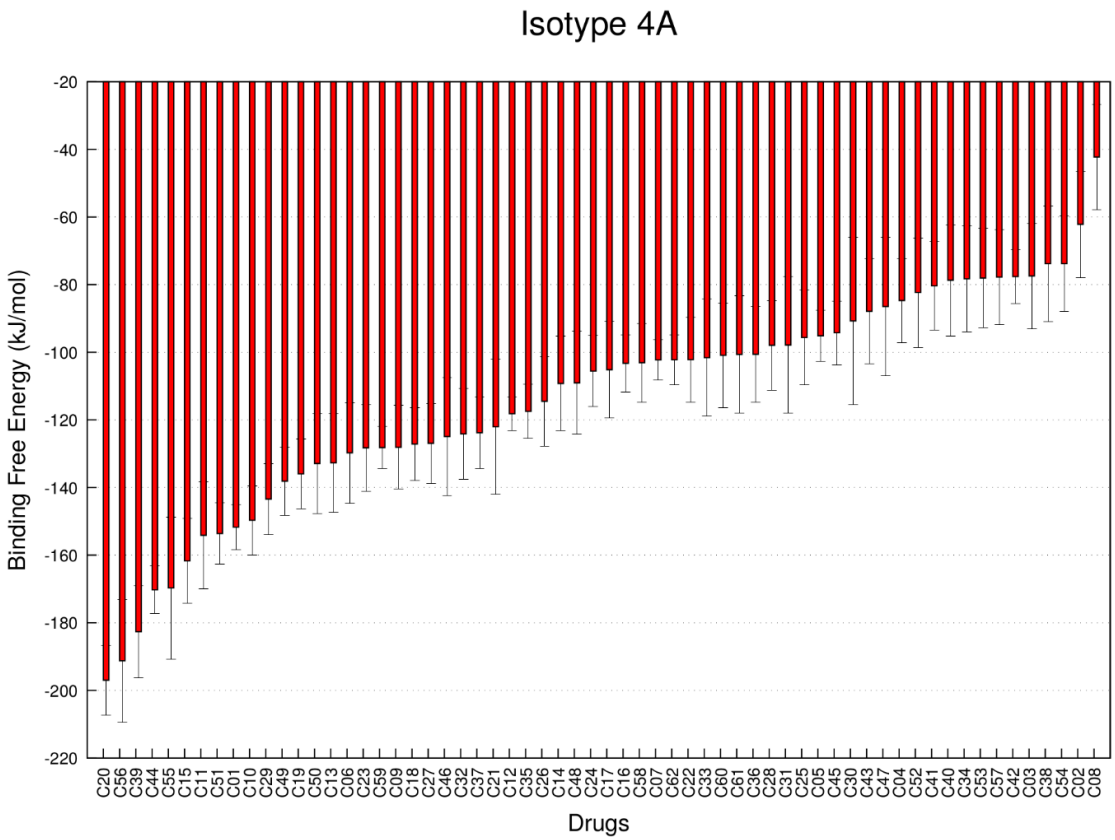


Supporting Figures 10-11-12: Ensemble Docking results for all the drugs over the three isotypes.

4.6.4 Binding Free Energies







Supporting Figures 13-14-15: Ranking of the binding free energy calculation for isotype 2A, 3 and 4A. The binding energies are lower for isotype 4A, compared to the others.

4.6.5 Drugs included in each cluster

Supporting Tables 1-2-3: Drugs included in each cluster for the three tubulin isotypes.

Isotype 2A	
Cluster 1	C01, C02, C03, C04, C05, C06, C07, C10, C11, C12, C13, C14, C15, C16, C17, C20, C21, C22, C23, C24, C25, C26, C29, C30, C31, C32, C33, C34, C35, C36, C39, C40, C44, C50, C52, C53, C56, C57, C59, C61, C62
Cluster 2	C08, C09, C18, C19, C27, C28, C37, C38, C43, C48, C49, C58, C60
Cluster 3	C41
Cluster 4	C42, C47, C51
Cluster 5	C45, C46, C54. C55

Isotype 3	
Cluster 1	C01, C02, C03, C04, C05, C06, C07, C10, C11, C12, C13, C14, C15, C16, C17, C20, C21, C22, C23, C24, C25, C26, C29, C30, C31, C32, C33, C34, C35, C36, C39, C40, C44, C50, C52, C53, C56, C57, C59, C61, C62
Cluster 2	C08, C09, C18, C19, C27, C28, C37, C38, C43, C48, C49, C58, C60
Cluster 3	C41
Cluster 4	C42, C47, C51
Cluster 5	C45, C46, C54, C55

Isotype 4A	
Cluster 1	C01, C02, C03, C04, C05, C06, C07, C10, C11, C12, C13, C14, C15, C16, C17, C21, C22, C23, C24, C25, C26, C29, C30, C31, C32, C33, C34, C35, C36, C40, C43, C45, C47, C52, C53, C54, C57, C61, C62
Cluster 2	C08, C09, C18, C19, C27, C28, C37, C38, C41, C48, C49, C50, C58, C59, C60
Cluster 3	C41, C42, C46
Cluster 4	C44, C51, C56
Cluster 5	C20, C39, C55

5. Molecular Dynamics and Local Effects on αβIII Tubulin Binders by colchicine derivatives.

In this chapter a study of the compound C19 based on molecular dynamics is presented. This drug has shown peculiar characteristics in isotype αβIII in the comparative study performed in the previous chapter: it is specific for this isotype, it has a different pose in the cleft compared to the colchicine and it has one of the highest binding energies to isotype αβIII. Long MD simulations are performed for the αβIII dimer bound either to the colchicine and to the novel compound, in order to confront their effects on the dimer. The results show that the compound C19 could be a good anticancer drug.

Abstract

The overexpression of the tubulin dimer isotype αβIII is believed to be one of the reasons of the drug-resistance in the tumoral cells[67]. Colchicine is among the most used drugs to treat the cancer, but it has the side effects of being quite toxic and not very effective on isotype αβIII[83]. The colchicine derivative C19, investigated in the previous section, may be effective to overcome colchicine's limitations. In particular C19's different position in the binding cleft and the high affinity might result in a higher efficiency. In this study, Molecular Dynamics (MD) simulations were carried out for comparing the effects of the colchicine and the compound C19 on the human tubulin isotype αβIII. The outcome of this work shows the differences between the two compounds and confirms that C19 could be an effective drug.

5.1 Introduction

Colchicine is one of the most employed microtubule destabilizing agents for cancer treatment. The lateral contacts are fundamental in the microtubule integrity[84], so a drug capable to reduce these contacts is a good destabilizer, able to enhance catastrophes in the microtubule structure. Despite this power, it is quite toxic for the organism, so there is the need to design and develop its derivatives in order to make them more specific for the tumoral cells and less poisonous for the normal cells. As the tubulin isotype αβIII is overexpressed in tumoral cells[52], it is interesting to produce compounds able to specifically target this tubulin. In the previous chapter, the performed analysis showed that, among all the compounds, the colchicine-derivative C19 could have the characteristic to be a better drug than the colchicine. In this study, the effects of the

colchicine and the compound C19 have been confronted on the human tubulin isotype αβIII by using Molecular Dynamics (MD) simulations and binding free energy calculation.

5.2 Materials and Methods

5.2.1 Molecular Dynamics simulations

The employed structures of both the dimer and the drugs were the ones built in the previous chapter. Molecular dynamics simulations were performed for the isotype αβIII bounded to both the colchicine and the compound C19.

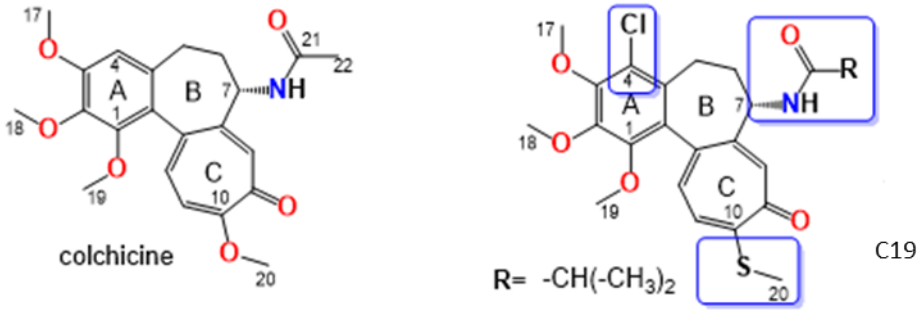


Figure 31: Colchicine and C19 2D structures.

The simulations were performed by using GROMACS 5.1.4 package. The same parameters have been used for both the systems. Periodic boundary conditions were applied along the xyz coordinates. The AMBER-ff99SB-ILDN force-field was used to define protein topology, while Mg^{2+}, GDP, GTP and colchicine topologies were defined using Antechamber program of the AMBER16 package, with general amber force field and bcc charge method, and ACPYPE. A triclinic box with 0.6 nm of minimum distance was used and all systems were solvated using TIP3P explicit water model, with the addition of chlorine and sodium ions to neutralize the charge, maintaining an ionic strength at 150 mM. A 1000-step energy minimization by steepest descent was then performed, with a maximum force of 100 $kJ\,mol^{-1}nm^{-1}$. After that, restraining potentials were applied on water molecules and sodium and chlorine ions and system equilibration in NVT and NPT ensembles was performed: NVT equilibrium phase was performed for 100 ps using velocity-rescale thermostat with the tau constant at 0.1 and the reference temperature at 300 K, while the NPT one for 300 ps using Berendsen barostat to maintain the pressure at 1 atm. Electrostatic interactions were treated by means of Particle Mesh Ewald (PME) approach, with 1.0 nm cutoff, Fourier spacing of 0.2 nm and interpolation order of 4. Lennard-Jones interactions were cut-off at a distance of 1 nm. Finally, each system was simulated for at least 100 ns without any restrain with a time step of 2 fs and coordinates saved

at every 2 ps. The Visual Molecular Dynamics (VMD) package was employed for the visual inspection of the simulated systems. Dedicated GROMACS tools were used for a quantitative analysis in terms of Root-Mean-Square Deviation (RMSD) and Root-Mean-Square Fluctuation (RMSF), while analysis of secondary structure was performed by applying the STRuctural IDEntification (STRIDE).

5.3.2 Binding free energy calculation

The binding free energies between protein and ligands were calculated by using the molecular mechanics Generalized Born surface area (MMGBSA) theory. The free energy was thus calculated with the software AMBER16. This calculation was performed using the trajectory frames from the last 50 ns MD simulations mentioned above and igb=8. In order to not reduce the predictive power[78], MMGBSA calculation has been performed separately on each ns of simulation and then mediated overall. The free energy was also decomposed on each residue.

5.3 Results

5.3.1 Molecular Dynamics simulation

The RMSD from the starting position for each compound was calculated for each drug during the 100 ns of simulation. The RMSD has been calculated without fitting on the drug's atoms, so it is a measure of the internal deformations in the drug (which are negligible) and the rigid roto-translation of the drug inside the cleft. Therefore, it can be used to evaluate the propensity of the drugs to go outside the cleft. Compound 19 shows a lower value of RMSD, indicating that it is more stable inside the cleft, while the colchicine has a major tendency to get out from the binding cleft.

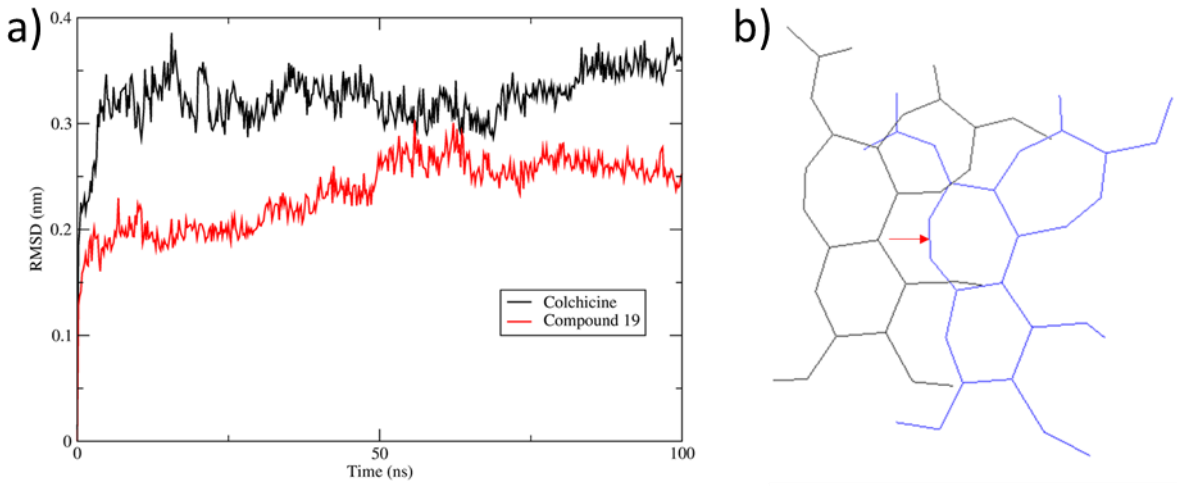


Figure 32. a) RMSD values from the starting position for the colchicine and C19. C19 has a lower value of RMSD, indicating that it is more stable inside the cleft. b) Variation of the colchicine's pose from the starting of the simulation (black) to end (blue). The colchicine tends to go outside the cleft more than C19.

Then, the secondary structure modifications in the colchicine binding site have been investigated in the last 50 ns of simulation, using the software STRIDE.

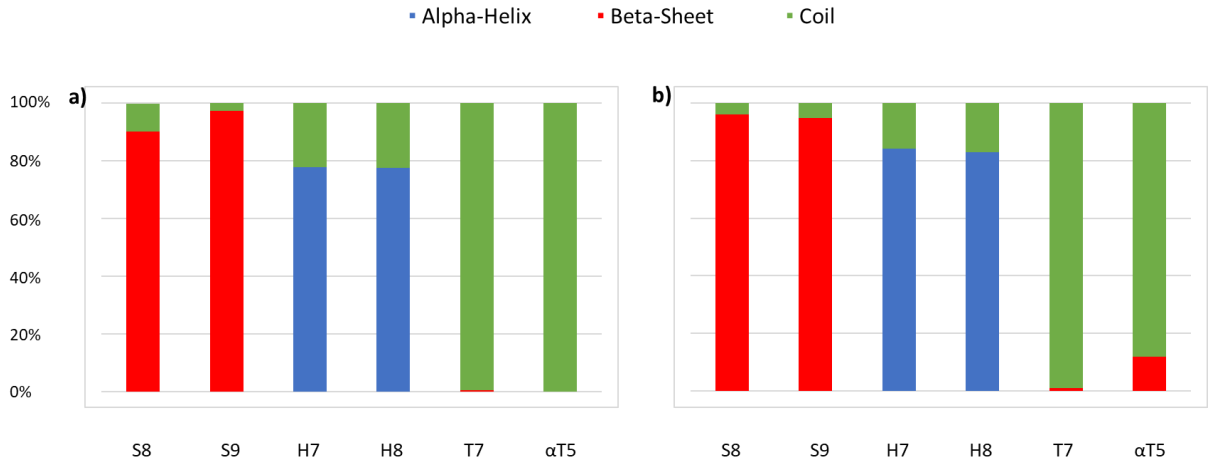


Figure 33: Secondary structure modifications in the colchicine binding site of isotype αβIII with the colchicine (a) and the compound C19 (b) in the last 50 ns of simulation. The secondary structures are quite stable during the simulation; the only difference is that the αT5 loop has a quite major tendency to form a sheet with the C19.

The secondary structures are stable in both the simulations. The only difference is in the α-T5 loop, which has a little tendency to form a Beta-sheet with the C19 (20%). This behaviour is absent with the colchicine. Following these results, the RMSF of the α-T5 loop were analysed in the last 50 ns of simulations. The results show a minor fluctuation with the compound C19, as it can be seen in the following figure:

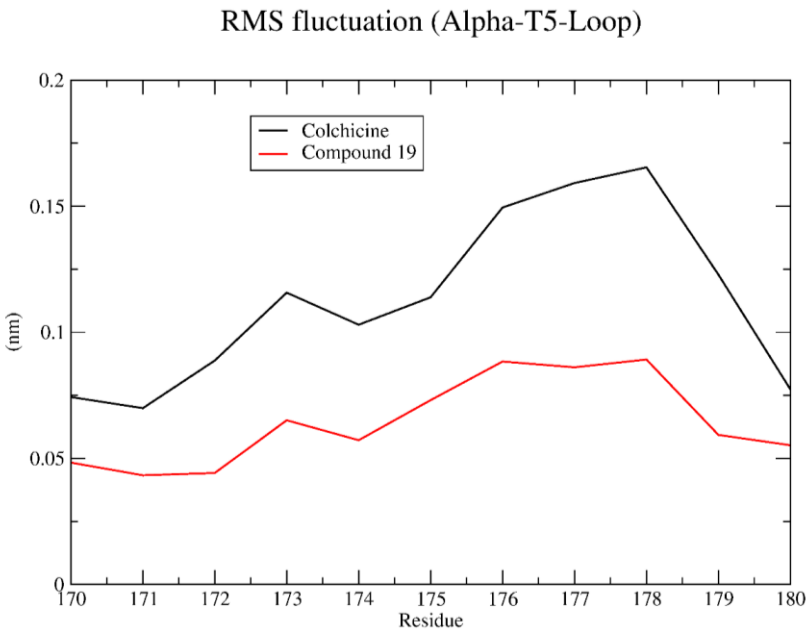


Figure 34: Root-Mean-Square-Fluctuations (RMSF) of the α-T5 loop for the isotype αβIII bound to the colchicine (black) and the compound C19 (red). It is evident that C19 reduces the fluctuations of this loop, compared to the colchicine.

5.3.2 Binding free energy calculation

Binding free energy calculation is quantified in the following figure:

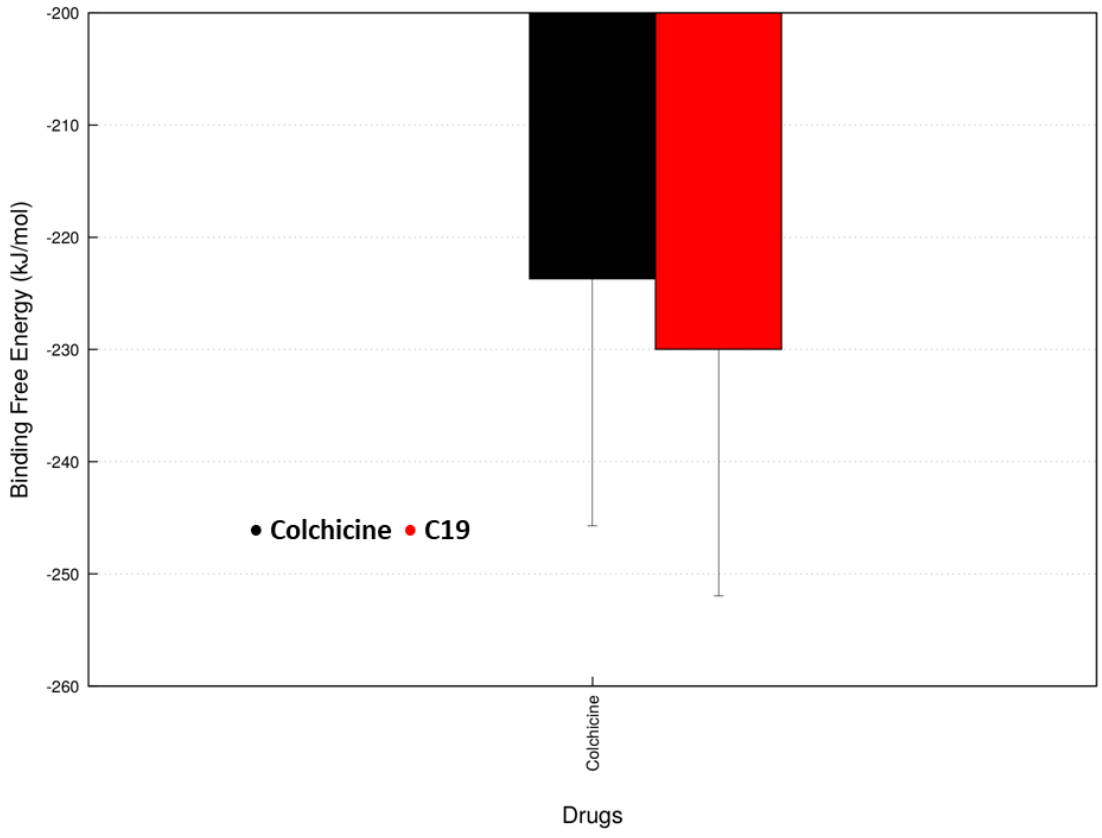


Figure 35: Binding Free Energy: the energy is mediated over the last 50 ns of simulation. Compound 19 shows a higher binding energy with isotype αβIII compared to the colchicine.

C19 shows a quite higher binding energy to isotype αβIII compared to the colchicine. To go deeper with the analysis, the energy decomposition over the residues in the cleft has been performed. The following figure shows the results for the residues in the cleft with the most significant energy contribution.

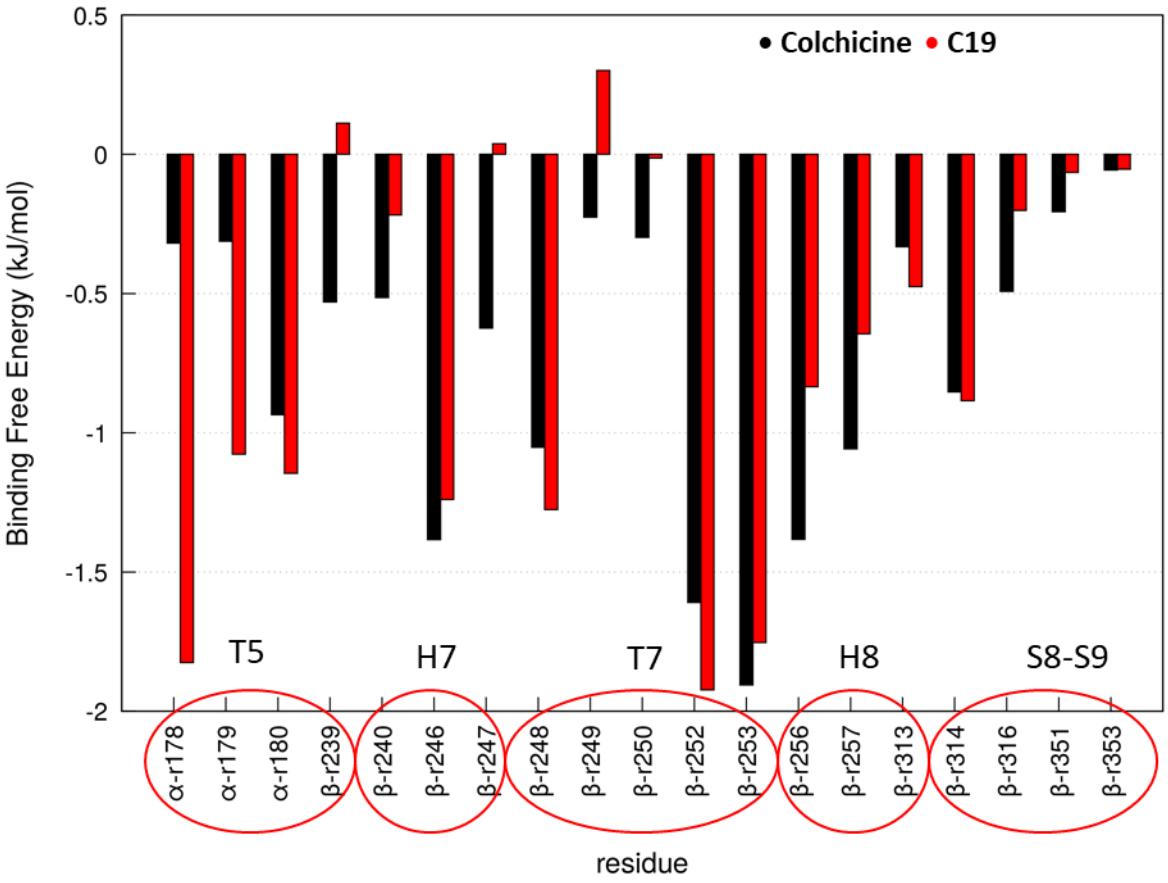


Figure 36: Binding Free Energy decomposition over the residues in the colchicine binding site. The residues reported are the ones with the higher energies. It is clear that compound C19 has a higher affinity for the α-T5-Loop, compared to the colchicine.

Residues from β-252 to β-257 (H8 helix) show a high binding energy for both the drugs, while the main difference can be observed in residues from α-178 to α-180 (T5 loop), where the compound C19 shows a higher binding energy compared to the colchicine. Looking at their positions in the cleft, it is clear that C19 pose favourites its interactions whit the T5 loop, while the colchicine is quite distant from it.

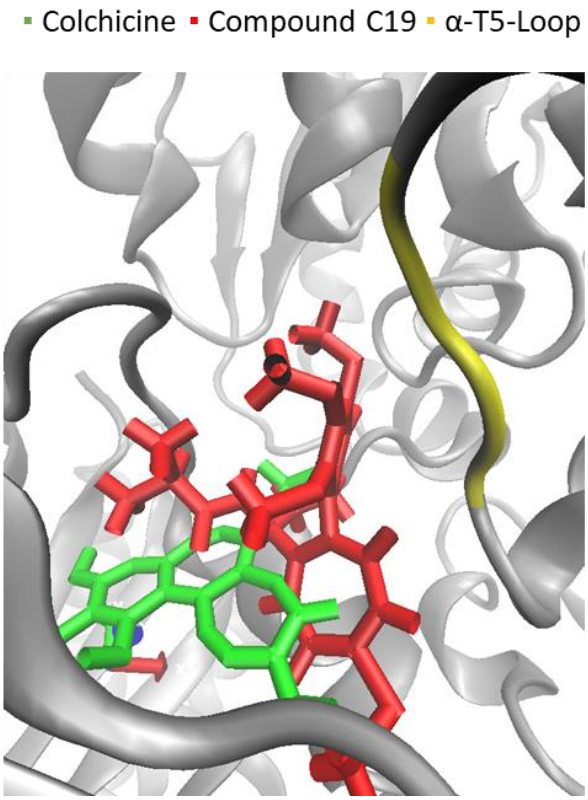


Figure 37: Colchicine and C19 positions inside the cleft. The colchicine is represented in green, while the compound C19 is coloured in red. It is evident that the different position of C19 makes it able to interact more with the α-T5-loop.

5.4 Discussion

One of the employed strategies to kill the cancerous cells is by using drugs able to selectively bind the microtubules of these cells, as they are fundamental in the cell's replication. The drugs have to be able to alter the dynamic instability[40] of the microtubule, moving the equilibrium either to enhanced polymerization or depolymerization. Among the depolymerizing compounds, colchicine has been one of the most studied during the years, but it has the side effects of being toxic and non-specific. In this study, a preliminary evaluation of one of its derivatives has been performed using Molecular Dynamics simulations, in order to understand if it could be a better drug than the colchicine. The compound, called C19, has been chosen among many derivatives after the analysis presented in the previous chapter. Tumoral cells show different tubulin isotypes in order to become more resistant to the drugs[85]. As they overexpress isotype αβIII compared to the non-tumoral cells[52], it is believed that this isotype could have a role in the drug-resistance[81], so in this study only isotype αβIII has been used. The simulations have been performed on the tubulin dimer bonded to both the colchicine and the new compound.

Figure 32 shows the Root-Mean-Square-Deviations (RMSDs) from their starting positions for both the drugs. A drug with a higher value of RMSD alters its position more than one with a lower value. Therefore, the colchicine has a major tendency to change its position and it indicates its tendency to go out the cleft, while the novel compounds may be more stable inside the pocket, has it shows a lower RMSD value. The different pose assumed by C19 (which can be seen in Figure 37) may help the drug to stay stable inside the cleft. This drug seems to interact more with the α-T5-loop than the colchicine (Figure 36) as its binding energy in this region is quite higher than the colchicine's one.

For what concerns the secondary structures analysis (Figure 33), the only difference is that the α-T5 loop has a little tendency to form a Beta-sheet, which is a more stable and ordered structure compared to a coil. This could be one of the reasons leading to the reduced Root-Mean-Square-Fluctuation (RMSF) of this region in the tubulin bonded to C19 (Figure 34). As longitudinal contacts are fundamental for the microtubule polymerization process[84] and α-T5 is fundamental in these contacts[85], these reduced fluctuations can enhance the microtubule depolymerization process, making this drug more effective than the colchicine (the fluctuation of the α-T5-loop are quite higher in the tubulin bonded to the colchicine). Plus, the Taxol seems to enhance the fluctuations of this loop[86]. As the Taxol is a microtubule stabilizer[55], the opposite effect of C19 on the α-T5-loop supports the thesis that it could be a powerful microtubule destabilizer.

5.5 Conclusions and Future Perspectives

In this chapter, the effects of the colchicine and the novel compound C19 have been analysed and compared on the human tubulin dimer αβIII. The aim was to test if C19 could have the potentiality to be a better anticancer drug than the colchicine. The results show that it has a higher destabilizing power on the microtubule, due to its interactions with the α-T5-loop.

Future studies may be needed to analyse the effect of this drug on other aspects of the microtubule dynamics, such as the bending angle and the fluctuations of other regions. It should also be tested on the other human tubulin isotypes in order to better prove its specificity.

6. Acknowledgments

I would like to show my gratitude to my supervisor prof. Marco Agostino Deriu for his constant support and help on this master thesis work. His positivity and enthusiasm motivated me during this work and increased my interest in the molecular modelling.

I would like to thank my co-supervisors Prof. Umberto Morbiducci and Prof. Jacek Adam Tuszynski for their precious advice and for their critical evaluation of my work.

I would also thank Gianvito Grasso for his advices and explanations, especially about the binding free energy calculation.

I am very thankful to my teammates, Stefano, Giacomo, Chiara, Simone and especially Lorenzo for sharing the bad and the good moments in this last year and for having built a solid and precious friendship.

I would like to thank all my friends: the ones in Ceccano, especially Christian, for supporting me even at long distance, my colleagues Maria, Mayra and Giovanni for being there for me since the beginning and my *"PokemonGo"* friends Melania, Marco, Danilo, Giulio and the others for constantly encouraging me and for the good time spent together.

I finally thank my family for having always believed in me and for providing me a great support during these years. In particular, I would like to thank my brother Pietro, for being my pillar of strength since we were born and for his constant help during this work.

Antonio Rocca

7. References

1. Crabbe, J. Molecular modelling: Principles and applications A.R. Leach. Adison Wesley Longman (1996). *Comput. Chem.* **21,** 185 (1997).

2. Combes, J. M., Duclos, P. & Seiler, R. The Born-Oppenheimer Approximation. in Rigorous Atomic and Molecular Physics 185–213 (Springer US, 1981). doi:10.1007/978-1-4613-3350-0_5. 185–213 (1981).

3. Vanommeslaeghe, K., Guvench, O. & MacKerell, A. D. Molecular Mechanics. *Curr. Pharm. Des.* **20,** 3281–3292 (2014).

4. Johnson, J. K., Zollweg, J. A. & Gubbins, K. E. The lennard-jones equation of state revisited. *Mol. Phys.* (1993). doi:10.1080/00268979300100411

5. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

6. Ding, H. Q., Karasawa, N. & Goddard, W. A. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97,** 4309–4315 (1992).

7. Tironi, I. G., Sperb, R., Smith, P. E. & Van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **102,** 5451–5459 (1995).

8. Fletcher, R. & Powell, M. J. D. A Rapidly Convergent Descent Method for Minimization. *Comput. J.* **6,** 163–168 (2012).

9. McCarthy, J. F. Block-conjugate-gradient method. *Phys. Rev. D* **40,** 2149–2152 (1989).

10. Verbeke, J. & Cools, R. The newton-raphson method. *Int. J. Math. Educ. Sci. Technol.* (1995). doi:10.1080/0020739950260202

11. Head, J. D. & Zerner, M. C. A Broyden-Fletcher-Goldfarb-Shanno optimization procedure for molecular geometries. *Chem. Phys. Lett.* (1985). doi:10.1016/0009-2614(85)80574-1

12. Grubmüller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized Verlet

Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. Mol. Simul. **6,** 121–142 (1991).

13. W.F., van G. & H.J.C.A., B. A leap-frog algorithm for stochastic dynamics. *Mol.Simul.* **1,** 173–185 (1998).

14. Martys, N. S. & Mountain, R. D. Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **59,** 3733–3736 (1999).

15. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3,** 198–210 (2013).

16. B.R. Brooks, C.L. Brooks, A.D. MacKerell Jr, L. Nilsson, R.J. Petrella, B., Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R., Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V., Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R. M. V. & H. L. Woodcock, X. Wu, W. Yang, D.M. York, and M. K. CHARMM: The Biomolecular Simulation Program. **30,** 1545–1614 (2010).

17. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91,** 43–56 (1995).

18. Berman, H. M. *et al.* The Protein Data Bank Helen. *Nucleic Acids Res.* **28,** 235–242 (2000).

19. Farady, C. J., Sellers, B. D., Jacobson, M. P. & Charles, S. Computational Drug Design. *Chem. Biol.* **6,** R227–R228 (1999).

20. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* (1997). doi:10.1006/jmbi.1996.0897

21. Krieger, E., Nabuurs, S. B. & Vriend, G. Homology Modeling. in *Structural Bioinformatics* (2005). doi:10.1002/0471721204.ch25

22. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual

screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery* (2004). doi:10.1038/nrd1549

23. Bolton, E. E., Wang, Y., Thiessen, P. A. & Bryant, S. H. Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. in *Annual Reports in Computational Chemistry* (2008). doi:10.1016/S1574-1400(08)00012-1

24. Hanwell, M. D. *et al.* Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* (2012). doi:10.1186/1758-2946-4-17

25. Castillo, O. & Aguilar, L. T. Genetic algorithms. in *Studies in Fuzziness and Soft Computing* (2019). doi:10.1007/978-3-030-03134-3_2

26. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science (80-.).* (1983). doi:10.1126/science.220.4598.671

27. Cervantes, M. The Monte Carlo Method. *Math. Sci. Eng.* (1972). doi:10.1016/S0076-5392(08)61352-1

28. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie - International Edition* (2002). doi:10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O

29. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10,** 449–461 (2015).

30. Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* (1994). doi:10.1016/S0959-440X(94)90321-2

31. Karplus, M. & Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* (1981). doi:10.1021/ma50003a019

32. Shirts, M. R., Mobley, D. L. & Brown, S. P. Free-energy calculations in structure-based drug design. in *Drug Design: Structure- and Ligand-Based Approaches* (2010).

33. Leach, A. R. *Molecular Modelling: Principles and Applications (2nd Edition)*.

Computers (2001). doi:10.1016/S0097-8485(96)00029-0

34. Clark Still, W., Tempczyk, A., Hawley, R. C. & Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* (1990). doi:10.1021/ja00172a038

35. Garrett M. Morris, David S. Goodsell, Michael E. Pique, William "Lindy" Lindstrom, Ruth Huey, Stefano Forli, William E. Hart, Scott Halliday, R. B. and A. J. O. AutoDock Version 4.2. *Citeseer* (2012).

36. Trott, O. & Olson, A. Autodock vina: improving the speed and accuracy of docking. *J. Comput. Chem.* **31,** 455–461 (2010).

37. Ewing, T. J. A., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided. Mol. Des.* (2001). doi:10.1023/A:1011115820450

38. Wolber, G. & Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **45,** 160–169 (2005).

39. Prota, A. E. *et al.* The novel microtubule-destabilizing drug BAL27862 binds to the colchicine site of tubulin with distinct effects on microtubule organization. *J. Mol. Biol.* (2014). doi:10.1016/j.jmb.2014.02.005

40. Mitchison, T., Kirschner, M. & Mitchison T, K. M. Dynamic instability of microtubule growth. *Nature* (1984). doi:10.1038/312237a0

41. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the αβ tubulin dimer by electron crystallography. *Nature* (1998). doi:10.1038/34465

42. Deriu, M. A. *et al.* Anisotropic elastic network modeling of entire microtubules. *Biophys. J.* **99,** 2190–2199 (2010).

43. Grafmüller, A., Noya, E. G. & Voth, G. A. Nucleotide-dependent lateral and longitudinal interactions in microtubules. *J. Mol. Biol.* (2013). doi:10.1016/j.jmb.2013.03.029

44. Buey, R. M., Fernando Díaz, J. & Andreu, J. M. The nucleotide switch of tubulin and microtubule assembly: A polymerization-driven structural change. *Biochemistry*

(2006). doi:10.1021/bi060334m

45. Gigant, B. *et al.* The 4 Å X-ray structure of a tubulin: Stathmin-like domain complex. *Cell* (2000). doi:10.1016/S0092-8674(00)00069-6

46. Löwe, J., Li, H., Downing, K. H. & Nogales, E. Refined structure of αβ-tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313,** 1045–1057 (2001).

47. Nettles, J. H. *et al.* The binding mode of epothilone A on α,β-tubulin by electron crystallography. *Science (80-.).* (2004). doi:10.1126/science.1099190

48. Ravelli, R. B. G. *et al.* Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. *Nature* (2004). doi:10.1038/nature02393

49. Gigant, B. *et al.* Structural basis for the regulation of tubulin by vinblastine. *Nature* (2005). doi:10.1038/nature03566

50. Alushin, G. M. *et al.* High-Resolution microtubule structures reveal the structural transitions in αβ-tubulin upon GTP hydrolysis. *Cell* (2014). doi:10.1016/j.cell.2014.03.053

51. Huzil, J. T., Ludueña, R. F. & Tuszynski, J. Comparative modelling of human β tubulin isotypes and implications for drug binding. *Nanotechnology* **17,** (2006).

52. Leandro-García, L. J. *et al.* Tumoral and tissue-specific expression of the major human β-tubulin isotypes. *Cytoskeleton* **67,** 214–223 (2010).

53. Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nature Reviews Cancer* (2004). doi:10.1038/nrc1317

54. Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & Mcphail, A. T. Plant Antitumor Agents.VI.The Isolation and Structure of Taxol, a Novel Antileukemic and Antitumor Agent from Taxus brevifolia2. *J. Am. Chem. Soc.* (1971). doi:10.1021/ja00738a045

55. Schiff, P. B. & Horwitz, S. B. Taxol stabilizes microtubules in mouse fibroblast cells. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.77.3.1561

56. Rao, S., Orr, G. A., Chaudhary, A. G., Kingston, D. G. I. & Horwitz, S. B. Characterization of the taxol binding site on the microtubule: 2-(m-

azidobenzoyl)taxol photolabels a peptide (amino acids 217-231) of β-tubulin. *J. Biol. Chem.* (1995). doi:10.1074/jbc.270.35.20235

57. Rao, S. *et al.* Characterization of the Taxol binding site on the microtubule. Identification of Arg282 in β-tubulin as the site of photoincorporation of a 7-benzophenone analogue of Taxol. *J. Biol. Chem.* (1999). doi:10.1074/jbc.274.53.37990

58. Rao, S. *et al.* 3'-(p-Azidobenzamido)taxol photolabels the N-terminal 31 amino acids of β-tubulin. *J. Biol. Chem.* (1994).

59. Snyder, J. P., Nettles, J. H., Cornett, B., Downing, K. H. & Nogales, E. The Binding Conformation of Taxol in β-tubulin: A Model Based on Electron Crystallographic Density. *Proc. Natl. Acad. Sci.* **98,** 5312–5316 (2001).

60. Moudi, M., Go, R., Yien, C. Y. S. & Nazre, M. Vinca alkaloids. *International Journal of Preventive Medicine* (2013).

61. Lee, C.-T., Huang, Y.-W., Yang, C.-H. & Huang, K.-S. Send Orders for Reprints to reprints@benthamscience.ae Drug Delivery Systems and Combination Therapy by Using Vinca Alkaloids. *Curr. Top. Med. Chem.* **15,** 1491–1500 (2015).

62. Cormier, A., Marchand, M., Ravelli, R. B. G., Knossow, M. & Gigant, B. Structural insight into the inhibition of tubulin by vinca domain peptide ligands. *EMBO Rep.* **9,** 1101–1106 (2008).

63. Spasevska, I. *et al.* Modeling the Colchicum autumnale tubulin and a comparison of its interaction with colchicine to human tubulin. *Int. J. Mol. Sci.* **18,** (2017).

64. Tripathi, S. *et al.* Insight into microtubule destabilization mechanism of 3,4,5-trimethoxyphenyl indanone derivatives using molecular dynamics simulation and conformational modes analysis. *J. Comput. Aided. Mol. Des.* **32,** 559–572 (2018).

65. Weick, J. K., Livingston, R. B. & Van Slyck, E. J. Colchicine in refractory chronic lymphocytic leukemia. *Invest. New Drugs* (2004). doi:10.1007/bf00177418

66. Dumont, R., Brossi, A., Chignell, C. F., Quinn, F. R. & Suffness, M. A Novel Synthesis of Colchicide and Analogues from Thiocolchicine and Congeners: Reevaluation of Colchicide as a Potential Antitumor Agent. *J. Med. Chem.* (1987).

doi:10.1021/jm00387a028

67. Lu, Y., Chen, J., Xiao, M., Li, W. & Miller, D. D. An overview of tubulin inhibitors that interact with the colchicine binding site. *Pharm. Res.* **29,** 2943–2971 (2012).

68. Kumbhar, B. V., Borogaon, A., Panda, D. & Kunwar, A. Exploring the Origin of Differential Binding Affinities of Human Tubulin Isotypes αβII, αβIII and αβIV for DAMA-Colchicine Using Homology Modelling, Molecular Docking and Molecular Dynamics Simulations. *PLoS One* **11,** e0156048 (2016).

69. Johnson, L. *et al.* Novel Colchicine Derivatives and their Anti-cancer Activity. *Curr. Top. Med. Chem.* **17,** 2538–2558 (2017).

70. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* (2006). doi:10.1016/j.jmgm.2005.12.005

71. Sousa Da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* (2012). doi:10.1186/1756-0500-5-367

72. Fiser, A. & Šali, A. MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Methods Enzymol.* (2003). doi:10.1016/S0076-6879(03)74020-8

73. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15,** 2507–2524 (2006).

74. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47,** D506–D515 (2019).

75. Heinig, M. & Frishman, D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* (2004). doi:10.1093/nar/gkh429

76. Sun, H. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* **20,** 14450–14460 (2018).

77. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods: II. The accuracy of ranking poses generated from

docking. *J. Comput. Chem.* **32,** 866–877 (2011).

78. Su, P. C., Tsai, C. C., Mehboob, S., Hevener, K. E. & Johnson, M. E. Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (FabI). *J. Comput. Chem.* **36,** 1859–1873 (2015).

79. Santoshi, S. & Naik, P. K. Molecular insight of isotypes specific β-tubulin interaction of tubulin heterodimer with noscapinoids. *J. Comput. Aided. Mol. Des.* **28,** 751–763 (2014).

80. Messaoudi, A., Belguith, H. & Ben Hamida, J. Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. *Theor. Biol. Med. Model.* **10,** 1–10 (2013).

81. Winter, P. *et al.* Quantitative analysis of the effect of tubulin isotype expression on sensitivity of cancer cell lines to a set of novel colchicine derivatives. *Mol. Cancer* **9,** 131 (2010).

82. Huzil, J. T., Chen, K., Kurgan, L. & Tuszynski, J. A. The roles of ??-tubulin mutations and isotype expression in acquired drug resistance. *Cancer Inform.* **3,** 159–181 (2007).

83. Banerjee, A. & Luduena, R. F. Kinetics of colchicine binding to purified ??-tubulin isotypes from bovine brain. *J. Biol. Chem.* **267,** 13335–13339 (1992).

84. Manka, S. W. & Moores, C. A. The role of tubulin–tubulin lattice contacts in the mechanism of microtubule dynamic instability. *Nat. Struct. Mol. Biol.* **25,** 1–9 (2018).

85. Natarajan, K. & Senapati, S. Understanding the basis of drug resistance of the mutants of αβ-tubulin dimer via molecular dynamics simulations. *PLoS One* **7,** (2012).

86. Mitra, A. & Sept, D. Taxol allosterically alters the dynamics of the tubulin dimer and increases the flexibility of microtubules. *Biophys. J.* **95,** 3252–3258 (2008).