

Politecnico di Torino

Department of Mechanical and Aerospace Engineering



Master's degree in biomedical engineering

Computational evaluation of a panel of combretastatin compounds as warheads for PROTAC design targeting select beta tubulin isoforms

Supervisors

Prof. J. A. TUSZYNSKI

Prof. M. A. DERIU

Candidate

Benedetta MORABITO

Academic Year 2021/2022

*“To my aunt Roberta,
for being the inspiration of my life.”*

Summary

A new strategy for chemotherapy-based treatment of various types of cancer is based on the design of PROTAC molecules, which are hybrid chemical structures aiming to degrade over-expressed proteins via ubiquitination pathways in cancer cells.

They consist of a warhead that binds to an over-expressed protein marked for degradation, a linker and a compound that binds to an E3 ligase enzyme.

A panel of recently synthesized Combretastatin A4 derivatives with attractive pharmacological properties have been investigated in this project. They were docked to the main β tubulin isotypes and their binding affinities were compared. The ADMET properties of these compounds were analyzed and used to search for a correlation with the biological activity of these derivatives, which were previously tested *in vitro* to assess their cytotoxicity. The best linear correlation was observed between the binding energy obtained by docking procedure and the logIC50 gained experimentally.

In a cell, all isotypes of β tubulin are present simultaneously, with different expression levels depending on tissue type and whether the tissue is cancerous or healthy. To carry out the regressions, an attempt was made to estimate the binding energy weighted on the isotypes expressed in that specific cell line.

Models were created for seven cancer cell lines whose experimental results had been provided. The results show that taking into account the correct expression of isotypes in different cell lines influences and improves the regression models, compared with the model created by considering a uniform distribution of the most common isotypes.

From the scores gained through docking, it seems that none of these compounds seems to exhibit selectivity for a specific isotype.

The two best compounds, MJ-CA4-I-005 and MJ-CA4-II-009, were chosen based on the results of the experimental assays and were evaluated as possible warheads for a PROTAC structure that aims to bind all β tubulin isotypes indiscriminately, thus as a general antimetabolic drug.

Nevertheless, from the computational molecular dynamics experiments it can be stated that the ligands, MJ-CA4-I-005 and MJ-CA4-II-009, bind the protein stably

over time, furthermore from the results obtained by calculating the binding free energy, ΔG_{bind} , using the MMGBSA technique, which is more time-consuming and computationally expensive, but more accurate it appears that one of the two ligands, namely the MJ-CA4-II-009, exhibits some selectivity on β III, thereby paving the way for a possible PROTAC structure targeting β III, which is over-expressed in some cancer cell types.

Acknowledgements

My sincere thankfulness goes to my supervisor, Prof. J. A. Tuszynski, for the wonderful opportunity he gave me and for his constant support.

A special thanks goes to Dr. Maral Aminpour and Dr. Mahshad Moshari for their valuable advice and kind availability.

Thanks to Paola for her willingness to help and support me in this project and in this last part of the journey.

Table of Contents

List of Tables	XI
List of Figures	XIII
1 Introduction	1
1.1 Computer-aided drug design (CADD)	1
1.2 Biological background	3
1.3 Antimitotic drugs	5
1.3.1 <i>Colchicine</i> and <i>Combretastatin A-4</i>	7
1.4 PROTAC	8
2 Materials and Methods	13
2.1 Novel Combretastatin A-4 derivatives	13
2.2 Expression of β tubulin isotypes	16
2.3 ADMET properties	18
2.3.1 Absorption	19
2.3.2 Distributions	20
2.3.3 Metabolism	22
2.3.4 Excretion	22
2.3.5 Toxicity	23
2.4 Experimental data	23
2.5 Docking	27
2.6 Regression	30
2.7 Molecular Dynamics Simulations	31
2.8 MMGB/PBSA	33
3 Results	35
3.1 Isotypes expression	35
3.2 ADMET properties analysis	37
3.3 Docking results and Regression models	49
3.4 PROTAC	81

3.4.1	Consensus docking	82
3.4.2	Molecular dynamics simulations	86
3.4.3	MMPB/GBSA	89
4	Conclusions and further developments	91
	Bibliography	93

List of Tables

2.1	Genes expressing isotypes of β -tubulin.	17
2.2	Cytotoxicity (IC50, μ M) of studied compounds estimated by MTT assay.	25
2.3	Cytotoxicity (IC50, μ M) of studied compounds estimated by SRB assay.	26
3.1	Properties affecting absorption.	38
3.2	Lipinski's rules of five color code.	40
3.3	Properties that influence the distribution of derivatives.	41
3.4	Metabolism Risk and Code. CYP Risk rule codes: 1A2= high 1A2 clearance, 2C19= high 2C19 clearance, 2C9= high 2C9 clearance, 2D6= high 2D6 clearance, 3A4= high 3A4 clearance, CL= high microsomal clearance.	43
3.5	Excretion related properties.	44
3.6	Toxicity properties.	45
3.7	Toxicity properties.	47
3.8	Overall indicator.	48
3.9	Autodock docking score and cluster's range.	50
3.10	Cluster size.	51
3.11	Binding energy weighted on β tubulin isotype expression for A549 cell line.	52
3.12	Binding energy weighted on β tubulin isotype expression for HaCaT cell line.	52
3.13	Binding energy weighted on β tubulin isotype expression for HepG2 cell line.	53
3.14	Binding energy weighted on β tubulin isotype expression for MDA cell line.	53
3.15	Binding energy weighted on β tubulin isotype expression for PC3 cell line.	54
3.16	Binding energy weighted on β tubulin isotype expression for SW480 cell line.	54

3.17	Binding energy weighted on β tubulin isotype expression for SW620 cell line.	55
3.18	Linear Regression models: R^2 value. BE_w : weighted binding energy, MLogP: logarithm of the partition coefficient, TPSA: Topological polar surface area, HBD: hydrogen bond donors, POL: polarizability.	80
3.19	Multivariate linear regression models: R^2 value	81
3.20	Consensus docking results.	82

List of Figures

1.1	Workflow of drug discovery [4].	2
1.2	Microtubule nucleation[8]	4
1.3	GTP cap models[9]	5
1.4	Antimitotic drug binding sites [11].	6
1.5	Structure of Colchicine [14].	8
1.6	Structure of Combretastatin-A4 [16].	8
1.7	Mechanism of operation of PROTAC [17]	9
1.8	Summary of recruited E3 ubiquitin ligases, ligase ligands, target protein, and ligands for targeted proteins. [20]	11
2.1	Novel Combretastatin A-4 derivatives.	16
2.2	β isotypes expression in normale tissues [22].	18
2.3	β isotypes expression in tumoral tissues [22].	18
2.4	Lipinski's rule of five [26]	21
2.5	Steps to perform docking [27].	27
2.6	Docking mode for flexible docking or rigid docking[29].	29
3.0	Isotypes expression of β tubulins.	37
3.1	Regression models between weighted binding energy and LogIC50.	58
3.2	Regression models between MlogP and LogIC50.	62
3.3	Regression models between TPSA and LogIC50.	65
3.4	Regression models between HBD and LogIC50.	69
3.5	Regression models between Polarizability and LogIC50.	72
3.6	Multivariate regression models: TPSA, Binding Energy and LogIC50.	76
3.7	Multivariate regression models: HBD, Binding Energy and LogIC50.	80
3.8	β IIa and MJ-CA4-I-005 interactions.	83
3.9	β IIa and MJ-CA4-II-009 interactions.	83
3.10	β III and MJ-CA4-I-005 interactions.	84
3.11	β III and MJ-CA4-II-009 interactions.	84
3.12	β IVb and MJ-CA4-I-005 interactions.	85
3.13	β IVb and MJ-CA4-II-009 interactions.	85

3.14	Molecular dynamics simulation results: β IIa and MJ-CA4-I-005. . .	86
3.15	Molecular dynamics simulation results: β IIa and MJ-CA4-II-009. . .	87
3.16	Molecular dynamics simulation results: β III and MJ-CA4-I-005. . .	87
3.17	Molecular dynamics simulation results: β III and MJ-CA4-II-009. . .	88
3.18	Molecular dynamics simulation results: β IVb and MJ-CA4-I-005. . .	88
3.19	Molecular dynamics simulation results: β IVb and MJ-CA4-II-009. . .	89
3.20	Binding free energy, ΔG_{bind} , calculated via the MMGBSA.	90

Chapter 1

Introduction

1.1 Computer-aided drug design (CADD)

The process involving the discovery and design of new drugs is time consuming, expensive and intricate. From idea to commercialization, the conventional drug development cycle takes 10 to 15 years and requires between 0.8 and 1.0 billion USD.

In order to enhance this process, several strategies have been created to shorten the research cycle, lower costs, and reduce failure risk in the search for novel drugs. One of the most efficient ways to accomplish these aims is through computer-aided drug design (CADD).

The term CADD is frequently used to refer to computational resources and techniques for the management, collection, investigation, and modelling of compounds. It encompasses a wide range of features of drug discovery, such as the creation of digital libraries of chemical compounds, tools for evaluating possible lead candidates and computer algorithms for examining chemical interactions [1].

The development of big data in biological, chemical, and pharmaceutical medicine has recently led to the optimization and application of a number of machine learning algorithms in the computer-aided drug design field. The efficiency of the drug design and discovery process can be significantly increased because of this integration. Only when reliable and accurate pre-processed data are combined with efficient computational methodologies and tools can advances in drug design, discovery, and development be successful.

With the development of computing and analytic approaches integrating big data and AI algorithms, the limits in the conventional drug research area brought on by the complexity of biological data may be computationally expressed and solved. Building automated models to assess protein three-dimensional structures, drug-receptor interactions and absorption, distribution, metabolism, excretion and

toxicity (ADMET) property prediction requires the integration of big data and artificial intelligence (AI) methodologies including pre-processing data, applications of AI algorithms, and statistical methods [2].

For several years, the process of drug design was based on empirical pharmacology, in which natural extracts were tested and if they proved to have the desired effect, allowed the identification of active chemicals and the eventual creation of new drugs. Regardless of the fact that this method has produced a number of drugs for pharma companies, it has significant limitations since it is sometimes difficult for drug optimization to increase a drug's tolerability, effectiveness and synergisms with other drugs when the biological target is unknown [3].

The fundamental techniques of CADD can be considered as the following:

- Homology modelling
- Molecular docking
- Molecular dynamics simulations
- Virtual screening (VS) or virtual high-throughput screening (vHTS)
- Quantitative structure-activity relationship (QSAR)
- (3D) Pharmacophore model (3D)

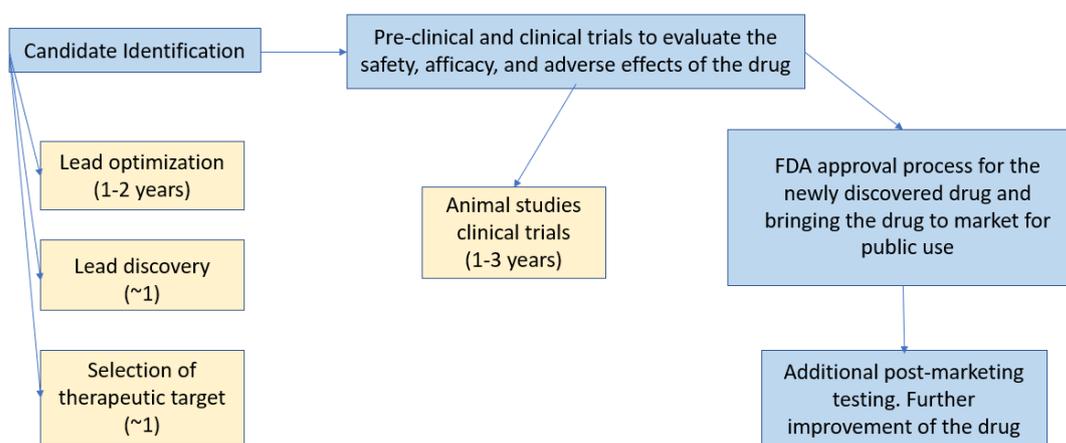


Figure 1.1: Workflow of drug discovery [4].

The workflow (Fig.1.1) shows the various phases of the clinical research cycle as well as the associated time needed to complete each step. The present advancement

in multidisciplinary research is another factor contributing to the rise in the adoption of the CADD method to drug development. Nevertheless, it is difficult to implement a standard methodology for using these techniques because of the variety and complexity of biological systems. For this reason, they are used in various combinations and at various phases of the research process, according to the intended outcome. The decision of which software programme or algorithm to employ is heavily influenced by a number of variables and the expected results since different tools can be used in every step of pre-clinical drug development. In order to evaluate and cross-validate the computational model, it is essential to calibrate theoretical findings against experimental data. The computational model or process under consideration may be modified if theoretical and experimental data do not correlate [5].

1.2 Biological background

One of the most important components of eukaryotic cells is the cytoskeleton, which is constituted of microtubules (MTs), intermediate filaments, and microfilaments. Among these elements, MTs have a role in chromosomal segregation during mitosis in addition to provide mechanical support, intracellular transport, and cell motility. MTs are typically metastable structures, during the cell cycle they can be quickly disassembled or also be strongly stabilised for a variety of purposes. Thus, within the cell, the degree of stability is carefully adjusted and modulates in time and space.

Several microtubules functions, such as chromosome segregation during mitosis, cell motion, and intracellular transport, depend on interactions between tubulin and other proteins; many of these interactions also have an influence on microtubules stability. A key research objective in the cytoskeleton area is to comprehend the interactions that play a role in the control of microtubules assembly and disassembly, both between tubulin molecules and between tubulin and its ligands.

Microtubules are made by self-assembly of $\alpha\beta$ -tubulin heterodimers, they are arranged head-to-tail to create protofilaments, which associate laterally to form a polymer with a cylindrical shape.

α and β tubulin are made up of about 450 amino acids and are broadly similar to each other, both monomers bind a guanine nucleotide. While the nucleotide that binds β tubulin is known as exchangeable, the nucleotide that binds α tubulin is known as non-exchangeable. The hydrolysis of the GTP on β tubulin, which is required for microtubule polymerization, enables the attachment of a dimer to the microtubule's end, making it non-exchangeable while the tubulin-GTP cap at the end preserves stability [6][7]. To regulate microtubule nucleation, polymerization, catastrophe, severing, stabilisation, and transportation, cells have

evolved sophisticated systems.

The nucleation centers of the microtubules consist of the γ tubulin assembly, called the γ -tubulin ring complexes (γ -TuRCs), this structure, formed by helically arranged gamma tubulin, influences the polarity of the microtubule and the *head-to-tail* arrangement of the α and β tubulin dimer, in fact, the α tubulin is always positioned in contact with the γ tubulin and constitutes the minus end and the outwardly exposed β tubulin constituting the plus end, furthermore, this arrangement promotes lateral interaction between the dimers during protofilament formation in the most energetically favorable manner [8].

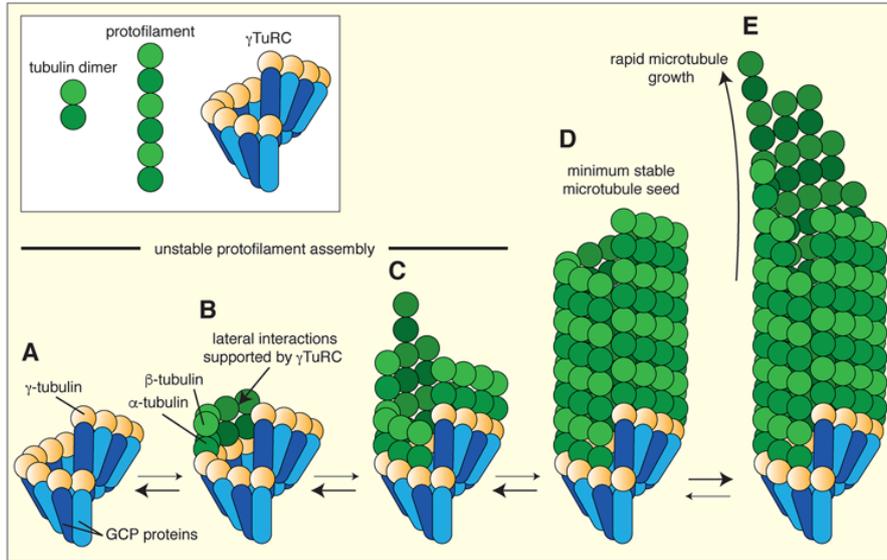


Figure 1.2: Microtubule nucleation[8]

In an attempt to understand how the tubulin-bound nucleotide regulates the structure and function of the protein, hence regulating microtubule dynamics, two theories have been considered so far. Both models show that GDP-bound tubulin dimers cannot polymerize but GTP-bound tubulin dimers may. GTP-tubulin is used to cap a developing microtubule, however the polymerization triggers the *GTPase* activity on tubulin and it progressively converts to GDP-tubulin.

The allosteric model suggests that GTP-tubulin dimers are straight or even more straight than GDP-tubulin dimers. The shape of the whole tubulin dimer is altered allosterically by the hydrolysis of the tubulin-bound GTP in the microtubule lattice, turning it from a straight to a curved minimal energy conformation. However, GDP-tubulin is unable to bend when it is inside a microtubule. Tubulin is constrained to an essentially straight form by the links between each subunit and its neighbours. As a result the microtubule wall with the GDP-tubulin is stressed and this causes instability. Removing the cap of GTP-tubulin microtubule depolymerization occurs

because of the GDP-tubulin at the end of the microtubule, which is able to relax to its curved minimal energy form, disintegrating the microtubule. The lattice model proposes that both GTP-bound and GDP-bound tubulins are supposed to be curved and their integration into the microtubule lattice is what causes them to become straight. According to this hypothesis, tubulin dimers' lateral or longitudinal linkages, flexibility, or bending stiffness are altered when the tubulin-bound GTP is hydrolyzed, which results in lower lattice stability [9].

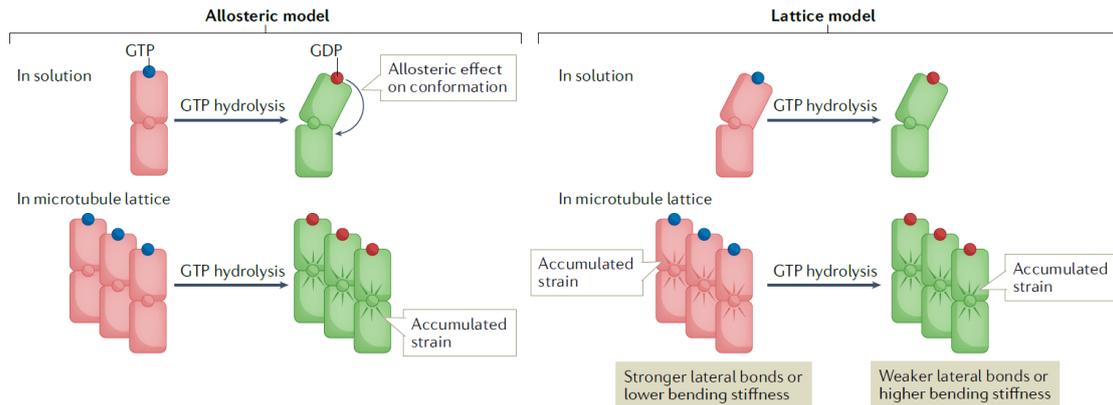


Figure 1.3: GTP cap models[9]

1.3 Antimitotic drugs

Cancer is a condition in which body's cells grow out of control and spread to other bodily regions. Human cells often divide among themselves to create new cells when the body requires them.

This process is known as cell growth and proliferation. New cells replace old ones when they die as a result of ageing or injury. Occasionally, this systematic process fails, causing damaged or dysfunctional cells to proliferate when they shouldn't. Tumors, which are tissue masses, can develop from these cells.

Some types of tumours can move to distant parts of the body to produce new tumours or invade neighbouring tissues, so this process is called metastasis [10].

One of the most important characteristics of tumors is their high proliferative potential, which can be considered a successful target for developing treatments, thus with this aim the antimitotic drugs were investigated.

This specific type of chemotherapeutic agents affects microtubule dynamics, since the mitotic spindle is made by microtubules, which are crucial for mitosis, tubulin has long been a typical target for chemotherapy.

Taxanes (such as *paclitaxel* and *docetaxel*) and *vinca alkaloids* (such as *vinblastine*

and *vincristine*) are the key components of chemotherapy strategies that target tubulin. Breast, ovarian, prostate, non-small-cell lung cancer, and Kaposi's sarcoma are all treated with *paclitaxel*.

Vinblastine is a drug used in the treatment lymphomas and leukaemia. *Colchicine* and, more recently, *combretastatin*, *laulimalide*, *peloruside*, and *noscipine* have also been shown to bind to tubulin [11].

The chemical compounds mentioned above differ from each other in their mechanism of action and in the position in which they bind the $\alpha\beta$ -tubulin dimer.

The figure below (Fig. 1.4) shows the binding sites on microtubules.

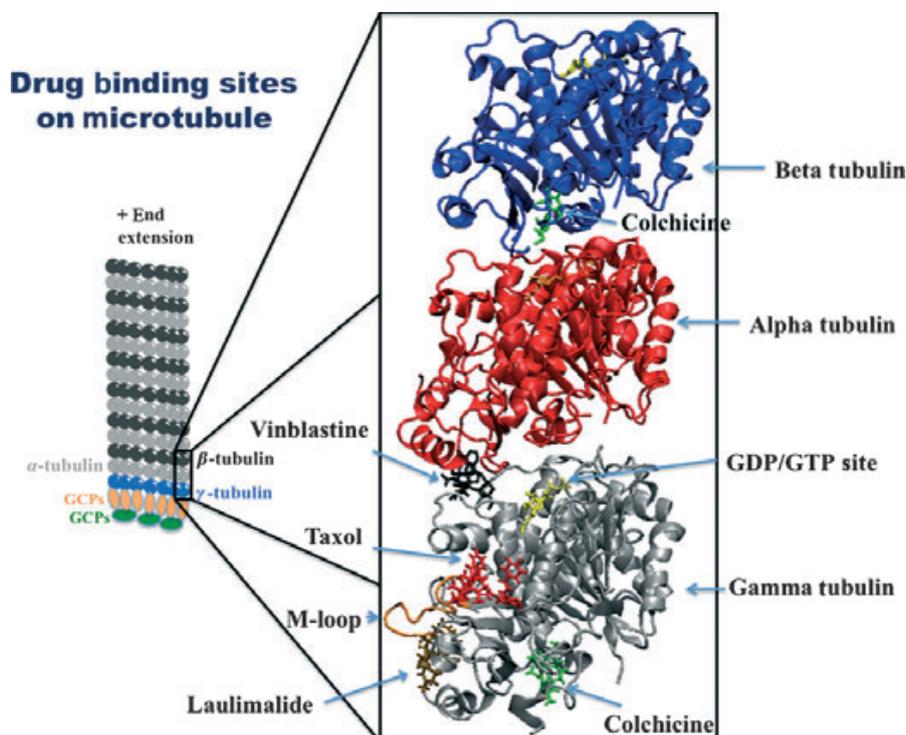


Figure 1.4: Antimitotic drug binding sites [11].

According to their mode of action, drugs that affect microtubules can be classified as either microtubule-destabilizing agents or microtubule-stabilizing agents.

When used in high concentrations, destabilising substances prevent microtubule polymerization, an example of this type of compounds is *Taxanes*. Drugs that, when given in high doses, promote microtubule polymerization maintain the structure of the tubules and prevent Ca^{2+} - or cold-induced depolymerization and subsequent disassembly, an example of this kind of agents is *Vinca alkaloids* and *Colchicine*. A challenge in both clinical and fundamental research to date is the cytotoxicity to

non-tumorigenic cells and the various cancer resistance established in response to these drugs[12].

1.3.1 *Colchicine and Combretastatin A-4*

Colchicine represents one of the groups of agents that interfere with microtubules, unlike the other substances, *Colchicine* binds β tubulin within the dimer, thus at the interface between α and β tubulin. The binding and consequently the presence of *Colchicine* represents a steric obstruction that results in an increased curvature of the dimer, which loses the ability to maintain a straight structure and thereby polymerize and form the microtubule.

Colchicine

Colchicine was the first tubulin destabilizing substance and was derived from the lethal meadow saffron *Colchicum autumnale L.* It was used for many years as an unlicensed drug to treat Behçet's disease, familial Mediterranean fever, gout, and pericarditis. It was approved by the U.S. Food and Drug Administration (FDA) in 2009 as a treatment option for acute gout and familial Mediterranean fever.

Since *Colchicine* can effectively inhibit mitosis and cancer cells undergo mitosis at a substantially higher rate than normal cells, making them more vulnerable to *Colchicine*.

On the other side, *Colchicine*'s low therapeutic index limits its effectiveness as a cancer treatment. Neutropenia, gastrointestinal distress, bone marrow injury, and anaemia are among its toxic side effects.

Despite the fact that *Colchicine* is not employed as an anticancer agent there have been several efforts to clinically produce Colchicine binding site drugs [13].

Combretastatin A-4

A group of stilbenoid phenols known as *Combretastatins* were discovered in the plant *Combretum caffrum*. The most powerful naturally occurring *Combretastatin* known in terms of both tubulin binding capacity and cytotoxicity is *Combretastatin A4* ((Fig. 1.6)[13].

Combretastatin A-4 (CA4) is the main compound in a relatively new class of vascular agents that affects existing cancers blood vessels and it might prevent angiogenesis. Angiogenesis, the process by which endothelial cells from pre-existing microvessels generate new blood vessels, is essential for the growth and spread of cancers. These new blood vessels develop inside the tumour and supply it with the nutrition, oxygen, and growth substances it needs to advance.

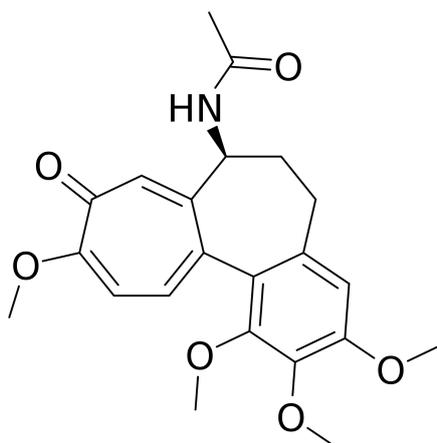


Figure 1.5: Structure of Colchicine [14].

Cancer treatment options include the potential therapeutic use of inhibiting angiogenesis. According to certain investigations, VEGF-induced proliferation, migration, and capillary-like tube formation were reduced by CA4. Microtubule-targeted drugs, including *taxanes*, *vinca alkaloids*, or *combretastatins*, have been shown to have anti-angiogenic effects. The tubulin-binding substance Combretastatin A4 (CA4) binds to the colchicine binding site. The effects of CA4 in preventing angiogenesis have lately attracted a lot of attention [15].

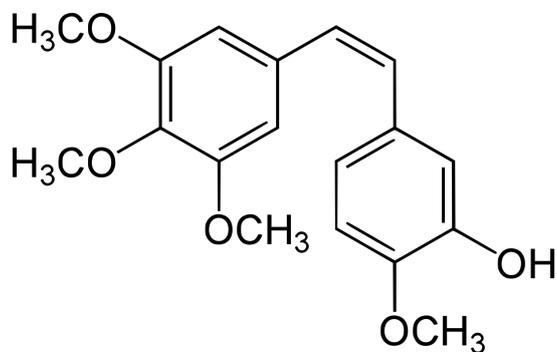


Figure 1.6: Structure of Combretastatin-A4 [16].

1.4 PROTAC

Because of their sensitivity to drug-resistant targets and different mode of action compared to those of standard inhibitors, proteolysis targeting chimeras (PROTACs)

have received a lot of attention from both academia and industry. Instead of blocking targets, PROTACs work by degrading the target protein. As a result, PROTACs might overcome several forms of resistance such target mutation or overexpression. Targeted therapy drug resistance has been a significant issue in contemporary research, particularly for cancer. Many characteristics are shared by the mechanisms of resistance to treatments that are intended to be selective for particular molecular targets, such as changes in the drug target, activation of prosurvival pathways, and unsuccessful induction of cell death.

PROTACs are a small bifunctional compounds that include a specific ligand for a desired target protein that is connected to a ligand for an E3 ubiquitin ligase by a linker.

It is the creation of a ternary complex, formed by a protein of interest (POI), PROTAC and E3 ubiquitin ligase, that enables ubiquitination and subsequent proteasome-mediated destruction of the target protein.

Actually just two degradants targeting the androgen receptor (AR) and estrogen receptor (ER), produced by Arvinas, are entering phase I clinical trials [17].

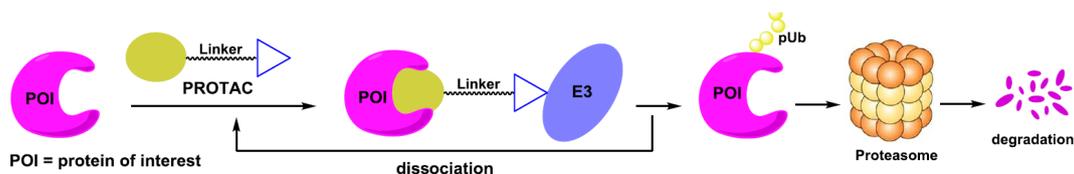


Figure 1.7: Mechanism of operation of PROTAC [17]

The major hurdles to using PROTACs in clinical settings continue to be their administration and bioavailability. The ideal drug would be able to target intracellular proteins as well, including the un-druggable proteome, have high oral bioavailability and selectivity, distribute well throughout a variety of tissues, possibly including the central nervous system (CNS), and have a catalytic mode of action that enables low exposures to be effective.

There are several reasons why the use of PROTAC can be considered advantageous over current treatments.

To date, very few proteins are treatable by current inhibitors, for the future of PROTACs, the possibility of interfering with proteins outside of the conventional target classes is particularly attractive. That is because the ligand for the protein of interest need not bind an active site of the protein and thus inactivate it, but can bind anywhere on the target, as long as it maintains the conformation that allows PROTAC to exert its function.

Thus, one of the major advantages of these new strategies is the ability to target the undruggable proteome.

In addition, it has been observed that a drug that binds to its target protein can lead to the accumulation of the latter because either the binding stabilizes the protein by extending its average lifetime or by activating an over-expression mechanism at the transcriptional level. Eliminating the target protein with a PROTAC is expected to be especially effective for proteins that could overexpress themselves or become protein stabilised in order to evade the effects of an inhibitor. While one of the most promising features of this technology is its capacity to target the undruggable proteome, additional features have emerged as a result of the design of these compounds. For example, it wasn't until these molecules began to exhibit a different degradation profile across proteins that bind the PROTAC that the opportunity to increase specificity by the extra lysine ubiquitination step was recognised. Where a ligand itself cannot be modified to be more selective, this discovery can be used to target a protein with selectivity [18].

PROTACs have a catalytic mechanism of action as opposed to conventional inhibitors, which drive target protein degradation through a competitive and occupancy-driven process, so it can induce target protein degradation at low concentration [19].

The potential, which are attributable to a PROTAC molecule's catalytic mode of action, is encouraging for the development of therapeutically useful PROTACs. The bioavailability and mode of administration are among the biggest obstacles that PROTACs must overcome. In the end, an orally accessible PROTAC would mark a significant advancement in the transformation of these molecules from a concept to a medicine. Understanding the therapeutic potential of PROTACs and developing them into effective medications are major research priorities [18].

Cerebion (CRBN) and VHL are now the most often used E3 ligases in the production of PROTACs, however MDM2, cIAP1, KEAP1, and RNF114 have also been used to varying degrees.

The theoretical safety risks associated with this modality should be considered as well. These risks include off-target degradation, intracellular accumulation of natural substrates for the ubiquitin proteasome system's E3 ligases, proteasome saturation by ubiquitinated proteins, and liabilities related to the "hook effect" of roteolysis-targeting chimeras.

For the POI to be degraded effectively, the PROTAC must first enter the cells and connect with the POI and the E3 ligase to form a ternary complex that causes the target to get ubiquitinated. The proteasome, one of the main cellular protein degradation mechanisms utilised by eukaryotic cells, degrades the POI once it has been ubiquitinated. It's significant to note that PROTAC molecules can engage in a novel process following the ubiquitination of the POI. This has substantial pharmacodynamic (PD) implications since PROTACs may now function catalytically, degrading several targets with only one PROTAC molecule. Additionally, the stability of PROTAC inside the cell and the speed of POI resynthesis will

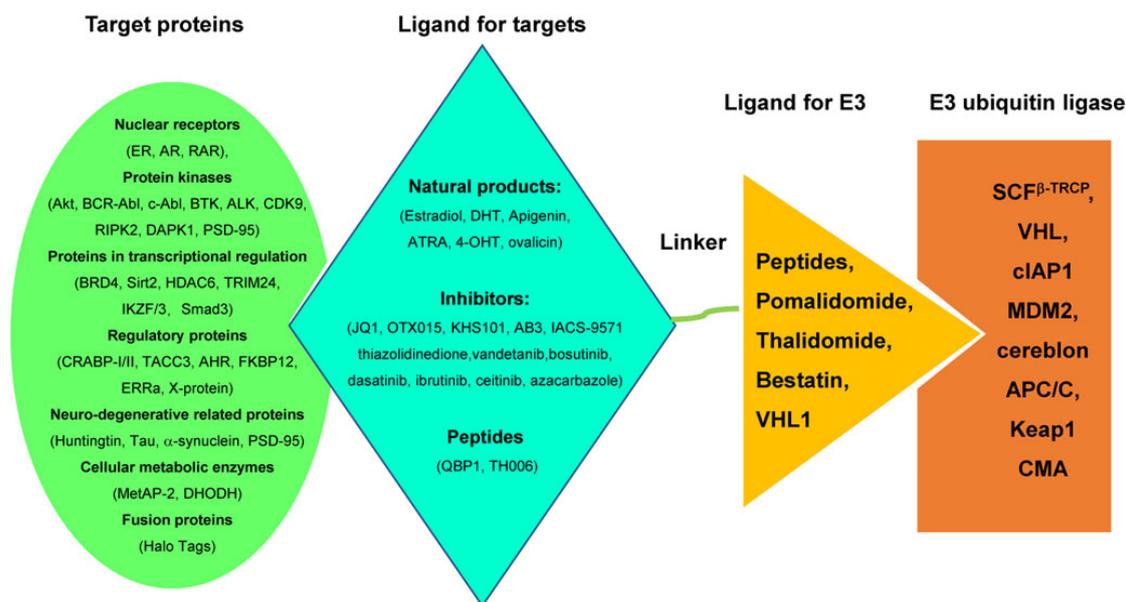


Figure 1.8: Summary of recruited E3 ubiquitin ligases, ligase ligands, target protein, and ligands for targeted proteins. [20]

determine how long the PROTAC mode of action will last. Any type of protein may be targeted by PROTAC provided that a small-molecule, low-MW ligand with the suitable affinity can be produced for the POI.

One of the problems related to the mode of operation of PROTAC is the degradation of other non-target proteins.

Another of the possible problems is that PROTAC-POI complexes compete with natural substrates for binding to the E3 ligase for ubiquitination and degradation, leading to accumulation of those substrates and possibly disrupting particular cellular pathways. PROTACs use an E3 ligase to initiate proteasomal degradation of a POI. Additionally, there is a chance that PROTACs will increase the amount of ubiquitinated proteins in the cell, saturating the proteasome and perhaps altering cellular homeostasis. Therefore, changes in proteasome activity and protein accumulation might have negative repercussions. The proteasome controls the cellular content of proteins that govern many aspects of cellular biology, including cell cycle, cell proliferation, immunological homeostasis, and metabolic activity. Moreover, When PROTAC concentrations are high, binding to the target and the E3 ligase become saturated, leading to the production of binary complexes rather than the useful ternary complex. This inhibits target ubiquitination and degradation, often known as the "hook" effect, which has been shown with numerous PROTACs.

By decreasing the hook effect, for instance, by enhancing the ternary complex's

cooperativity or protein-protein interaction, this problem might be avoided. The development of safer and more effective PROTAC molecules would be very useful for the use of this technique for non-oncologic or less severe diseases, where the required safety margins are much higher [21].

Chapter 2

Materials and Methods

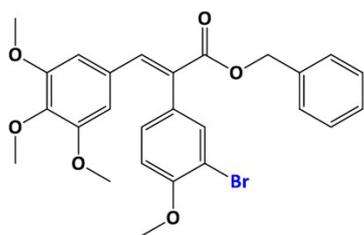
2.1 Novel Combretastatin A-4 derivatives

With the aim of evaluating new compounds, belonging to the category of microtubule interfering agents, with the intention that they will be more effective and less toxic than the parent compounds or other drugs already in use for cancer treatment, for the studies conducted in this thesis compounds derived from Combretastatin A-4 (CA4) were provided by M. Jedrzejczyk and A. Huczynski (private communication, 2022).

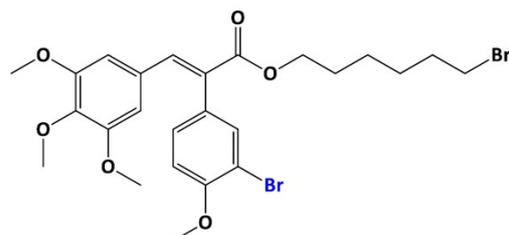
One of the most significant pockets for putative tubulin polymerization destabilizers is the colchicine binding site, to which CA4 interacts. The trimethoxyphenyl moiety, which is necessary for interaction with tubulin, is present in it.

The derivatives differ from the original structure of Combretastatin A4 by the Br in place of the carboxyl group in the second benzene ring and by the addition of different moieties in the part that connects the two benzene rings.

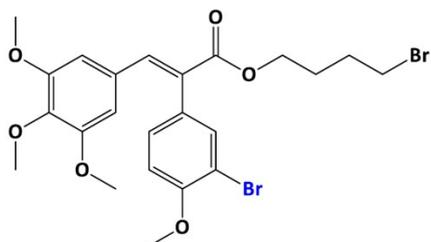
To create the 3D structure of the new compounds, CA4 derivatives were sketched in MOE using the *builder* tool, and the structures have been minimized and washed.



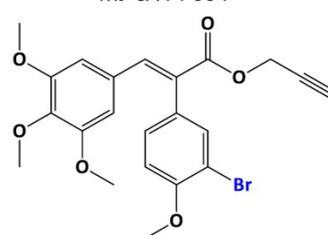
MJ-CA4-I-001



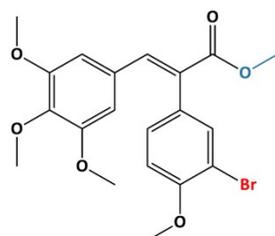
MJ-CA4-I-004



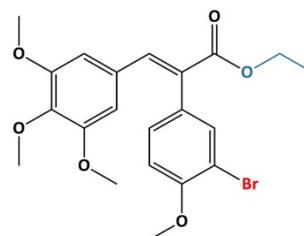
MJ-CA4-I-005



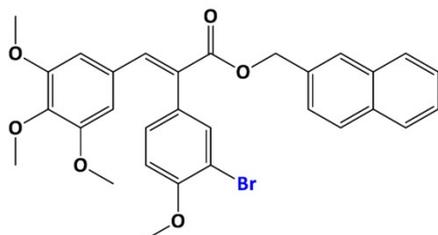
MJ-CA4-I-006



MJ-CA4-I-007



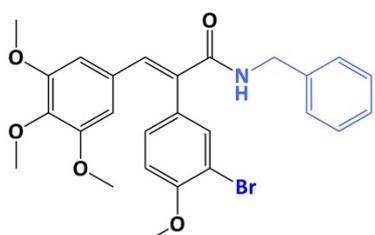
MJ-CA4-I-008



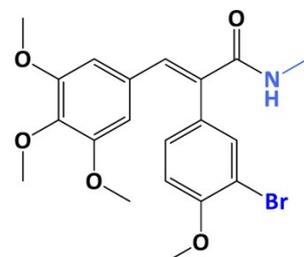
MJ-CA4-I-009



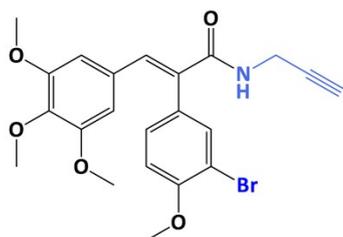
MJ-CA4-I-010



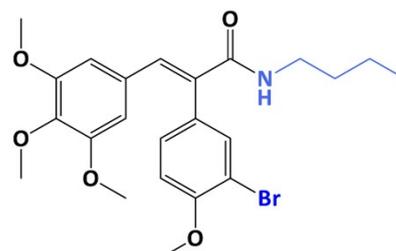
MJ-CA4-II-001



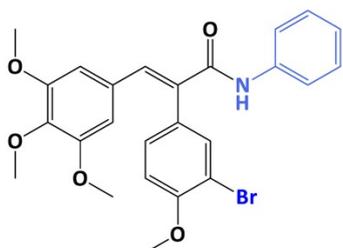
MJ-CA4-II-002



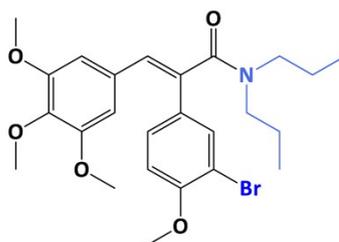
MJ-CA4-II-003



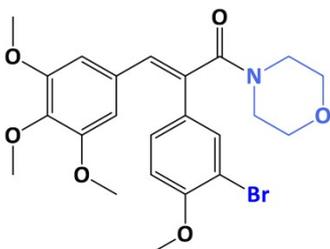
MJ-CA4-II-004



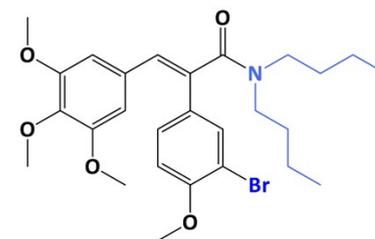
MJ-CA4-II-005



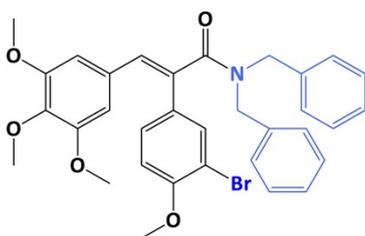
MJ-CA4-II-007



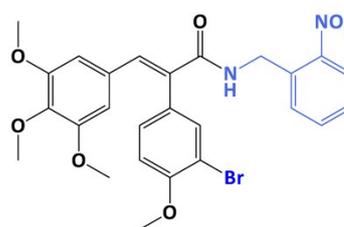
MJ-CA4-II-008



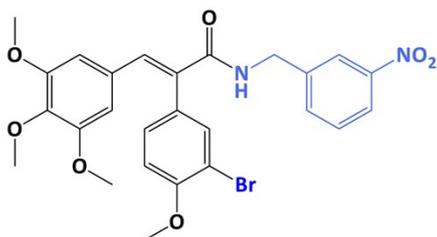
MJ-CA4-II-009



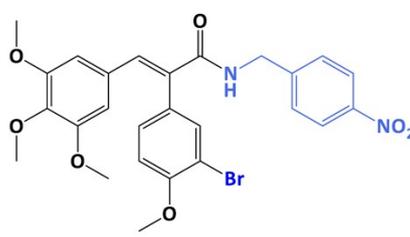
MJ-CA4-II-010



MJ-CA4-II-011



MJ-CA4-II-012



MJ-CA4-II-013

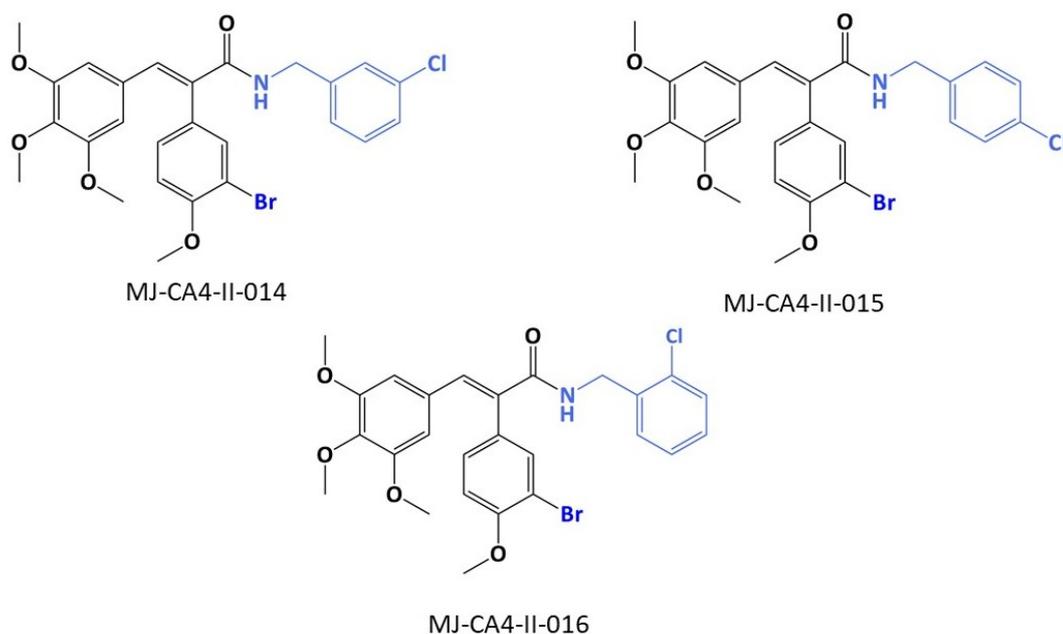


Figure 2.1: Novel Combretastatin A-4 derivatives.

2.2 Expression of β tubulin isotypes

The β tubulins are encoded by a multigene family that results in slight distinct proteins with elaborate expression patterns. Numerous used anticancer drugs work by attaching to β tubulin, changing microtubule dynamics and preventing cell division. It may be extremely important for the success of treatment if these pharmacological targets are expressed in different ways in tumour cells and normal cells.

The table below shows which genes express the different isotypes of β tubulin.

Gene name	Protein name
TUBB	β I
TUBB2A	β IIa
TUBB2B	β IIb
TUBB3	β III
TUBB4	β IVa
TUBB2C	β IVb
TUBB6	β V
TUBB1	β VI
TUBB8	β VIII

Table 2.1: Genes expressing isotypes of β -tubulin.

Each tissue showed a unique pattern in the distribution of the various isotypes to the overall amount of β tubulin, most isotypes in cancerous tissue showed differential expression in relation to tumoral features.

For example, in most cancers, TUBB3 expression increased significantly whereas TUBB6 levels significantly reduced. This different distribution of β tubulin isotypes may influence the toxic profile of the potential drug being investigated.

The various isotypes differ mainly in the area of C-terminal amino acids, that is, the part involved in interaction with microtubule-associated proteins (MAPs). Some of these isotypes can be considered constitutive, others, however, may be considered more specific to some particular tissues.

It has been demonstrated that various β tubulin isotypes exhibit altered expression in cancerous cells. Multiple lines of evidence indicate that variable expression of β tubulin isotypes is implicated in anticancer treatment resistance, even if the roles of each isotype have not yet been completely clarified.

In the figure shown (Fig.2.2 and Fig.2.3), it is possible to see the percentage of expression of each of the isotypes for several normal tissues and some tumor tissues [22].

The complex expression of β tubulins in the cell lines under investigation was taken into account in this research study. To perform this, the CellMinerTM tool was used and data found in the research work of Luis J. Leandro-García et al.[22].

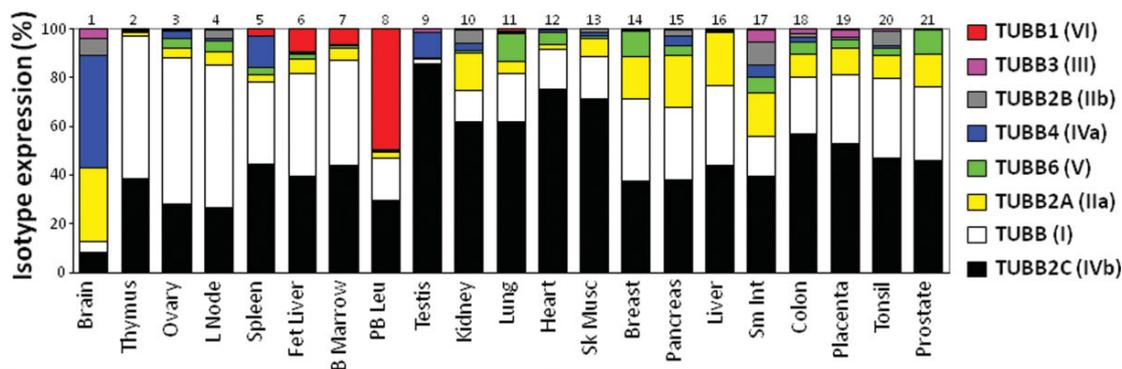


Figure 2.2: β isotypes expression in normale tissues [22].

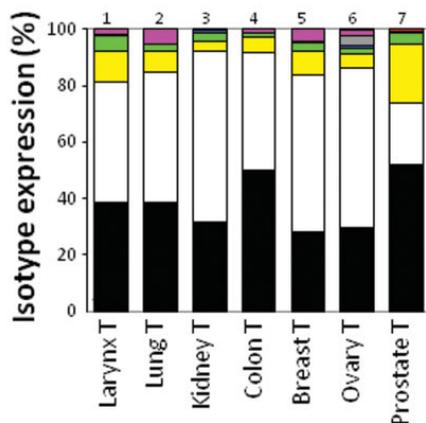


Figure 2.3: β isotypes expression in tumoral tissues [22].

2.3 ADMET properties

The most common ways to deliver a drug, approved by the U.S. Food and Drug Administration (FDA), are oral, intravenous, intramuscular, topical, inhalational and intranasal, but oral treatment is the most common way out of all of them due to its ease of use.

Pharmaceutics is the branch of pharmacy that deals with creating a dosage form that will allow a pharmaceutical molecule to be delivered to a human efficiently and safely.

The four processes of a drug — absorption, distribution, metabolism (biotransformation), and elimination — begin as soon as the drug enters into the body (commonly called as ADME).

Another important aspect to consider is the toxicity of a molecule, so the acronym

by which we refer to the analysis of all the characteristics of a potential drug becomes, in its most complete form, ADMET:

- Absorption
- Distribution
- Metabolism
- Excretion
- Toxicity

An effective therapeutic molecule must also have a low toxicity profile, which is typically stated as therapeutic index.

The therapeutic index (TI), also known as the margin of safety, is calculated by dividing the lowest concentration (dose) of a drug that causes toxicity by the minimal concentration (dose) of the same substance that results in therapeutic effects. The safer the drug, the higher the therapeutic index[23].

2.3.1 Absorption

By orally administering a drug, it transits through the gastrointestinal tract, in which it is subjected to an environment with different pH values, the latter of which can vary from 1 a 8 going from the stomach to the colon. In addition, to exert its therapeutic function, it needs to cross the cell membrane, which consists of the phospholipid bilayer.

Water molecules can pass through the membrane quite easily, also if a molecule is hydrophobic it is more likely to cross the membrane instead polar or large charged molecules do not diffuse through the phospholipid bilayer. Otherwise, there are membrane proteins or channels through which these molecules can be brought into the cell. In determining the dosage to be administered, it should be taken into consideration that only drug molecules not bound to plasma proteins are able to pass inside the cell.

There are two types of transport possible to cross the plasma membrane, passive transport that does not require energy and active transport that does require energy for molecules to pass through. Passive transport is suitable for small, uncharged and non-polar molecules while active transport is used for larger, polar and charged molecules. An example of passive transport is diffusion due to the concentration gradient from the higher concentration region to the lower concentration zone, while as for active transport an example is the Na/K pump that allows transport against the gradient.

The partition coefficient may be used to calculate the lipophilicity/hydrophobicity.

It is the ratio of a compound's equilibrium concentrations in a combination of two immiscible phases. As a result, this ratio represents the variation in the compound's solubility between these two phases. It is determined by the equation shown below:

$$P = \frac{[Drug]_{octanol}}{[Drug]_{aqueous}} \quad (2.1)$$

The compound is more hydrophobic the higher the partition coefficient (P) is. The majority of drugs are weak acids or bases that exist in solution in a non-ionized state. These compounds easily dissolve in lipids. P typically has a value between 10^1 and 10^6 . The partition coefficient is most conveniently expressed as $\log P$, which has a range of 1-6 for the majority of pharmacological compounds [23].

Another important feature for the passive transport of a molecule is the Polar Surface Area (PSA), it is a molecular descriptor employed to evaluate intestinal absorption and blood-brain barrier (BBB) permeability. It considers the contribution of all the polar atoms and their hydrogens, but the necessity to choose and then precisely compute the proper 3D molecular geometry or ensemble of geometries for each molecule under study complicates its calculation. Topological polar surface area (TPSA), which was established by researchers to solve this problem, it is easier and faster to be computed and it is now widely used in medicinal chemistry for virtual screening and for the prediction of ADME features [24].

Also water solubility of a molecule (LogS) is to be taken into account, typically the logarithm of the molar concentration ($\log \text{mol/L}$) is used to determine a compound's water solubility.

The first area where absorption occurs is the intestine, if the drug is administered orally, so it is important to understand how much absorption at the intestinal level of a substance is, if it is less than 30 percent that molecule is considered to have little chance of being absorbed and thus reaching the site of action.

Regarding active transport, transmembrane transport protein has been mentioned above, some of these proteins, however, perform the opposite function, that is, they tend to throw xenobiotic substances out of the cell. The most relevant protein in case is P-gp and thanks to some models it is possible to predict whether a molecule is a substrate or an inhibitor of it [25].

2.3.2 Distributions

The amount of drug that reaches the circulatory system is relevant to know to achieve the desired effect, since if the drug is administered by intravenous injection the maximum dose corresponds to the dosage of the drug itself, but if the drug is administered orally the actual dose that reaches the circulatory system will be less, due to the rate of absorption and from the metabolism that the molecule

undergoes.

To facilitate the development of orally accessible drugs, several criteria have been defined, one of the most widely used of which are the *Lipinski's rules of five*. These empirical rules suggest the following [26]:

- Molecular weight less than 500 g/mol
- Log P less than 5
- Hydrogen bond donors (HBD) less than 5
- Hydrogen bond acceptors (HBA) less than 10
- Rotable bonds (RB) less than 5
- Polar surface area (PSA) less than 140Å

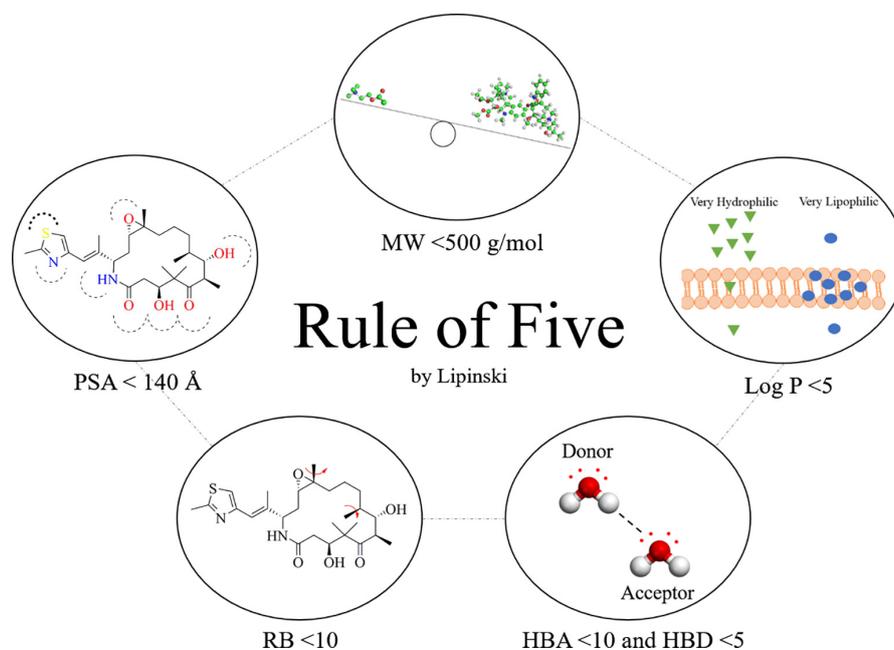


Figure 2.4: Lipinski's rule of five [26]

The volume of distribution (VD_{ss}) is the theoretical volume that the total dose of a drug should distribute equally to achieve the same concentration in the blood plasma. The higher the VD is, the more a drug is distributed in the tissues rather than in the plasma. The logarithm of the VD_{ss} of a given compound (log L/Kg) is considered. Low VD_{ss} is defined as less than 0.71 L/Kg (log VD_{ss} < -0.15) while high VD_{ss} is defined as more than 2.81 L/Kg (log VD_{ss} > 0.45).

The drugs in plasma will balance between being in an unbound form or being bound to serum proteins. A drug's effectiveness may be influenced by how tightly it binds to blood proteins; the more tightly it is attached, the less effectively the drug can disperse or cross cellular membranes. Calculations will be made to determine the expected proportion of a particular molecule that would be unbound in plasma. The brain is shielded from exogenous substances by the blood-brain barrier (BBB), it is important to know how a drug interacts with the BBB in order to predict its side effects and possible toxicity or to improve the efficacy of drugs that are supposed to act in the brain. Again, the logarithm (\log_{BB}) is considered; it is the logarithm of the ratio of the concentration of the drug in the brain to the concentration in the plasma.

A molecule with $\log_{BB} > 0.3$ can be considered to easily pass the BBB, while one with $\log_{BB} < -1$ does not easily reach the brain [25].

2.3.3 Metabolism

The metabolism is the process by which substances not belonging to the body are processed and made more polar and more soluble in water. During metabolism, enzymes carry out a series of reactions that can be divided into two phases. In phase 1, oxidation, reduction, hydrolysis and carboxylation reactions take place, Cytochrome P450 plays a major role in this phase, which oxidizes xenobiotics to facilitate their elimination. It can also activate or deactivate a drugs and it is located in the liver. Drug metabolism may be affected by this enzyme's inhibitors, which are contraindicated. Analyzing a compound's capacity to inhibit cytochrome P450 is crucial. A given chemical will be evaluated by the predictors to determine if it is probable to be a cytochrome P450 inhibitor or a substrate [25].

2.3.4 Excretion

An essential function in the renal clearance of pharmaceuticals and endogenous compounds is played by the renal absorption transporter Organic Cation Transporter 2. OCT2 substrates and concurrently given OCT2 inhibitors may interact negatively. A candidate's ability to be carried by OCT2 may be evaluated, and this information is important for determining both the candidate's clearance and any potential drawbacks. If a specific molecule is likely to be an OCT2 substrate or inhibitor, the predictor will determine this.

The proportionality constant CL_{tot} is used to evaluate the amount of drug clearance, which mostly is a result of renal clearance (excretion via kidneys) and hepatic clearance (liver metabolism and biliary clearance). It is relevant to bioavailability and crucial for figuring out the dosage rate needed to reach steady-state concentrations. A particular compound's predicted total clearance $\log(CL_{tot})$ is expressed in

$\log(\text{ml}/\text{min}/\text{kg})$ [25].

2.3.5 Toxicity

It has become increasingly relevant to be able to predict whether newly synthesized compounds are toxic to the body and to what extent they are. The toxicity of a compound can be assessed in several ways; in this study, the first parameter taken into consideration was the Maximum Recommended Tolerated Dose (MRTD), it gives an idea of the dangerous dose threshold for humans for various substances. A MRTD of less than or equal to $0.477 \log(\text{mg}/\text{kg}/\text{day})$ is classified as low for a particular compound, whereas one of higher than $0.477 \log(\text{mg}/\text{kg}/\text{day})$ is regarded as high [25].

Another aspect to consider is hepatotoxicity, since all drugs sooner or later pass through the liver. The predictor seeks to predict whether the test substance can cause a change in normal liver function.

Cardiotoxicity also needs to be carefully evaluated, and it would be very useful to have a model that could reliably predict it. To assess this type of toxicity, an attempt is made to predict whether the compound is a substrate or inhibitor of the hERG-encoded potassium channel, since inhibition of this channel is the main cause of long QT syndrome leading to arrhythmia that can sometimes be mortal. Regarding the topical use of a drug, it is important to evaluate skin sensitization, as it is a potential side effect and the predictor tries to assess the possibility.

2.4 Experimental data

The biological activity of the new Combretastatin A4 derivatives was analyzed *in vitro*, and the results were kindly provided by M. Jedrzejczyk and A. Huczynski (private communication, 2022). Cytotoxicity, calculated in terms of IC₅₀, was assessed by two experimental assays, the MTT and SRB. The tetrazolium assay (MTT) is a colorimetric assay to estimate the metabolic activity of living cells, while the sulforhodamine B (SRB) assay based on the assessment of cell density considering protein content. The half maximal inhibitory concentration (IC₅₀) measures a substance's ability to effectively block a certain biological or metabolic activity. The IC₅₀ test quantifies the amount of an inhibitory substance (such as a drug) required to inactivate 50% of a certain biological process or biological component *in vitro*.

The data are shown as the mean value and standard deviation of at least three independent experiments, and the IC₅₀ (μM) represents the concentration of the compound that, after 48 hours of culture with the particular compound, results in

a 50% growth inhibition of the cell line relative to the control.

The experiments were conducted on seven distinct cancer cell lines, the following:

- A549 \Rightarrow human lung cancer cell line
- HaCaT \Rightarrow human immortalized keratinocyte cell line
- HepG2 \Rightarrow human liver cancer cell line
- MDA \Rightarrow human breast cancer cell line
- PC3 \Rightarrow human prostate cancer cell line
- SW480 \Rightarrow human primary colon cancer cell line
- SW620 \Rightarrow human metastatic colon cancer cell line

The tables below present the results obtained through the MTT and SRB methods, respectively, in the experiments performed.

Compound	HaCaT	SW480	SW620	PC3	HepG2	MDA	A549
MJ-CA4	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-001	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-004	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-005	13.1±2.0	10.3±0.7	8.93±1.59	11.0±5.3	15.5±2.4	13.9±1.8	13.1±2.7
MJ-CA4-I-006	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-007	57.06±7.0	49.8±11.7	42.0±16.8	45.8±14.0	55.4±12.6	54.2±16.3	56.3±16.7
MJ-CA4-I-008	42.9±12.2	45.7±14.1	40.1±19.3	39.8±6.7	53.2±10.9	49.3±5.4	49.2±8.1
MJ-CA4-II-001	88.2±15.5	83.6±21.8	69.8±22.7	73.6±4.5	71.9±11.3	63.1±12.8	69.0±15.4
MJ-CA4-II-002	33.5±11.7	34.9±13.7	40.5±12.8	37.3±7.8	53.8±12.3	42.4±6.7	38.3±2.1
MJ-CA4-II-003	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-004	26.4±15.6	30.5±9.0	27.0±5.9	26.0±10.4	51.2±4.7	42.1±5.6	40.6±4.1
MJ-CA4-II-005	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-007	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-008	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-009	19.5±5.9	17.6±5.3	15.7±3.4	22.7±6.2	29.7±11.6	26.3±12.3	22.9±9.6
MJ-CA4-II-010	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-011	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-014	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-015	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-016	84.4±9.9	>100	>100	94.3±4.9	>100	67.5±8.3	86.4±7.9

Table 2.2: Cytotoxicity (IC₅₀, μ M) of studied compounds estimated by MTT assay.

Compound	HaCaT	SW480	SW620	PC3	HepG2	MDA	A549
MJ-CA4	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-001	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-004	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-005	11.4±3.0	12.1±1.5	9.03±0.58	7.51±1.63	8.20±1.00	9.85±1.78	11.2±5.1
MJ-CA4-I-006	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-I-007	68.3±19.6	57.6±9.4	61.8±11.4	65.9±16.1	63.7±12.6	55.2±17.3	53.0±10.8
MJ-CA4-I-008	45.0±7.2	73.6±8.4	57.5±3.9	43.2±10.2	42.1±10.2	45.2±5.5	46.6±12.1
MJ-CA4-II-001	68.5±4.8	59.8±23.0	62.4±3.1	55.8±23.7	56.2±22.2	59.8±18.8	82.1±13.9
MJ-CA4-II-002	55.3±17.0	52.2±17.9	46.0±11.4	44.9±12.1	50.5±10.5	55.5±20.3	49.3±16.3
MJ-CA4-II-003	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-004	58.6±22.2	50.2±11.2	50.8±14.5	67.2±18.6	50.5±17.5	61.4±21.8	65.1±17.3
MJ-CA4-II-005	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-007	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-008	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-009	35.0±12.6	37.0±17.9	35.0±14.8	32.7±9.4	32.9±12.8	28.7±12.9	32.8±12.6
MJ-CA4-II-010	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-011	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-014	>100	>100	>100	>100	>100	>100	>100
MJ-CA4-II-015	>100	>100	>100	>100	89.6±24.4	66.7±6.4	62.2±8.6
MJ-CA4-II-016	>100	>100	>100	>100	>100	99.8±3.9	92.8±7.7

Table 2.3: Cytotoxicity (IC_{50} , μM) of studied compounds estimated by SRB assay.

2.5 Docking

Protein-ligand or protein-protein docking is an important element in current drug discovery for predicting the orientation of the ligand when it binds to a protein receptor or enzyme by quantifying its shape and electrostatic interactions.

The forces that come into play in determining this interaction are Van der Waals, Coulomb and hydrogen bond interactions, the sum of all of which approximately represents the docking score, which represents the binding potential between the ligand and receptor.

The behaviour of smaller molecules at the binding site of a target protein is investigated by molecular docking methods. As more protein structures are identified experimentally by nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography.

It also becomes feasible for proteins whose structures are unknown to dock to homology modelled targets. For subsequent lead optimization procedures, the druggability of the compounds and their specificity against a certain target may be determined using the docking techniques.

Programs for molecular docking use a search method in which the ligand's conformation is evaluated repeatedly until convergence to the minimal energy is achieved. In order to rank the candidate poses as the sum of the electrostatic and van der Waals energies, an affinity scoring function is used.

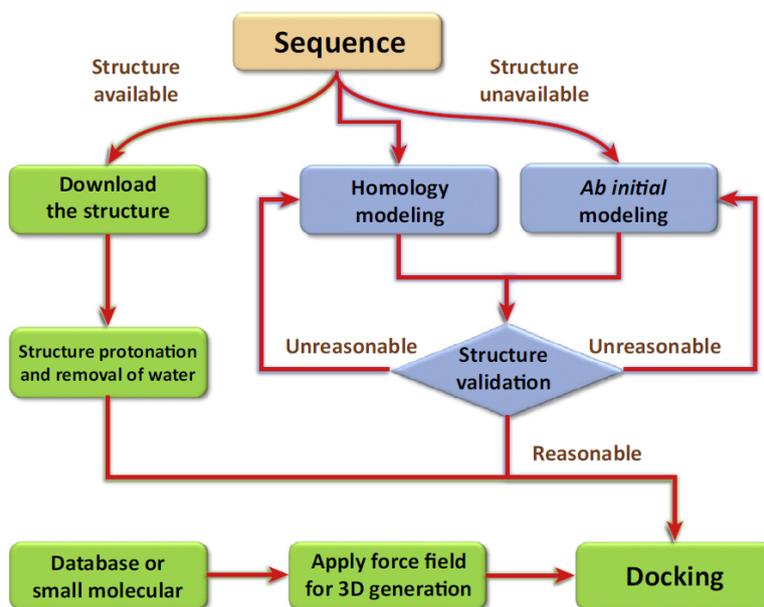


Figure 2.5: Steps to perform docking [27].

These particular interactions in biology are motivated by complementarities between the electrostatics and shape of the binding site surfaces and the ligand or substrate [28].

The most effective "binding affinity" can then be selected as the powerful ligand for more biochemistry research and development.

When a protein and its ligand are docked rigidly, their bond angles and lengths cannot be altered. Although it needs a lot more time and computing power, flexible docking, which allows for conformational alterations, is commonly utilised today. Changing the solvation, the pH, and docking with or without water are other docking possibilities.

The first "*lock-and-key model*", which refers to the rigid docking of receptors and ligands to determine the best orientation for the "*key*" to unlock the "*lock*," is depicted in Figure A. The significance of geometric complementarity is emphasised by this model.

Unfortunately, the real docking procedure is so adaptable that ligands and receptors must alter their conformation to match one another properly. Consequently, researchers create a "*induced fit model*" (Figure B). The energy complementarity and pre-organization, which are based on geometric complementarity, ensure that receptors and ligands will get the most stable structure in a way that minimises free energy [29]. There are several scoring functions; theoretically, a complex's lower Gibbs free energy implies the stability of the protein-ligand interaction. However, alternative methods for these services are created for users with varying needs. Choosing the scoring function that is best for a certain target protein might be challenging.

The docking procedure and its results should always be used with care, since one of the main disadvantages of this procedure is accuracy, so the results may not be reliable. The target protein's incorrect binding site, screening with the wrong small-molecule database, the choice of the docking pose, a good dock score (binding affinity) but a failed Molecular Dynamics simulation or docking results that don't match bioassays are common issues to be care of [27].

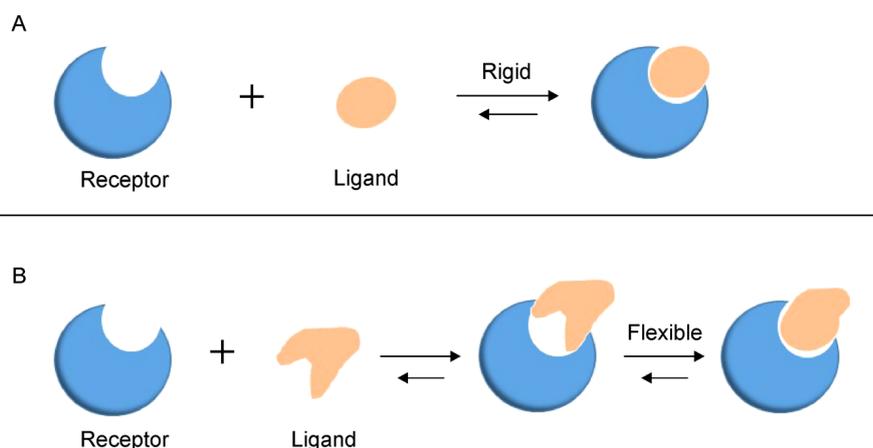


Figure 2.6: Docking mode for flexible docking or rigid docking[29].

Three-dimensional models of β tubulin isotypes were used to dock Combretastatin derivatives; since there are no 3D structures of human β tubulin isotypes, the models were created by homology models.

The idea behind homology modelling is that a similar sequence implies a comparable structure and this can help predict the tertiary structure of a protein whose spatial disposition is unknown.

To effectively model a protein whose structure is unknown, do an alignment research between the sequence without a 3D structure and other sequences with known 3D structures, which show better similarity to the sequence may serve as a template to model the sequence with no structure.

Once the structure has been rebuilt, it may be evaluated for quality using a variety of tools.

The models, created through homology model, already produced in another thesis are used in this work (P.Vottero, private communication 2022). The choice of template was made on the basis of resolution, number of missing residues, and publication date. The PDB entry 5EYP was chosen primarily because of its higher resolution, 1.90Å and it is referred to free tubulin, regarding the amino acid sequences to model of the different isotypes, the FASTA file was downloaded from the UniProt website [30][31]. All models employed the tubulin α IA chain as the α subunit.

Models described above are the proteins used as targets in subsequent docking procedures.

To obtain the binding energy values on individual isotypes, docking was carried out using AutoDock4. A semi-empirical free energy force field is used by AutoDock to assess conformations during docking simulations. A significant number of protein-inhibitor complexes with known structures and inhibition constants, or K_i , were used to parameterize the force field.

Knowing that Combretastatin binds in the same binding site as Colchicine due to

the trimethoxyphenyl portion, the grid box was centered in the following coordinates for the X, Y and Z respectively -16, 13, -23. The dimensions of the box along the three dimensions are 40x40x40, with a spacing of 0.375Å. The poses were clustered, and the binding energy considered was taken from the most populated cluster.

To evaluate the promising derivatives, evidenced by the experimental results, consensus docking was used to obtain the poses and energies on the representative proteins. DockBox was used to make the consensus docking

A python wrapper package called DockBox was created to make it easier to utilise common docking programmes individually or in combination. Additionally, DockBox gives users the option to rescore the generated docking poses using several widely used scoring algorithms and to examine the outcomes using various consensus docking/scoring methods.

The programs that have been used are AutoDock 4, AutoDock Vina and MOE, while rescoring was done with Autodock Vina.

2.6 Regression

The correlation between a dependent variable, y , and one or more independent variables, X , is expressed by a linear regression model. The response variable is another name for the dependent variable. Explanatory or predictor variables are other names for independent variables. Categorical predictor variables are also known as factors, whereas continuous predictor variables are also known as covariates. The design matrix is typically referred to as the matrix X of observations on predictor variables.

A model of multiple linear regression is:

$$y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_p X_{i,p} + \epsilon_i, \quad i = 1, \dots, n \quad (2.2)$$

wherein

- y_i is the i th response variable,
- β_k is the k th coefficient, where β_0 is the constant term in the model,
- $X_{i,j}$ is the i th observation on the j th predictor variable, $j = 1, \dots, p$,
- ϵ_i is the i th noise term, that is, random error.

To carry out the linear regression considering one or more predictor variables, MATLAB's "regress" function was used, which taking as input the vector with the response variable and the vector or matrix with the predictor variables returns the vector of coefficients of the linear regression and the value of R^2 .

R^2 is a statistic that measures the proportion of variance for a dependent variable that is explained by one or more independent variables in a regression model. While correlation reveals the strength of the relationship between independent and dependent variables, R-squared reveals the measurement by which the variation of one variable reveals the variation of the second variable. Therefore, if a model's R^2 is 0.50, then roughly half of the observed variation may be explained by the model's input.

A model's coefficient of determination rises in value as more explanatory variables are included. Increasing the number of variables makes the model better, but it may also imply that the regression line and observed data are now closer together, increasing R^2 [32].

2.7 Molecular Dynamics Simulations

In biology, molecular dynamics makes it possible to simulate the motion of a system of particles, such as the fluctuation of atoms in a protein as a function of time.

The ability to look at these phenomena allows a more accurate description of biological events that are sometimes difficult to explain by experimental assays. Molecular dynamics is also used to provide a more accurate definition of the structure of a protein from models produced by other techniques.

On timescales range from femtoseconds to milliseconds, molecular dynamics (MD) enables the atomistic analysis of biological and chemical processes. It supports experiment while also providing a tool to track down processes that are challenging to identify using experimental methods. There are several software programmes available for doing MD simulations, with Amber being one of the most popular and it is used in this case study to carry out the molecular dynamics analysis [33]. Simulations of molecular dynamics start with an understanding of the system's energy as a function of the atomic coordinates. The stabilities of the various potential stable or metastable structures are determined by the potential energy surface. By resolving Newton's equations of motion for the atoms as a function of time, it is possible to utilise the forces acting on the system's atoms, which are correlated to the first derivatives of the potential with respect to the atom locations, to determine the dynamic behaviour of the system. For proteins and their surroundings, however, only empirical energy functions may supply this information; whereas quantum mechanical computations can provide potential surfaces for tiny molecules.

The energy functions employed for proteins are often made up of non-bonding components like van der Waals interactions and electrostatic contributions and bonding terms like bond lengths, bond angles, and torsional angles. One often used

equation is:

$$E(R) = \frac{1}{2} \sum_{bonds} K_b(b - b_0)^2 + \frac{1}{2} \sum_{bonds\ angles} K_\Theta(\Theta - \Theta_0)^2 + \frac{1}{2} \sum_{torsional} K_\phi[1 + \cos(n\phi - \delta)] + \sum_{non-bonding\ pairs} \left(\frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_1 q_2}{Dr} \right) \quad (2.3)$$

In the formula given above E represents energy and is a function of Cartesian coordinates, R, that is, the positions of all atoms from which internal coordinates for bond lengths (b), bond angles (Θ), dihedral angles(ϕ), and interparticle distance (r) were calculated.

A Hooke's law (harmonic) potential's instantaneous displacements from the ideal bond length, b_0 , are represented by the first term in equation (2.3). The energy of a bond as a function of length is initially approximated by a harmonic potential of this kind. The bond's flexibility is determined by the bond force constant, K_b .

The second term in the equation is associated with the bond angles and is also expressed by a harmonic potential, while the third term in the equation describes the torsional angles, and this contribution is considered periodic and modeled by a cosine. The last term describes the contributions from non-bonding interactions and is composed of three parts: a repulsive term that prevents atoms from interpenetrating each other, an attractive term that considers the London dispersion forces between atoms, and an electrostatic contribution that takes into account the charges and their sign. Based on the atoms considered, parameters A and B are determined. A Coulomb potential, with D representing the effective dielectric function for the medium and the separation between the two charges, is used to model electrostatic interactions between pairs of atoms.

The potential energy is also referred to as force field (FF), that is a system of equations and parameters. There are numerous ways to approach the dynamics given a potential energy function. Molecular dynamics simulations, in which Newton's equations of motion are solved for the system's atoms and any surrounding solvent, offer the most precise and comprehensive information. Average structural and dynamic parameters may be established for a basic homogeneous system, such as a box of water molecules with periodic boundary conditions, in simulations taking only a few picoseconds [34].

Each change in the internal coordinates of the system results in a change in the potential energy function and a change in the system's position in the Potential Energy Surface (PES), which is a multidimensional surface that represents the potential energy function. The PES is comprised of a number of stable states, and any departure from these configurations results in a higher energy state. These points are known as local minima, and the global minimum is the point with the lowest energy. An energy minimization (EM) can decrease the system's potential

energy and help it reach a stable state; EM is typically essential before beginning the simulation, especially for complex systems, to prevent potential failures.

During the simulation, the trajectories of the interacting particles are determined, and the average characteristics of the system may be obtained by solving the Newton’s equation of motion numerically.

In this thesis, molecular dynamics simulations were carried out on the three targets with the two ligands separately, for a total of six simulations. The parameters used during the simulations are the same so that the results can be compared.

The ligands normally present in the $\alpha\beta$ tubulin dimer were considered in the dynamics, so GTP and GDP are also present, in addition there is also the magnesium ion. These ligands were maintained because they are important at the structural level as explained earlier.

The simulations were run for 100ns at 298K with an octahedral-shaped box.

2.8 MMGB/PBSA

Finding a new pharmaceutical molecule that binds to a macromolecular receptor is one of the aims of structure-based drug design, the strength of the binding is governed by the binding free energy, or ΔG_{bind} . The most popular computational techniques in drug design are docking and scoring, which estimate the drug’s binding mechanism measuring its affinity with the receptor. These techniques are effective but not very precise; they can be used to detect binding modes and distinguish between binders and non-binders, but they often cannot discriminate between drugs with affinities that differ by less than one order of magnitude, or by 6 kJ/mol in ΔG_{bind} . Alchemical perturbation (AP) approaches, which are on the other extreme and, in theory, extremely accurate, are developed from statistical mechanics. However, they are computationally costly due to the fact that they are based on Monte Carlo or molecular dynamics (MD) simulations and need considerable sampling of the complex and the free ligand in solution, as well as of unphysical intermediate states.

There is a range of approaches with intermediate performance between these two categories. They likewise rely on sampling, but only of the final conditions, i.e., the complex and possibly the free receptor and ligand. Thus, end point methods refer to these techniques. In comparison to AP, they were designed to be less expensive but more precise than the scoring functions.

In the most popular method, known as MM/PBSA (molecular mechanics (MM) with Poisson–Boltzmann (PB) and surface area solvation), the binding free energy is estimated by the following equation:

$$\Delta G_{bind} = \langle G_{PL} \rangle - \langle G_P \rangle - \langle G_L \rangle \quad (2.4)$$

In the equation given above, in which P means unbounded protein, L free ligand and PL complex protein-ligand, the free energy of a state is predicted using the following sum of terms:

$$G = E_{bnd} + E_{el} + E_{vdw} + G_{pol} + G_{np} - TS \quad (2.5)$$

where the first three components are the typical MM energy contributions from van der Waals, electrostatic, and bound (bond, angle, and dihedral) interactions. The polar and non-polar contributions to the solvation free energy are G_{pol} and G_{np} , respectively. G_{pol} can be obtained by solving the PB equation or by Generalized Born (GB) model (giving the MM/GBSA approach), whereas the non-polar term is approximated from a linear relationship to the solvent accessible surface area (SASA). The absolute temperature, T, multiplied by the entropy, S, calculated by performing a normal-mode analysis on the vibrational frequencies, is the final term in the equation above [35].

Furthermore, MM/PBSA and MM/GBSA enable free energy decomposition analysis of the contributions from specific residues or energy terms, providing detailed information on the energetic contributions made by each residue to the system binding, identifying the dominant interactions involved in the binding process, and facilitating the development of customised drugs.

The MM/GBSA approach is substantially faster than the PB method since it provides an analytical equation for the polar solvation energy, so MM/GBSA was used to calculate the free energy of binding in this work and was calculated using AMBER tools.

Amber supports two types of decomposition: pairwise and per-residue. By adding up all of the interactions between individual residues in the system, *per-residue* decomposition determines the energy contribution of each individual residue. The interaction energy between pairs of residues in the system is determined via *pairwise* decomposition.

Chapter 3

Results

3.1 Isotypes expression

Within the cell, all isotypes of beta tubulin are simultaneously present, in different percentages depending on the type of cell considered and whether it is a healthy or cancerous cell.

In the case under consideration seven cancer cell lines whose *in vitro* cytotoxicity results are available were considered, and the following analysis was performed to find the expression levels of each of isotypes.

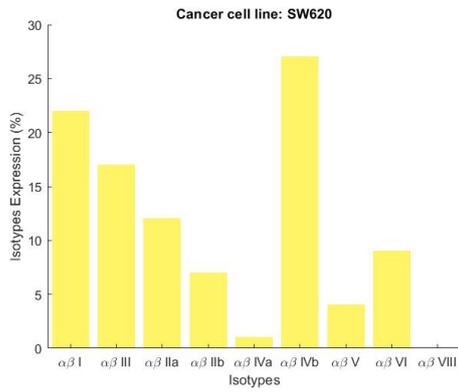
Concerning the SW620, MDA, and PC3 cell lines, it was possible to have the exact gene expression values, since these cell lines are present in the CellMiner database, which has a tool for evaluating gene expression values in RNA sequences.

In contrast, with regard to the other cell lines, approximations had to be made to estimate the expression rates of β tubulin isotypes. For the SW480 and A549 cell lines, the percentages were calculated by considering the average gene expression values of all cancer cells belonging to that particular tissue in the CellMiner database, colon and lung cancer cells, respectively.

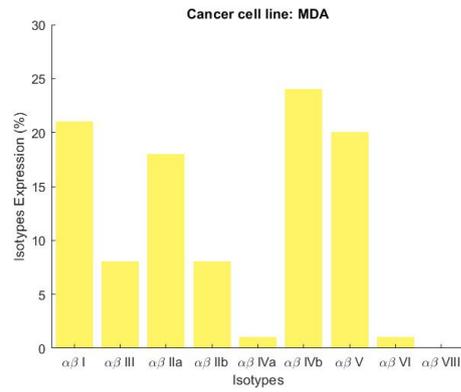
For the HepG2 cell line, which is a liver cancer cell line, there are no cancer cells of the same organ in the CellMiner database, so it was decided to use the gene expression level of healthy cells of that tissue, the data were taken from the work of Gracia et al.[22].

For the HaCaT cell line, i.e., human immortalized keratinocyte cell line, it was not possible to find sufficient data to estimate a precise expression of the isotypes, so in order to subsequently calculate the binding energy weighted on the percentage of isotypes expressed, a strong approximation was made, considering a uniform distribution over the most expressed isotypes usually in the cells, and these values were also taken from the work of Garcia et al. [22].

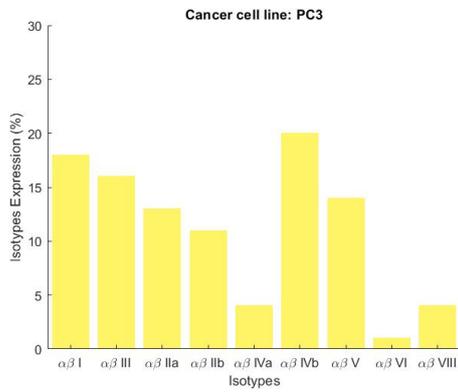
Results



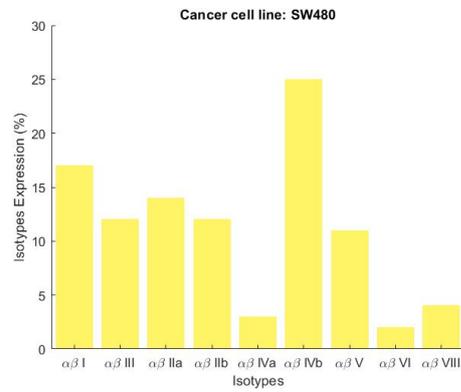
(a)



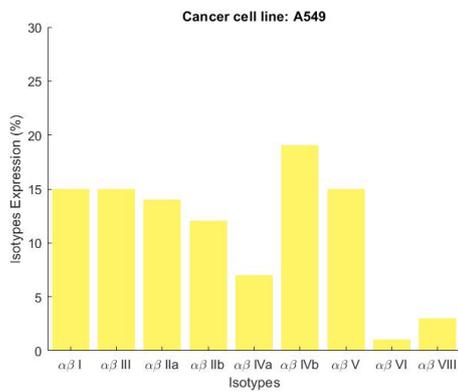
(b)



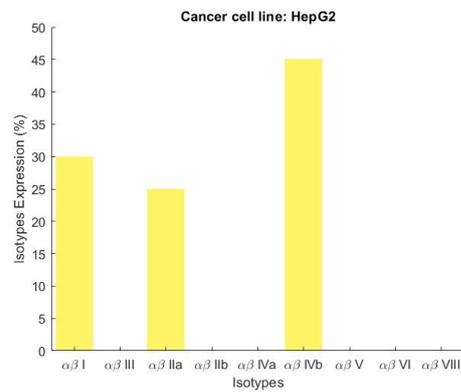
(c)



(d)

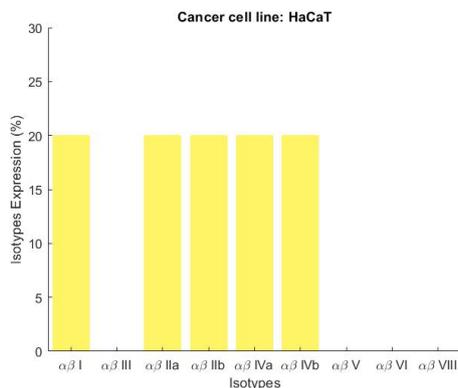


(e)



(f)

The graphs above show the distribution of beta tubulin isotypes for each cell line under investigation in this work, it is interesting to note that βIVb and βI are those most highly expressed in all cell lines, while $\beta VIII$ can be considered negligible in



(g)

Figure 3.0: Isotypes expression of β tubulins.

all cell lines, while β III is definitely more expressed in cancer cells than in healthy cells, as can be seen from the approximations based on normal cells, where β III is not even considered.

3.2 ADMET properties analysis

Physical and chemical characteristics influence the pharmacokinetics of drugs, so to optimize the process of development of new substances that could potentially be used as drugs, it is useful to have tools capable of predicting these properties without having to test them individually, a process that would be costly both in terms of money and development time.

SimulationsPlus, specifically the ADMET predictor tool, was used in this work to gain access to all the features to be taken into account when evaluating new compounds.

SimulationsPlus is the state-of-the-art machine learning algorithms for the prediction of ADMET properties and allows these properties to be obtained quickly and accurately simply from the 2D structure of molecules.

Compounds	MlogP	Water solubility (mg/mL)	TPSA (\AA^2)	Pgp substrate	Pgp inhibitor
CA4	2,572	0,006	57,15	Yes	Yes
MJ-CA4-I-001	3,436	0,004	63,22	Yes	Yes
MJ-CA4-I-004	3,756	0,004	63,22	Yes	Yes
MJ-CA4-I-005	3,357	0,007	63,22	Yes	Yes
MJ-CA4-I-006	2,551	0,012	63,22	Yes	Yes
MJ-CA4-I-007	2,41	0,019	63,22	Yes	Yes
MJ-CA4-I-008	2,627	0,016	63,22	Yes	Yes
MJ-CA4-I-009	4,066	0,0002069	63,22	Yes	Yes
MJ-CA4-I-010	1,868	0,107	83,45	Yes	Yes
MJ-CA4-II-001	3,436	0,013	66,02	Yes	Yes
MJ-CA4-II-002	2,41	0,059	66,02	Yes	Yes
MJ-CA4-II-003	2,551	0,047	66,02	Yes	Yes
MJ-CA4-II-004	3,049	0,031	66,02	Yes	Yes
MJ-CA4-II-005	3,236	0,002	66,02	Yes	Yes
MJ-CA4-II-007	3,458	0,034	57,23	Yes	Yes
MJ-CA4-II-008	1,898	0,137	66,46	Yes	Yes
MJ-CA4-II-009	3,854	0,017	57,23	Yes	Yes
MJ-CA4-II-010	4,55	0,004	57,23	Yes	Yes
MJ-CA4-II-011	3,287	0,007	111,84	Yes	Yes
MJ-CA4-II-012	3,287	0,004	111,84	Yes	Yes
MJ-CA4-II-013	3,287	0,002	111,84	Yes	Yes
MJ-CA4-II-014	3,632	0,007	66,02	Yes	Yes
MJ-CA4-II-015	3,632	0,004	66,02	Yes	Yes
MJ-CA4-II-016	3,632	0,007	66,02	Yes	Yes

Table 3.1: Properties affecting absorption.

Looking at the Tab. 3.1 is noticeable that all the new compounds derived from combretastatin A4 are substrates and inhibitors of P-glycoprotein, which means that while they interact with the protein as a substrate, they can also inhibit its activity and thus not be expelled from the cell and eventually explicate their therapeutic function.

In addition, compounds can be categorised based on their solubility values (LogS); those with solubility values of 0 and above are very soluble, those between 0 and 2 are soluble, those between 2 and 4 are marginally soluble, and compounds less than 0 are insoluble [36]. The results about the solubility in the Tab. 3.1 show that the derivatives exhibit good solubility.

Lipinski's rules of five, while not mandatory to comply with, are a good indicator of whether a compound has good pharmacokinetic properties and therefore can be administered orally effectively.

The Table 3.2 shows all the physical and chemical characteristics taken into account by the Lipinski's rules, and to have an immediate visual reading, the following color code was used: green if the indicator value meets the rule, yellow if the value is borderline, red if the value is outside the ranges expressed by the Lipinski's rules. It can be seen from the table that the largest compounds have a high molecular weight, and in addition, the number of rotatable bonds is met only by Combretastatin A4, while all derivatives have a number of rotatable bonds higher than the value imposed as the maximum.

In contrast, regarding the value of the partition coefficient, the number of Hydrogen bond donors and Hydrogen bond acceptors, and Topological Polar surface area, the values perfectly comply with Lipinski's rules.

These empirical rules are a suggestion to facilitate the development of orally accessible drugs, other parameters must be considered regarding the distribution of a substance within the human body.

Compounds	Weight	MlogP	HBD	HBA	N° rotatable bonds	TPSA
CA4	316,356	2,572	1	5	4	57,15
MJ-CA4-I-001	513,393	3,436	0	6	7	63,22
MJ-CA4-I-004	586,331	3,756	0	6	11	63,22
MJ-CA4-I-005	558,277	3,357	0	6	9	63,22
MJ-CA4-I-006	447,29	2,551	0	6	6	63,22
MJ-CA4-I-007	437,295	2,41	0	6	5	63,22
MJ-CA4-I-008	451,322	2,627	0	6	6	63,22
MJ-CA4-I-009	563,454	4,066	0	6	7	63,22
MJ-CA4-I-010	467,321	1,868	1	7	7	83,45
MJ-CA4-II-001	512,409	3,436	1	6	7	66,02
MJ-CA4-II-002	436,31	2,41	1	6	5	66,02
MJ-CA4-II-003	446,305	2,551	1	6	6	66,02
MJ-CA4-II-004	478,391	3,049	1	6	8	66,02
MJ-CA4-II-005	498,381	3,236	1	6	6	66,02
MJ-CA4-II-007	506,445	3,458	0	6	9	57,23
MJ-CA4-II-008	492,375	1,898	0	7	5	66,46
MJ-CA4-II-009	534,499	3,854	0	6	11	57,23
MJ-CA4-II-010	602,534	4,55	0	6	9	57,23
MJ-CA4-II-011	557,406	3,287	1	8	7	111,84
MJ-CA4-II-012	557,406	3,287	1	8	7	111,84
MJ-CA4-II-013	557,406	3,287	1	8	7	111,84
MJ-CA4-II-014	546,854	3,632	1	6	7	66,02
MJ-CA4-II-015	546,854	3,632	1	6	7	66,02
MJ-CA4-II-016	546,854	3,632	1	6	7	66,02

Table 3.2: Lipinski's rules of five color code.

Compounds	Volume of distribution (L/Kg)	Percent unbound to blood plasma proteins	BBB Filter	LogBB
CA4	1,243	6,309	High	0,194
MJ-CA4-I-001	1,028	3,395	High	0,481
MJ-CA4-I-004	1,217	3,392	High	0,543
MJ-CA4-I-005	1,123	3,909	High	0,505
MJ-CA4-I-006	1,144	5,286	High	0,179
MJ-CA4-I-007	1,065	6,175	High	0,286
MJ-CA4-I-008	1,102	5,69	High	0,316
MJ-CA4-I-009	1,037	2,506	High	0,626
MJ-CA4-I-010	0,933	8,204	High	-0,142
MJ-CA4-II-001	0,705	3,933	High	0,142
MJ-CA4-II-002	0,77	7,763	High	-0,037
MJ-CA4-II-003	0,817	6,016	Low	-0,094
MJ-CA4-II-004	0,805	5,608	High	0,157
MJ-CA4-II-005	0,724	3,954	High	0,265
MJ-CA4-II-007	0,924	5,686	High	0,207
MJ-CA4-II-008	0,78	8,838	High	-0,207
MJ-CA4-II-009	1,041	4,895	High	0,378
MJ-CA4-II-010	0,975	3,364	High	0,249
MJ-CA4-II-011	0,643	4,372	Low	0,092
MJ-CA4-II-012	0,654	4,17	Low	0,13
MJ-CA4-II-013	0,67	3,968	Low	0,13
MJ-CA4-II-014	0,789	3,692	High	0,255
MJ-CA4-II-015	0,801	3,606	High	0,296
MJ-CA4-II-016	0,795	3,805	High	0,3

Table 3.3: Properties that influence the distribution of derivatives.

Among the properties that influence the distribution in the Tab. 3.3 the volume of distribution (VD_{ss}) is the ideal volume across which the entire dose of a drug should be evenly dispersed to produce the same blood plasma concentration. The higher the volume of distribution, the more a substance is distributed in the tissues rather than in the blood.

As mentioned before, low VD_{ss} is defined as less than 0.71 L/Kg ($\log \text{VD}_{ss} < -0.15$) while high VD_{ss} is defined as more than 2.81 L/Kg ($\log \text{VD}_{ss} > 0.45$). The results in the table give the value of Volume of distribution (L/kg) in human at steady state, not the logarithm, so they seem to present an intermediate value regarding this feature. It is interesting to note that the bulkier derivatives also have a smaller VD_{ss} value.

The percentage of unbound substance in human blood plasma proteins is important to know since only unbound compounds can reach the site of therapeutic interest. From the tabulated values in Tab. 3.3 it can be seen that only a few derivatives improve this percentage, while the other compounds show a lower percentage of compound not bound to plasma proteins.

If a substance is to be used in the treatment of diseases affecting the brain, a key characteristic of that substance must be the ability to cross the blood-brain barrier (BBB).

BBB filter predicts whether or not a compound can penetrate the Blood Brain Barrier, and from the results obtained, most derivatives are able to cross it, with the exception of 4 compounds that have low penetrative ability. The last column of the Tab.3.3 shows the values of LogBB, which is the logarithm of the ratio of the concentration of the drug in the brain to that in the plasma. A molecule with $\log \text{BB} > 0.3$ can be considered to easily pass the BBB, while one with $\log \text{BB} < -1$ does not easily reach the brain. Some outcomes appear to contradict each other when compared between the BBB filter and the logBB, in particular for the derivative MJ-CA4-II-008.

In terms of metabolism, Cytochrome P450, which oxidises xenobiotics to promote their expulsion, is crucial. It is found in the liver and has the ability to both activate and deactivate drugs. The inhibitors of this enzyme are contraindicated because they may alter drug metabolism. It is essential to examine a substance's ability to inhibit cytochrome P450. The predictors will assess a given chemical to determine whether it is more likely to be a cytochrome P450 inhibitor or a substrate.

SimulationsPlus also calculates a risk indicator related to cytochrome P450, CYP Risk by scoring on the basis of risks related to P450 oxidation with a score in the range 0-6 indicating the number of potential problems a compound could have due to metabolism by one or more of the five major cytochromes P450, while CYP Code provides a code indicating which cytochrome in the family to attribute risk

Compounds	CYP Risk	CYP Code
CA4	3	1A2; 2C9; CL
MJ-CA4-I-001	3,189	1A2; 2C9; 3A4; CL
MJ-CA4-I-004	4,861	1A2; 2C9; 2D6; 3A4; CL
MJ-CA4-I-005	3,253	1A2; 2C9; 3A4; CL
MJ-CA4-I-006	3	1A2; 2C9; CL
MJ-CA4-I-007	3	1A2; 2C9; CL
MJ-CA4-I-008	3	1A2; 2C9; CL
MJ-CA4-I-009	3,331	1A2; 2C9; 3A4; CL
MJ-CA4-I-010	2	1A2; 2C9
MJ-CA4-II-001	2,72	2C9; 3A4; CL
MJ-CA4-II-002	3	1A2; 2C9; CL
MJ-CA4-II-003	3	1A2; 2C9; CL
MJ-CA4-II-004	3,889	1A2; 2C9; 2D6; 3A4; CL
MJ-CA4-II-005	3,078	1A2; 2C9; 3A4; CL
MJ-CA4-II-007	4	1A2; 2C9; 3A4; CL
MJ-CA4-II-008	1,823	1A2; CL
MJ-CA4-II-009	4,086	1A2; 2C9; 2D6; 3A4; CL
MJ-CA4-II-010	3	2C9; 3A4; CL
MJ-CA4-II-011	1,62	1A2; 3A4; CL
MJ-CA4-II-012	1,607	1A2; 3A4; CL
MJ-CA4-II-013	2,942	1A2; 2C9; 3A4; CL
MJ-CA4-II-014	3,315	1A2; 2C9; 3A4; CL
MJ-CA4-II-015	3,339	1A2; 2C9; 3A4; CL
MJ-CA4-II-016	3,289	1A2; 2C9; 3A4; CL

Table 3.4: Metabolism Risk and Code. CYP Risk rule codes: 1A2= high 1A2 clearance, 2C19= high 2C19 clearance, 2C9= high 2C9 clearance, 2D6= high 2D6 clearance, 3A4= high 3A4 clearance, CL= high microsomal clearance.

and clearance to.

Organic Cation Transporter 2 (OCT2) plays a central function in the renal clearance of pharmaceuticals and endogenous compounds, it is known that OCT2 substrates and concurrently given OCT2 inhibitors may interact negatively.

The outcomes obtained in the Tab. 3.5 reveals that all the compounds are substrate but not inhibitor of the Organic Cation Transporter 2.

Furthermore, by examining the diverse clearance mechanisms to which the derivatives under consideration may be subjected, according to the predictor they are all subject to metabolism clearance and no to renal or hepatic clearance.

Compounds	OCT2	OCT2	Clearance Metabolism	Clearance Renal	Clearance hepatic uptake	Clearance Mechanism
	substrate	inhibitor				
CA4	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-001	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-004	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-005	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-006	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-007	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-008	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-009	Yes	No	Yes	No	No	Metabolism
MJ-CA4-I-010	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-001	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-002	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-003	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-004	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-005	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-007	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-008	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-009	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-010	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-011	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-012	No	No	Yes	No	No	Metabolism
MJ-CA4-II-013	Yes	No	Yes	No	No	Metabolism
MJ-CA4-II-014	No	No	Yes	No	No	Metabolism
MJ-CA4-II-015	No	No	Yes	No	No	Metabolism
MJ-CA4-II-016	No	No	Yes	No	No	Metabolism

Table 3.5: Excretion related properties.

Compounds	MRTD	hERG Filter	hERG pIC50	Skin sensitivity	Respiratory sensitivity	Reproductive Toxicity
CA4	Above_3.16	No	4,904	Sensitizer	Sensitizer	Toxic
MJ-CA4-I-001	Above_3.17	Yes	6,321	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-I-004	Above_3.18	Yes	6,741	Sensitizer	Nonsensitizer	Toxic
MJ-CA4-I-005	Above_3.19	No	6,514	Sensitizer	Nonsensitizer	Toxic
MJ-CA4-I-006	Above_3.20	No	6,013	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-I-007	Above_3.21	No	5,843	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-I-008	Above_3.22	No	6,005	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-I-009	Below_3.16	Yes	6,351	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-I-010	Above_3.16	No	5,77	Nonsensitizer	Nonsensitizer	Toxic
MJ-CA4-II-001	Above_3.17	No	5,982	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-002	Above_3.18	No	5,538	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-003	Above_3.19	No	5,677	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-004	Above_3.20	No	5,993	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-005	Above_3.21	No	5,912	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-007	Above_3.22	No	6,263	Nonsensitizer	Nonsensitizer	Toxic
MJ-CA4-II-008	Above_3.23	No	6,072	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-009	Above_3.24	Yes	6,472	Nonsensitizer	Nonsensitizer	Toxic
MJ-CA4-II-010	Above_3.25	Yes	6,216	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-011	Above_3.26	No	5,626	Nonsensitizer	Nonsensitizer	Toxic
MJ-CA4-II-012	Above_3.27	No	5,683	Nonsensitizer	Nonsensitizer	Toxic
MJ-CA4-II-013	Above_3.28	No	5,674	Nonsensitizer	Nonsensitizer	Toxic
MJ-CA4-II-014	Below_3.16	No	6,299	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-015	Above_3.16	No	6,334	Nonsensitizer	Sensitizer	Toxic
MJ-CA4-II-016	Below_3.16	No	6,233	Nonsensitizer	Sensitizer	Toxic

Table 3.6: Toxicity properties.

In relation to toxicity of compounds, many aspects need to be attended to, SimulationPlus provides several indicators to assess these characteristics.

The Table 3.6 presents Maximum Recommended Tolerated Dose (MRTD), cardiotoxicity, Skin sensitivity, Respiratory sensitivity and Reproductive Toxicity.

While the Table 3.8 presents a general indicator that takes into account all possible risks associated with that compound.

The full ADMET Risk score is a value between 0 and 22 that represents the total number of potential ADMET issues that a compound may experience. Absorption contributes up to 8, distribution up to 2, CYP metabolism up to 6, and Toxicity up to 6.

The rules code of the full ADMET risk is the following: Size, RotB=rotatable bonds, HBD=H-bond donors, HBA=H-bond acceptors, ch=charge, Kow=lipophilicity, Peff=permeability, Sw=water solubility, fu=fraction unbound, Vd=volume of distribution, hERG=hERG inhibition, rat=acute rat toxicity, Xr=carcinogenicity in rat, Xm=carcinogenicity in mice, HEPX=hepatotoxicity, MUT=likely Ames positive; 1A2=high clearance by CYP 1A2, etc., CL=high microsomal clearance.

Compounds	Toxicity risk	Toxicity code
CA4	0	
MJ-CA4-I-001	1	hERG
MJ-CA4-I-004	2	hERG; HEPX
MJ-CA4-I-005	1	HEPX
MJ-CA4-I-006	0	
MJ-CA4-I-007	0	
MJ-CA4-I-008	0	
MJ-CA4-I-009	1	hERG
MJ-CA4-I-010	0	
MJ-CA4-II-001	1	HEPX
MJ-CA4-II-002	1	HEPX
MJ-CA4-II-003	0	
MJ-CA4-II-004	1	HEPX
MJ-CA4-II-005	1	HEPX
MJ-CA4-II-007	1	HEPX
MJ-CA4-II-008	1,063	rat; HEPX
MJ-CA4-II-009	2	hERG; HEPX
MJ-CA4-II-010	1,5	hERG; HEPX
MJ-CA4-II-011	1	HEPX
MJ-CA4-II-012	1	HEPX
MJ-CA4-II-013	2	HEPX; MUT
MJ-CA4-II-014	1	HEPX
MJ-CA4-II-015	1	HEPX
MJ-CA4-II-016	1	HEPX

Table 3.7: Toxicity properties.

Compounds	ADMET Risk	ADMET Code
CA4	4,233	Kow; Sw; 1A2; 2C9; CL
MJ-CA4-I-001	8,189	Size; Kow; Sw; fu; hERG; 1A2; 2C9; 3A4; CL
MJ-CA4-I-004	11,861	Size; RotB; Kow; Sw; fu; hERG; HEPX; 1A2; 2C9; 2D6; 3A4; CL
MJ-CA4-I-005	8,276	Size; RotB; Kow; Sw; fu; HEPX; 1A2; 2C9; 3A4; CL
MJ-CA4-I-006	4,357	Kow; fu; 1A2; 2C9; CL
MJ-CA4-I-007	4	Kow; 1A2; 2C9; CL
MJ-CA4-I-008	4,181	Size; Kow; fu; 1A2; 2C9; CL
MJ-CA4-I-009	8,331	Size; Kow; Sw; fu; hERG; 1A2; 2C9; 3A4; CL
MJ-CA4-I-010	2,346	Size; 1A2; 2C9
MJ-CA4-II-001	6,72	Size; Kow; fu; HEPX; 2C9; 3A4; CL
MJ-CA4-II-002	4	HEPX; 1A2; 2C9; CL
MJ-CA4-II-003	3	1A2; 2C9; CL
MJ-CA4-II-004	6,653	Size; Kow; fu; HEPX; 1A2; 2C9; 2D6; 3A4; CL
MJ-CA4-II-005	8,046	Size; Kow; Sw; fu; HEPX; 1A2; 2C9; 3A4; CL
MJ-CA4-II-007	7,657	Size; RotB; Kow; fu; HEPX; 1A2; 2C9; 3A4; CL
MJ-CA4-II-008	3,734	Size; rat; HEPX; 1A2; CL
MJ-CA4-II-009	9,639	Size; RotB; Kow; fu; hERG; HEPX; 1A2; 2C9; 2D6; 3A4; CL
MJ-CA4-II-010	9	Size; RotB; Kow; Sw; fu; hERG; HEPX+; 2C9; 3A4; CL
MJ-CA4-II-011	6,121	Size; Kow; Sw; fu; HEPX; 1A2; 3A4; CL
MJ-CA4-II-012	6,522	Size; Kow; Sw; fu; HEPX; 1A2; 3A4; CL
MJ-CA4-II-013	8,942	Size; Kow; Sw; fu; HEPX; MUT; 1A2; 2C9; 3A4; CL
MJ-CA4-II-014	7,83	Size; Kow; Sw; fu; HEPX; 1A2; 2C9; 3A4; CL
MJ-CA4-II-015	8,339	Size; Kow; Sw; fu; HEPX; 1A2; 2C9; 3A4; CL
MJ-CA4-II-016	7,799	Size; Kow; Sw; fu; HEPX; 1A2; 2C9; 3A4; CL

Table 3.8: Overall indicator.

3.3 Docking results and Regression models

To estimate the affinity between the new Combretastatin A4 derivatives and β tubulin isotypes, the docking procedure was carried out using Autodock4 software, and the results obtained are presented in Tab. 3.9 and the values are expressed in Kcal/mol.

Poses were grouped into clusters based on their similarity, and for each chosen cluster, the range of energy variation within that cluster was also provided. Scores were chosen based on the most representative cluster. The amounts, in percent, of the cluster size have been reported in the Tab. 3.10.

From a first analysis, it is possible to state that these compounds do not exhibit selectivity on isotypes, because although they have different binding energies, the difference is roughly in the range of 2 kcal/mol and cannot be considered a significant difference since it can be confused with thermal fluctuations.

With the above values, it was possible to calculate the weighted binding energies on the different distributions of β tubulin isotypes in the seven tumor cell lines available to perform subsequent analyses more accurately.

The reason for this operational choice, i.e., to consider precisely the different expression of beta isotypes in the calculation of binding affinity, is due to the desire to find a correlation between experimental results and computational predictions. Within the cell all isotypes coexist simultaneously as well as a whole range of other possible interactions to which the tested compounds might be subjected, so the more faithfully the system attempts to simulate reality the more reliable the results are likely to be.

One of the main purposes of this study is to seek an explanation, through linear models, for the biological behavior of the new derivatives.

A possible correlation was sought between the physical-chemical features of the molecules predicted by the SimulationPlus software, the binding affinity calculated by docking procedures, and the experimental results, particularly the IC50 values provided by A. Huczynski's research group.

The following will show all the model plots created by separately considering the docking scores and physical-chemical characteristics that could explain the observed IC50 values. Then through multivariate regression it tried to improve the models by considering more than one variable thus taking into account more than one characteristic at the same time.

	$\alpha\beta\text{I}$	Range	$\alpha\beta\text{IIa}$	Range	$\alpha\beta\text{IIb}$	Range	$\alpha\beta\text{III}$	Range	$\alpha\beta\text{IVa}$	Range
MJ-CA4-I-005	-7,83	1,29	-7,66	1,64	-7,57	1,26	-7,85	2,18	-7,79	1,4
MJ-CA4-I-007	-8,69	1,09	-8,35	0,24	-8,4	0,22	-8,67	1,9	-8,63	0,31
MJ-CA4-I-008	-9,32	0,65	-8,96	0,19	-8,95	0,33	-9,16	1,11	-9,21	0,35
MJ-CA4-II-001	-8,67	1,37	-8,48	1,03	-8,16	2,38	-8,99	1,09	-8,85	0,59
MJ-CA4-II-002	-8,56	0,3	-8,51	0,26	-8,53	0,24	-8,34	1,03	-8,78	0,18
MJ-CA4-II-004	-9	1,78	-8,82	1,65	-8,84	2,13	-8,86	1,17	-9,16	1,51
MJ-CA4-II-009	-7,94	1,95	-7,59	1,85	-7,57	1,45	-7,76	1,72	-7,98	2,18
MJ-CA4-II-015	-8,52	2,31	-9,63	2,09	-9,36	1,78	-9,36	2,18	-9,57	2,24
MJ-CA4-II-016	-9,56	1,29	-8,91	2,04	-9,23	2,25	-8,95	2,3	-9,47	2,32
	$\alpha\beta\text{IVb}$	Range	$\alpha\beta\text{V}$	Range	$\alpha\beta\text{VI}$	Range	$\alpha\beta\text{VIII}$	Range		
MJ-CA4-I-005	-7,72	1,3	-7,12	0,99	-7,98	2,26	-7,9	0,96		
MJ-CA4-I-007	-8,52	1,2	-8,21	0,33	-8,81	0,68	-8,9	0,87		
MJ-CA4-I-008	-9,1	0,28	-8,8	1,89	-9,5	2,12	-9,46	0,72		
MJ-CA4-II-001	-8,49	1,46	-8,47	1,2	-8,92	0,84	-8,98	0,87		
MJ-CA4-II-002	-8,63	0,18	-8,34	0,66	-8,44	1,7	-8,58	0,34		
MJ-CA4-II-004	-9,14	1,4	-8,82	1,14	-8,89	1,4	-8,92	1,34		
MJ-CA4-II-009	-8,13	2,65	-8,06	1,62	-7,82	1,6	-7,59	2,38		
MJ-CA4-II-015	-9,73	2,08	-8,92	1,19	-9,28	2,71	-9,2	2,11		
MJ-CA4-II-016	-9,32	2,32	-9,12	2,89	-9,49	2,14	-8,98	3,71		

Table 3.9: Autodock docking score and cluster's range.

	$\alpha\beta\text{I}$	$\alpha\beta\text{IIa}$	$\alpha\beta\text{IIb}$	$\alpha\beta\text{III}$	$\alpha\beta\text{IVa}$	$\alpha\beta\text{IVb}$	$\alpha\beta\text{V}$	$\alpha\beta\text{VI}$	$\alpha\beta\text{VIII}$
MJ-CA4-I-005	58%	51%	60%	57%	69%	73%	31%	54%	40%
MJ-CA4-I-007	84%	72%	80%	64%	72%	76%	62%	72%	79%
MJ-CA4-I-008	85%	78%	73%	54%	70%	74%	50%	68%	74%
MJ-CA4-II-001	63%	55%	51%	53%	68%	54%	54%	71%	62%
MJ-CA4-II-002	74%	81%	90%	95%	78%	80%	72%	88%	75%
MJ-CA4-II-004	75%	74%	68%	79%	86%	86%	72%	92%	87%
MJ-CA4-II-009	57%	50%	44%	61%	46%	62%	59%	62%	77%
MJ-CA4-II-015	53%	49%	65%	66%	61%	72%	37%	71%	64%
MJ-CA4-II-016	55%	54%	42%	44%	38%	43%	49%	56%	44%

Table 3.10: Cluster size.

A549	Binding Energy
MJ-CA4-I-005	-7,65
MJ-CA4-I-007	-8,08
MJ-CA4-I-008	-8,41
MJ-CA4-II-001	-8,46
MJ-CA4-II-002	-8,47
MJ-CA4-II-004	-8,55
MJ-CA4-II-009	-8,45
MJ-CA4-II-015	-8,55
MJ-CA4-II-016	-8,63

Table 3.11: Binding energy weighted on β tubulin isotype expression for A549 cell line.

HaCaT	Binding Energy
MJ-CA4-I-005	-7,69
MJ-CA4-I-007	-8,08
MJ-CA4-I-008	-8,41
MJ-CA4-II-001	-8,42
MJ-CA4-II-002	-8,44
MJ-CA4-II-004	-8,52
MJ-CA4-II-009	-8,40
MJ-CA4-II-015	-8,50
MJ-CA4-II-016	-8,58

Table 3.12: Binding energy weighted on β tubulin isotype expression for HaCaT cell line.

HepG2	Binding Energy
MJ-CA4-I-005	-7,75
MJ-CA4-I-007	-8,14
MJ-CA4-I-008	-8,48
MJ-CA4-II-001	-8,51
MJ-CA4-II-002	-8,51
MJ-CA4-II-004	-8,58
MJ-CA4-II-009	-8,47
MJ-CA4-II-015	-8,54
MJ-CA4-II-016	-8,62

Table 3.13: Binding energy weighted on β tubulin isotype expression for HepG2 cell line.

MDA	Binding Energy
MJ-CA4-I-005	-7,62
MJ-CA4-I-007	-8,04
MJ-CA4-I-008	-8,38
MJ-CA4-II-001	-8,42
MJ-CA4-II-002	-8,44
MJ-CA4-II-004	-8,52
MJ-CA4-II-009	-8,43
MJ-CA4-II-015	-8,53
MJ-CA4-II-016	-8,61

Table 3.14: Binding energy weighted on β tubulin isotype expression for MDA cell line.

PC3	Binding Energy
MJ-CA4-I-005	-7,67
MJ-CA4-I-007	-8,09
MJ-CA4-I-008	-8,43
MJ-CA4-II-001	-8,47
MJ-CA4-II-002	-8,48
MJ-CA4-II-004	-8,55
MJ-CA4-II-009	-8,46
MJ-CA4-II-015	-8,56
MJ-CA4-II-016	-8,63

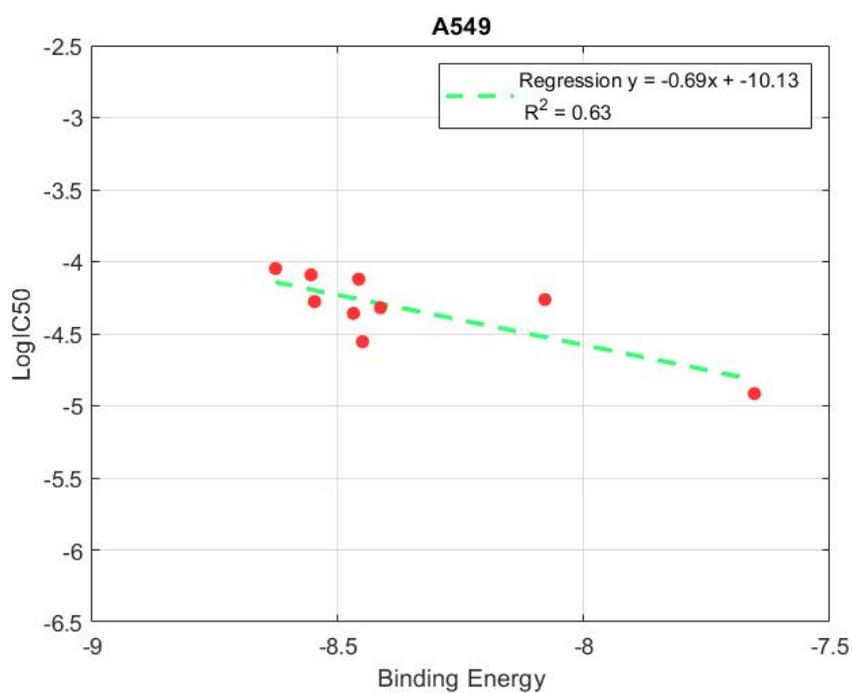
Table 3.15: Binding energy weighted on β tubulin isotype expression for PC3 cell line.

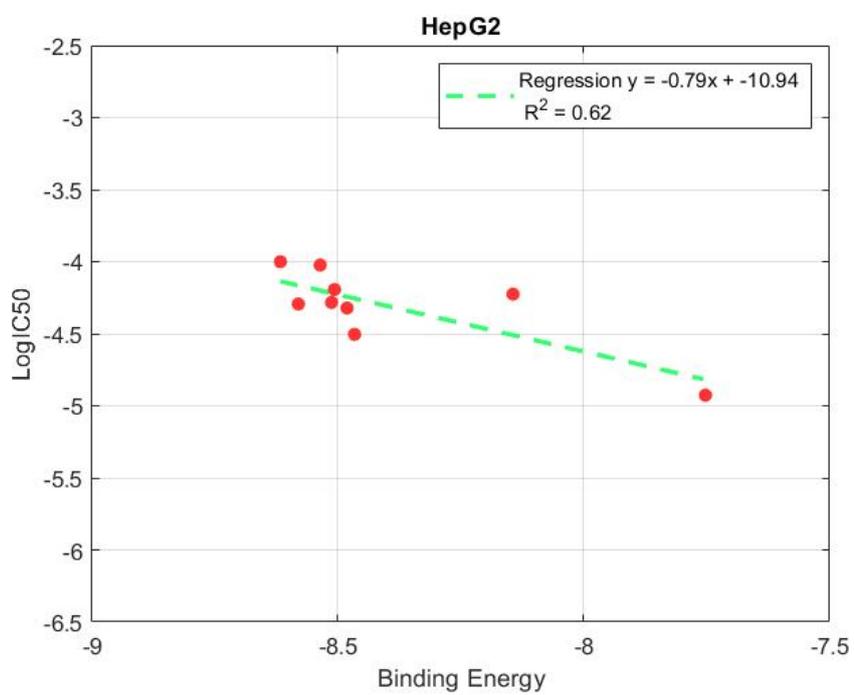
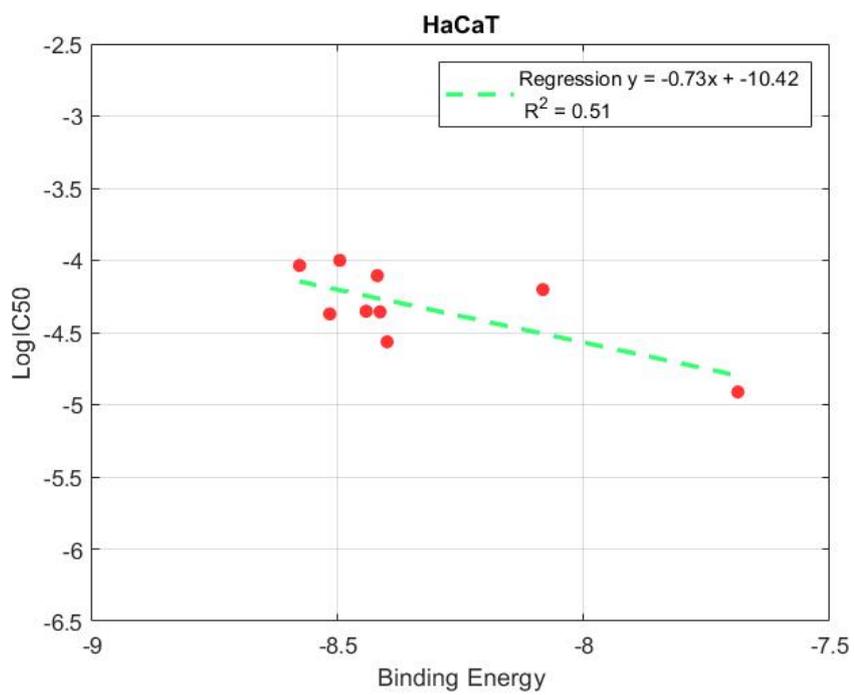
SW480	Binding Energy
MJ-CA4-I-005	-7,68
MJ-CA4-I-007	-8,10
MJ-CA4-I-008	-8,43
MJ-CA4-II-001	-8,47
MJ-CA4-II-002	-8,48
MJ-CA4-II-004	-8,56
MJ-CA4-II-009	-8,46
MJ-CA4-II-015	-8,56
MJ-CA4-II-016	-8,64

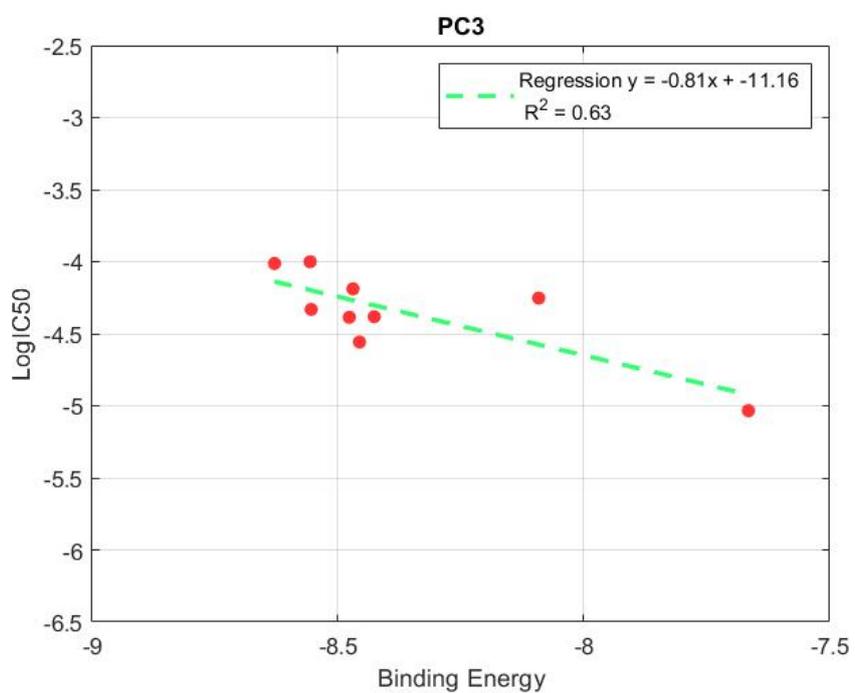
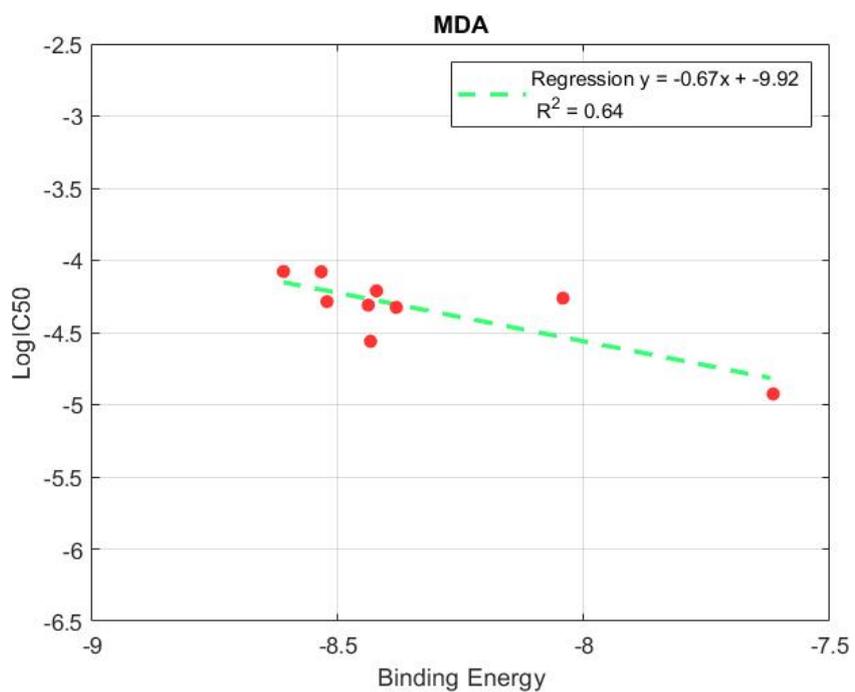
Table 3.16: Binding energy weighted on β tubulin isotype expression for SW480 cell line.

SW620	Binding Energy
MJ-CA4-I-005	-7,75
MJ-CA4-I-007	-8,16
MJ-CA4-I-008	-8,49
MJ-CA4-II-001	-8,53
MJ-CA4-II-002	-8,52
MJ-CA4-II-004	-8,60
MJ-CA4-II-009	-8,50
MJ-CA4-II-015	-8,60
MJ-CA4-II-016	-8,67

Table 3.17: Binding energy weighted on β tubulin isotype expression for SW620 cell line.







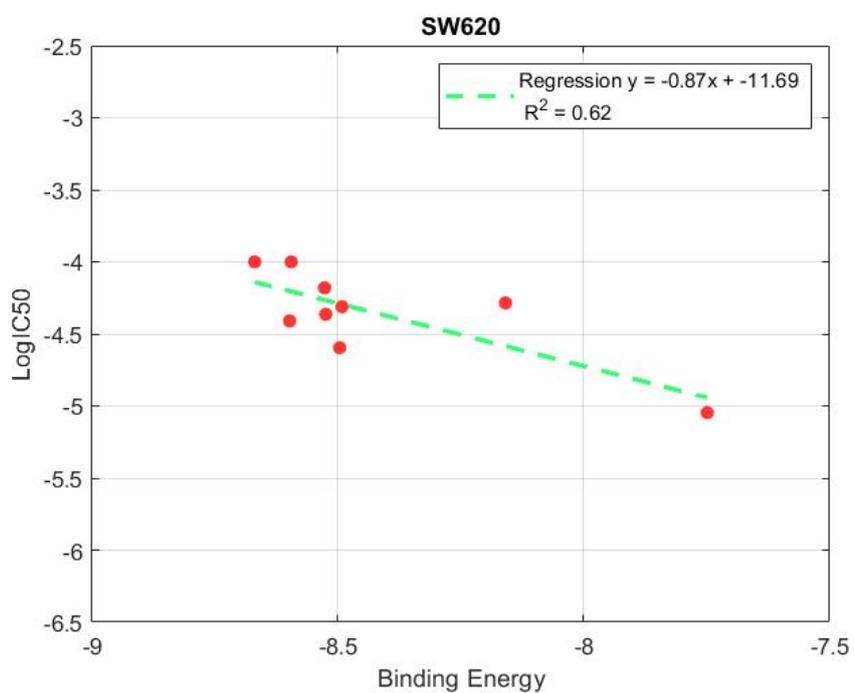
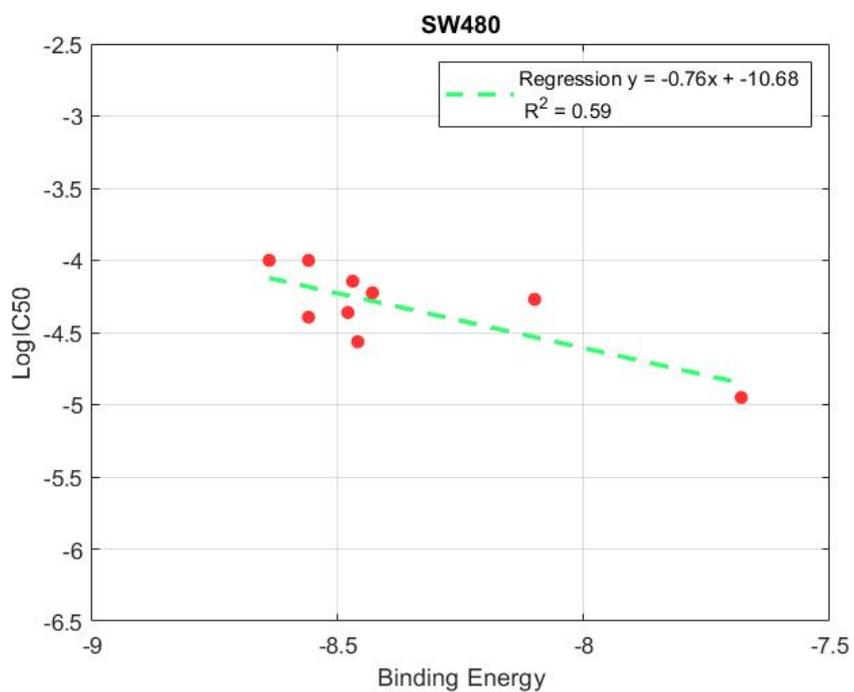
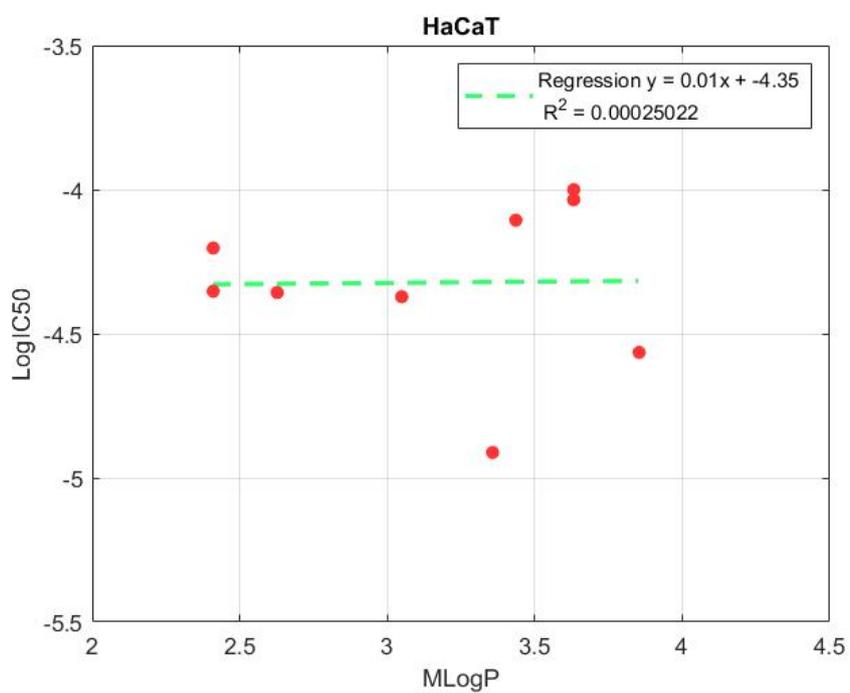
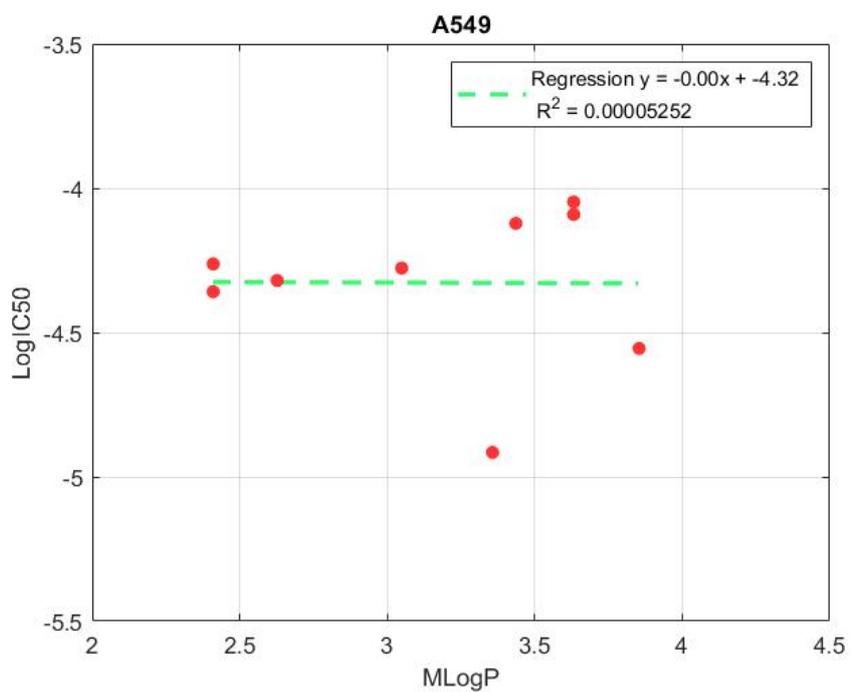
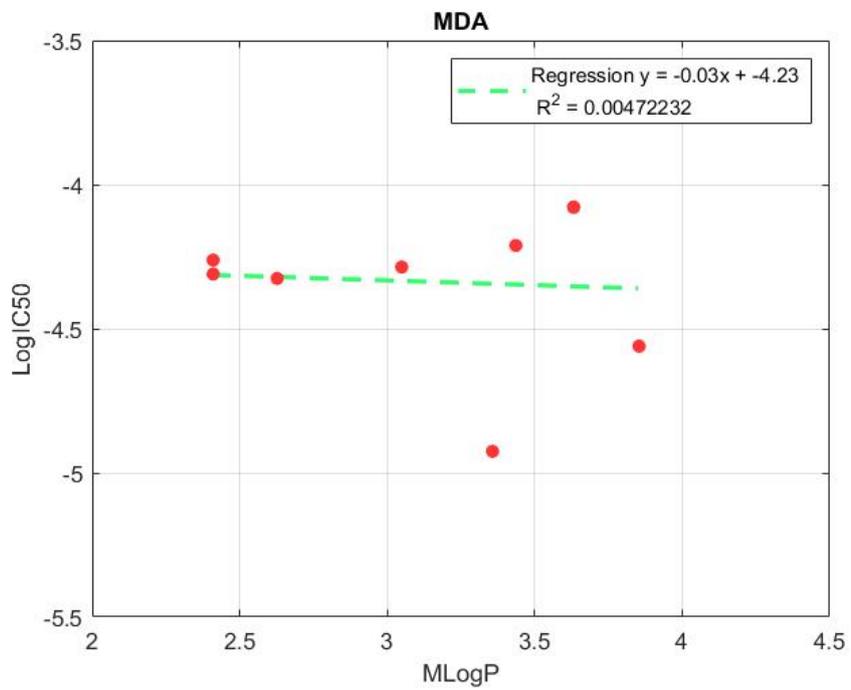
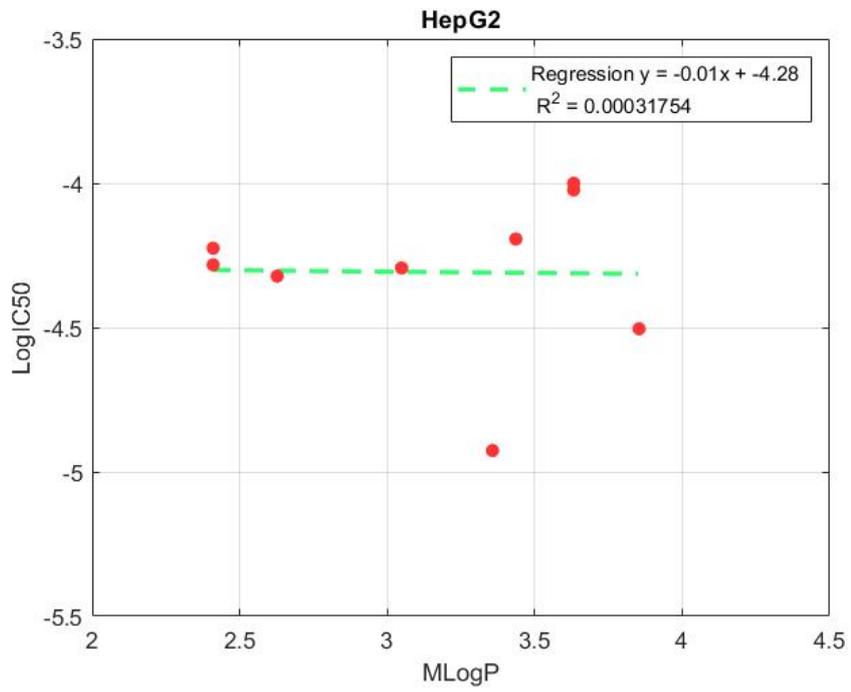
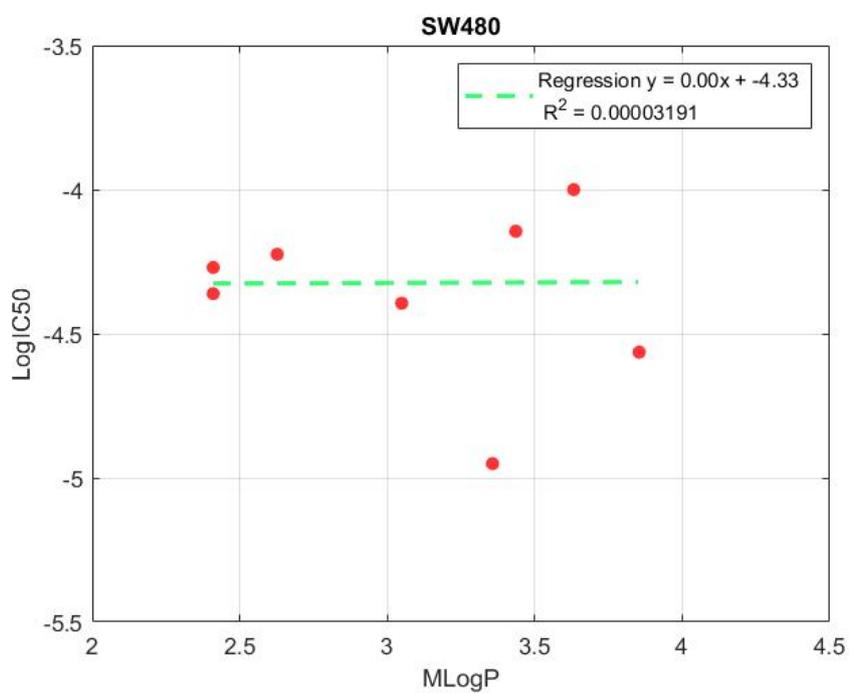
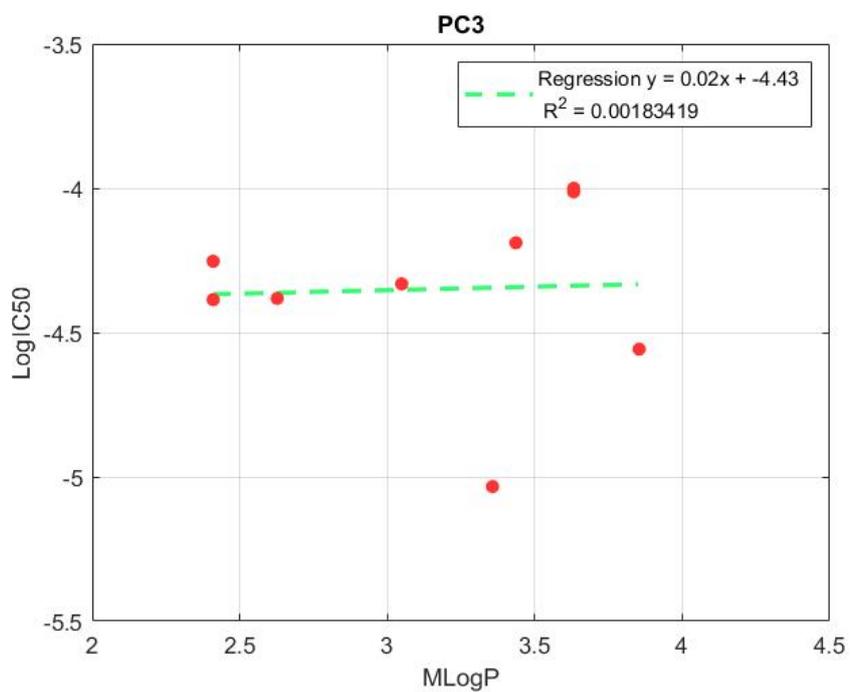


Figure 3.1: Regression models between weighted binding energy and LogIC50.







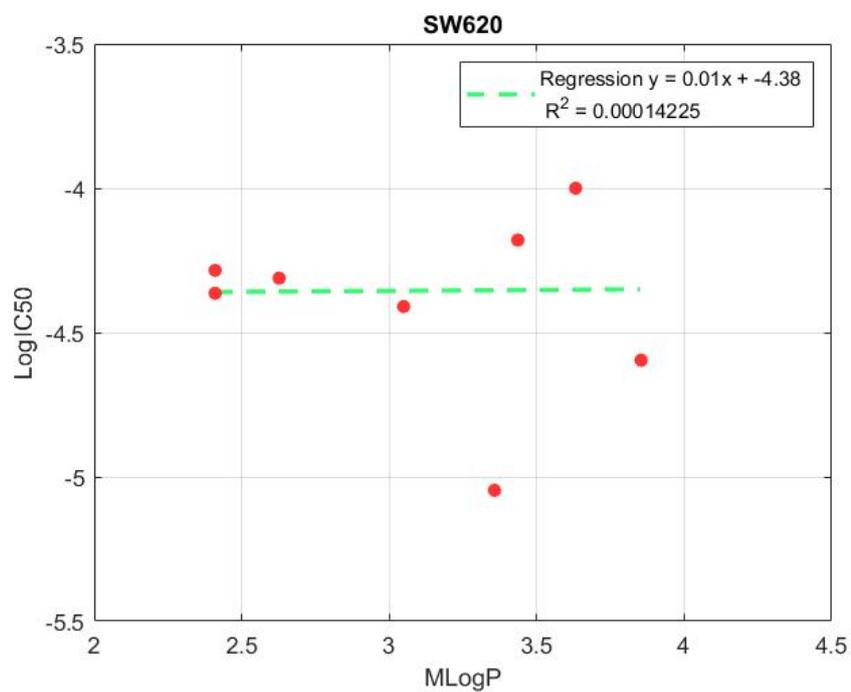
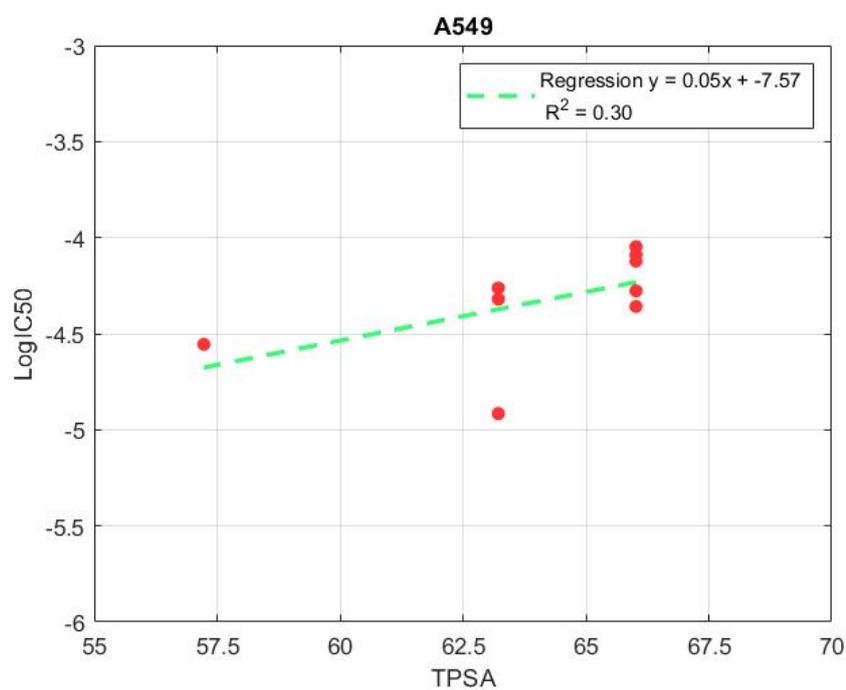
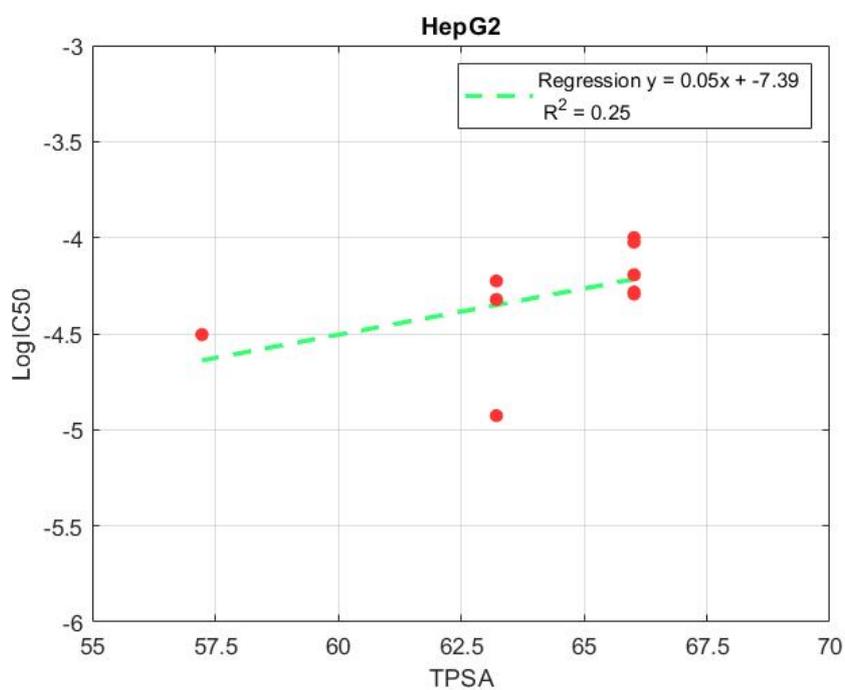
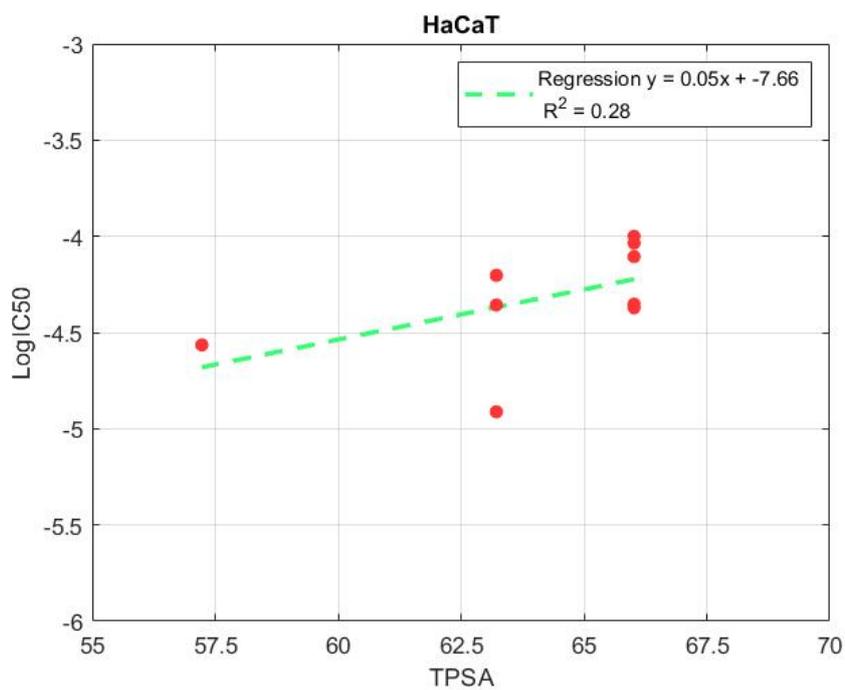
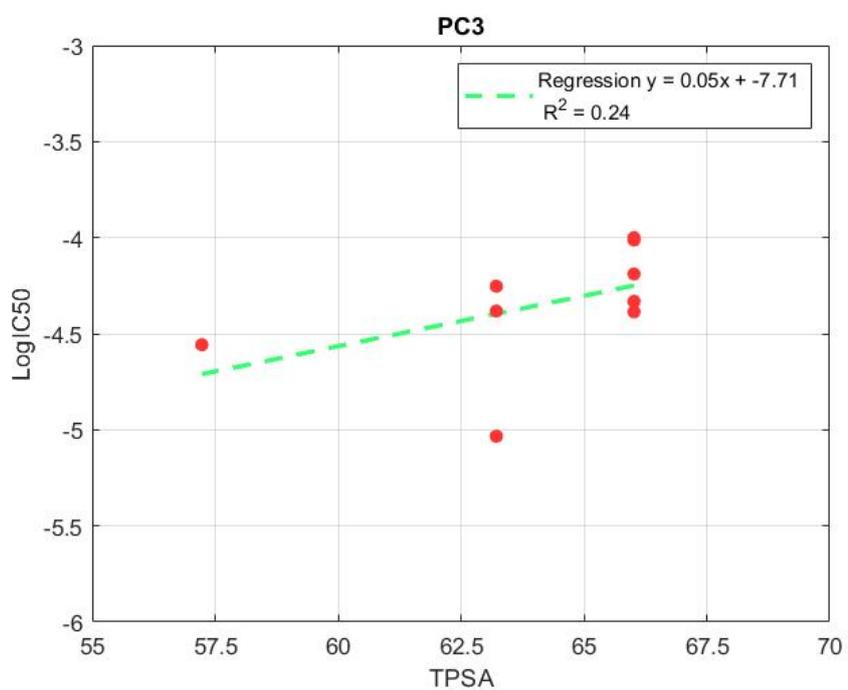
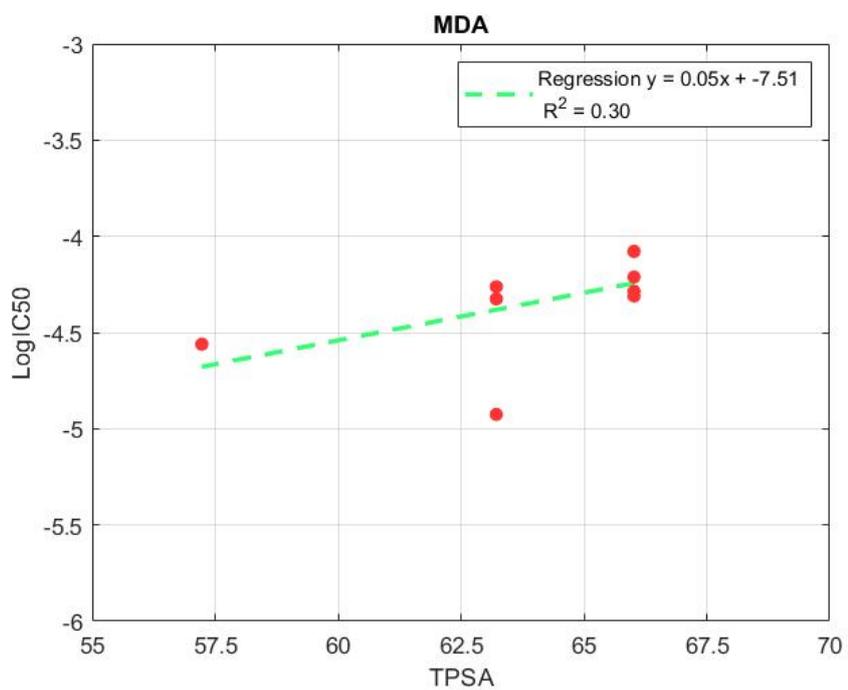


Figure 3.2: Regression models between MlogP and LogIC50.







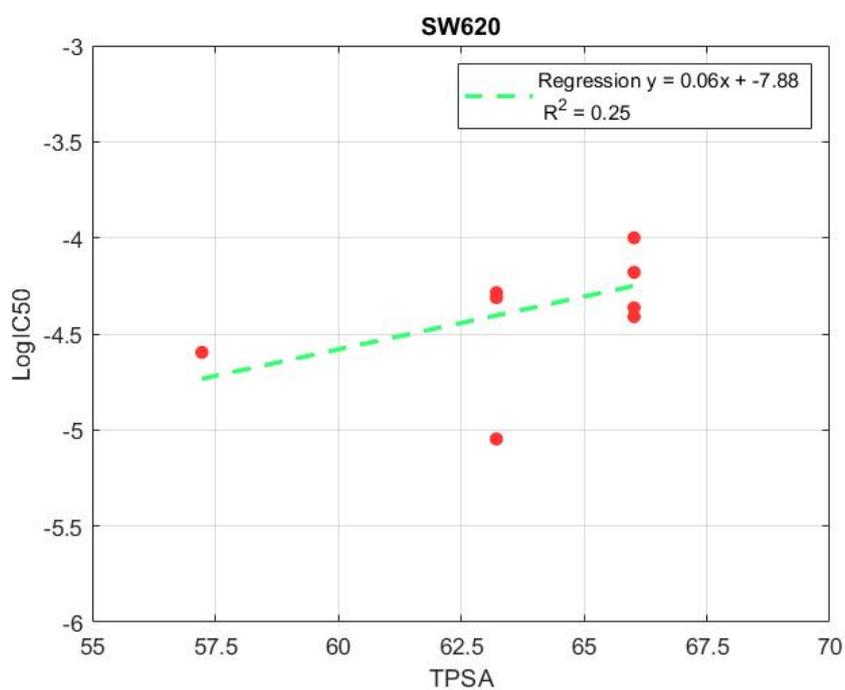
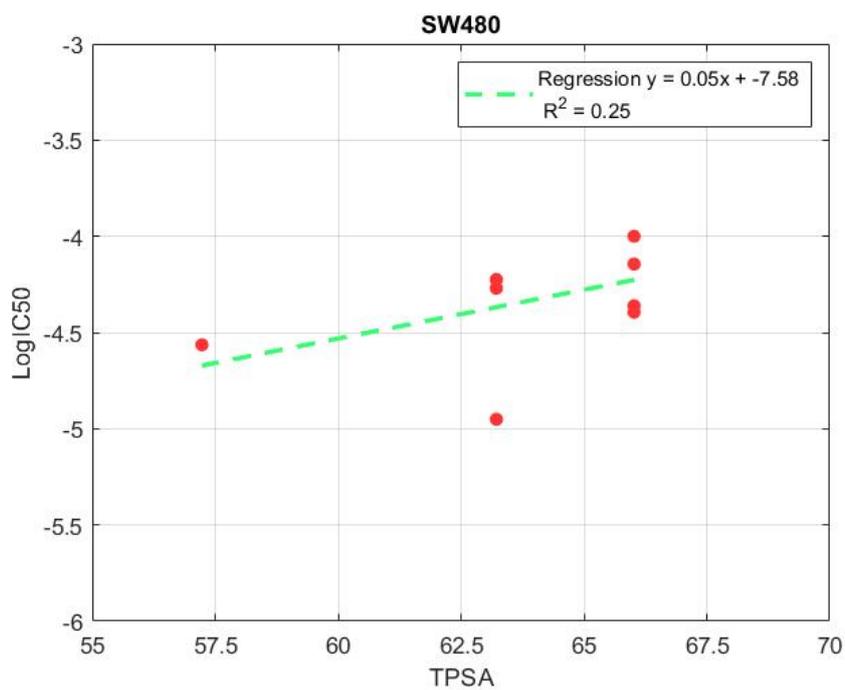
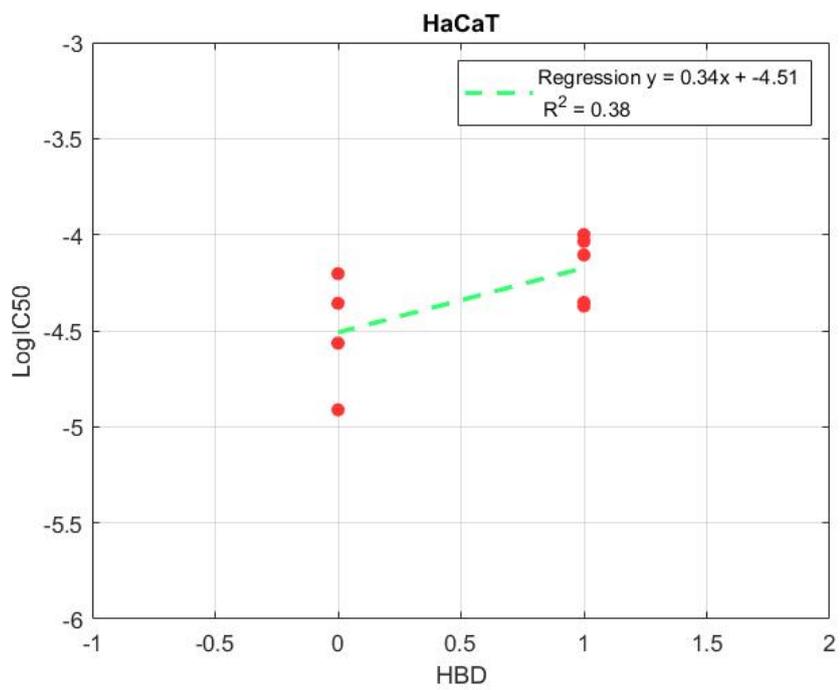
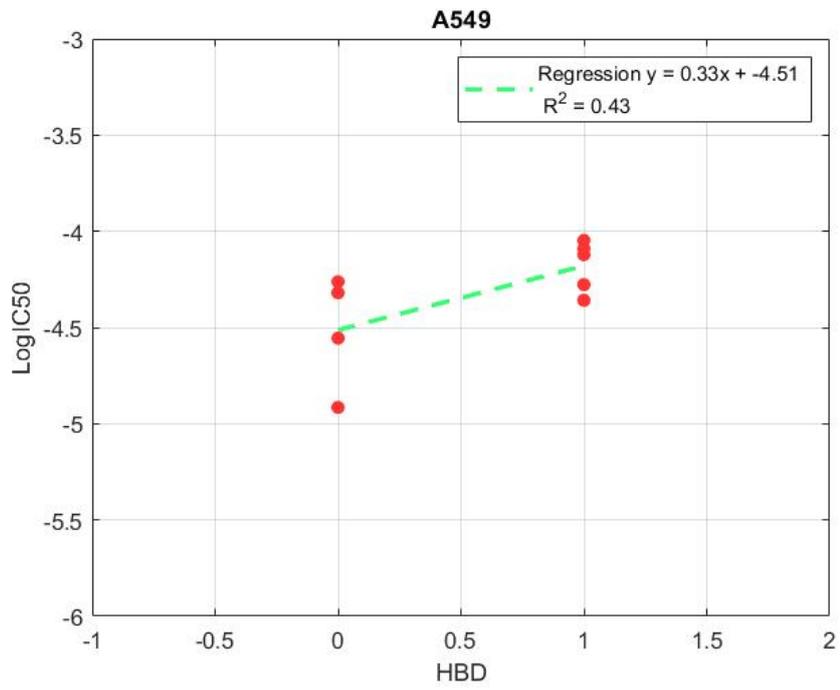
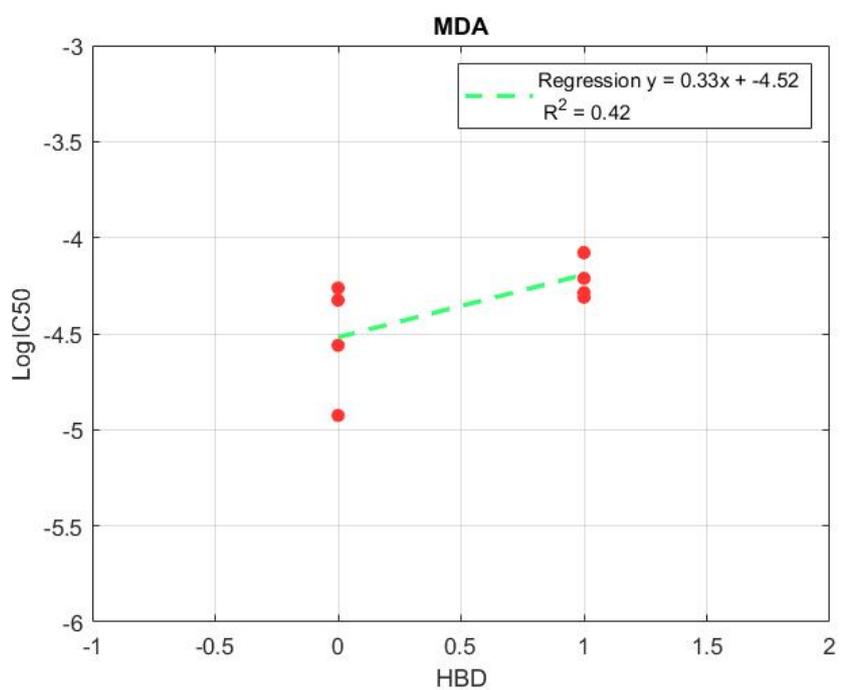
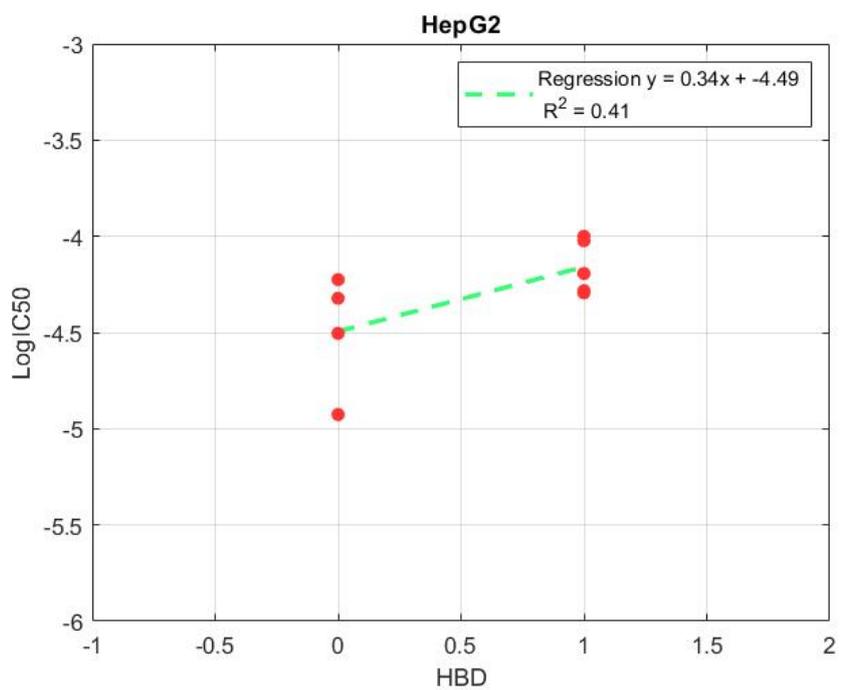
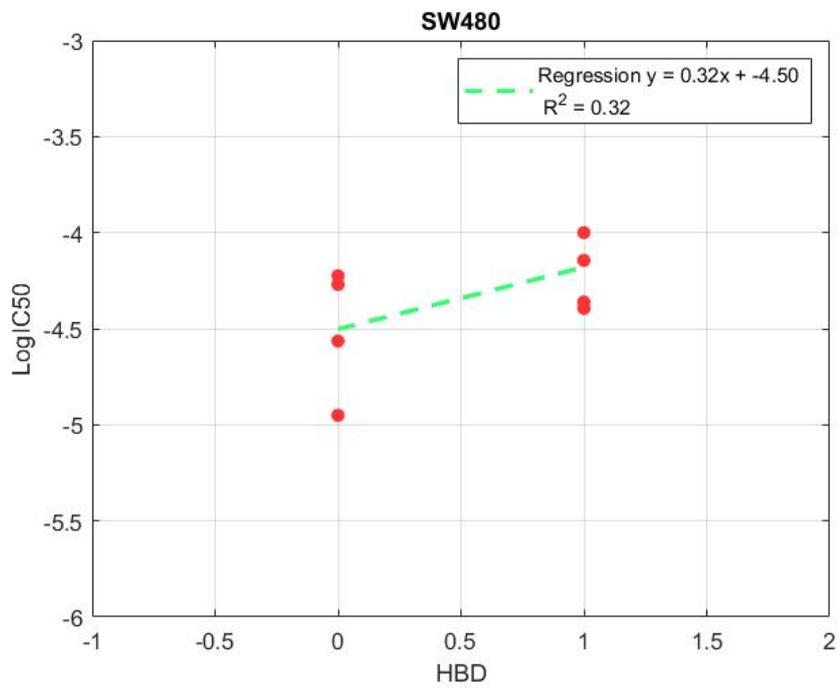
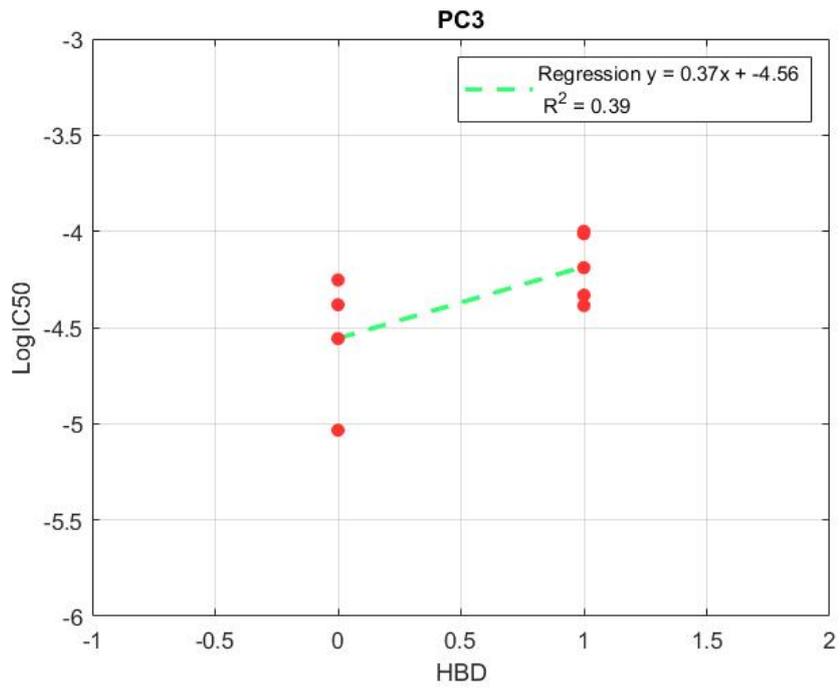


Figure 3.3: Regression models between TPSA and LogIC50.







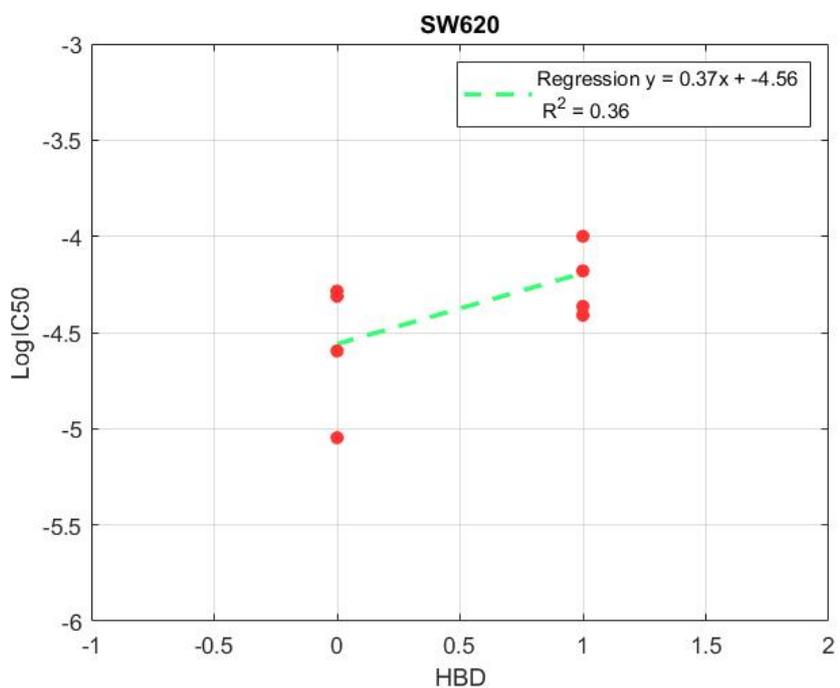
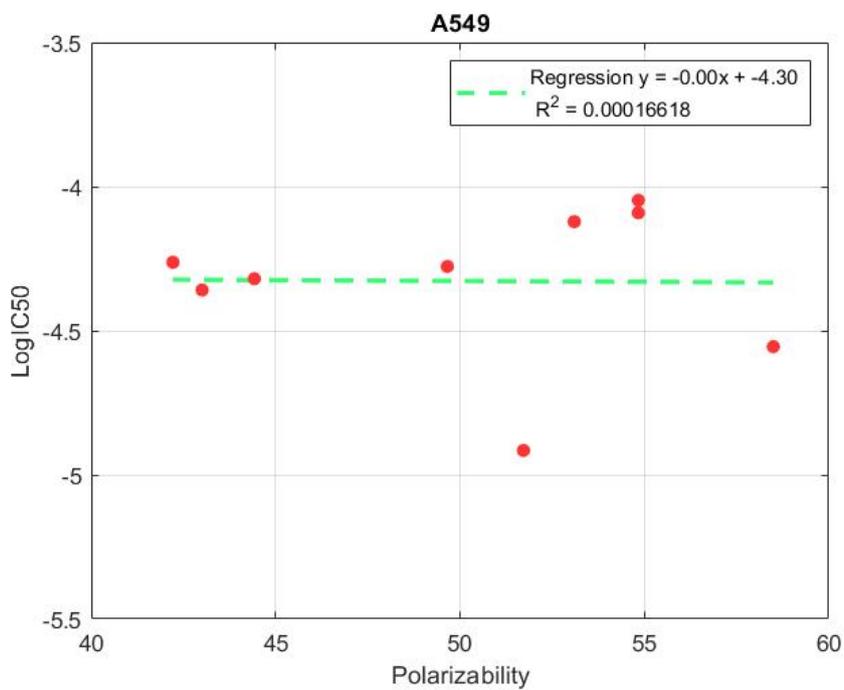
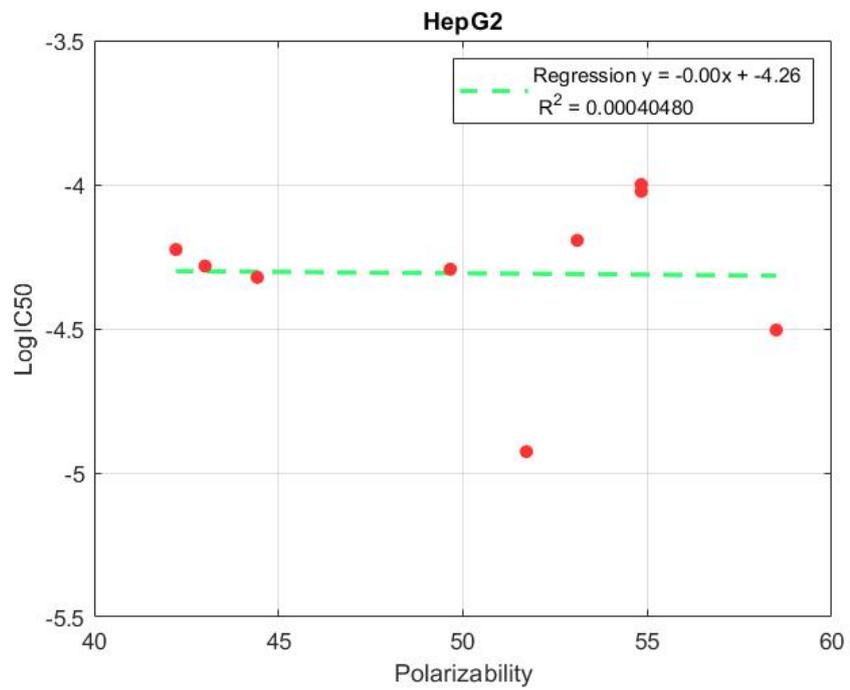
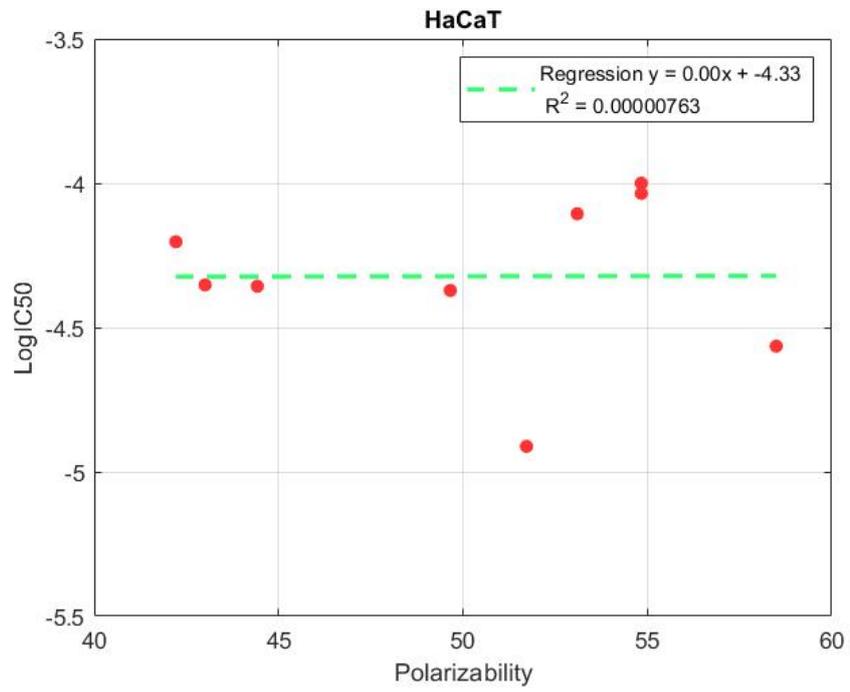
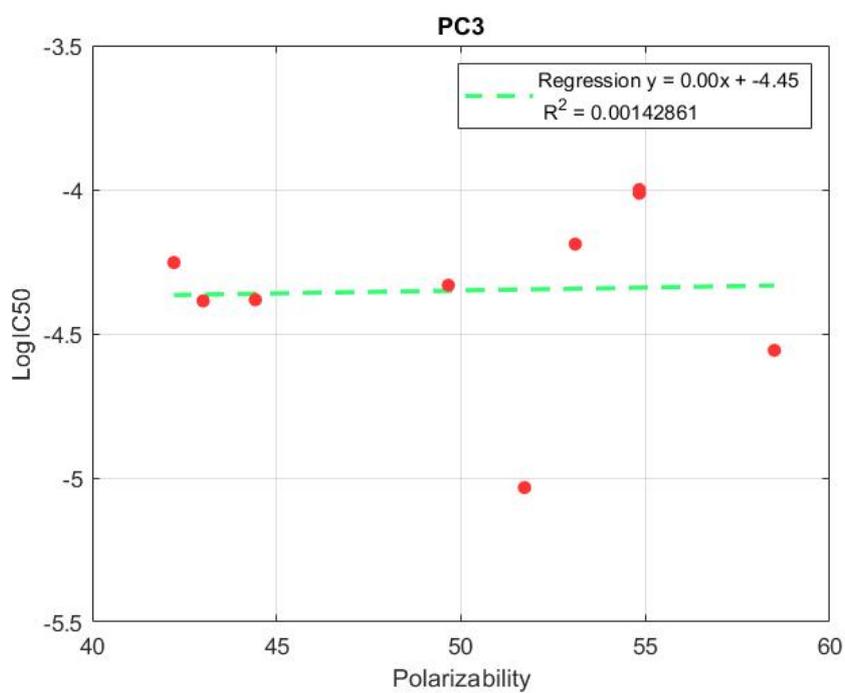
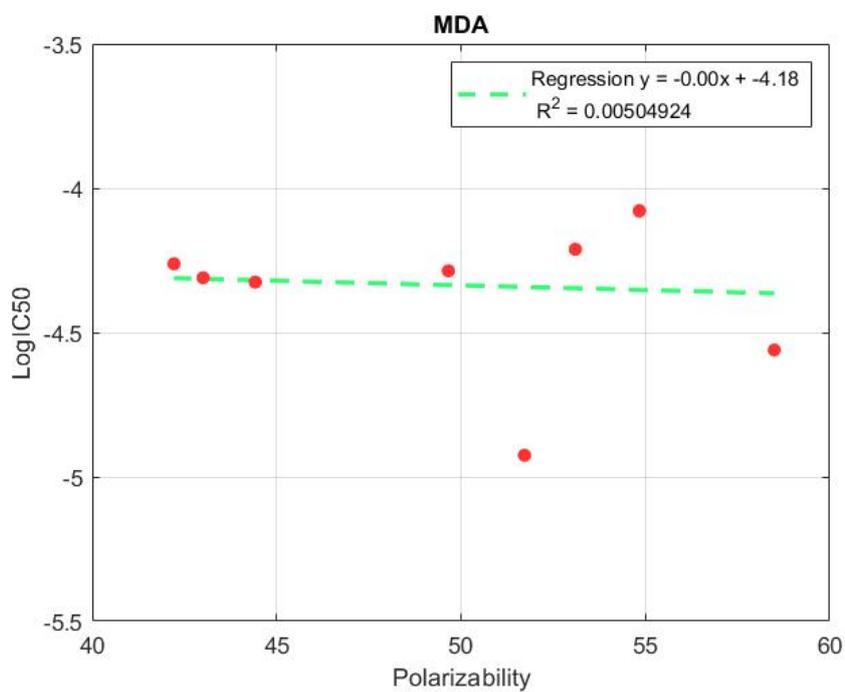


Figure 3.4: Regression models between HBD and LogIC50.







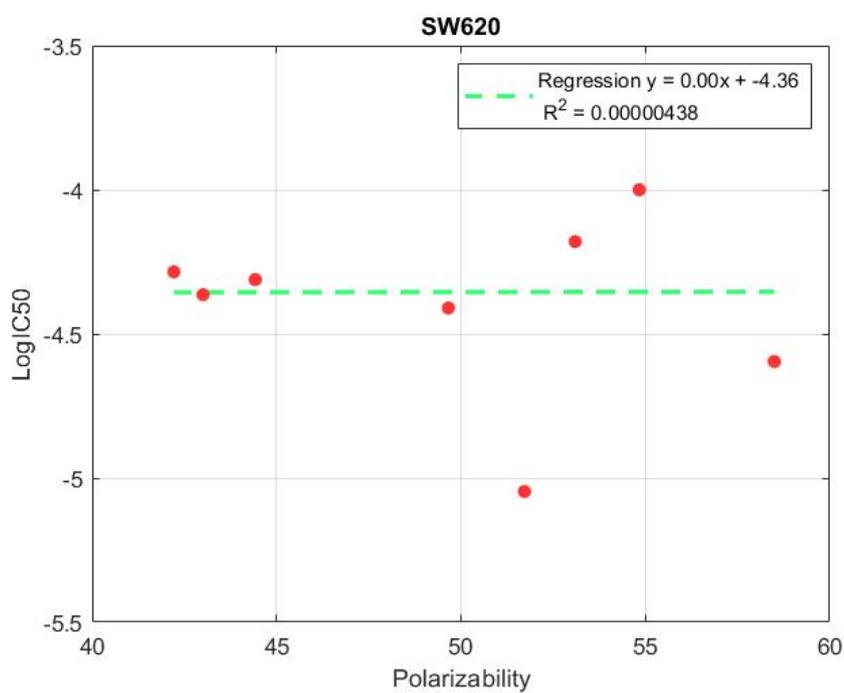
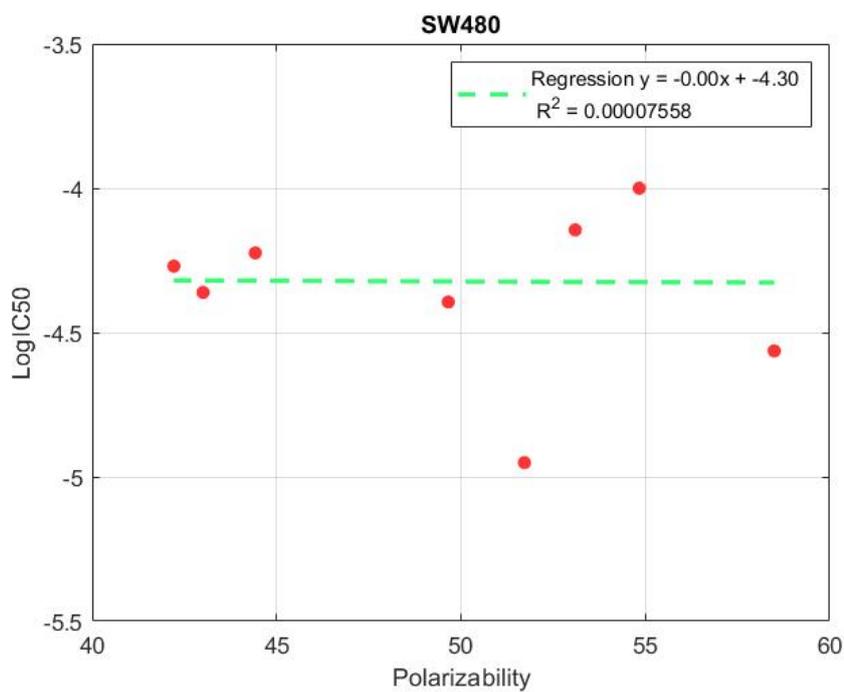
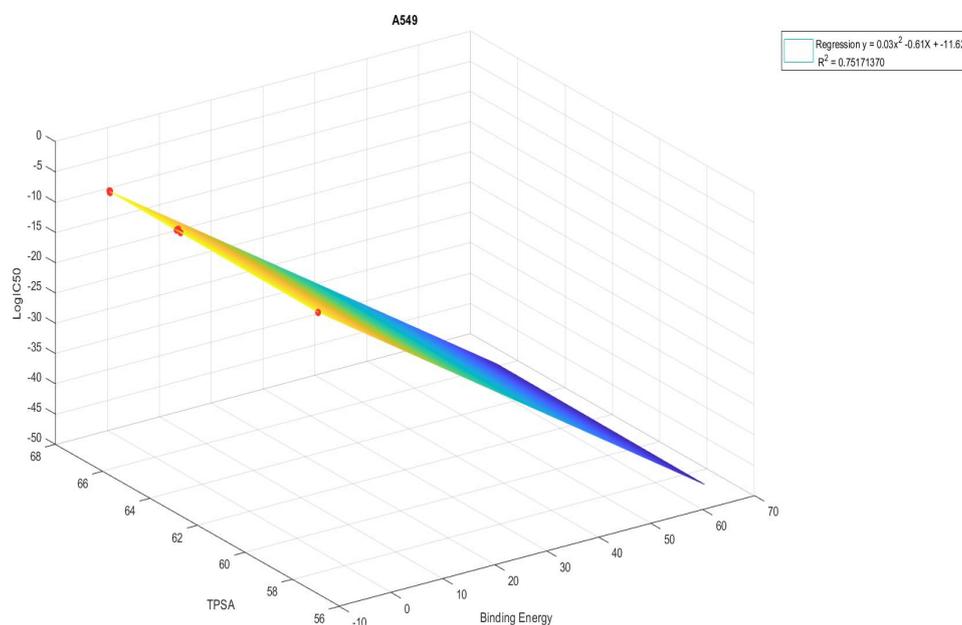


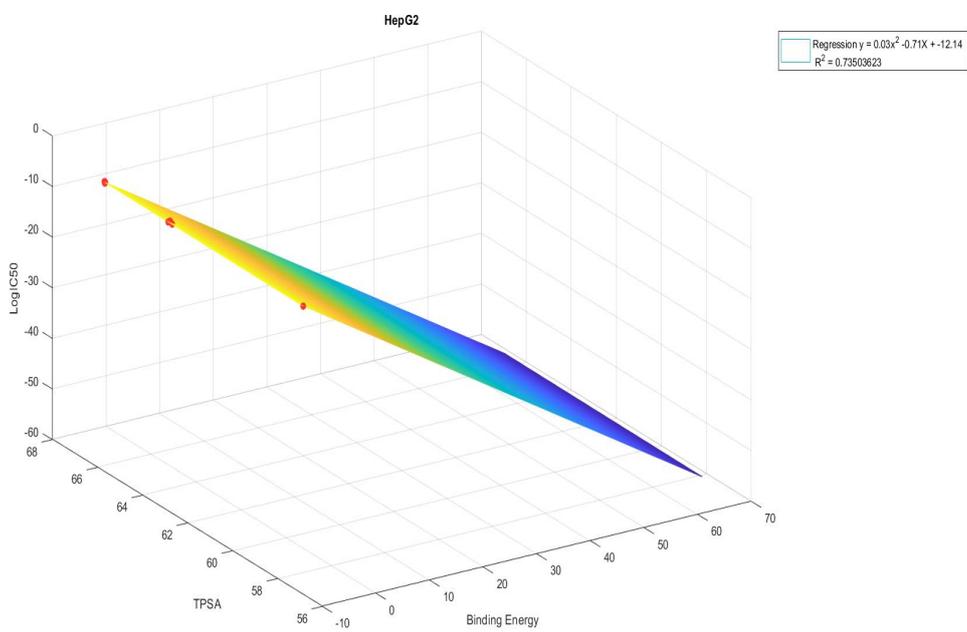
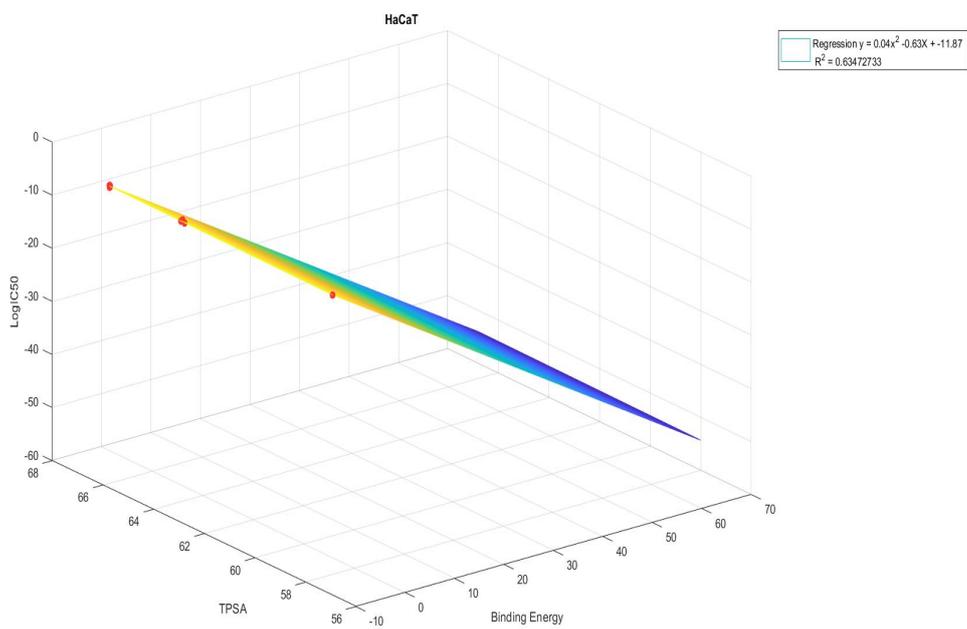
Figure 3.5: Regression models between Polarizability and LogIC50.

The results achieved show that binding affinity appears to be the reason that best explains the biological behavior of the compounds, while the partition coefficient (MlogP) and the polarizability of the molecule appear to show no correlation, or at least not a linear correlation.

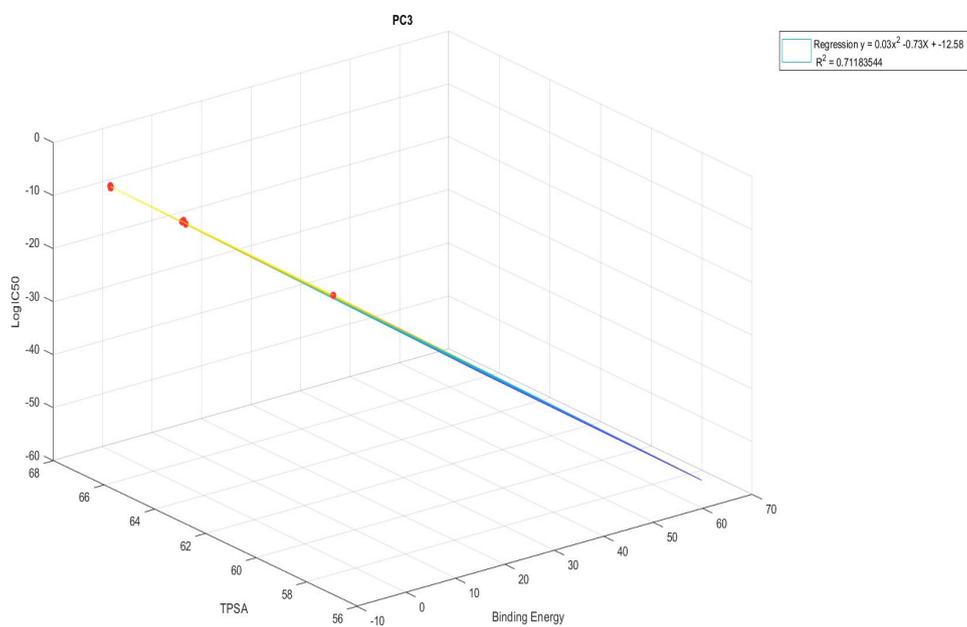
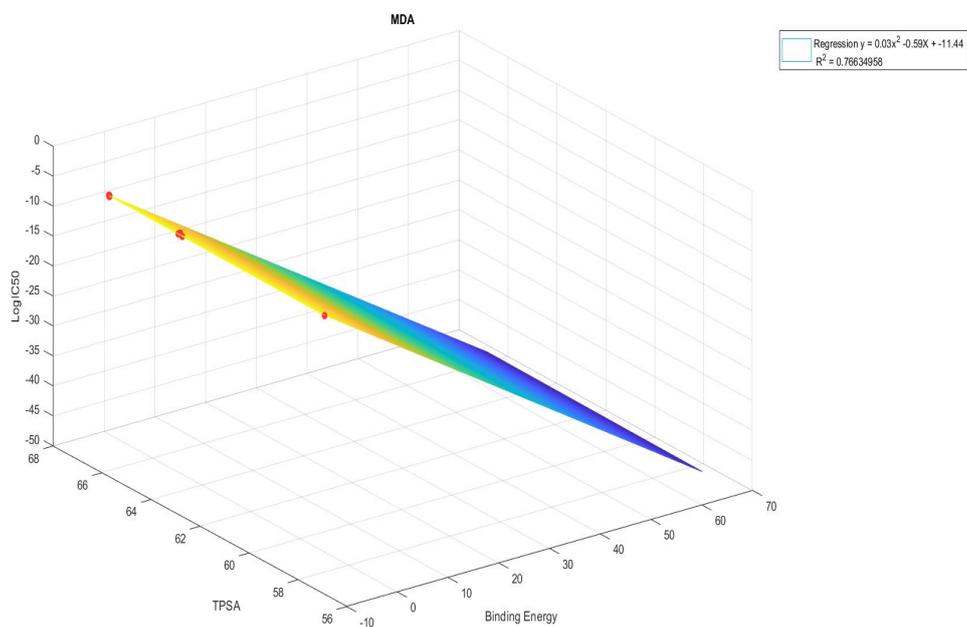
To improve the model, a multivariate regression was performed considering both β tubulin binding energy and Topological polar surface area (TPSA) as independent variables and LogIC₅₀ as the dependent variable, to understand how closely these two characteristics together describe biological behavior. The same thing was done considering the hydrogen bond donors (HBD) instead of TPSA, and all results are reported below.



Results



3.3 – Docking results and Regression models



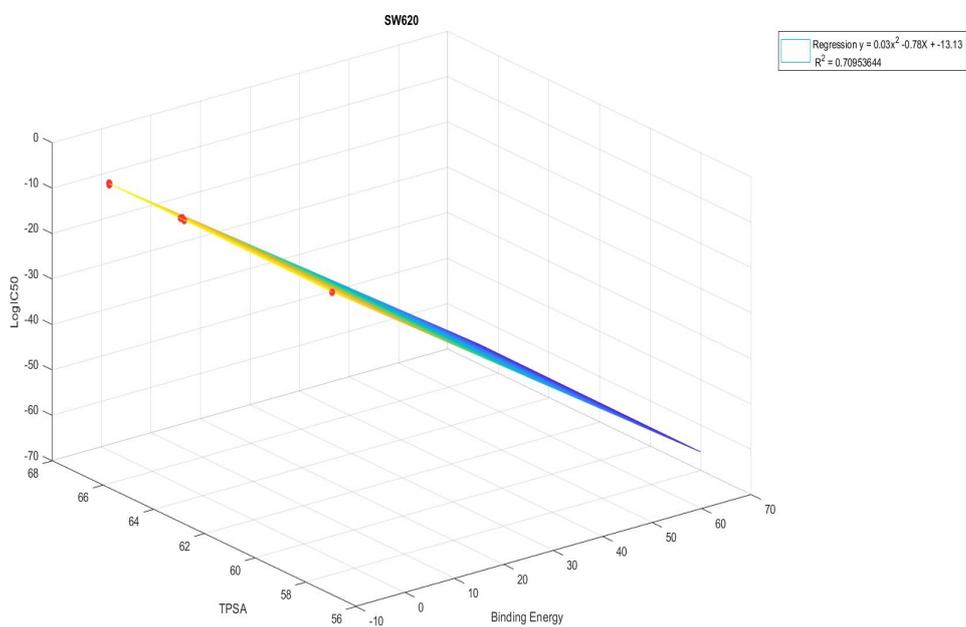
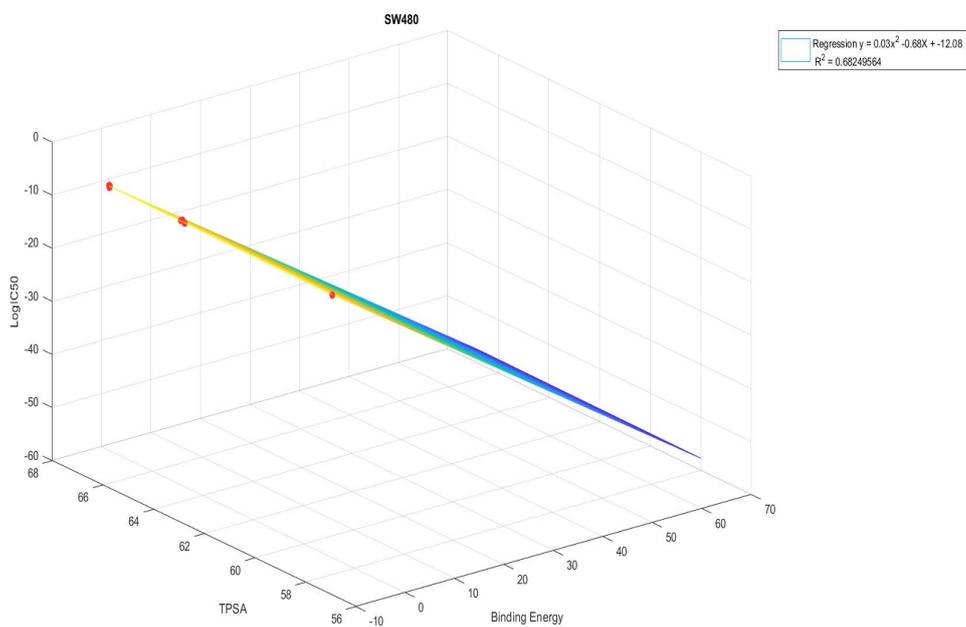
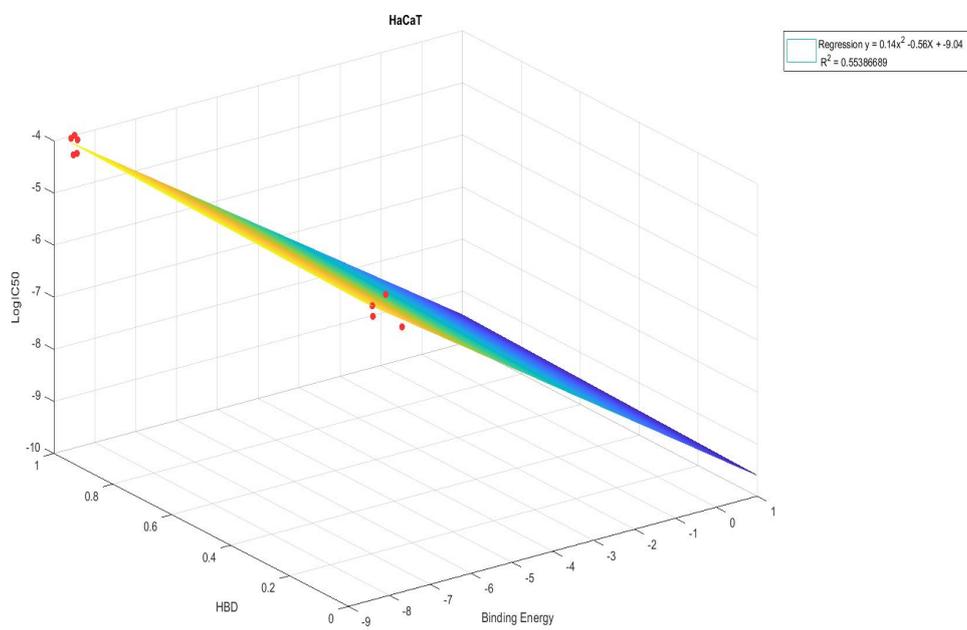
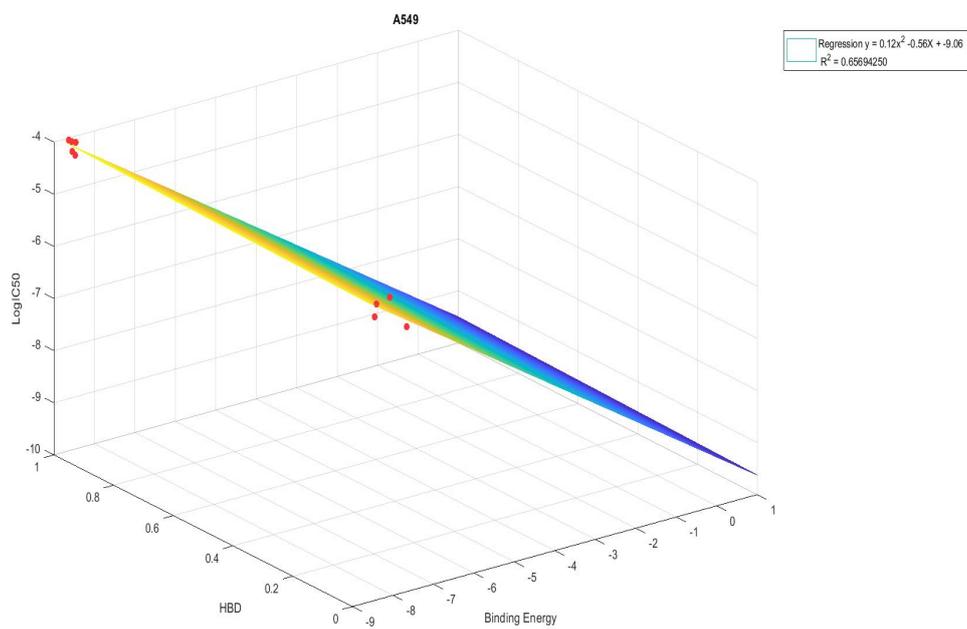
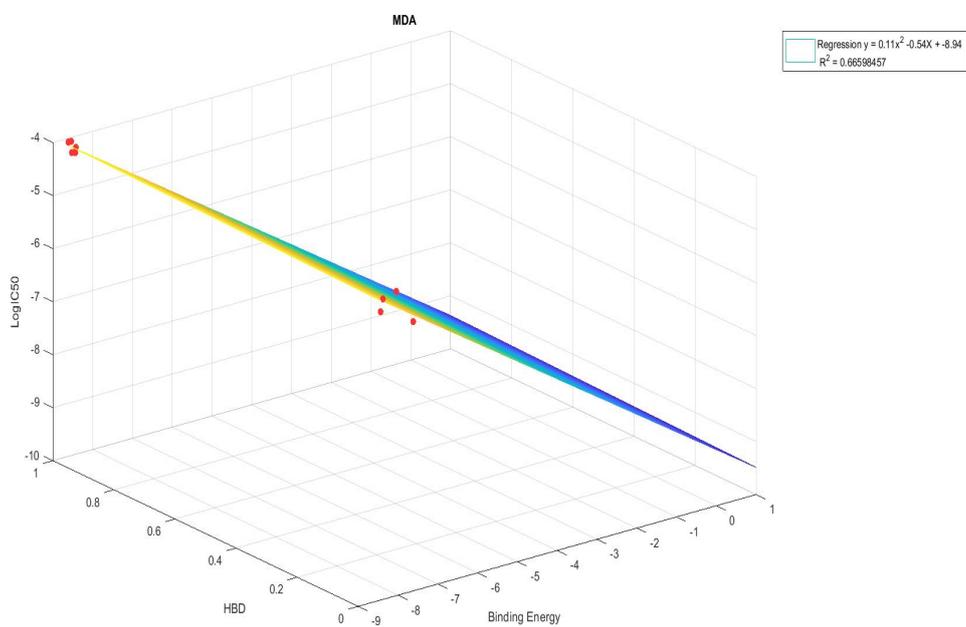
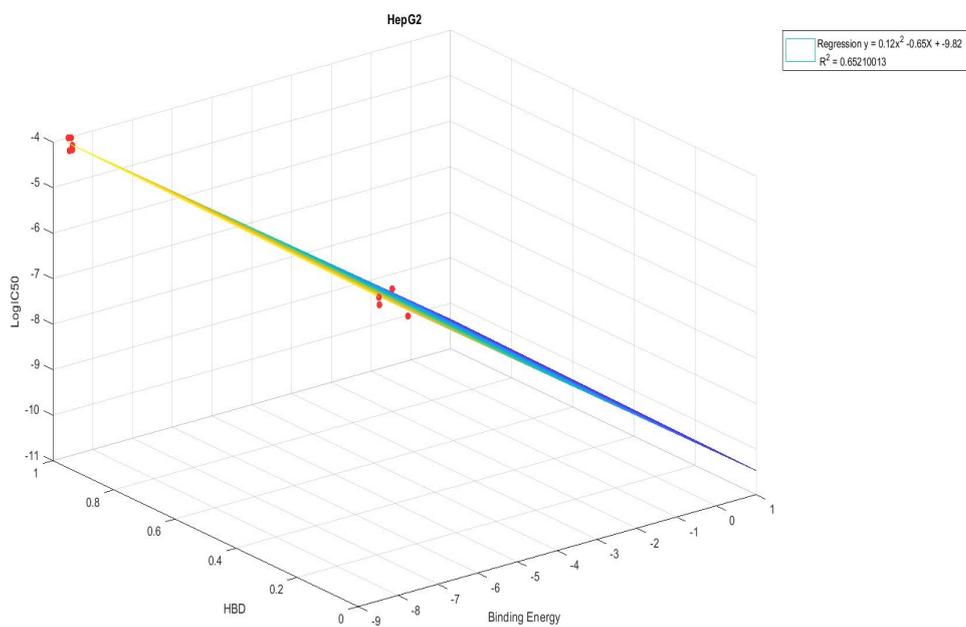


Figure 3.6: Multivariate regression models: TPSA, Binding Energy and LogIC50.

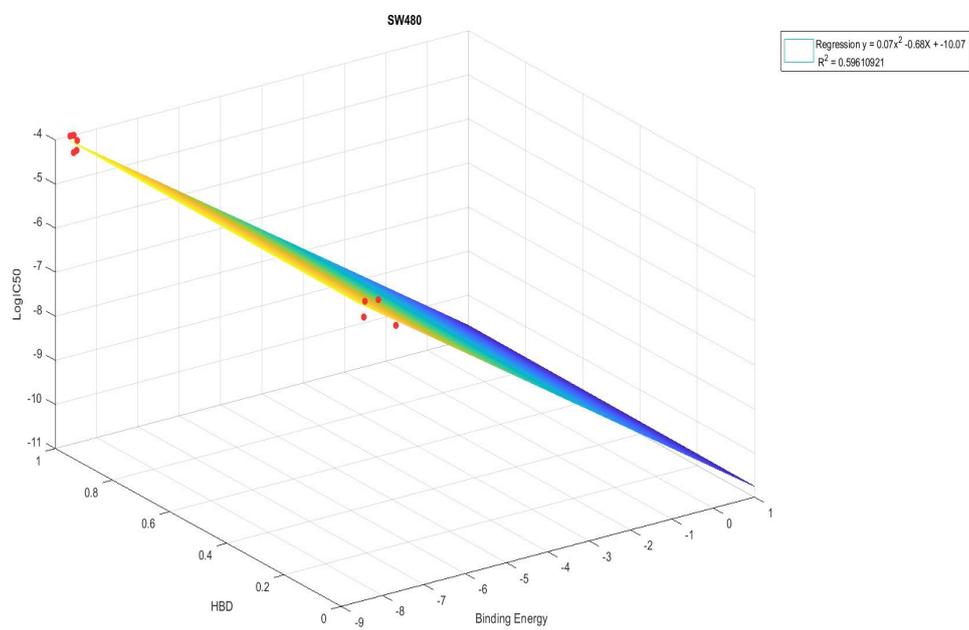
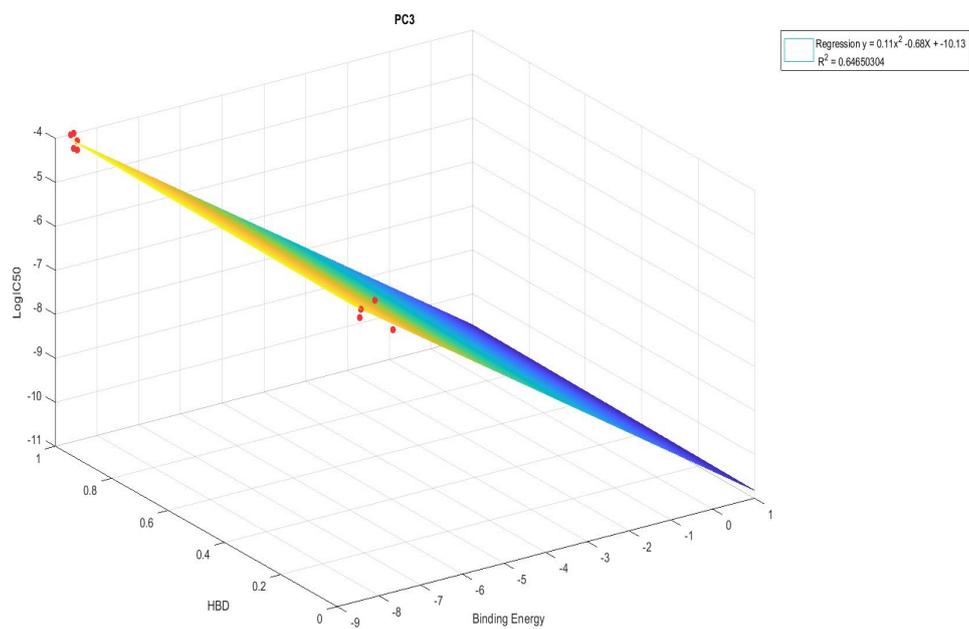
3.3 – Docking results and Regression models



Results



3.3 – Docking results and Regression models



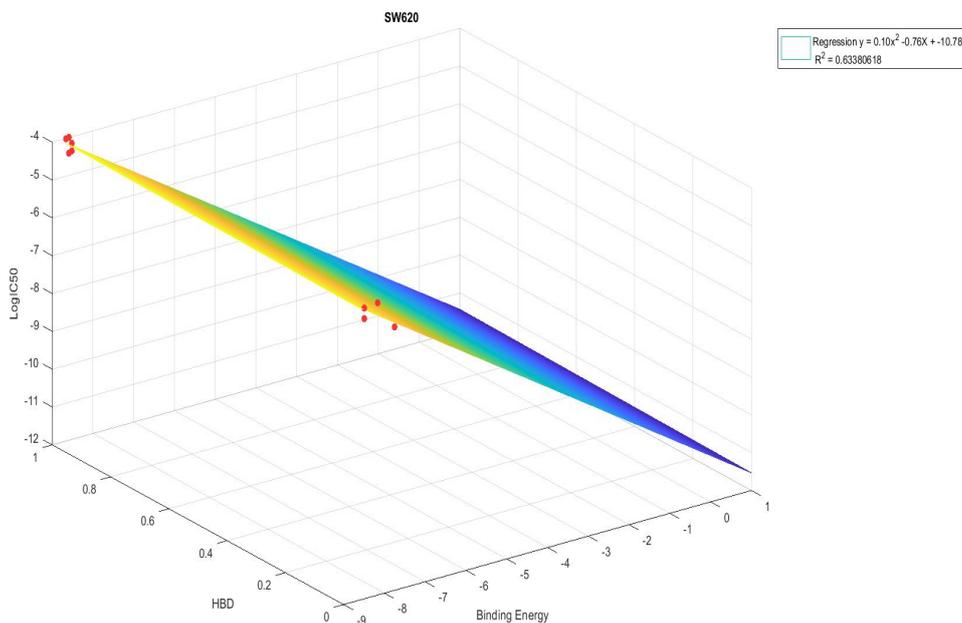


Figure 3.7: Multivariate regression models: HBD, Binding Energy and LogIC50.

R^2	A549	HaCaT	HepG2	MDA
BE_w	0.63	0.51	0.62	0.64
MLogP	5.2×10^{-5}	2.5×10^{-4}	3.1×10^{-4}	4.7×10^{-3}
TPSA	0.30	0.28	0.25	0.30
HBD	0.43	0.38	0.41	0.42
POL	1.6×10^{-4}	7.6×10^{-6}	4×10^{-4}	5×10^{-3}
	PC3	SW480	SW620	
BE_w	0.63	0.59	0.62	
MLogP	1.8×10^{-3}	3.1×10^{-5}	1.4×10^{-4}	
TPSA	0.24	0.25	0.25	
HBD	0.39	0.32	0.36	
POL	1.4×10^{-3}	7.5×10^{-5}	4.4×10^{-6}	

Table 3.18: Linear Regression models: R^2 value. BE_w : weighted binding energy, MLogP: logarithm of the partition coefficient, TPSA: Topological polar surface area, HBD: hydrogen bond donors, POL: polarizability.

R ²	A549	HaCaT	HepG2	MDA	PC3	SW480	SW620
BE _w -TPSA	0.75	0.63	0.74	0.76	0.71	0.68	0.71
BE _w -HBD	0.66	0.55	0.65	0.67	0.65	0.60	0.63

Table 3.19: Multivariate linear regression models: R² value

From the R square values calculated on the resulting models, it is easily seen that the weighted binding energy is the best correlated variable with logIC50, thus with the biological activity of the molecules. Moreover, the multivariate model that takes into account both TPSA as well shows even higher correlation values, indicating that these two features, in this case seem to be the best explain the *in vitro* behavior of the new compounds.

One point to note is that the weighted binding energies, calculated by considering the closest to the real expression of β tubulin isotypes in those cell lines also show higher correlation values, while the weighted energy over a uniform distribution of isotypes, the HaCaT cell line for which the exact percentage distribution could not be found, has the lowest R² value, both in the model with only one variable, the weighted binding energy, and in the model that takes into account both TPSA and weighted binding energy.

3.4 PROTAC

In this section, the goal is to evaluate potentially applicable molecules as warheads for a PROTAC construct. Compounds MJ-CA4-I-005 and MJ-CA4-II-009 were the best in the entire list, because they exhibit higher cytotoxicity, while for the other compounds there was a problem with solubility.

From the binding affinities obtained through Autodock, it appears that the derivatives show no selectivity on the different isotypes, for this reason the subsequent evaluations were carried out considering β IIa, β III and β IVb as targets because by clustering the isotypes with reference to the Colchicine binding site structural similarity they are to be considered representative of all possible variations for the residues forming the pocket.

It is possible to categorize the structures as follows [37]:

- Type I = $\alpha\beta$ I, $\alpha\beta$ IVa, $\alpha\beta$ IVb,
- Type II = $\alpha\beta$ IIa, $\alpha\beta$ IIb,
- Type III = $\alpha\beta$ III, $\alpha\beta$ V.

Therefore, the docking procedure was carried out on three targets with two possible ligands each with a total of six combinations.

3.4.1 Consensus docking

To obtain the best poses and binding energies, consensus docking was performed in this case, this computational procedure involves docking ligands using several docking techniques and comparing the binding orientations predicted by the various methods for the same ligand.

The root-mean-square deviation of the multiple docking results gathered for each ligand is often calculated in order to determine the number of docking techniques that produce the same binding posture. Three different programs were considered, specifically Autodock, Vina, and Moe.

The results obtained are shown in the table below:

Target	Ligand	Score	Rescore
β IIa	MJ-CA4-I-005	-6.800	-7.091
β IIa	MJ-CA4-II-009	-9.116	-7.343
β III	MJ-CA4-I-005	-9.079	-7.208
β III	MJ-CA4-II-009	-8.736	-7.229
β IVb	MJ-CA4-I-005	-8.399	-7.002
β IVb	MJ-CA4-II-009	-9.507	-6.934

Table 3.20: Consensus docking results.

A crucial step in the construction of a PROTAC molecule is the site selection of the linker attachment point, as concerns the ligand for the protein of interest the most important aspect to consider is that the linker should be attached at a point on the ligand that does not interfere with or weaken the interaction between ligand and receptor. For this reason, poses were analyzed and the best anchor point was found by choosing the most exposed area.

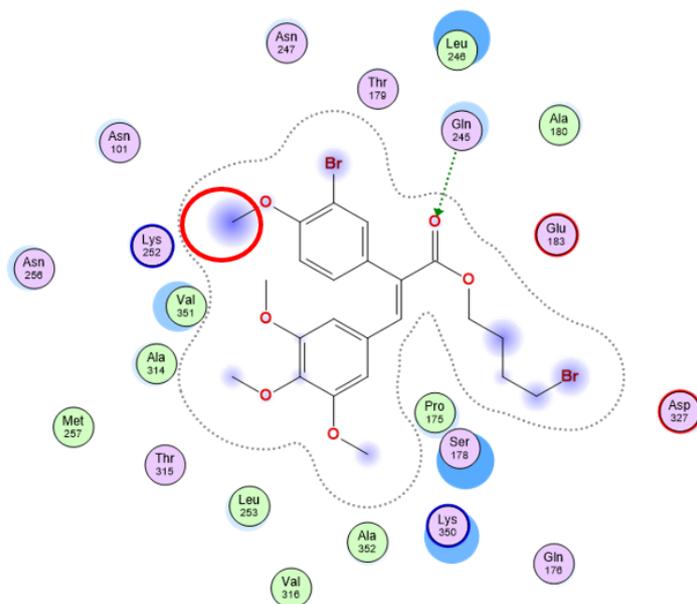


Figure 3.10: β III and MJ-CA4-I-005 interactions.

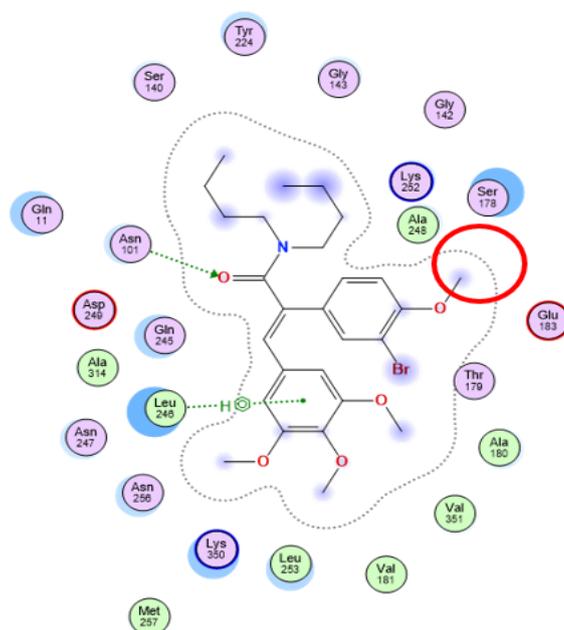


Figure 3.11: β III and MJ-CA4-II-009 interactions.

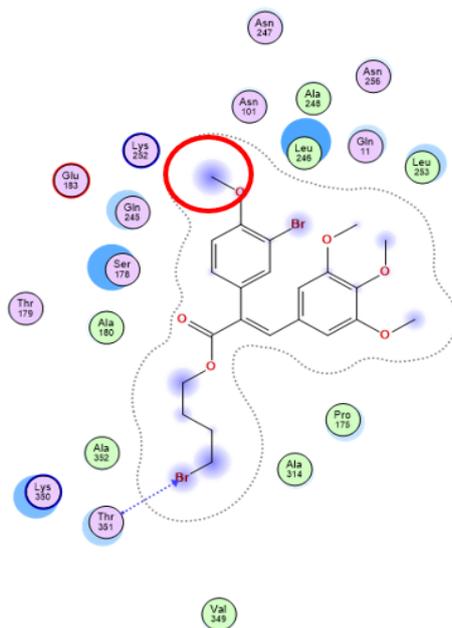


Figure 3.12: β IVb and MJ-CA4-I-005 interactions.

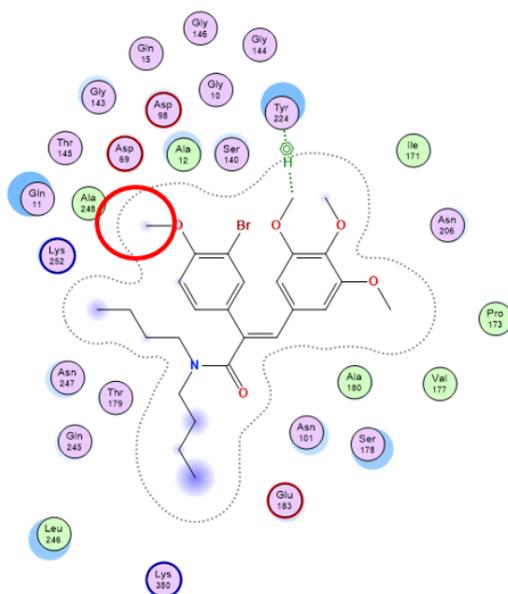


Figure 3.13: β IVb and MJ-CA4-II-009 interactions.

3.4.2 Molecular dynamics simulations

It is relevant to assess the binding stability between the protein of interest and the ligands chosen as possible PROTAC warheads, to accomplish this, molecular dynamics simulations were conducted. The system was simulated for 100ns at a temperature of 298K with an octahedral-shaped box.

The average distance between atoms of superimposed molecules, the backbone atoms, is measured by the root-mean-square-deviation (RMSD) of atomic locations. Subsequently, the RMSD value of the trajectories per thousand frames was plotted, specifically in red for the protein and in blue for the ligand. In the graph, the scale by which the value of RMSD is reported is in Angstroms.

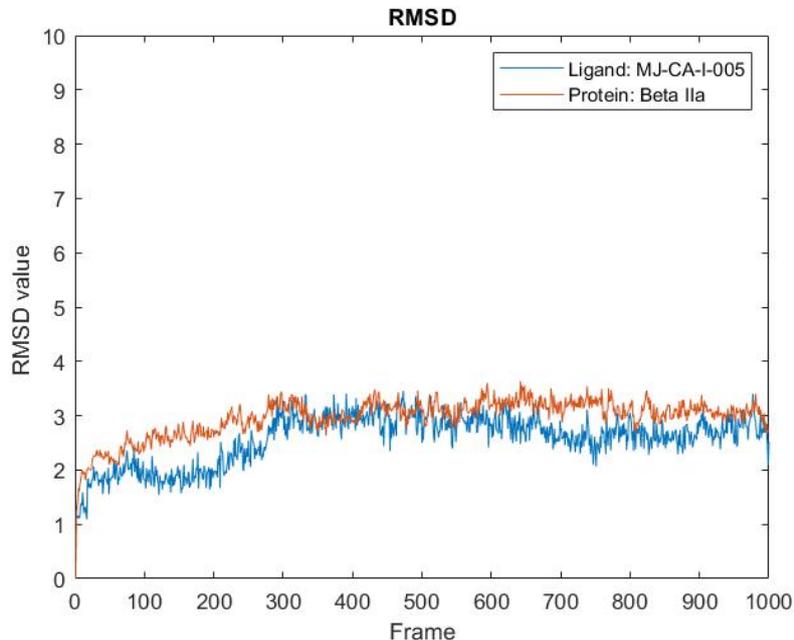


Figure 3.14: Molecular dynamics simulation results: β IIa and MJ-CA4-I-005.

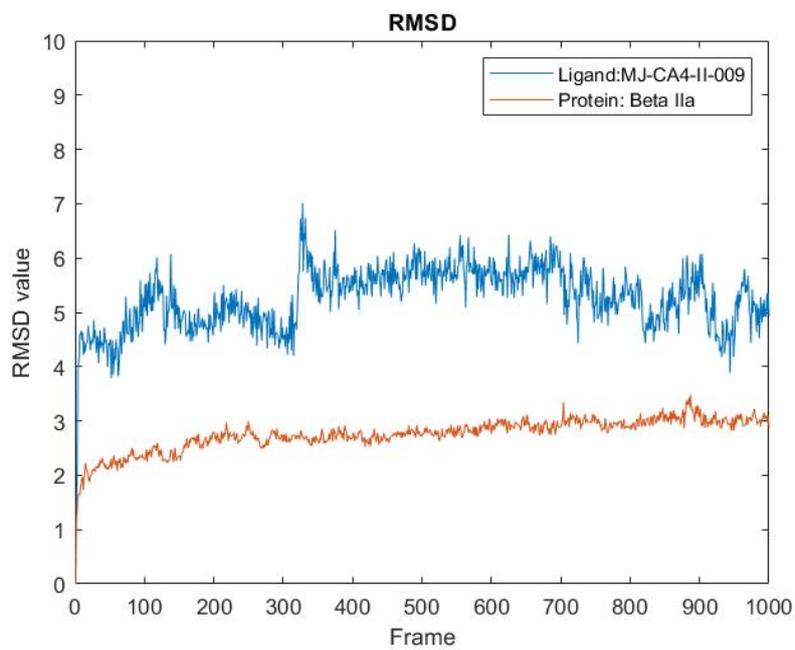


Figure 3.15: Molecular dynamics simulation results: β IIa and MJ-CA4-II-009.

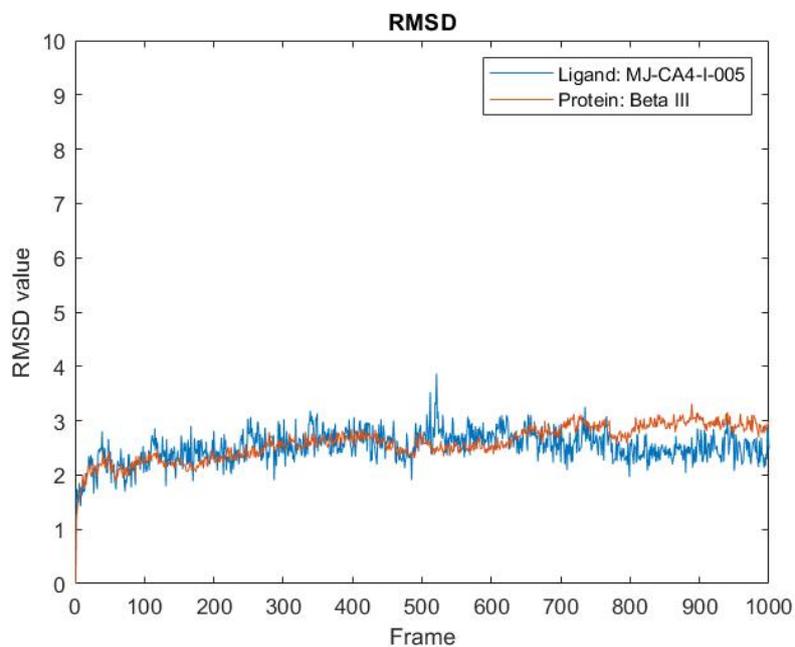


Figure 3.16: Molecular dynamics simulation results: β III and MJ-CA4-I-005.

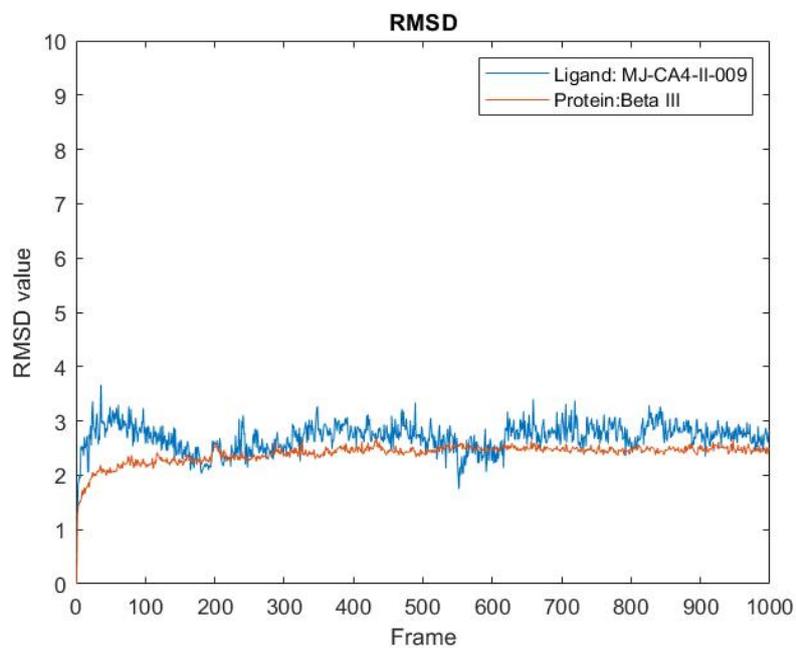


Figure 3.17: Molecular dynamics simulation results: β III and MJ-CA4-II-009.

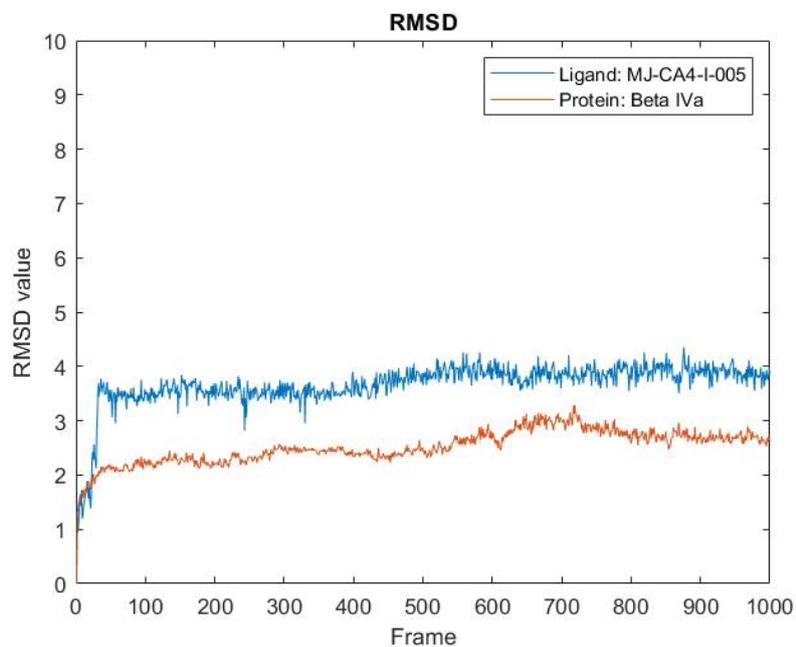


Figure 3.18: Molecular dynamics simulation results: β IVb and MJ-CA4-I-005.

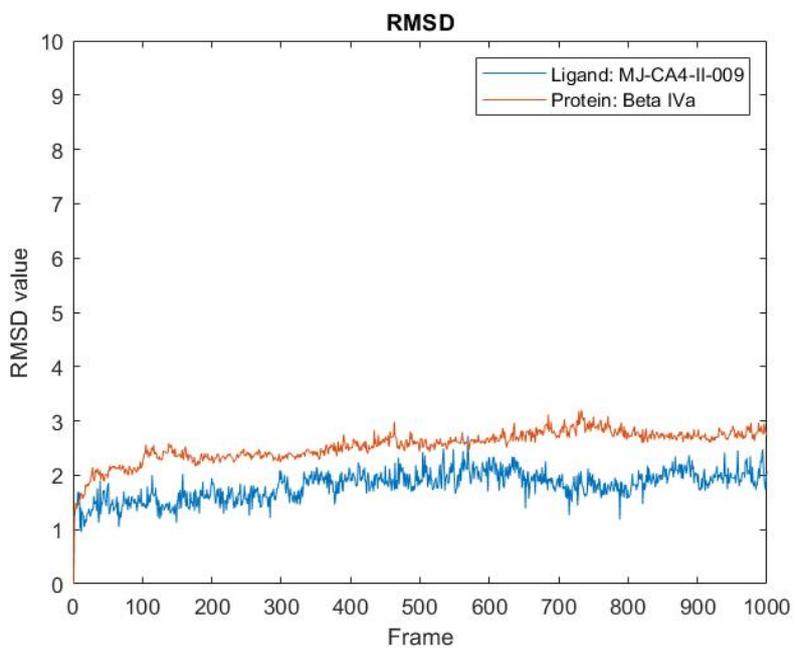


Figure 3.19: Molecular dynamics simulation results: β IVb and MJ-CA4-II-009.

3.4.3 MMPB/GBSA

To assess the binding strength between receptors and ligands, the binding energy was calculated using MMGBSA, which by taking into account multiple contributions, especially the contributions of all residues and polar and non-polar interactions, provides more accurate results.

In Figure 3.20, the results are shown of the binding free energy, or ΔG_{bind}

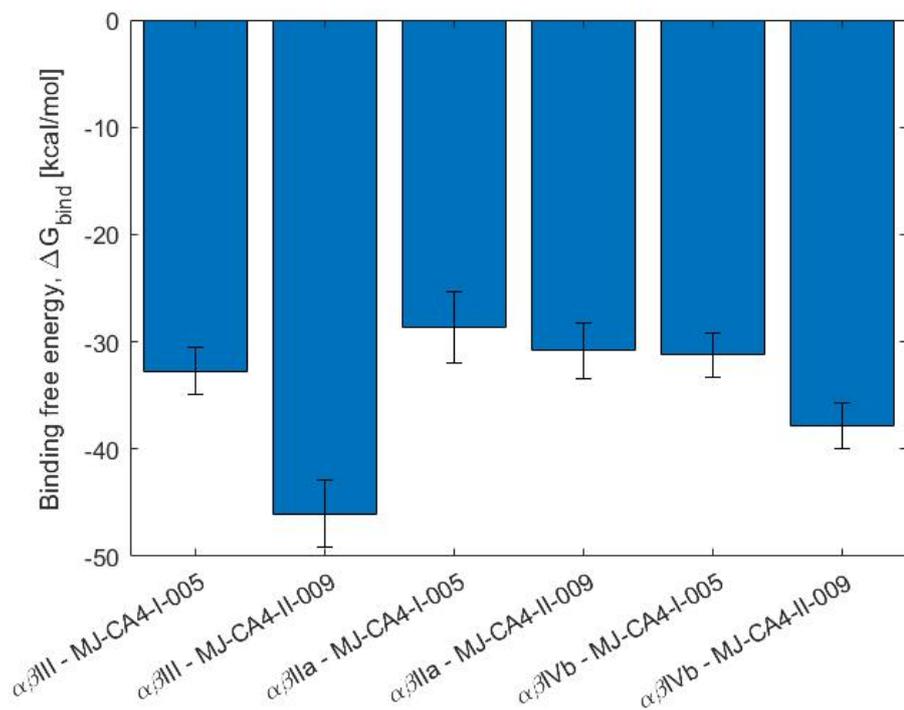


Figure 3.20: Binding free energy, ΔG_{bind} , calculated via the MMGBSA.

From the results shown in the figure, it appears that one of the two ligands, MJ-CA4-II-009, has some affinity for βIII , this could be a good starting point to create a PROTAC structure, with a strong affinity for βIII , which also appears to be overexpressed in some cancer cell types.

Chapter 4

Conclusions and further developments

According to the research done on the novel Combretastatin A4 derivatives, which were created in an effort to discover new therapeutics based on molecules with a similar mechanism of action to Colchicine but with less side effects, it can be summed up that the presented compounds have some toxicity, but less than Colchicine.

Comparing the results between the new molecules and the Colchicine, they present a higher IC₅₀ value than colchicine, being less toxic, with an IC₅₀ in the micromolar range; moreover, many derivatives present a higher cytotoxic action on cells than Combretastatin A4 itself, which means that the chemical modifications made to the initial structure enhance the biological activity.

With the intention of looking for a correlation between *in vitro* results and parameters calculated by computational models, the results that most justify the *in vitro* behavior are the binding affinities calculated by the docking technique, while among the other features that are the ADMET properties of the molecules, none seem to have a linear correlation with the experimental results.

The only parameter that improves the linear regression model between binding energy weighted on the expression of β tubulin isotypes and LogIC₅₀ is the topological polar surface area (TPSA).

Considering this variable, the coefficient of determination (R^2) reaches values greater than 0.7 in certain cell lines.

One of the most interesting facts to point out is that taking into account the percentage of expression of the different isotypes affects the model, the greater the precision with which the percentages were calculated the better the determination coefficient and consequently the model.

While the model made by considering a uniform distribution of the generally more expressed isotypes leads to deterioration of the model and lower R^2 values.

It is crucial to keep in mind that these drugs do not appear to exhibit any selectivity on β tubulin isotypes based on computational studies and binding affinity comparisons when considering them as potential PROTAC warhead. All the β tubulins must thus be viewed as targets in the scenario where they would be used to build a PROTAC structure.

The two compounds, MJ-CA4-I-005 and MJ-CA4-II-009, were chosen because they produced the best results *in vitro*, even though the PROTAC design has a tendency to favor molecules that already inhibit the protein. However, with this strategy, no inhibitory activity would be necessary, but just stable binding between the protein's ligand and the target protein.

After picking the two derivatives, it was possible to determine a potential attachment point for the linker from the analysis of the consensus docking results. It was also possible to determine from the analysis of the rmsd and visualization of the trajectories that the ligands bind on beta tubulin at the same location as the Colchicines, that is, at the interface zone between α and β tubulin, and that they are stably bound over time.

Closer analysis performed after molecular dynamics simulations shows that one of the ligands, MJ-CA4-II-009, appears to exhibit some selectivity for β III; this could be a selective targeting tool, since many types of tumors show overexpression of β III compared with the corresponding healthy tissue.

A possible future development of this work could involve the evaluation of ligands with linkers, and thereafter the complete ternary structure of PROTAC.

Since the choice of the linker represents a critical moment in the design process of PROTAC because although it does not have to interact with the external structures, i.e., POI and E3 ubiquitin ligase, it is essential for the correct spatial positioning of the complete molecule and for the interaction between the two proteins.

Bibliography

- [1] Si-sheng Ou-Yang, Jun-yan Lu, Xiang-qian Kong, Zhong-jie Liang, Cheng Luo, and Hualiang Jiang. «Computational drug discovery». In: *Acta Pharmacologica Sinica* 33.9 (2012), pp. 1131–1140 (cit. on p. 1).
- [2] Jai Woo Lee, Miguel A Maria-Solano, Thi Ngoc Lan Vu, Sanghee Yoon, and Sun Choi. «Big data and artificial intelligence (AI) methodologies for computer-aided drug design (CADD)». In: *Biochemical Society Transactions* 50.1 (2022), pp. 241–252 (cit. on p. 2).
- [3] Igor José dos Santos Nascimento, Thiago Mendonça de Aquino, and Edeildo Ferreira da Silva-Júnior. «The New Era of Drug Discovery: The Power of Computer-aided Drug Design (CADD)». In: *Letters in Drug Design & Discovery* 19.11 (2022), pp. 951–955 (cit. on p. 2).
- [4] Deepika Purohit, Manish Makhija, Parijat Pandey, Sunil Kumar, Sahil Kumar, Rohit Dutt, Deepak Kaushik, Pawan Kumar, and Sanjiv Kumar. «Role of computer-aided drug design in the discovery and development of new medicinal agents a review». In: () (cit. on p. 2).
- [5] Victor T Sabe, Thandokuhle Ntombela, Lindiwe A Jhamba, Glenn EM Maguire, Thavendran Govender, Tricia Naicker, and Hendrik G Kruger. «Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review». In: *European Journal of Medicinal Chemistry* 224 (2021), p. 113705 (cit. on p. 3).
- [6] Kenneth H Downing and Eva Nogales. «Tubulin structure: insights into microtubule properties and functions». In: *Current opinion in structural biology* 8.6 (1998), pp. 785–791 (cit. on p. 3).
- [7] Kenneth H Downing and Eva Nogales. «Tubulin and microtubule structure». In: *Current opinion in cell biology* 10.1 (1998), pp. 16–22 (cit. on p. 3).
- [8] Corinne A Tovey and Paul T Conduit. «Microtubule nucleation by γ -tubulin complexes and beyond». In: *Essays in biochemistry* 62.6 (2018), pp. 765–780 (cit. on p. 4).

- [9] Nikita B Gudimchuk and J Richard McIntosh. «Regulation of microtubule dynamics, mechanics and function through the growing tip». In: *Nature Reviews Molecular Cell Biology* 22.12 (2021), pp. 777–795 (cit. on p. 5).
- [10] URL: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (cit. on p. 5).
- [11] Douglas E Friesen, Khaled H Barakat, Valentyna Semenchenko, Rolando Perez-Pineiro, Bruce W Fenske, Jonathan Mane, David S Wishart, and Jack A Tuszynski. «Discovery of small molecule inhibitors that interact with γ -tubulin». In: *Chemical biology & drug design* 79.5 (2012), pp. 639–652 (cit. on p. 6).
- [12] Rustelle Janse van Vuuren, Michelle H Visagie, Anne E Theron, and Annie M Joubert. «Antimitotic drugs in the treatment of cancer». In: *Cancer chemotherapy and pharmacology* 76.6 (2015), pp. 1101–1112 (cit. on p. 7).
- [13] Yan Lu, Jianjun Chen, Min Xiao, Wei Li, and Duane D Miller. «An overview of tubulin inhibitors that interact with the colchicine binding site». In: *Pharmaceutical research* 29.11 (2012), pp. 2943–2971 (cit. on p. 7).
- [14] URL: <https://en.wikipedia.org/wiki/Colchicine> (cit. on p. 8).
- [15] Min Su et al. «The anti-angiogenic effect and novel mechanisms of action of Combretastatin A-4». In: *Scientific reports* 6.1 (2016), pp. 1–11 (cit. on p. 8).
- [16] URL: https://en.wikipedia.org/wiki/Combretastatin_A-4 (cit. on p. 8).
- [17] Xiuyun Sun and Yu Rao. «PROTACs as potential therapeutic agents for cancer drug resistance». In: *Biochemistry* 59.3 (2019), pp. 240–249 (cit. on p. 9).
- [18] Taavi K Neklesa, James D Winkler, and Craig M Crews. «Targeted protein degradation by PROTACs». In: *Pharmacology & therapeutics* 174 (2017), pp. 138–144 (cit. on p. 10).
- [19] Hongying Gao, Xiuyun Sun, and Yu Rao. «PROTAC technology: opportunities and challenges». In: *ACS medicinal chemistry letters* 11.3 (2020), pp. 237–240 (cit. on p. 10).
- [20] Yutian Zou, Danhui Ma, and Yinyin Wang. «The PROTAC technology in drug development». In: *Cell biochemistry and function* 37.1 (2019), pp. 21–30 (cit. on p. 11).
- [21] Kevin Moreau, Muireann Coen, Andrew X Zhang, Fiona Pachl, M Paola Castaldi, Goran Dahl, Helen Boyd, Clay Scott, and Pete Newham. «Proteolysis-targeting chimeras in drug development: a safety perspective». In: *British Journal of Pharmacology* 177.8 (2020), pp. 1709–1718 (cit. on p. 12).

- [22] Luis J Leandro-Garcia, Susanna Leskelä, Iñigo Landa, Cristina Montero-Conde, Elena López-Jiménez, Rocío Letón, Alberto Cascón, Mercedes Robledo, and Cristina Rodríguez-Antona. «Tumoral and tissue-specific expression of the major human β -tubulin isotypes». In: *Cytoskeleton* 67.4 (2010), pp. 214–223 (cit. on pp. 17, 18, 35).
- [23] Haizhen A Zhong. «ADMET properties: overview and current topics». In: *Drug design: principles and applications* (2017), pp. 113–133 (cit. on pp. 19, 20).
- [24] S Prasanna and RJ Doerksen. «Topological polar surface area: a useful descriptor in 2D-QSAR». In: *Current medicinal chemistry* 16.1 (2009), pp. 21–41 (cit. on p. 20).
- [25] Douglas EV Pires, Tom L Blundell, and David B Ascher. «pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures». In: *Journal of medicinal chemistry* 58.9 (2015), pp. 4066–4072 (cit. on pp. 20, 22, 23).
- [26] Caroline Manto Chagas, Sara Moss, and Laleh Alisaraie. «Drug metabolites and their effects on the development of adverse reactions: Revisiting Lipinski’s Rule of Five». In: *International journal of pharmaceutics* 549.1-2 (2018), pp. 133–149 (cit. on p. 21).
- [27] Yu-Chian Chen. «Beware of docking!» In: *Trends in pharmacological sciences* 36.2 (2015), pp. 78–95 (cit. on pp. 27, 28).
- [28] Nataraj S Pagadala, Khajamohiddin Syed, and Jack Tuszynski. «Software for molecular docking: a review». In: *Biophysical reviews* 9.2 (2017), pp. 91–102 (cit. on p. 28).
- [29] Jiyu Fan, Ailing Fu, and Le Zhang. «Progress in molecular docking». In: *Quantitative Biology* 7.2 (2019), pp. 83–89 (cit. on pp. 28, 29).
- [30] URL: <https://www.uniprot.org/help/uniprotkb> (cit. on p. 29).
- [31] URL: <https://www.rcsb.org/> (cit. on p. 29).
- [32] URL: <https://datascience.eu/it/matematica-e-statistica/definizione-r-squared/> (cit. on p. 31).
- [33] Romelia Salomon-Ferrer, David A Case, and Ross C Walker. «An overview of the Amber biomolecular simulation package». In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.2 (2013), pp. 198–210 (cit. on p. 31).
- [34] Martin Karplus and Gregory A Petsko. «Molecular dynamics simulations in biology». In: *Nature* 347.6294 (1990), pp. 631–639 (cit. on p. 32).

- [35] Samuel Genheden and Ulf Ryde. «The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities». In: *Expert opinion on drug discovery* 10.5 (2015), pp. 449–461 (cit. on p. 34).
- [36] Murat Cihan Sorkun, Abhishek Khetan, and Süleyman Er. «AqSolDB, a curated reference set of aqueous solubility and 2D descriptors for a diverse set of compounds». In: *Scientific data* 6.1 (2019), pp. 1–8 (cit. on p. 39).
- [37] Lorelei Johnson et al. «Novel colchicine derivatives and their anti-cancer activity». In: *Current topics in medicinal chemistry* 17.22 (2017), pp. 2538–2558 (cit. on p. 81).