Federal University of Uberlandia
Institute of Biotechnology
Biotechnology Course

Virtual Screening, Selection and Molecular Dynamics of Protein Inhibitors N-myristoyltransferase (NMT) in *Plasmodium vivax*

Pedro Serafim Marques

Monograph presented to the Biotechnology Course Coordination, at the Federal University of Uberlandia, in order to obtain the Bachelor's degree in Biotechnology.

Uberlandia – MG
December - 2018

Federal University of Uberlandia
Institute of Biotechnology
Biotechnology Course

Virtual Screening, Selection and Molecular Dynamics of Protein Inhibitors N-myristoyltransferase (NMT) in *Plasmodium vivax*

Pedro Serafim Marques

Nilson Nicolau Junior

Monograph presented to the Biotechnology Course Coordination, at the Federal University of Uberlandia, in order to obtain the Bachelor's degree in Biotechnology.

Uberlandia – MG
December - 2018

Federal University of Uberlandia
Institute of Biotechnology
Biotechnology Course

Virtual Screening, Selection and Molecular Dynamics of Protein Inhibitors N-myristoyltransferase (NMT) in *Plasmodium vivax*

Pedro Serafim Marques

Nilson Nicolau Junior
Institute of Biotechnology

Approved by the Coordination of the Biotechnology Course in __/__/__

Edgar Silveira Campos

Uberlandia – MG
December – 2018

Federal University of Uberlandia
Institute of Biotechnology
Biotechnology Course

Virtual Screening, Selection and Molecular Dynamics of Protein Inhibitors N-myristoyltransferase (NMT) in *Plasmodium vivax*

Pedro Serafim Marques

Approved by the Examination Board in: / / Grade: ____

Name and signature from the Examination Board president

Uberlandia, de de

Acknowledgement

I would like to thank God first, for placing each of these people in my life, starting with my family, which is the greatest good I have, especially my mother who has always supported me all the way, invested and trusted me, and always took pride in the person I became. I thank my brother, who always gave good advice and cheered for me from the beginning. Thank you very much to my friends, who have always been there to welcome me and always one helping the other to endure the difficulties, and thank all the Fridays that provided me with our trick and so on. Finally, I am grateful for everything I have experienced and thank the university for showing me all this!

Abstract

Malaria is one of several neglected diseases characterized by parasitic infection of global importance present in tropical countries. Malaria is caused by protozoa, belonging the genus *Plasmodium* and the vector for the transmission of the disease are female of the *Anopheles* mosquitoes infected with the *Plasmodium ssp.* According the World Health Organization (WHO), there was 219 million malaria cases worldwide in 2017, and approximately 435 000 malaria deaths in the same year. The increasing resistance to treatment has posed a major problem since the beginning of the 21st century, and with this, new research has been done with the intention of finding a new drug to be used for the eradication of the disease. A new protein has been studied as a potential target drug in malaria, N-myristoyltransferase (NMT), and the NMT acts on myristoylation of proteins, facilitating the binding of these to the plasma membrane and also contributes to the stabilization of protein-protein interactions. By inhibiting NMT it has already been established that the cell of the parasite can be killed. Therefore, in order to predict and detect potential inhibitors against *Plasmodium* NMT, the bioinformatic tools were used in this work, encompassing the virtual screening, docking, validation and molecular dynamics. With the analysis performed during the work it was possible to obtain compounds with great similarity to the original ligand, thus being good potential inhibitors of *Plasmodium* NMT.

Key words: Malaria, NMT, *Plasmodium*

Summary

1. Introduction

1.1. Malaria

Malaria is one of several neglected diseases characterized by parasitic infection of global importance present in tropical countries. It is also considered endemic mainly in low-income populations (Suh, Kain e Keystone, 2004). It is caused by protozoa belonging to the genus *Plasmodium*, which has five species capable of causing disease in humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale* and *Plasmodium knowlesi* (Bannister e Sherman, 2009). The cases of disease that led to death are related to *P. falciparum*, the other species of the genus *Plasmodium* cause a milder form of the disease (ASHLEY et al., 2006). There are reports of *P. ovale* and *P. knowlesi* in some regions of the African continent and in Southeast Asia. In Brazil there are reports only of the other species, being the *P. vivax* the most prevalent species (Lee e Vythilingam, 2013) (WHO, 2018).

The vectors responsible for the transmission of the disease are females of *Anopheles* mosquitoes infected with *Plasmodium* spp. The bite of this mosquito introduces the parasites into the skin then reach bloodstream of the individual which they migrate to the liver, become mature and reproduce (Beier, 1998). In the asexual cycle of the parasite, the mosquito vector transmits the parasite through the bit in the sporozoite stage. The sporozoite invade liver cells, where replication happens dividing into merozoites, thereafter, the liver cell that was infected suffers rupture, releasing the merozoites into the bloodstream, occurring invasion of the red blood cells, thus configuring the beginning of the asexual blood stage, which is the symptomatic phase of disease. Individuals perceive the symptoms of the disease in a period of 8 to 25 days after the vector inoculation. However, the signs of malaria are nonspecific, with fever being the main sign, making it difficult to distinguish easily from another diseases that also cause fever (Burchard, 2014). Fever occurs when the infected red blood cells suffer lyses and release a

massive endotoxin. (PHILLIPS et al., 2017) (Figure 1). Another very common symptom is paroxysm (cyclic occurrence of sudden coldness followed by tremors and subsequently fever and sweating) in *P. vivax* infections which is the main parasite of this work (TOUZE, JE. *et al.*, 1998).

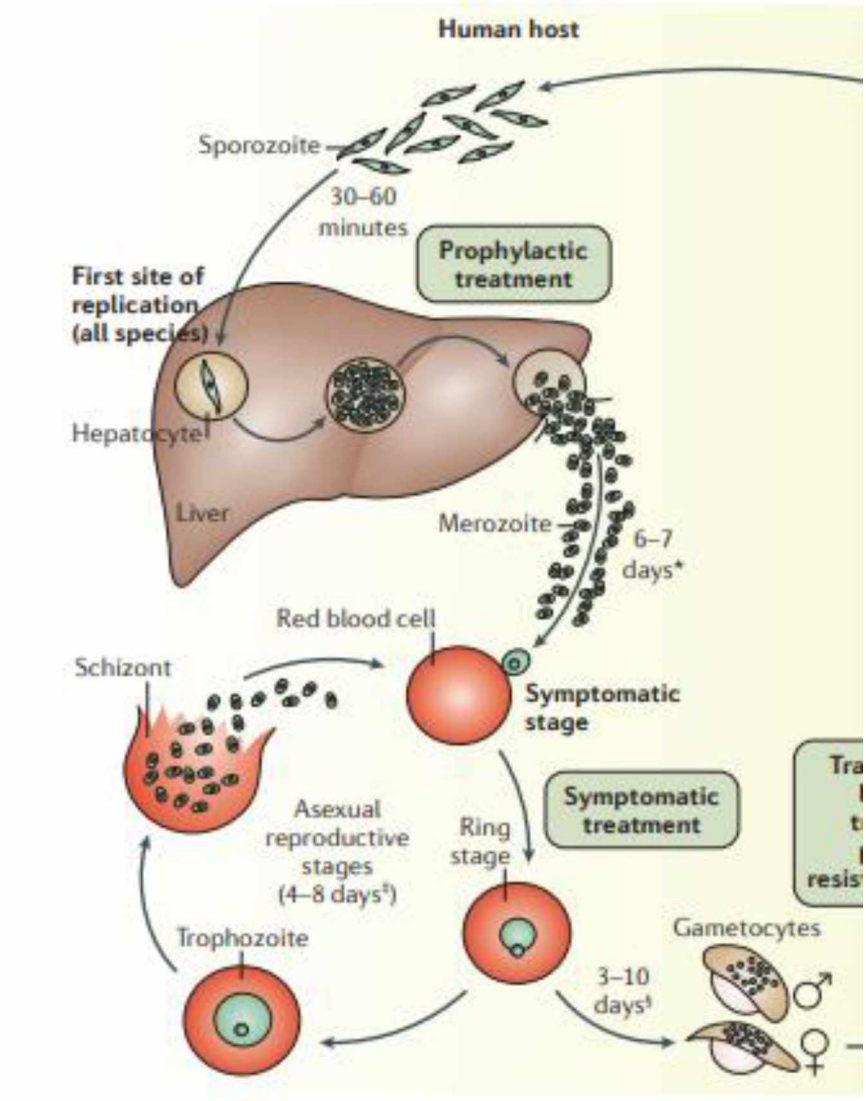

Figure 1. Asexual cycle of the *Plasmodium ssp.* Adapted by: "Malaria" (Phillips *et al.*, 2017).

Usually, the disease is diagnosed through light microscopy of thick and thin stained blood smears and rapid antigen-based tests. The treatment of choice will depend on the severity of the disease. This treatment involves the use of artemisinin-based therapies combined with conventional antimalarials, such as chloroquine, sulfadoxine-pyrimethamine and amodiaquine

(White *et al.*, 2014). The increasing resistance to the current treatment has posed a serious problem of public health in endemic, but new potential drugs have emerged to delay the spread of the disease (Greenwood *et al.*, 2014).

Recently the protein N-myristoyltransferase (NMT) has been studied as a potential target drug in malaria (Brannigan *et al.*, 2014; Price *et al.*, 2003). The NMT catalyzes the transfer of myristoyl-CoA myristate fatty acid to glycine from the amino terminal portion of a set of proteins (Sonnenburg e Gordon, 2013). Myristoylation of proteins facilitates the binding of these to the plasma membrane and also contributes to the stabilization of protein-protein interactions. There is high sequence homology between human (HsNMT) and parasite protozoa NMT (PvNMT), and myristoyl-CoA binding sites are highly conserved among species. There is, however, a divergence between the binding sites of the peptide substrates between the human species and parasitic protozoa that can be exploited to search for specific inhibitors for pathogens (Resh, 2012). The NMT was successfully used as a drug target in *Trypanosoma brucei* by (Frearson *et al.*, 2010), which found inhibitors that bind with high affinity to the peptide substrate site, inhibiting N-myristoylation in trypanosome. NMT is thus an appropriate target for the search for therapies against parasitic protozoa including those responsible for malaria.

Currently, many research groups have focused their studies on the search for new therapeutic agents, with an emphasis on cheaper compounds that have lower toxicity when compared to frequently used drugs. One of the strategies in this sense is the search for potential drugs in libraries of compounds already available, provided by the companies producing these new molecules. This search can be performed by means of virtual sorting of compounds with the aid of computational methods such as modeling of pharmacophores and molecular docking.

1.2. Bioinformatics

1.2.1. Pharmacophores

Pharmacophores are, by their most current definition, a set of steric and electronic characteristics that are necessary to ensure optimal supramolecular interactions with a specific biological target, activating or blocking their biological response (Wermuth *et al.*, 1998). The central pharmacophore concept involves the notion that the molecular recognition of a biological target by a group of compounds can be attributed to a small set of common characteristics that interact with complementary regions in the biological target, these being quite general characteristics: hydrogen bonds, hydrogen bond acceptors, positively and negatively charged groups, and hydrophobic regions (Leach *et al.*, 2010).

There are basically two methods of modeling pharmacophores, the structure-based and the ligand-based. Structure-based uses information on protein-binding or protein-only interactions, obtained from an experimentally derived 3D structure, to generate a model. The ligand-based method extracts information from the model only from the ligand, obtaining common chemical characteristics in active compound assemblies or only the use of a single compound that serves as a template for the elucidation of the pharmacophore (McGann, 2011).

Pharmacophore modeling is a well-established *in silico* technique that generates many benefits for initial research with drug designs. Its advantages are the visualization of the chemical characteristics supposedly important in the protein-ligand interaction as well as the obtaining of structurally diverse bioisosteric compounds that would not have been discovered by systematic derivatization of existing compounds (Markt, Schuster e Langer, 2011). The growing range of applications of pharmacophores, together with cases of success in drug discovery, has allowed the enrichment of the concept of pharmacophore by promoting the development and application of new approaches (Yang, 2010). Other strategy similar to pharmacophore modeling is the shape-based models. Shape-based modeling aims to extract not

only the common chemical characteristics but also the shape of the ligands used as query, thus, producing very refined models.

1.2.2. Molecular Docking

Molecular docking involves the search for modes of interaction between two molecules, such as protein-protein, protein-ligand, or protein-DNA interactions. The macromolecular protein-ligand docking system consists of two steps: a conformational search algorithm that involves all degrees of freedom of the ligand (translational, rotational, and conformational) and a scoring function that ranks the probable spatial positions for a compound or several from a large library (Rognan, 2011).

1.2.3. Molecular Dynamics

Molecular dynamics represents the computational approach to statistical mechanics and is used to estimate equilibrium and dynamic properties of complex systems that cannot be calculated analytically (Shlick, 2010). Understanding the movements of proteins has been an important field of research, since the dynamics are inexorably linked to their function (Samish, Gu e Klein, 2009).

1.2.4. Justification

Protein-ligand docking was one of the pioneer methodologies in the 1980s in bioinformatics analyzes (Kuntz *et al.*, 1982) and when only the structure of a target and its binding site are available, high performance docking is the first choice as a tool for identifying drug candidates. Although docking and scoring functions are based on approximations, the application of these techniques during the optimization of lead compounds together with other computational methods adds up to traditional approaches in the design of drugs based on

structure (Kitchen *et al.*, 2004). Currently, the dominant technique for identifying new leading compounds in drug discovery is the physical prospection of large substance libraries against a biological target. An alternative approach, known as virtual screening, is to search computationally large libraries of compounds that target a known structure, and experimentally test those that have good interactions (Shoichet, 2004).

Since malaria is an endemic and tropical disease that affects mainly low (African continent and part of the Asian continent) and middle income countries, such as Brazil, and several cases of death due to this disease worldwide have been reported, this research was conducted in order to contribute in the proposal of a practical and economical approach in the search of effective drugs against Pv-NMT protein. To do this, we used databases of proteins and ligands available to the scientific community, as well as computational strategies and tools to search for compounds with pharmacological potential against malaria.

2. Objectives

2.1. General Objectives

Search for potential ligands for the three-dimensional structure of the *Plasmodium vivax* N-myristoyltransferase enzyme through the construction of a shape-based model, virtual screening, molecular docking and molecular dynamics.

2.2. Specific Objectives

o Construct a shape-based model based on the structure of the ligands extracted from a complex with the N-myristoyltransferase protein;

o Perform a shape-based search through virtual screening in a conformer library of small ligands.

o Analyze the receptor-ligand interaction between the *P. vivax* N-myristoyltransferase and the selected compounds in the virtual screening, through the molecular docking technique.

o Ranking molecular docking results of the protein-ligand complexes using specific scoring functions.

o Evaluate the potential toxic, *in silico*, properties of the selected compounds.

o Perform the molecular dynamics in the protein-binding complex of the selected compounds, in order to measure the stability of the interaction.

3. Material and Methods

For a better understanding, according to Figure 2, the ligands in complex with the PvNMT of the Protein data bank site were initially found, thus allowing the alignment of these molecules together with the generation of the active and decoys molecules, which are fundamental for the generation of the validated pharmacophoric model. With the generated model and the prepared compound libraries, it was possible to perform the virtual screening, molecular anchoring, pkCSM analysis and the Molecular Dynamics. Each topic will be covered in specific throughout the text.

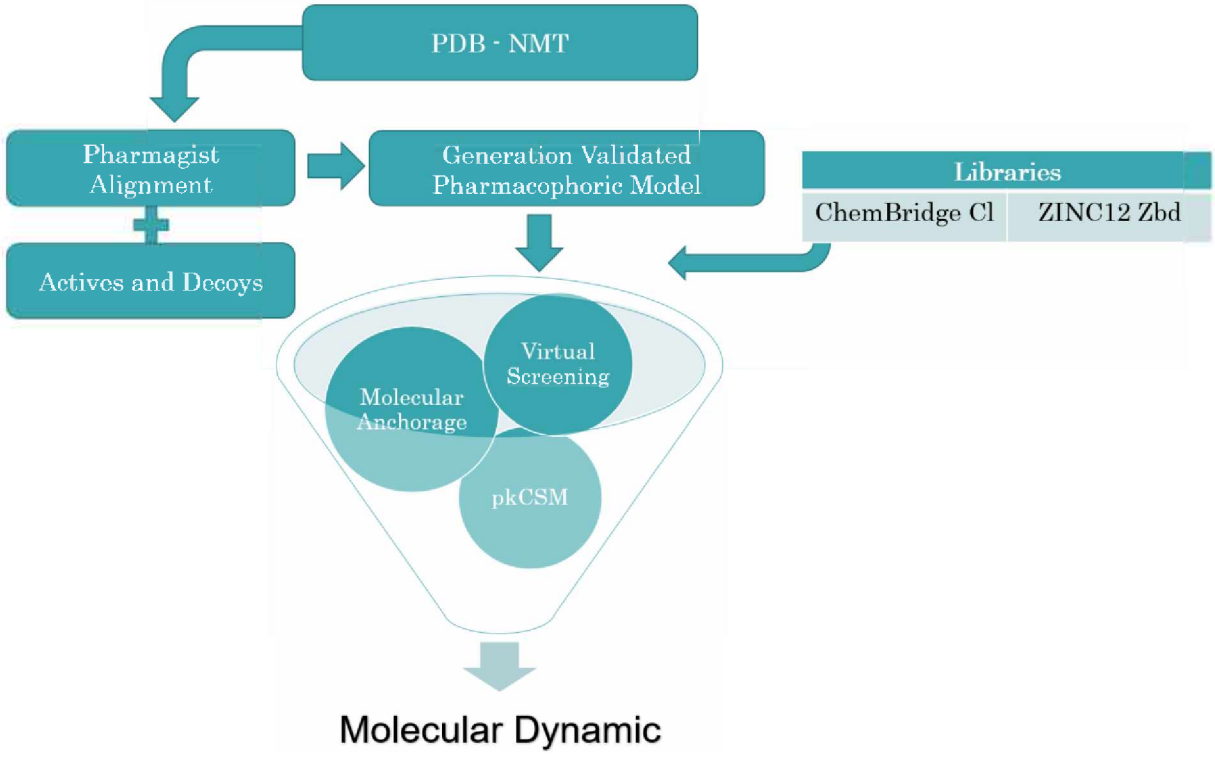

Figure 2. General outline of the methodology steps

3.1. Generation of shape-based model based on the ligand form

In order to perform a virtual screening against the ligand libraries, a shape-based model was generate based in ligands, extracted from NMT structural complexes, present on the Protein Data Bank (PDB) - PDB IDs: 4B12, 4B13, 4B14, 4BBH, 4CAE, 4CAF. This step was developed with the help of PharmaGist webserver (Schneidman-Duhovny *et al.*, 2008) and ROCS 3.2.0.4 program, OpenEye Scientific Software, Santa Fe, NM (Hawkins, Skillman e Nicholls, 2007). PharmaGist was used to align the known inhibitors structures and ROCs was used to construct shape-based models based on the alignment. Also, ROCs is shape comparison method, based on the idea that molecules have similar shape if their volumes overlap properly. Although ROCS is primarily a shape-based method, the user can define for specific regions of the model, chemical characteristics described as: hydrogen donors, hydrogen acceptors, anions, cations, hydrophobic and regions of aromatic rings. In this way, the models that are formed

during the overlapping process facilitate the identification of the compounds which are similar in shape and in chemistry.

3.2. Conformers libraries generation

The shape-based model generated through the ROCS program was used in the virtual screening of compounds libraries originating from two databases. The first database belongs to Chembridge Corporation, San Diego, California, more specifically the DIVERSet ™ Cl, and this library has a total of 50.000 compounds of diversity *druglike* structures that are carefully selected to provide the widest coverage of the pharmacophoric space. The second library belongs to ZINC12 database (Irwin e Shoichet, 2005) more specifically the Zbd subset of natural products. These libraries were submitted to the FILTER program of OMEGA 2.5.1.4, OpenEye Scientific Software, Santa Fe, NM (Hawkins *et al.*, 2010). FILTER is a rapid filtering and compound selection program that uses a combination of physical property calculations and functional group knowledge. It was used prior to experimental analysis to remove undesirable compounds such as compounds with toxic functions, with low oral bioavailability, with high probability of forming covalent bonds with the protein target and compounds with probability of interference in the experimental trial.

The compounds of the library filtered through FILTER were submitted to the program QUACPAC 1.6.3.1, OpenEye Scientific Software, Santa Fe, NM (http://www.eyesopen.com) in order to generate, for each molecule, precise charges at pH 7.4 and the probable tautomers. Finally, the compound libraries were submitted to the generation of conformers through the OMEGA program. OMEGA generates bioactive composite multi-conformer databases with high speed and reliability. The generation of conformers is an essential step for virtual screening using shape-based models, since these rely on bioactive conformations to achieve satisfactory results.

3.3. Virtual screening and Validation

For accurate understanding, interpretation and comparison of virtual screening, ROCS contain methods of statistical validation of the results achieved in the screening. The ROC curve (receiver operating characteristic) and AUC (area under the curve) was used in this validation. The ROC curve represents the comparison of active compounds versus decoys (non-target protein binding ligands and potentially inactive) and AUC, extracted from the ROC curve graph, is simply the probability that a randomly selected active compound scores higher than an inactive compound chosen at random. For the construction of the ROCS curve and obtaining the AUC value, biologically active ligands found in the literature against the N-myristoyltransferase protein was used (Bell *et al.*, 2012) . The decoys were generated on the DUD-E online platform (Mysinger *et al.*, 2012) which uses active ligands to generate a set of decoys. In this way, it was possible to evaluate the quality of the shape-based model generated from the ligands (4B12, 4B13, 4B14, 4BBH, 4CAE and 4CAF) by overlapping and comparing the results between active ligands and decoys.

The shape-based model previously generated was used in the screening of Cl and Zbd libraries. Each screened libraries generated a subset of the best 500 molecules or the best 500 hits, that is, the molecules with high overlap score to the model, called TanimotoCombo score.

3.4. Molecular Docking and Evaluation

The best 500 molecules, obtained from the shape-based model screening, were submitted to a molecular docking. For this purpose, the FRED program, OpenEye Scientific Software, Santa Fe, NM (McGann, 2011) was used. FRED performs a systematic and non-stochastic examination of all possible protein-ligand positions where these was ranked using the Chemgauss4 scoring function. Chemgauss4 measures the complementarity of the positions of

the ligand within the active site. To do so, it recognizes the following ways of interaction: shape; hydrogen bonds between ligand and protein; interactions of hydrogen bonds with implicit solvent and chelator interactions.

3.5. Analysis of toxic properties *in silico*

Following screening and molecular docking, an *in silico* approach was used to evaluate the toxicity properties of the selected compounds. To predict the pharmacokinetic properties of the compounds, a software online called pkCSM (Pires, Blundell e Ascher, 2015) based on graph-based signatures was used. The pkCSM signatures were successfully used in the five major classes of pharmacokinetic properties to develop predictive regression and classification models. This step is important to verify and predict if the potential inhibitors presents the acceptable parameters to drug selection and a possible administration in humans.

The results were judged by positive and negative pharmacokinetic characteristics, and in case of this work, was choose some in specific like AMES toxicity, Max tolerated dose (human), hERG I inhibitor, hERG II inhibitor, Oral Rate Acute Toxicity (LD50), Oral Rat Chronic Toxicity (LOAEL), Hepatotoxicity, Skin Sensitization and the Minnow toxicity.

The AMES toxicity parameter is a widely employed method to assess a potential mutagenic compound using in bacteria, so if the test is positive, that indicates that the compound in question is mutagenic and may act as a carcinogen.

The Maximum Tolerated Dose (human) parameter estimate the toxic dose threshold of chemical in humans (mg/kg/day), being less or equal to 0.477(mg/kg/day) considered as low and high if greater than 0.477(mg/kg/day).

The hERG I/II parameter analyses the long QT syndrome (LQTS), generated by the inhibition of the potassium channels and it can be a complication in the electrical cycle of heart, resulting is risks of sudden death, according if the result is positive or negative, but even if the

result is positive to the inhibition of these channel, depends for what the medicament will be used.

The Oral Rate Acute Toxicity (LD50) parameters refers to the lethal dosage values given all at once (LD50) that is necessary to cause the death of 50% of a group of test animals.

The Hepatotoxicity parameters refers a liver associated side effects observed in humans, being able to cause or not.

The Skin Sensitization parameters is a potential adverse effect for dermally applied products, being able to cause allergic response.

And as a last parameter, the Minnow toxicity refers at the lethal concentration values (LC50) necessary to cause the death of 50% of the "Flathead minnows" (species of fish used as an indicator of toxicity), and values below 0.5 mM are considered as high acute toxicity.

3.6. Molecular Dynamics

The ligands selected from the molecular docking results and the pharmacokinetic properties analysis were submitted to receptor-ligand molecular dynamics (DM) using the GROMACS program (Abraham *et al.*, 2015). The parameters to generate the topology for each ligand were performed using SwissParam (Zoete *et al.*, 2011), using the CHARMM force field. The MD of Protein-ligand complex were performed on GROMACS (Abraham *et al.*, 2015), using CHARMM27 force-field, which is the recommended to use the TIP3P as water model. The unit cell was defined as triclinic shape, water and ions were added and energy minimization were performed and the solutions were subjected to 10 ns of simulation during the production phase of the dynamics under NTP conditions. The result of the dynamics was evaluated through tools provided by the GROMACS program. In this way, the deviation of the quadratic mean (RMSD) of the solution and the number of hydrogen bonds and free energy between the ligand and the binding site along the dynamics was evaluated.

For molecular dynamics, a molecule with the best predictions in pkCSM was selected. The first step in DM was the preparation of the ligand and the topology generation of the protein. The preparation of the ligand and the protein was done by treating the molecules, where undesirable compounds (ions, cations, salts) have been removed and added hydrogens.

Once these basic treatments were done, the dynamics began, which consists of some steps for a simulation of this protein-ligand interaction within a biological system. Then after the topology generation, a water box was created to encompass the molecules to study the interaction. Thus, a solvated system was obtained, and ions were then added to that system in order to balance charges in it. Then, the last step began with a energy minimization procedure.

With the system minimized its balance was realized. Two conditions exist to carry out the equilibrium: restriction of the ligand and treatment of the temperature of the complex. The balance consisted of two types, the NVT balance and the NPT balance. At the end of both equilibrium situations, a fully balanced system was obtained at correct temperature and pressure conditions and the final production of the molecular dynamics was performed.

3.7. Generation of interaction diagrams 2D protein-ligand

In order to analyze the Dynamic Molecular results, 2D diagrams were generated using LigPlot+ (Wallace, Laskowski e Thornton, 1995). It helped to demonstrate the interactions present in the protein-ligand complexes.

4. Results and Discussion

4.1. Generation of shape-based model

The molecular interactions between the ligand in complex with the Pv-NMT protein, observed in Figure 3, were taken from PDB file with Id: 4B13, enabled the construction of a virtual screening model based on the way shown in Figure 4. The 4B13 complex was chosen because all the molecules present in the PDB in complex with the *P. vivax* NMT protein have similar resolutions, allowing the any of these molecules to be used as a basis in the generation of the model, being a completely arbitrary choice. In Figure 3, it is important to note that the model has some important points of interaction, such as hydrogen bridges, salt bridges and pi interactions, indicated by black dashed lines. Also noted are some hydrophobic interactions indicated by the green solid line.

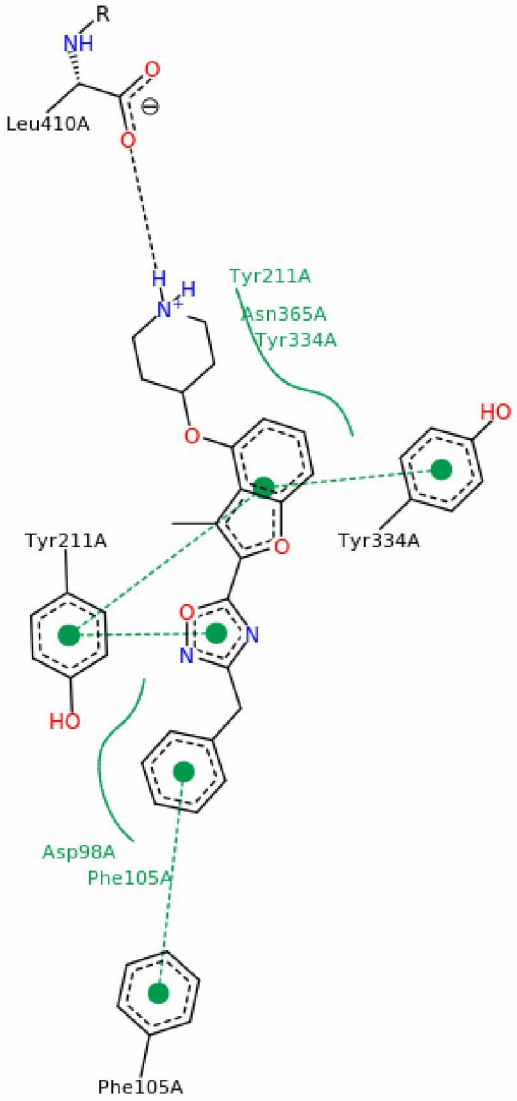

Figure 3. Interaction of 4B13 ligand in complex with the NMT *P.vivax*. The black dashed lines indicate the hydrogen bonds, salt bridges and pi interaction; and the green solid lines represents the hydrophobic interactions. Fonte: https://www.rcsb.org/structure/4B13.

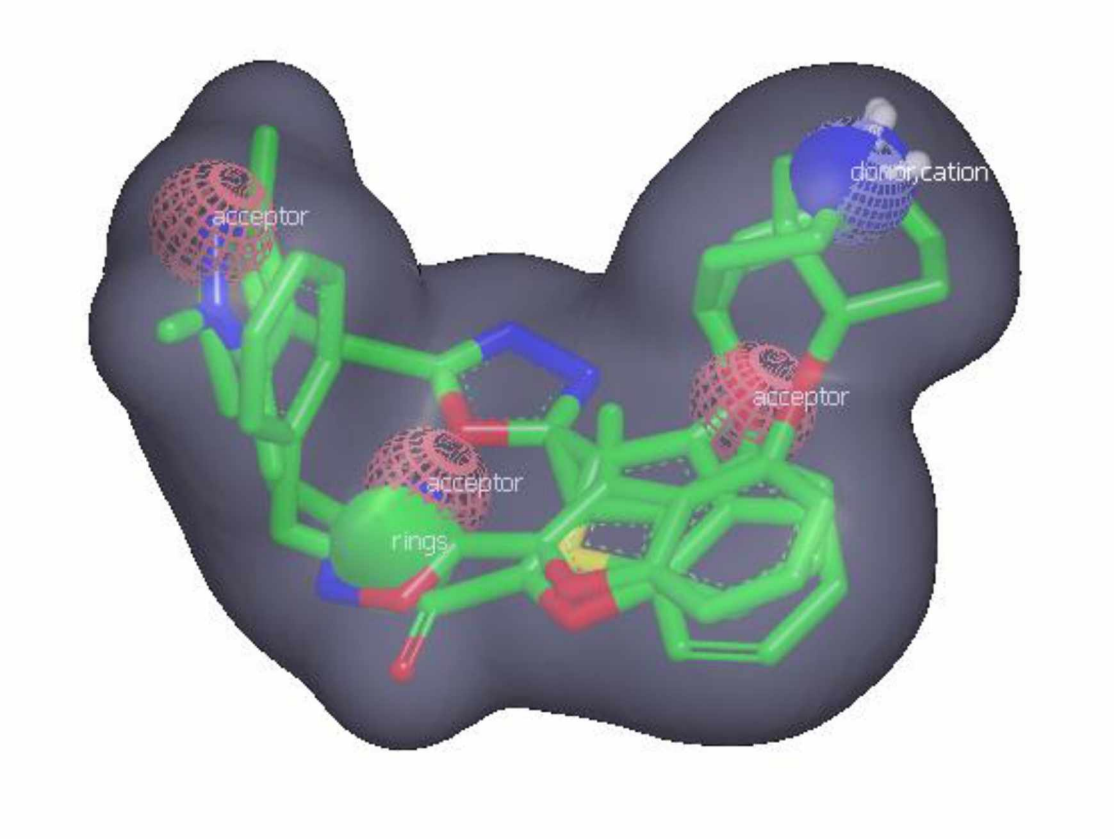

Figure 4. Shape-based model built with the help of vROCS ligand model builder tool. This was the generated model based in the interactions presents in Figure 3, that is in focus the donors of electrons (blue spheres) that represents hydrogen bonds, acceptors (red spheres) and rings (green sphere) that represents the hydrophobic interactions.

4.2. Generation of ligand conformer databases

Both database the DIVERSet™ Cl and the Zbd were properly treated through some steps like the filtering of undesirable compounds, protonation and deprotonation state with adjusts in pH and generation of tautomers and conformers. Thus, it was guaranteed that just compounds with desirable features for drugs was selected.

4.3. Virtual screening in compound database and validation

According to the 2D structure of inhibitors, six PDB files were chosen, which have already been validated in complex with PvNMT: 4B12, 4B13, 4B14, 4BBH, 4CAE and 4CAF. PharmaGist webserver was used to provide the most suitable pharmacophores for virtual prospecting along with their alignment. The chosen result exhibited score 36,000 and 6 ligands. The known inhibitors were based on (Bell *et al.*, 2012), accounting ten structures which are listed in the (Figure 4), and all the structures were design using PubChem Draw Structure tool (Kim *et al.*, 2016).

To predict the decoys through known inhibitors, was used the DUD-E webtool (Mysinger *et al.*, 2012) (Figure 5). Decoys are structures that have similar proprieties with ligand but different chemical structures. To realize the docking, both decoys and actives structures were submitted using Omega (Hawkins *et al.*, 2010) and QUACPAC, which consisted in removing undesirable molecules, generating tautomer, correcting the pKa to physiological pH 7.4. After that, 3D structures were generated using vRocs (Hawkins, Skillman e Nicholls, 2007). Sixteen queries were generated, being possible the manual editing of them, and validation of the model using decoys, actives and libraries. All models are listed in the Table 1.

According the higher AUC (area under the curve), which means a simply probability that a randomly chosen active ligand bind instead of inactive ligand, was selected the model 13 (Table 1). The ROC curve plot is shown in Figure 6.

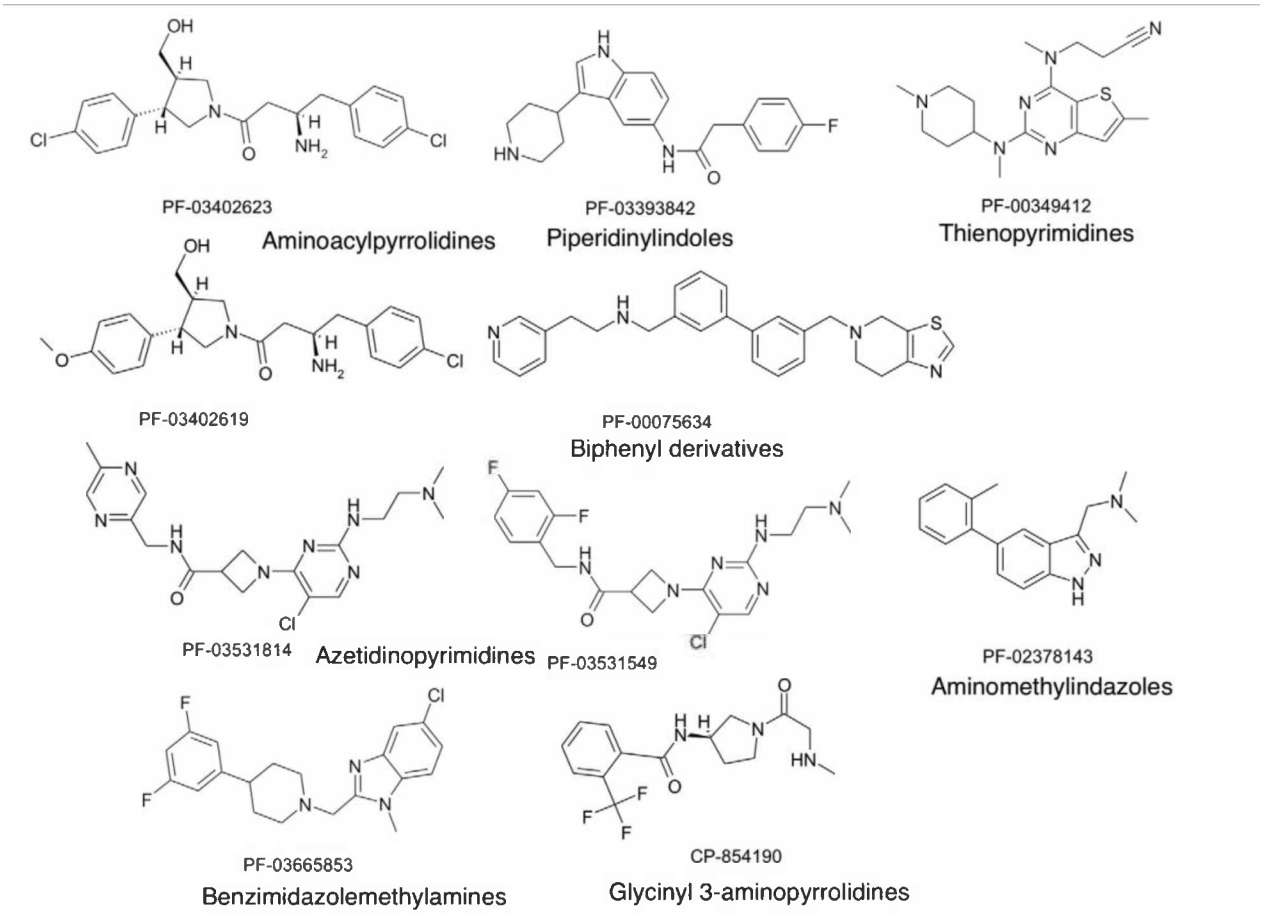

Figure 5. Known inhibitors design using PubChem Draw tool. Extracted by "Selective Inhibitors of Protozoan Protein N-myristoyltransferases as Starting Points for Tropical Disease Medicinal Chemistry Programs" (Bell *et al.*, 2012).

Query number	Pharmacophore points	AUC
1	5	**0.714**
2	5	**0.764**
3	5	**0.727**
4	5	**0.797**
5	5	**0.849**
6	5	**0.749**
7	5	**0.755**
8	5	**0.604**
9	6	**0.739**
10	6	**0.780**
11	6	**0.682**
12	6	**0.727**
13	**6**	0.880
14	6	**0.842**
15	6	**0.845**
16	**6**	**0.743**

Table 1. Number of queries generated, pharmacophore points and AUC of the models.

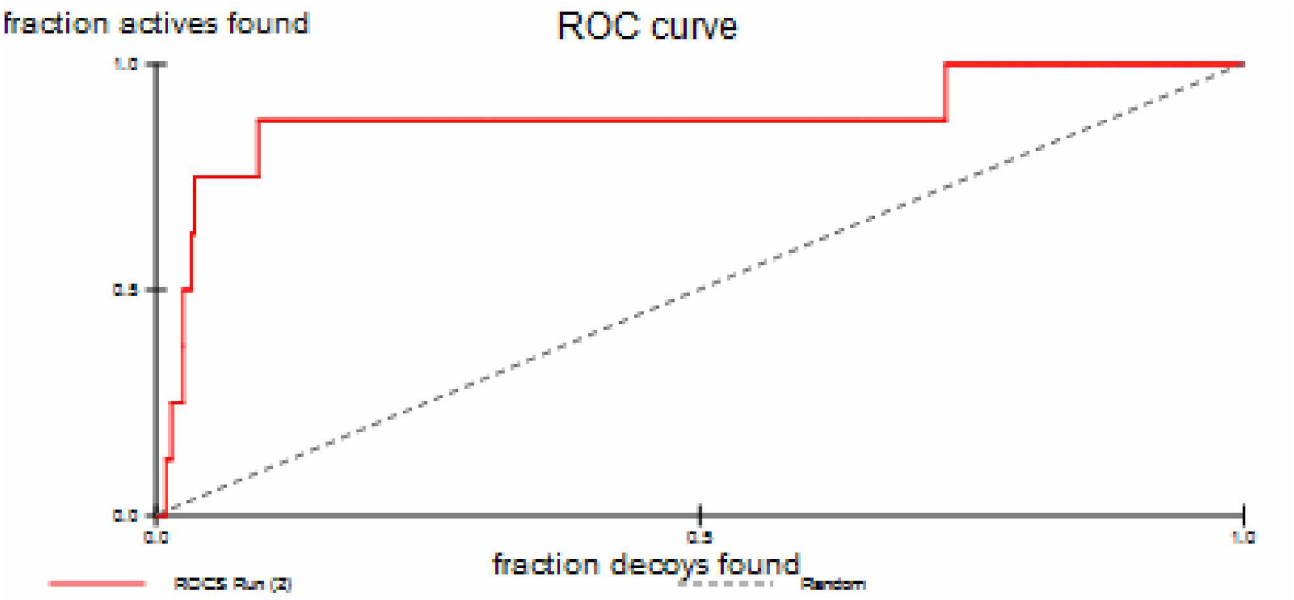

Figure 6. ROC curve obtained in the validation of the model

4.4. Molecular Anchorage and Evaluation

The best 500 compounds according to *TanimotoCombo score* were obtained from the virtual screening performed on vROCS. FRED (McGann, 2011) docking program was used to perform the molecular docking between PvNMT enzyme and the best 500 ligands of each library (cl and zbd). After docking, only the 1% best Chemgauss4 scored compounds were selected and submitted in the pkCSM analysis (Table 2 and 3).

ChemBridge - Cl

Ranking	Molecule ID	Chemgauss4 Score
1	19371731	-18.08
2	**24675937**	**-16.81**
3	27045314	-15.93
4	77009166	-15.54
5	30306998	-15.11
6	19532149	-15.08
7	99428690	-15.08
8	48788186	-14.94

Table 2. Selected compounds with best score of the ChemBridge Cl library after molecular docking.

ZINC12 - Zbd

Ranking	Molecule ID	Chemgauss4 Score
1	ZINC20358414	-17.82
2	ZINC20589172	-17.28
3	ZINC19370024	-17.20
4	ZINC20358910	-16.73
5	ZINC20394331	-16.63
6	**ZINC20391476**	**-15.82**
7	ZINC19361186	-15.17

Table 3. Selected compounds with the best score of the ZINC12 Zbd library after molecular docking.

4.5. Analysis of toxic properties in silico

PKCSM was used to evaluate the principal pharmacokinetics characteristics (ADMET) of the compounds. The importance of this step it's that through this analysis, it can be predicted if the potential inhibitors presents the suitable parameters to drug selection, aiming a possible administration in human organism. To analyze the Cl database, was selected eight molecules, according to the Table 2, and in the Zbd data base was selected seven molecules, according to the Table 3. The results after the processing in software is present in Table 4 e Table 5, being the Cl and Zbd database respectively. The results were judged by positive and negative pharmacokinetic characteristics, and in case of this work, was choose some in specific like AMES toxicity, Max tolerated dose (human), hERG I inhibitor, hERG II inhibitor, Oral Rate Acute Toxicity (LD50), Oral Rat Chronic Toxicity (LOAEL), Hepatotoxicity, Skin Sensitization and the Minnow toxicity.

Through these parameters, was possible to select the possible best compound in each library, being of Cl the second compound ID: 24675937 (2-isopropyl-N-{[2-(4-methylpiperazin-1-yl)pyridin-3-yl]methyl}pyrimidine-4-carboxamide) in Table 4 and Zbd library was selected the sixth compound, id: ZINC20391476 (7-hydroxy-3-(3-methoxyphenoxy)-8-[(4-methylpiperazin-1-yl)methyl]-2 (trifluoromethyl)chromen-4-one) in Table 5.

ChemBridge ID	AMES toxicity	Max.tolerated dose (human)	hERG I inhibitor	hERG II inhibitor	Oral Rat Acute Toxicity (LD50)	Oral Rat Chronic Toxicity (LOAEL)	Hepatotoxicity	Skin Sensitisation	Minnow toxicity
19371731	Yes	-0.75	No	Yes	2.61	2.08	Yes	No	1.45
24675937	No	-0.13	No	Yes	2.81	1.54	Yes	No	2.14
27045314	No	-0.28	No	Yes	2.81	1.44	Yes	No	2.09
77009166	No	0.41	No	Yes	2.28	1.68	Yes	No	1.49
30306998	No	0.22	No	Yes	2.63	2.68	Yes	No	0.59
19532149	No	-0.65	No	Yes	2.68	0.34	Yes	No	2.12
99428690	Yes	0.54	No	Yes	2.02	0.6	No	No	-0.64
48788186	No	0.02	No	Yes	2.59	0.89	Yes	No	0.79

Table 4. Results generated in pkCSM analysis of the Cl library after docking and evaluation.

ZINC ID	AMES toxicity	Max.tolerated dose (human)	hERG I inhibitor	hERG II inhibitor	Oral Rat Acute Toxicity (LD50)	Oral Rat Chronic Toxicity (LOAEL)	Hepatotoxicity	Skin Sensitisation	Minnow toxicity
ZINC20358414	No	-0.377	No	Yes	2.915	1.108	Yes	No	3.521
ZINC20589172	No	-0.69	No	Yes	2.986	0.488	Yes	No	3.201
ZINC19370024	No	-0.387	No	Yes	3.001	1.084	Yes	No	3.42
ZINC20358910	No	-0.919	No	Yes	3.173	0.661	Yes	No	2.897
ZINC20394331	No	-0.856	No	Yes	3.222	0.822	Yes	No	2.913
ZINC20391476	No	0.235	No	Yes	2.517	1.78	No	No	3.538
ZINC19361186	No	0.355	No	Yes	3.111	1.707	No	No	3.086

Table 5. Results generated in pkCSM analysis of the Zbd library after docking and evaluation.

4.6. Molecular Dynamics

The molecular dynamics simulation was performed on two selected compounds extracts of pkCSM analyzes (ChemBridge: 24675937 and ZINC20391476) for a further and important validation to this work. With this step, three parameters were observed: Vacuum minimum energy, Hydrogen-bonding analysis and RMSD of the complex involved in dynamics (Kumar, Garg e Bharatam, 2015). As a first parameter, in the free energy calculation, both of compounds have shown a good result, showing free energy very similar to the original ligand (Figure 7 and 8). That result showed that both of compounds present balance and stability.

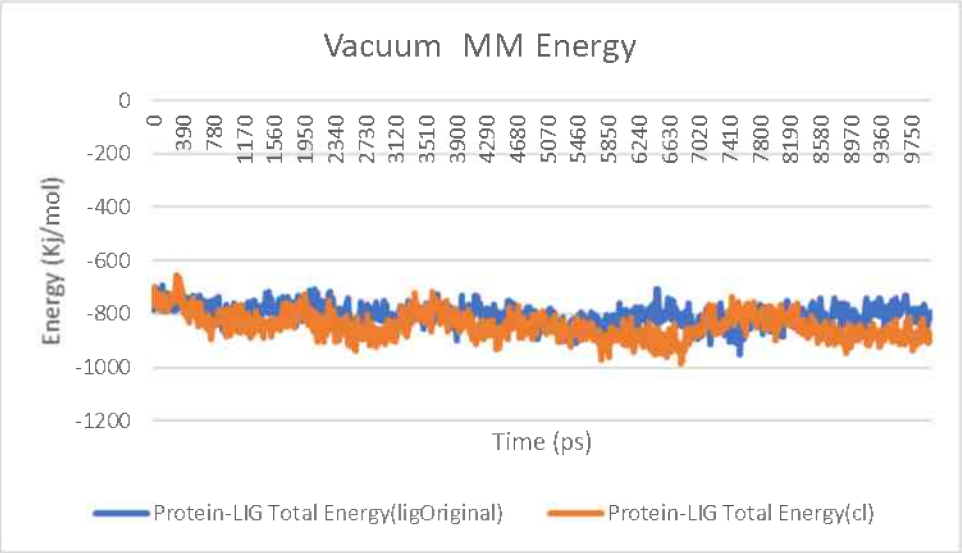

Figure 7. The minimization energy plots that demonstrate the comportment of the Cl library compound (ChemBridge: 24675937) in relation of the original ligand in complex with the NMT to verify the stability of the system during the MD simulation.

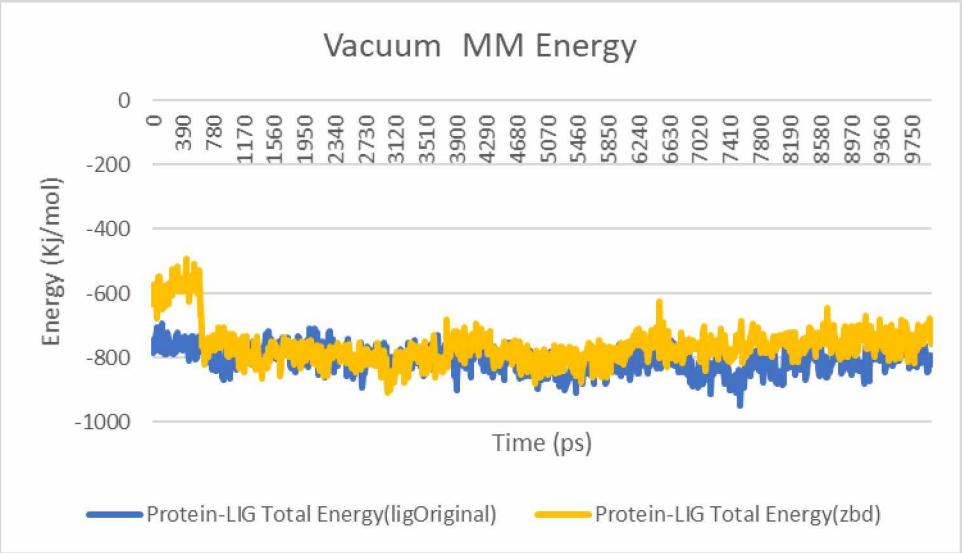

Figure 8. The minimization energy plots that demonstrate the comportment of the Zbd library compound (ZINC20391476) in relation of the original ligand in complex with the NMT to verify the stability of the system during the MD simulation.

About the hydrogen-bonding interactions, the results were very satisfactory. During the simulation the ligands have shown highly conserved interactions with the active site residues and both remain bound of active site of PvNMT. Both the compounds demonstrated a very similar hydrogen-bonding interaction to the original ligand of NMT (Figure 9 to Figure 12).

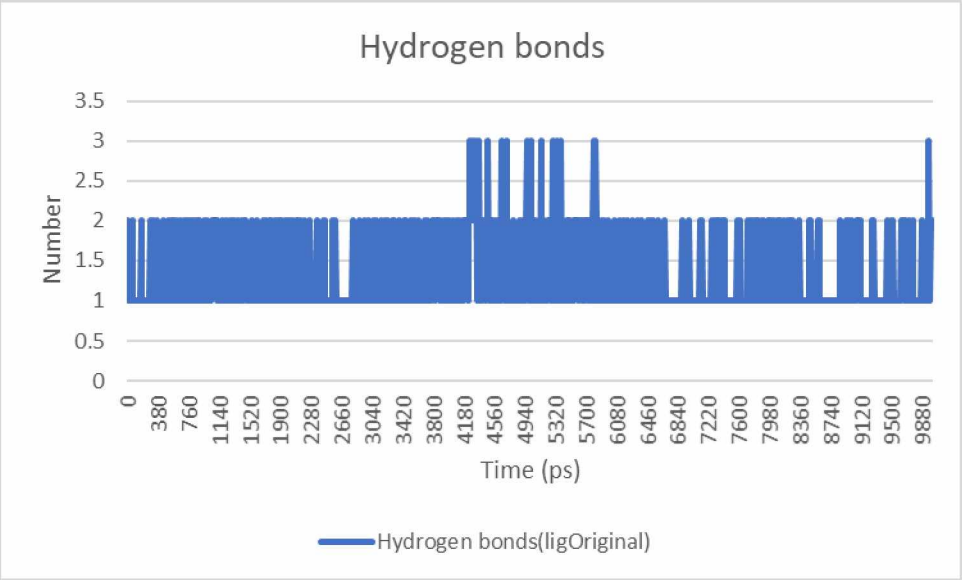

Figure 9. Numbers of hydrogen bounds formed during the simulation that was formed into the active site of NMT. In this plot was the number of hydrogen bounds of the ligand original in complex with the NMT.

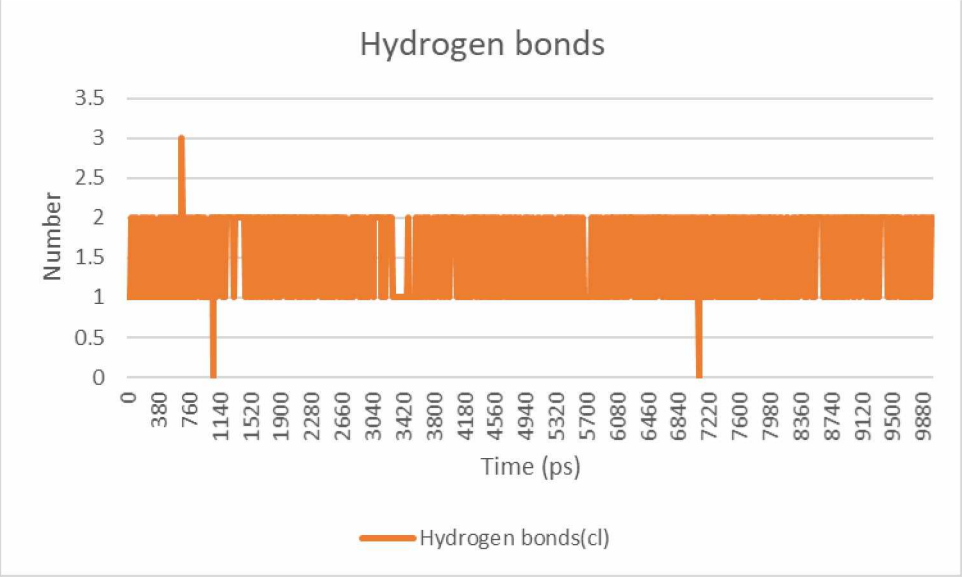

Figure 10. Numbers of hydrogen bounds formed during the simulation that was formed into the active site of NMT. In this plot was the number of hydrogen bounds of (ChemBridge: 24675937) in complex with the NMT.

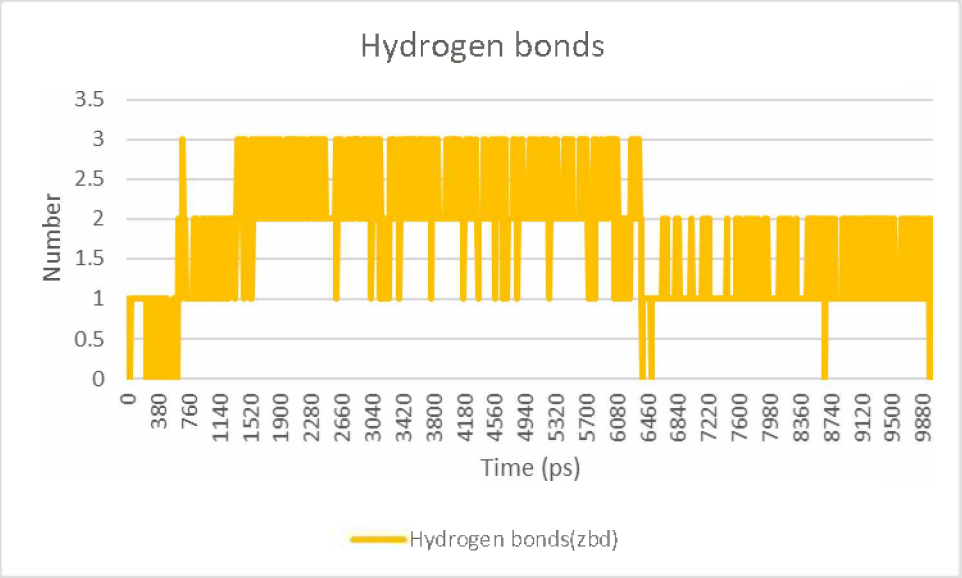

Figure 11. Numbers of hydrogen bounds formed during the simulation that was formed into the active site of NMT. In this plot was the number of hydrogen bounds of (ZINC20391476) in complex with the NMT.

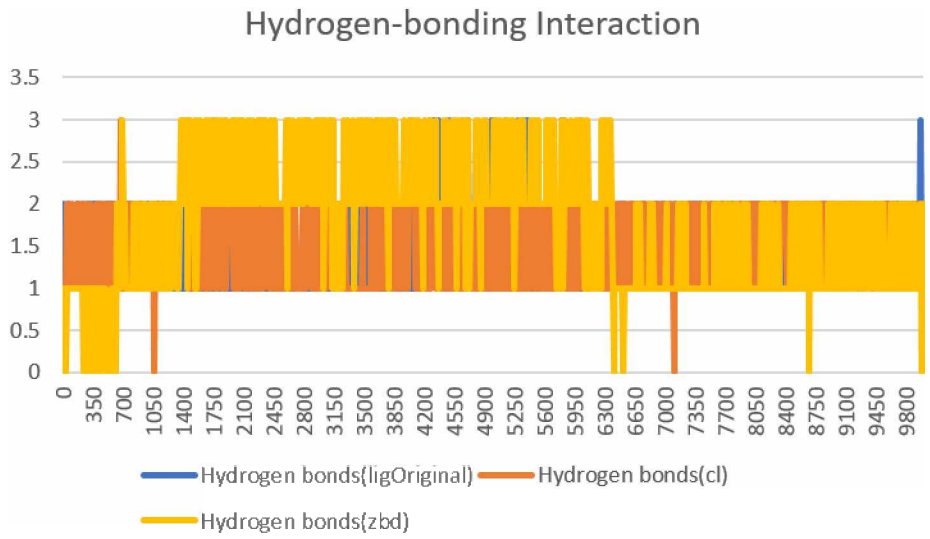

Figure 12. Numbers of hydrogen bounds formed during the simulation that was formed into the active site of NMT. In this plot was compared the number of hydrogen bounds between the ligand original in complex with the NMT and the (ChemBridge: 24675937) and (ZINC20391476) during the MD simulation.

RMSD, the root-mean-square-deviation of atomic positions, is one of most used to identify the similarity criteria in structural biology, having direct relation to the stability analysis of the compound at the active site of the protein (Neveu *et al.*, 2018). The results of this work

demonstrated that the compound (ChemBridge: 24675937) have shown a similar and stable comportment in relation of the original ligand and the protein only, in contrast the compound (ZINC20391476) have shown a higher deviation when compared to the other compounds, but even with this deviation it was still stable (Figure 13 to Figure 16).

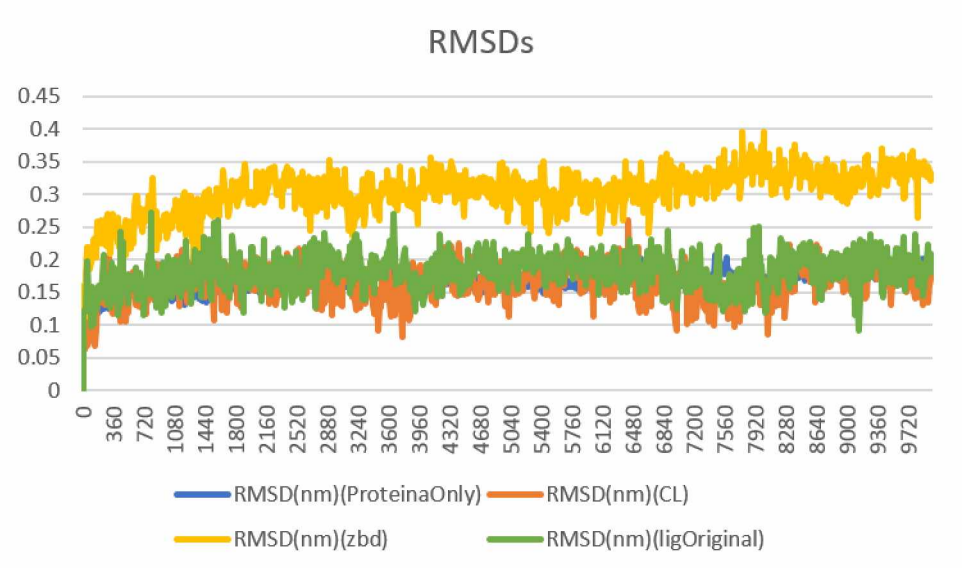

Figure 13. The RMSD plot comparing the root-mean-square deviation of atomic position. This plot demonstrated how was the RMSD of the original ligand, the protein only, the (ChemBridge: 24675937) and (ZINC20391476) during the MD simulation.

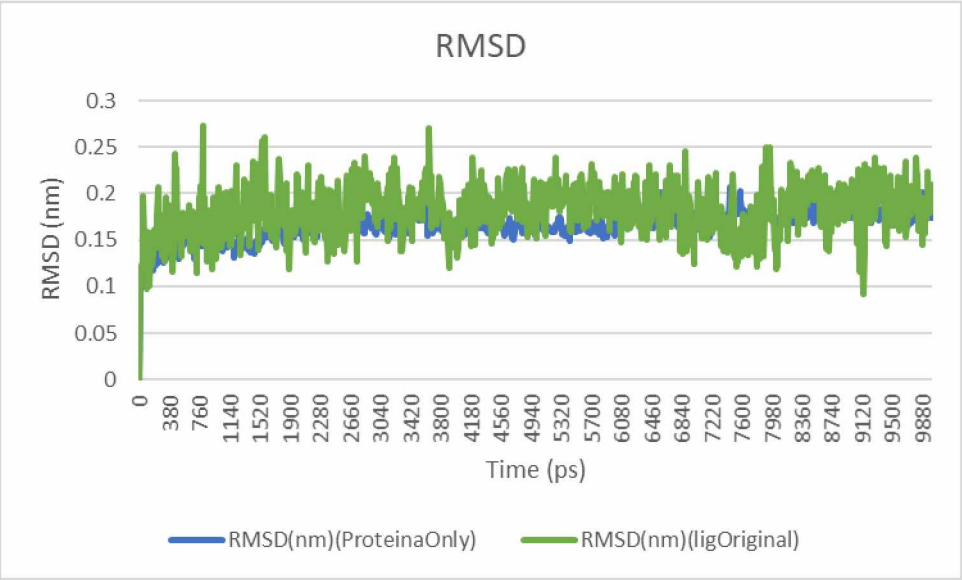

Figure 14. The RMSD plot comparing the root-mean-square deviation of atomic position. This plot demonstrated how was the RMSD of the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) and the protein only (NMT) during the MD simulation.

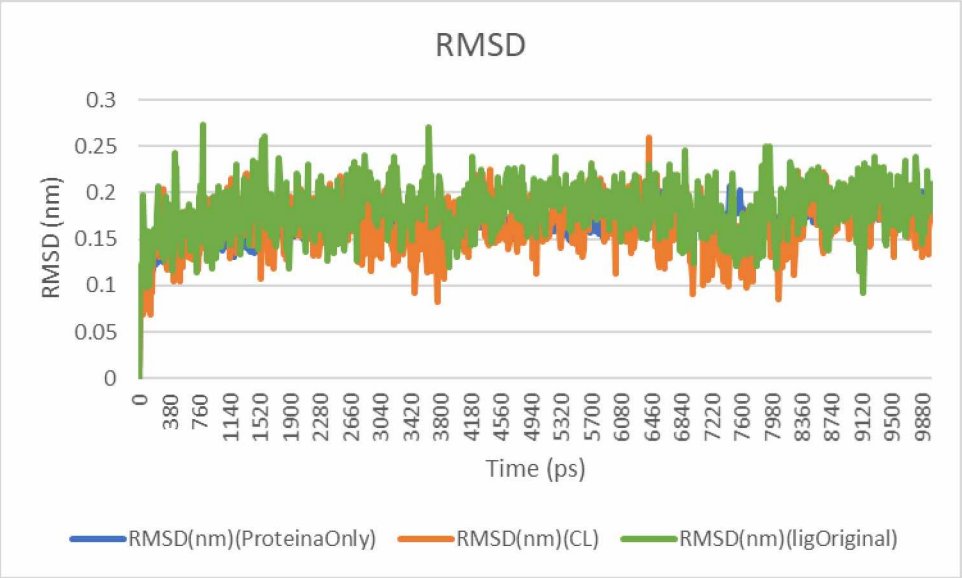

Figure 15. The RMSD plot comparing the root-mean-square deviation of atomic position. This plot demonstrated how was the RMSD of the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine), the protein only (NMT) and the (ChemBridge: 24675937) during the MD simulation.

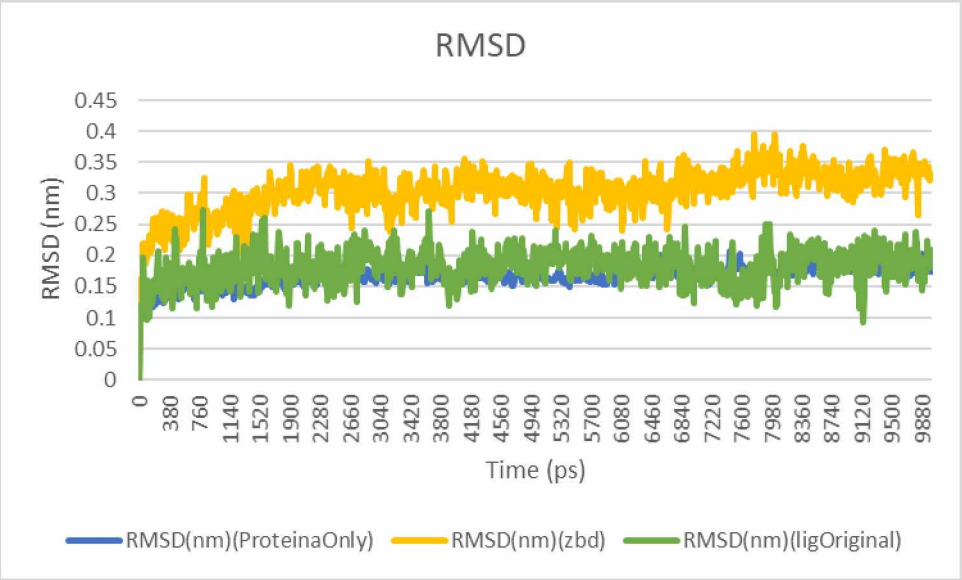

Figure 16. The RMSD plot comparing the root-mean-square deviation of atomic position. This plot demonstrated how was the RMSD of the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine), the protein only (NMT) and the (ZINC20391476) during the MD simulation.

4.7. Generation of interaction diagrams 2D protein-ligand

The LigPlot + 2D tool was used to generate 2D interaction patterns of the ligand-protein complex resulting from molecular dynamics. During the molecular dynamics simulation, about 1000 frames were generated per tested NMT complex (Cl and Zbd), and to demonstrate how the ligand interacted with the protein during the simulation, three main frames were saved: frame 0, demonstrating how the interaction was at the beginning of the dynamics, frame 500 demonstrating how the interaction was in the midst of the dynamics and the frame 1000, observing how the interaction carried in the final part of the dynamics. To have a valid comparison parameter, the same was done with the original ligand ((4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) in complex with the PvNMT.

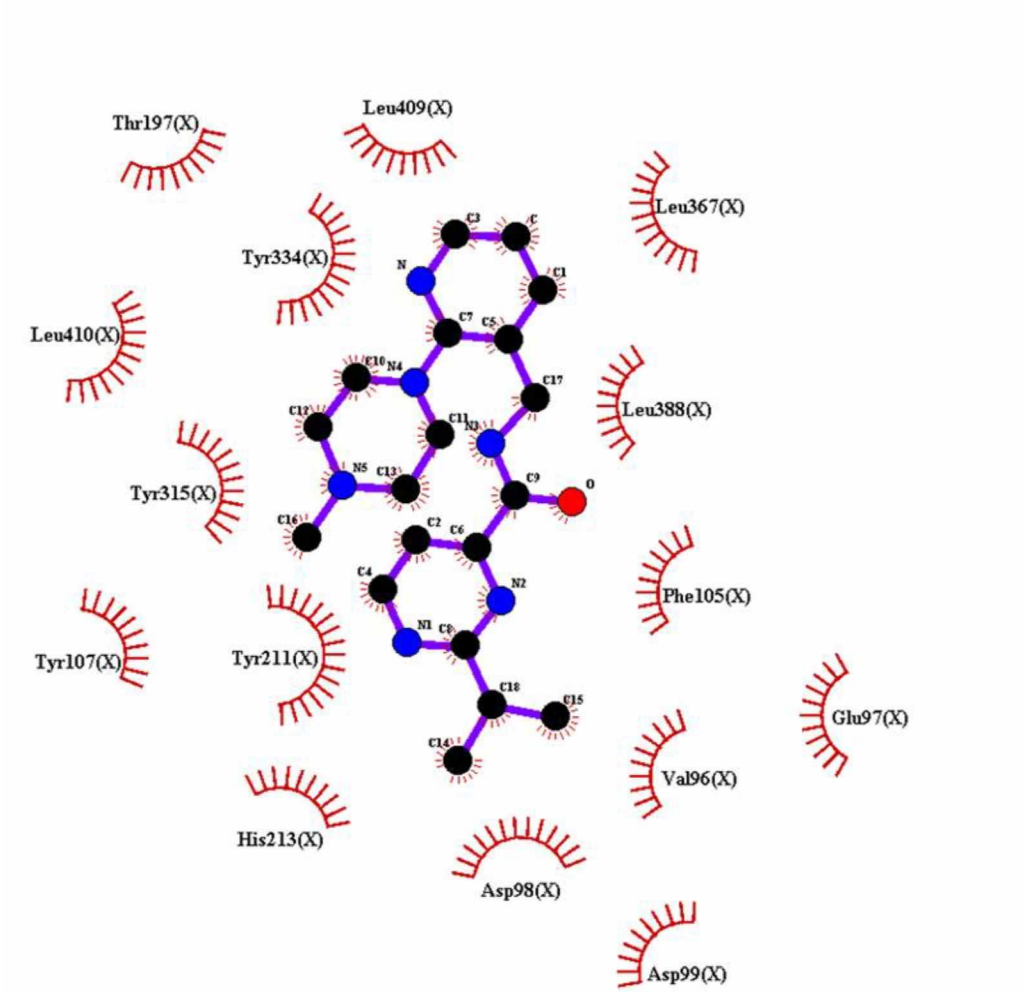

Figure 17. This plot shown the interaction in the frame 0 between the ligand (ChemBridge: 24675937) with the NMT. The ligand is represented by the purple color in the protein active site, and the shapes in red brick represents the hydrophobic interactions during the MD.

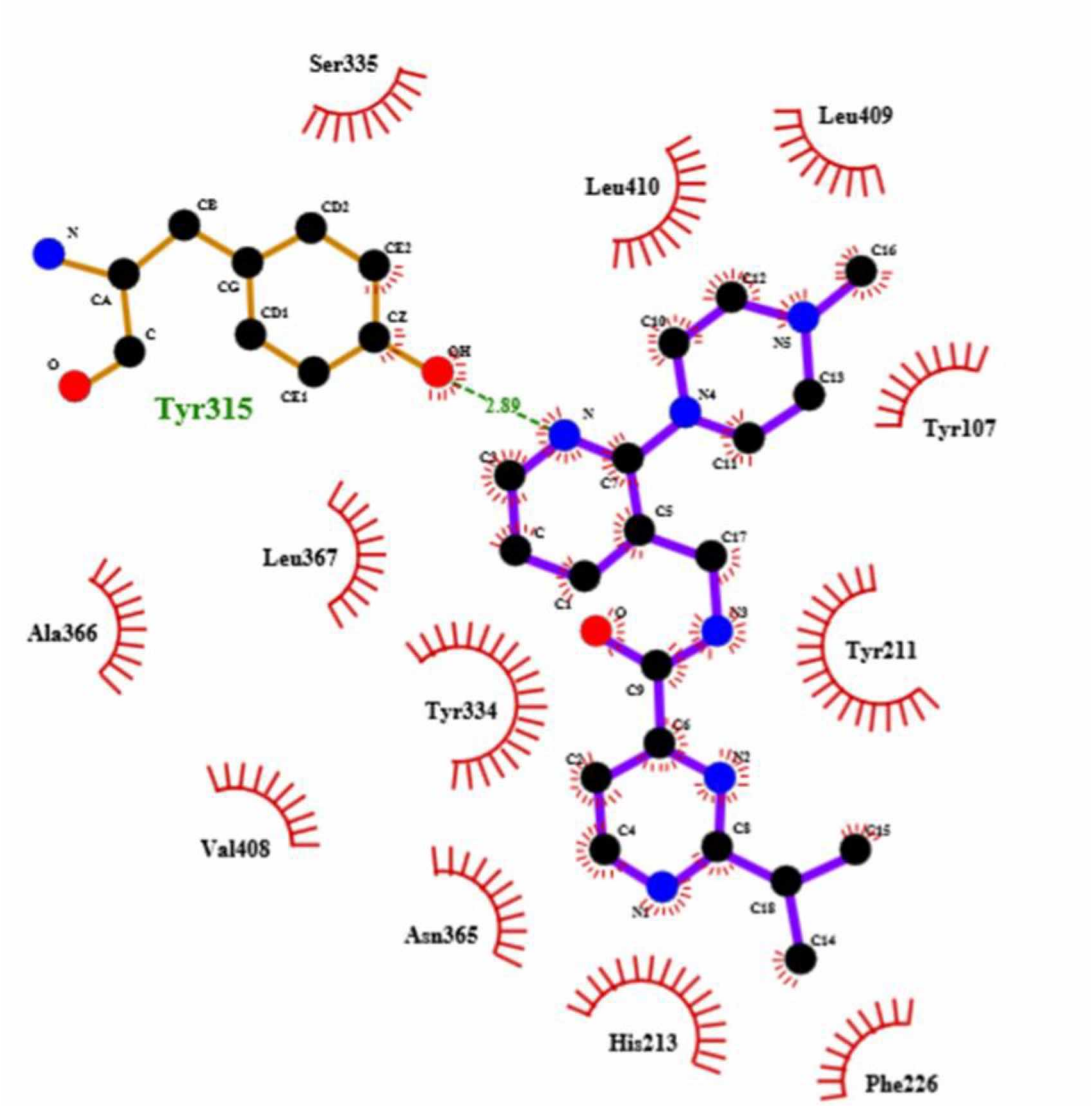

Figure 18. This plot shown the interaction in the frame 500 between the ligand (ChemBridge: 24675937) with the NMT. The ligand is represented by the purple color in the protein active site, the brown color represents the amino acids, the shapes in red brick represents the hydrophobic interactions and the green dashed lines represents the hydrogen bonds during the MD.

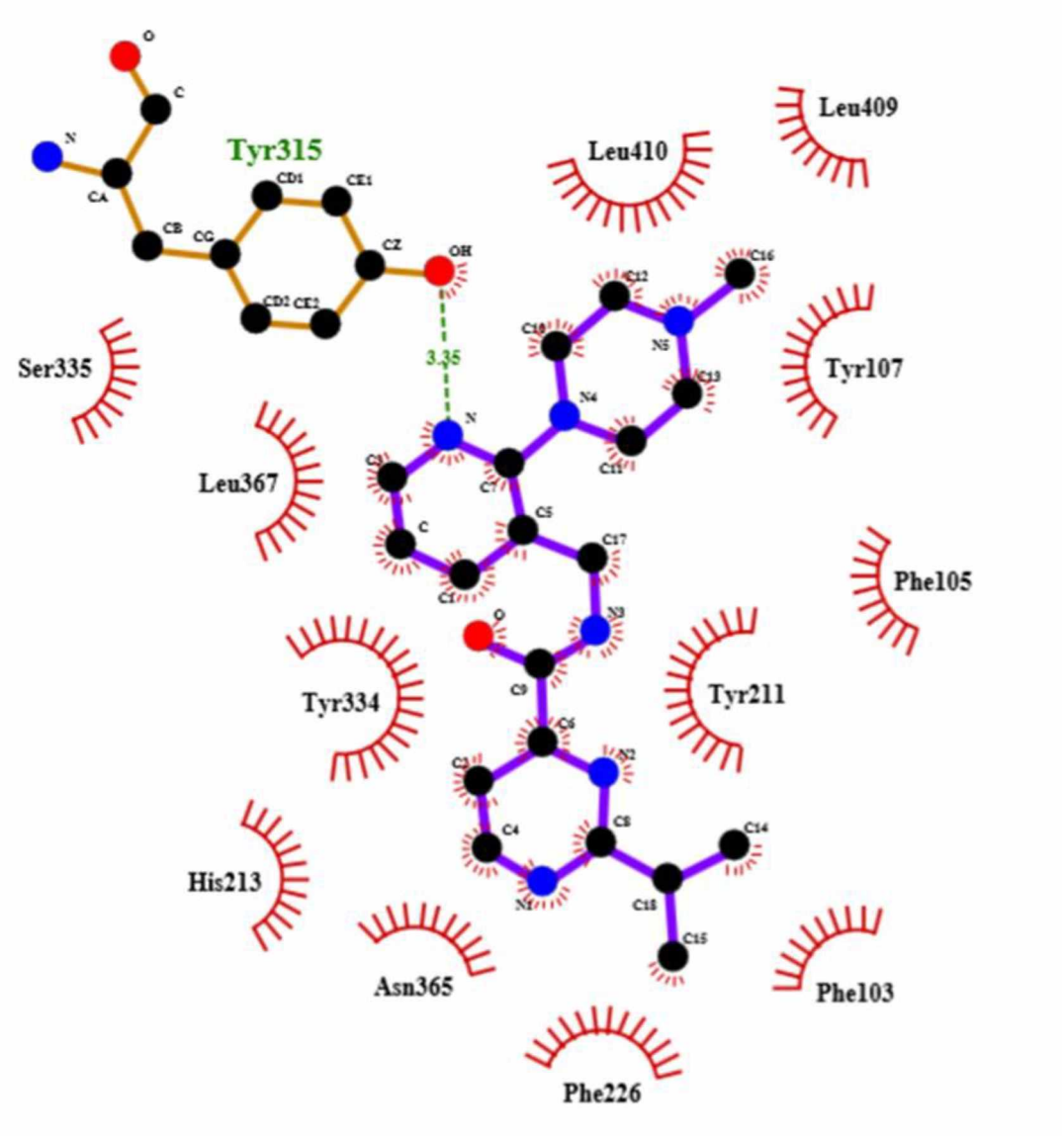

Figure 19. This plot shown the interaction in the frame 1000 between the ligand (ChemBridge: 24675937) with the NMT. The ligand is represented by the purple color in the protein active site, the brown color represents the amino acids, the shapes in red brick represents the hydrophobic interactions and the green dashed lines represents the hydrogen bonds during the MD.

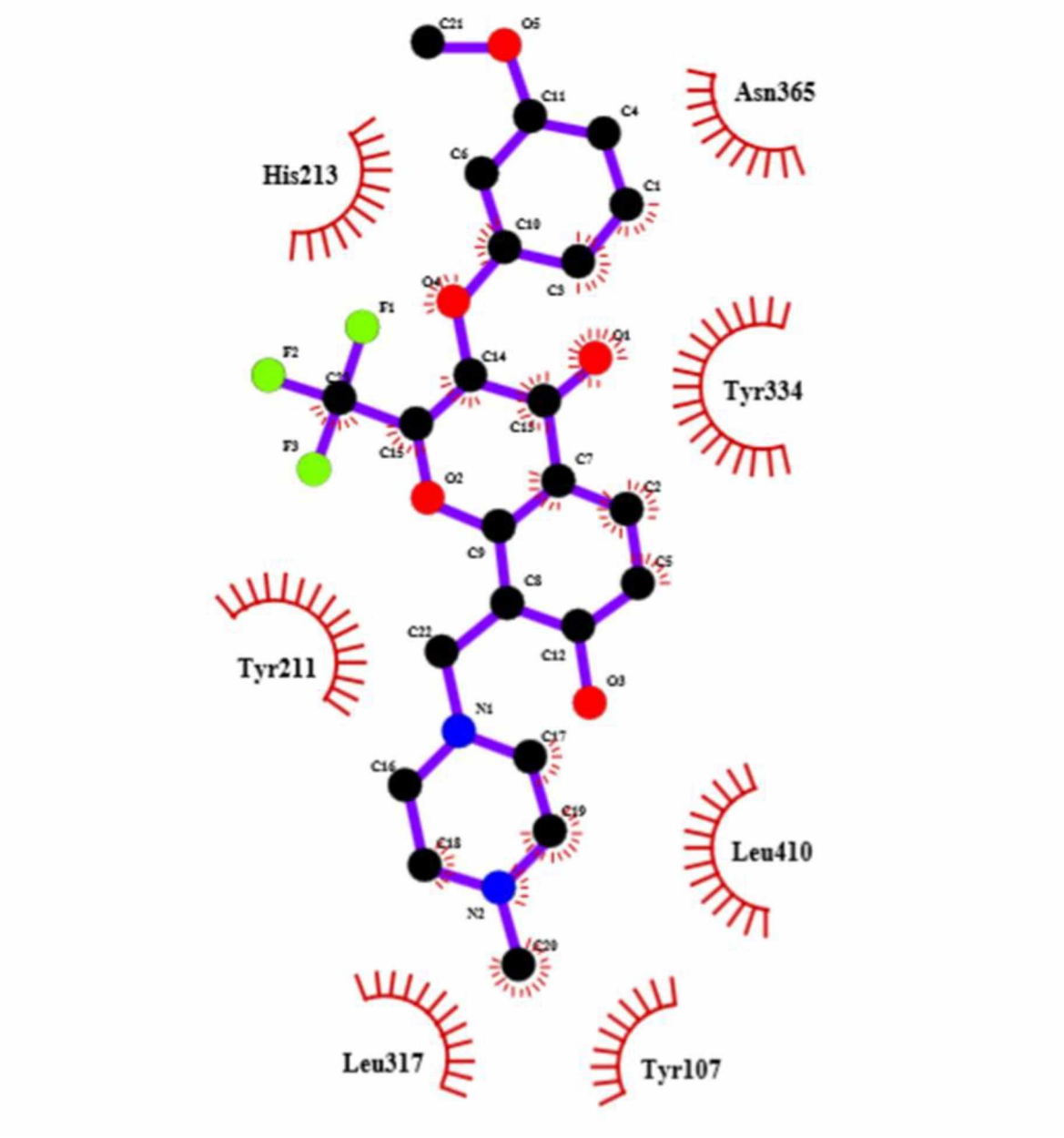

Figure 20. This plot shown the interaction in the frame 0 between the ligand (ZINC20391476) with the NMT. The ligand is represented by the purple color in the protein active site, and the shapes in red brick represents the hydrophobic interactions during the MD.

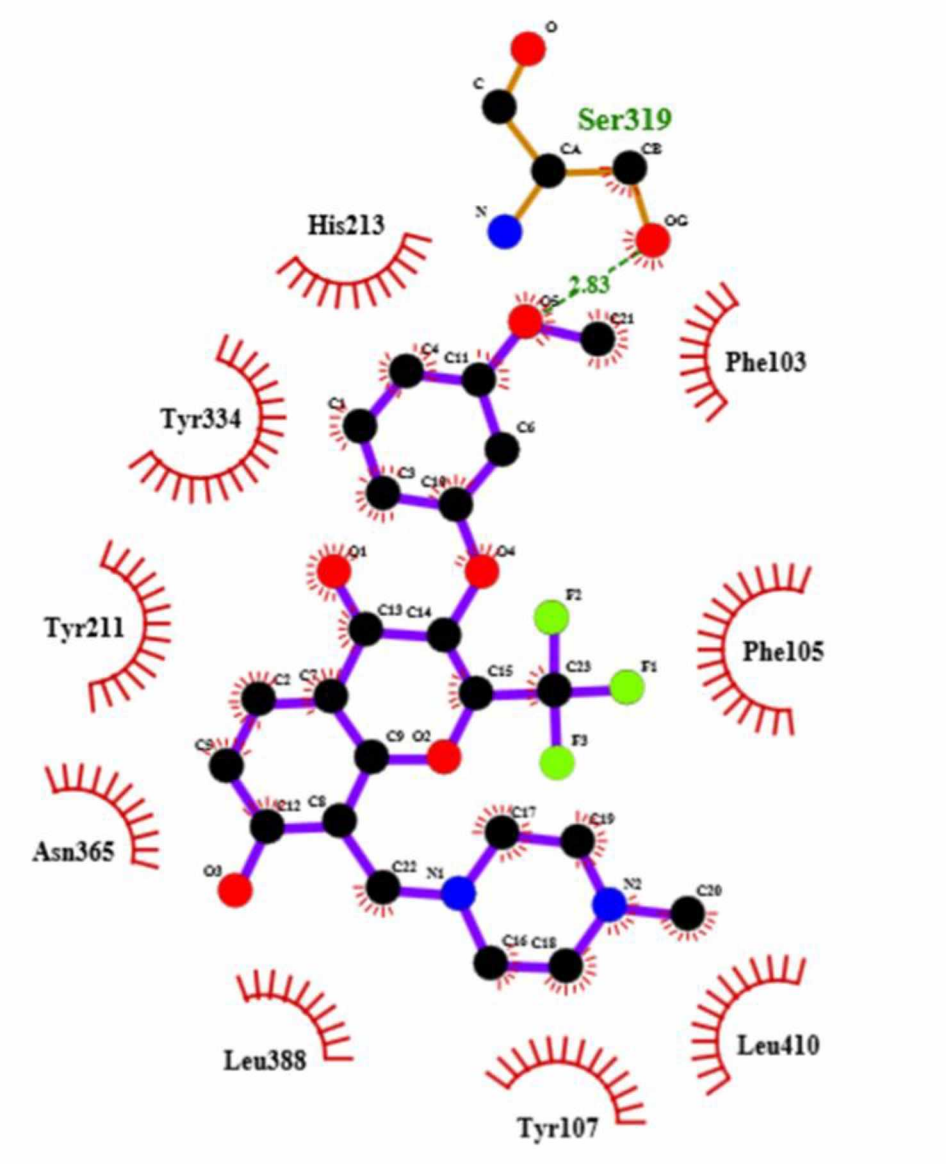

Figure 21. This plot shown the interaction in the frame 500 between the ligand (ZINC20391476) with the NMT. The ligand is represented by the purple color in the protein active site, the brown color represents the amino acids, the shapes in red brick represents the hydrophobic interactions and the green dashed lines represents the hydrogen bonds, during the MD.

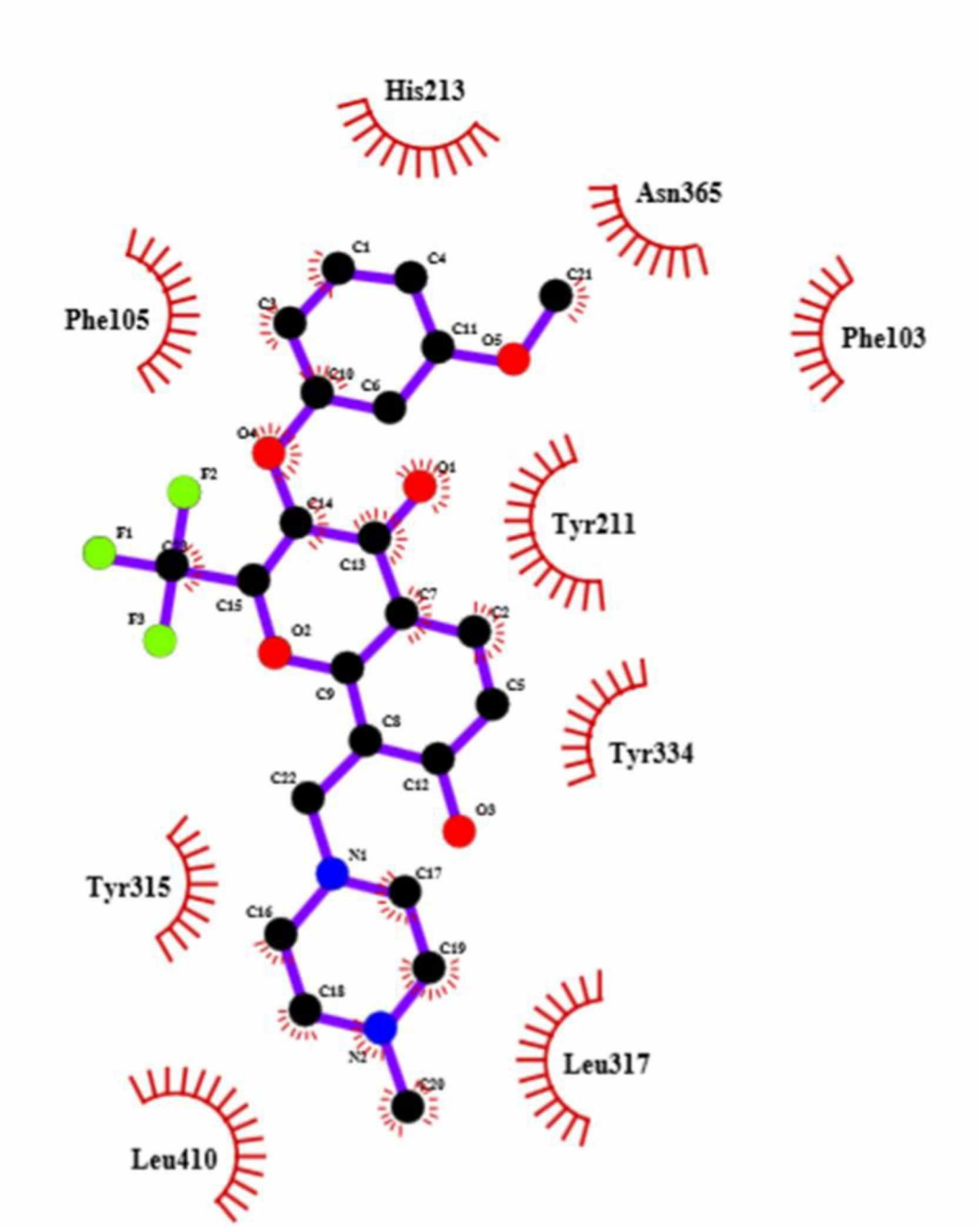

Figure 22. This plot shown the interaction in the frame 1000 between the ligand (ZINC20391476) with the NMT. The ligand is represented by the purple color in the protein active site and the shapes in red brick represents the hydrophobic interactions during the MD.

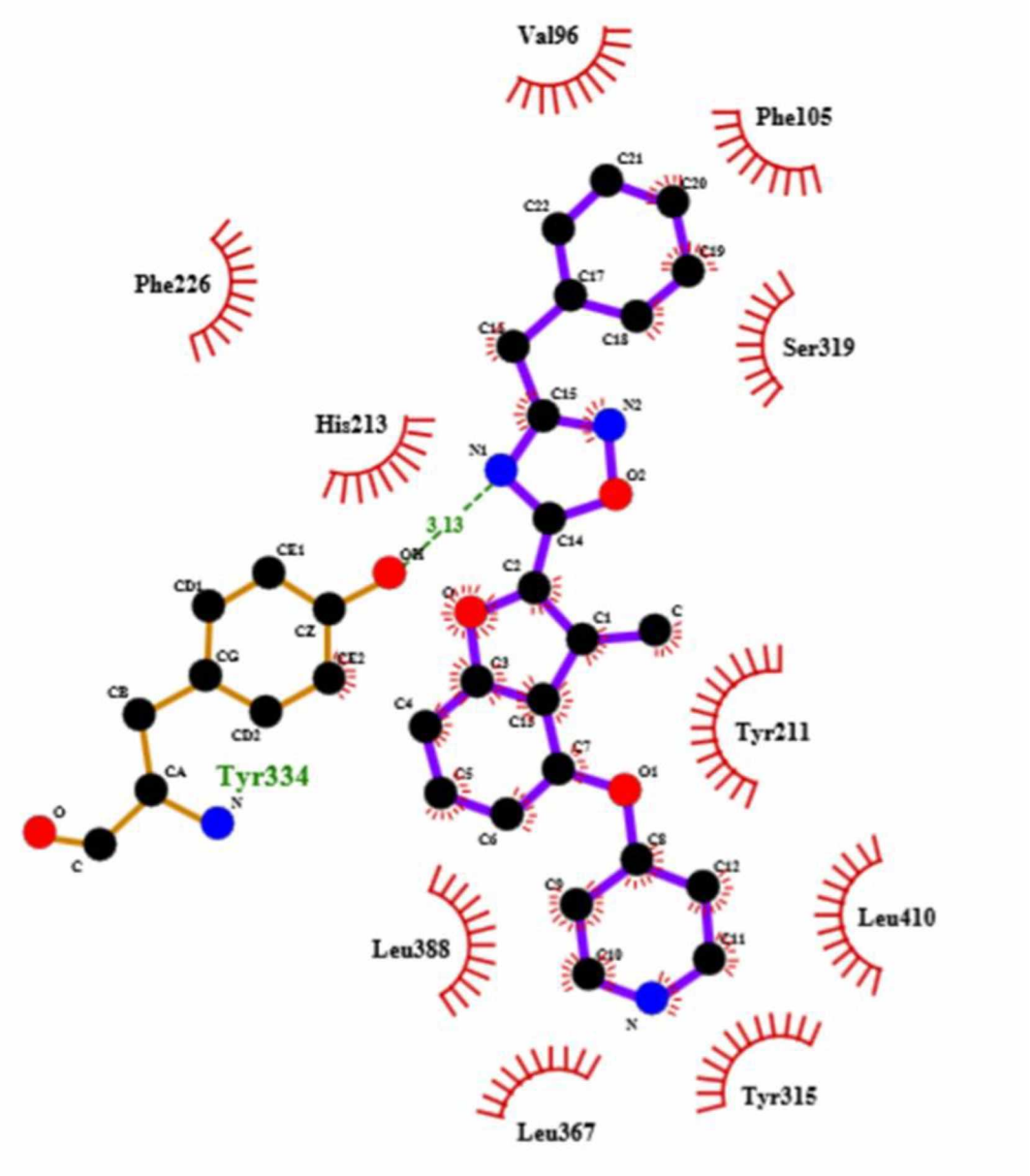

Figure 23. This plot shown the interaction in the frame 0 between the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) with the NMT. The ligand is represented by the purple color in the protein active site, the brown color represents the amino acids, the shapes in red brick represents the hydrophobic interactions and the green dashed lines represents the hydrogen bonds, during the MD.

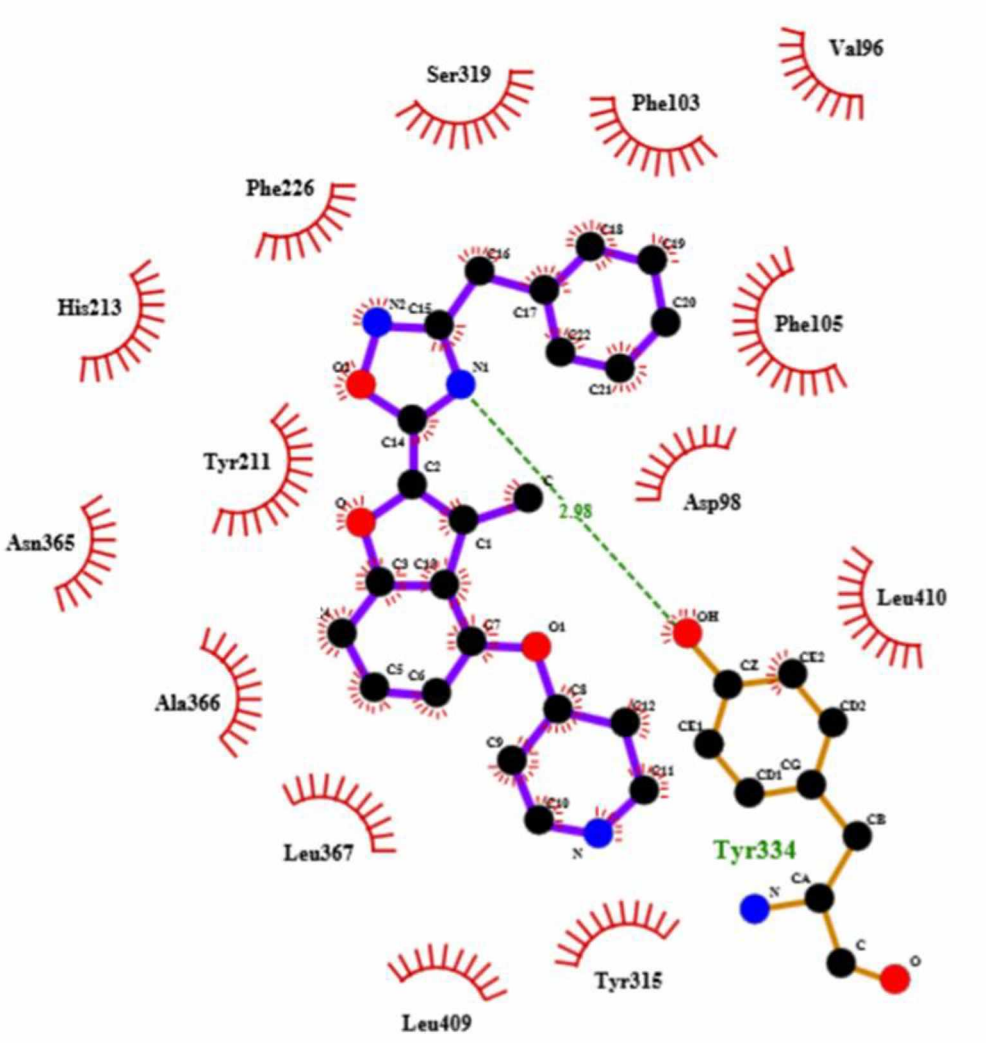

Figure 24. This plot shown the interaction in the frame 500 between the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) with the NMT. The ligand is represented by the purple color in the protein active site, the brown color represents the amino acids, the shapes in red brick represents the hydrophobic interactions and the green dashed lines represents the hydrogen bonds, during the MD.

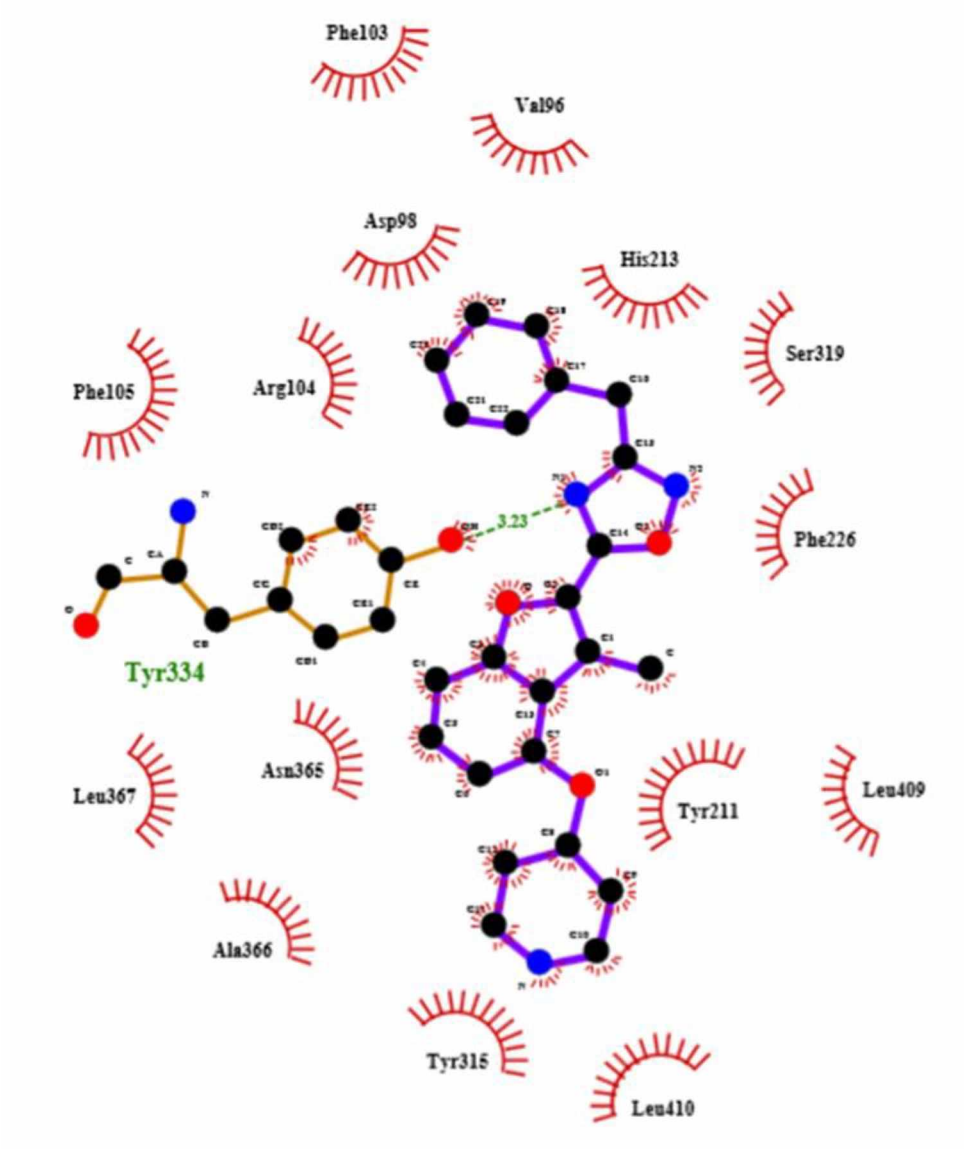

Figure 25. This plot shown the interaction in the frame 1000 between the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) with the NMT. The ligand is represented by the purple color in the protein active site, the brown color represents the amino acids, the shapes in red brick represents the hydrophobic interactions and the green dashed lines represents the hydrogen bonds, during the MD.

When analyzing the interactions in each frame, it was visualized that the compound (ChemBridge: 24675937) shown more hydrogen bonds interaction, both in frame 500 and 1000 (Figures 18 and 19), while the compound (ZINC20391476) demonstrated this interaction just in frame 500 (Figure 21). Then, compared to the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) which demonstrated hydrogen bonds in all frames (Figures 23-25), it can be said that the compound (ChemBridge: 24675937) during the MD was more stable and it resembled to the original ligand, while the (ZINC20391476) was shown stable but with greater variation.

In order to improve the results, a possible alternative would be to further reduce the minimization energy in the vacuum, increasing the duration of the simulation in the system, in order to observe if there is a greater stability of the molecule within the complex with a longer time.

5. Conclusion

In this work it was possible to create a shape-based model based on the structure of the ligands from a complex with the PvNMT, carry out virtual screening and molecular docking using ligand libraries and perform toxicity analyses on the best ligands. The toxicity analyses has showed two potential compounds to be used as inhibitors against *Plasmodium vivax* NMT.

The MD was performed to simulate a biological system and to show the atomic interactions between the protein-ligand inside a water box, in controlled pressure, temperature and pH. Thus, it was possible to obtain a balanced system, allowing for more detailed analyzes on the specificity and stability of the selected ligand in the active site of the protein.

By analyzing all the results presented, it was possible to obtain good indications that the ligands tested could act as inhibitors in complex with NMT. The Cl library compound (ChemBridge: 24675937) was better in relation to the Zbd library compound (ZINC20391476), having greater similarity with the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine) and its interactions. Compound (ChemBridge: 24675937) demonstrated good stability at the active site of the protein, with relevant hydrogen bonding interactions, which made the compound even more stable in the active site of the protein. Although (ZINC20391476) does not show similar results to that of the original ligand (4-{[2-(3-benzyl-1,2,4-oxadiazol-5-yl)-3-methyl-1-benzofuran-4-yl]oxy}piperidine), it is still a potential molecule to act as an inhibitor in complex with NMT, because it also reached certain stability during the simulation, requiring some adjustments such as: decrease of minimizing energy in vacuum and improvement in root-mean-square-deviation of atomic positions parameter.

The tools used in the area of bioinformatics are extremely important and can optimize some analyzes, but these separately are not enough to fully validate such methodologies. In order to have a complete closure and full validation, it would be necessary to proceed to another step, which would be *in vitro* and *in vivo* tests. This work achieved the objective of finding possible effective inhibitors, and if synthesized and tested could help on the development of new drugs against malaria.

6. References

ABRAHAM, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, 2015.

BANNISTER, L. H.; SHERMAN, I. W. **Plasmodium**, 2009. (Nota técnica).

BEIER, J. C. Malaria parasite development in mosquitoes. **Annual Review of Entomology**, v. 43, n. 1, p. 519–543, 1998.

BELL, A. S. *et al.* Selective inhibitors of protozoan protein N-myristoyltransferases as starting points for tropical disease medicinal chemistry programs. **PLoS Neglected Tropical Diseases**, v. 6, n. 4, 2012.

BRANNIGAN, J. A. *et al.* Diverse modes of binding in structures of Leishmania major N-myristoyltransferase with selective inhibitors. **IUCrJ**, 2014.

BURCHARD, G. D. [Malaria]. **Der Internist**, 2014.

FREARSON, J. A. *et al.* N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. **Nature**, v. 464, n. 7289, p. 728–732, 2010.

GREENWOOD, B. M. *et al.* Malaria. **Lancet**, 2014.

HAWKINS, P. C. D. *et al.* Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. **Journal of Chemical Information and Modeling**, v. 50, n. 4, p. 572–584, 2010.

HAWKINS, P. C. D.; SKILLMAN, A. G.; NICHOLLS, A. Comparison of shape-matching and docking as virtual screening tools. **Journal of Medicinal Chemistry**, 2007.

IRWIN, J.; SHOICHET, B. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. v. 45, n. 1, p. 177–182, 2005.

J.-E., T. *et al.* **Diagnostic and treatment of malaria attackRevue du Praticien**, 1998.

KIM, S. *et al.* PubChem substance and compound databases. **Nucleic Acids Research**, 2016.

KITCHEN, D. B. *et al.* **Docking and scoring in virtual screening for drug discovery: Methods and applicationsNature Reviews Drug Discovery**, 2004.

KUMAR, R.; GARG, P.; BHARATAM, P. V. Shape-based virtual screening, docking, and molecular dynamics simulations to identify Mtb -ASADH inhibitors. **Journal of Biomolecular Structure and Dynamics**, v. 33, n. 5, p. 1082–1093, 2015.

KUNTZ, I. D. *et al.* A geometric approach to macromolecule-ligand interactions. **Journal of Molecular Biology**, v. 161, n. 2, p. 269–288, out. 1982.

LEACH, A. R. *et al.* **Three-dimensional pharmacophore methods in drug discoveryJournal of Medicinal Chemistry**, 2010.

LEE, K. S.; VYTHILINGAM, I. Plasmodium knowlesi: Emergent human malaria in Southeast Asia. *In*: **Parasites and their vectors: A special focus on Southeast Asia**. [s.l: s.n.]. .

MARKT, P.; SCHUSTER, D.; LANGER, T. Pharmacophore Models for Virtual Screening. *In*: **Virtual Screening**. [s.l: s.n.]. p. 115–152.

MCGANN, M. FRED pose prediction and virtual screening accuracy. **Journal of Chemical Information and Modeling**, 2011.

MYSINGER, M. M. *et al.* Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. **Journal of Medicinal Chemistry**, v. 55, n. 14, p.

6582–6594, jul. 2012.

NEVEU, E. *et al.* RapidRMSD: rapid determination of RMSDs corresponding to motions of flexible molecules. **Bioinformatics**, 2018.

PHILLIPS, M. A. *et al.* Malaria. **Nature Reviews Disease Primers**, v. 3, 2017.

PIRES, D. E. V.; BLUNDELL, T. L.; ASCHER, D. B. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. **Journal of Medicinal Chemistry**, 2015.

PRICE, H. P. *et al.* Myristoyl-CoA:Protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. **Journal of Biological Chemistry**, 2003.

RESH, M. D. **Targeting protein lipidation in diseaseTrends in Molecular Medicine**, 2012.

ROGNAN, D. Docking Methods for Virtual Screening : Principles and Recent Advances. *In*: **Virtual Screening**. [s.l: s.n.]. p. 153–176.

SAMISH, I.; GU, J.; KLEIN, M. Protein Motion: Simulation. *In*: **Structural Bioinformatics**. 2nd. ed. [s.l.] Wiley-Blackwell, 2009. p. 907–936.

SCHNEIDMAN-DUHOVNY, D. *et al.* PharmaGist: a webserver for ligand-based pharmacophore detection. **Nucleic acids research**, v. 36, n. Web Server issue, p. W223–W228, jul. 2008.

SHLICK, T. Molecular Dynamics: Basics. *In*: **Molecular Modeling and Simulation: An Interdisciplinary Guide**. 2nd. ed. [s.l: s.n.]. p. 425–462.

SHOICHET, B. K. **Virtual screening of chemical librariesNature**, 2004.

SONNENBURG, E. D.; GORDON, J. I. Protein N-Myristoylation. *In*: **Encyclopedia of Biological Chemistry: Second Edition**. [s.l: s.n.]. .

SUH, K. N.; KAIN, K. C.; KEYSTONE, J. S. Malaria. **CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne**, 2004.

WALLACE, A. C.; LASKOWSKI, R. A.; THORNTON, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. **Protein engineering**, 1995.

WERMUTH, C. G. *et al.* Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). **Pure and Applied Chemistry**, 1998.

WHITE, N. J. *et al.* Malaria. **The Lancet**, v. 383, n. 9918, p. 723–735, 2014.

YANG, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug Discovery Today**, 2010.

ZOETE, V. *et al.* SwissParam: A fast force field generation tool for small organic molecules. **Journal of Computational Chemistry**, 2011.